GAUHATI UNIVERSITY
Institute of Distance and Open Learning

OPERATIONS RESEARCH AND COMPUTER IN BUSINESS

M Com • Second Semester • 2.8

www.idolgu.in
E-Portal: www.bodhidroom.idolgu.org

Course Coordinator

Prof. Kandarpa Das	Director, GUIDOL
Prof. H.C. Gautam	H.O.D, Dept. of Commerce, GU.
Dipankar Saikia	Editor, Study Materials, GUIDOL
Gitartha Goswami	Asstt. Director, GUIDOL
Rajen Chetry	Academic Consultant, GUIDOL

Contributors [Authors]

CR Kothari, Ex-principal, University Commerce College, Jaipur

Units (1.3-1.7, 1.9-1.14, 2.0-2.3, 2.8, 4.0-4.3, 4.4-4.6) © CR Kothari, 2014

S Kalavathy, Professor, Department of Mathematics, RMD Engineering College, Kavaraipettai, Thiruvallur

Units (1.0-1.2, 1.8, 2.4-2.7, 2.9-2.13, 3, 4.3.1-4.3.2) © Reserved, 2014

Rohit Khurana, CEO and Founder, ITL Education Solutions, New Delhi

Units (5.4.3, 5.6 - 5.6.3, 6.0-6.7) © ITL Education Solutions, 2014

Saurabh Sharma, Assistant Professor, MCA Department and Coordinator, Cisco Courses

Vidya College of Engineering, Meerut

Units (6.8-6.15) © Reserved, 2014

Vikas® Publishing House: Units (4.7 - 4.12, 5.0-5.4.2, 5.4.4-5.5.6, 5.6.4-5.11) © Reserved, 2014

Editorial Team [Content & Language Editing]

Dr. Pranjal Sarma	Guest Faculty, Dept. of Commerce, GU, Asstt. Prof., L.C.B. College, Guwahati
Dipankar Saikia	Editor Study Material, GUIDOL
Rajen Chetry	Academic Consultant, GUIDOL

Reviewed By

Dr. Pranjal Sarma	Guest Faculty, Dept. of Commerce, GU., Asstt. Prof., L.C.B. College, Guwahati

All rights reserved. No part of this publication which is material protected by this copyright notice may be reproduced or transmitted or utilized or stored in any form or by any means now known or hereinafter invented, electronic, digital or mechanical, including photocopying, scanning, recording or by any information storage or retrieval system, without prior written permission from the Publisher.

Information contained in this book has been published by VIKAS® Publishing House Pvt. Ltd. and has been obtained by its Authors from sources believed to be reliable and are correct to the best of their knowledge. However, the Publisher & its Authors shall in no event be liable for any errors, omissions or damages arising out of use of this information and specifically disclaim any implied warranties or merchantability or fitness for any particular use.

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT LTD
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: 7361, Ravindra Mansion, Ram Nagar, New Delhi - 110 055
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Operations Research and Computer in Business

Syllabi	Mapping in Book
Group - A: Operations Research	
Unit I: Operations Research: An Introduction Origin and Development of Operations Research (OR), Phases of OR, Characteristics of OR Model, Role of OR in Managerial Decision-Making, Uses and Limitations of OR.	**Unit 1:** Operations Research: An Introduction **(Pages 3–26)**
Unit II: Linear Programming Problem (LPP) Meaning, Uses and Limitations, Formulation of General LPP, Solutions of LPP, Solutions of LPP by Graphic Method and Simplex Method, Artificial Variable Technique, Duality in Linear Programming, Economic Interpretations of Dual Variables.	**Unit 2:** Linear Programming Problem (LPP) **(Pages 27–82)**
Unit III: Assignment and Transportation Problems Concept of Assignment Problem; Transportation Problem: Definition, Mathematical Formulation of Transportation Problem, To Find the Initial Basic Feasible Solution, To Derive Optimal Solution by MODI Method, Difference between a Transportation Problem and an Assignment Problem.	**Unit 3:** Transportation and Assignment Problems **(Pages 83–129)**
Unit IV: Theory of Games and Investment Analysis **Theory of Games:** Origin, Two Person and Zero-Sum (Rectangular) Game, Strategy, Maximin and Minimax Criteria of Optimality, Saddle Point, Solution of a Rectangular Game With and Without Saddle Point, Limitations of Games in Competition. **Investment Analysis:** Methods of Investment Analysis, Break-Even Analysis, Payback Period Method, Average (Accounting) Rate of Return Method, Time Adjusted or Discounted Cash Flow (DCF) Method, Discounted Rate, Certainty-Equivalent Approach, Expected Monetary Value and Hillier and Hertz's Models.	**Unit 4:** Theory of Games and Investment Analysis **(Pages 131–172)**
Group - B: Computer in Business	
Unit V: System Analysis and Design **System Analysis and Design:** Concept of System, Characteristics of a System, Elements of a System, Types of Systems, Systems Models, Open or Closed Systems, Man-Made Information System, Computer Based Information System. **Information System Development:** The Process of System Development, System Life Cycle, System Development Methodologies. **System Analysis:** System Planning and Initial Investigation, Information Gathering Tools: Interview, Questionnaire; Tools Used in Structured Analysis: Data Flow Diagram (DFD), Data Dictionary, Structured English, Decision Trees, Decision Tables. **System Design:** The Process and Stages of System Design, Logical and Physical Design, Design Methodologies, Input/Output and Forms Design; File Organization and Database Design. **System Implementation:** System Testing, Quality Assurance, Hardware/Software Selection, System Security.	**Unit 5:** System Analysis and Design **(Pages 173–239)**
Unit VI: Introduction to E-Commerce Meaning and Concept; Scope and Promise; E-Commerce and E-Business; Lure of E-Commerce; Different Categories of E-Commerce; Encryption and Authentication; Elements of Encryption, Encryption Keys; Different Devices of Authentication, Digital Signatures, Categories of Digital Signature, Cryptographic Algorithms, Digital Certificates, Secure Socket Layer (SSL), Secure Electronic Transaction (SET), Biometrics; Future of E-Commerce.	**Unit 6:** Introduction to E-Commerce **(Pages 241–287)**

CONTENTS

INTRODUCTION 1–2

UNIT 1: OPERATIONS RESEARCH: AN INTRODUCTION 3–26

1.0 Introduction
1.1 Objectives
1.2 Operations Research: Meaning, Nature and Origin
  1.2.1 Nature of Operations Research
1.3 Development of Operational Research
1.4 Operations Research in India
1.5 Operations Research as a Tool in Decision-Making
1.6 Operations Research and Management
  1.6.1 Significance of Operations Research
  1.6.2 Operations Research and Modern Business Management
1.7 Features and Methodology of Operations Research and Phases of Operations Research Study
  1.7.1 Methodology of Operations Research
1.8 Models in Operations Research and Methods of Deriving the Solution
  1.8.1 Advantages of a Model; 1.8.2 Classification of Models
1.9 Limitations of Operations Research
1.10 Summing Up
1.11 Key Terms
1.12 Answers to 'Check Your Progress'
1.13 Questions and Exercises
1.14 References and Suggested Readings

UNIT 2: LINEAR PROGRAMMING PROBLEM (LPP) 27–82

2.0 Introduction
2.1 Objectives
2.2 Linear Programming: Meaning and Uses
  2.2.1 Meaning of Linear Programming
  2.2.2 Fields Where Linear Programming can be Used
2.3 Concepts, Notations and General Form of Linear Programming Model
  2.3.1 Basic Concepts and Notations
  2.3.2 General Form of the Linear Programming Model
2.4 Applications and Limitations of Linear Programming Problems
2.5 Formulation of Linear Programming Problem
  2.5.1 Graphic Solution
  2.5.2 General Formulation of Linear Programming Problem
  2.5.3 Matrix Form of Linear Programming Problem
2.6 Solution of Linear Programming Problem: Graphical Solution and Simplex Method
  2.6.1 Graphical Solution
  2.6.2 Some Important Definitions
  2.6.3 Canonical or Standard Forms of LPP
  2.6.4 Simplex Method
2.7 Artificial Variable Techniques
  2.7.1 The Big *M* Method
  2.7.2 The Two-Phase Simplex Method
2.8 Duality
  2.8.1 Duality in Linear Programming
2.9 Summing Up
2.10 Key Terms

2.11 Answers to 'Check Your Progress'
2.12 Questions and Exercises
2.13 References and Suggested Readings

UNIT 3: TRANSPORTATION AND ASSIGNMENT PROBLEMS **83–129**

3.0 Introduction
3.1 Objectives
3.2 Transportation Problem: Different Initial Allocation Methods
  3.2.1 Basics of Transportation Problem
  3.2.2 Applications of Transportation Problem
  3.2.3 Solutions of Transportation Problem
3.3 Optimality and MODI Method of Solving Transportation Problems
3.4 Assignment Problem
  3.4.1 Basics of Assignment Problem
  3.4.2 Job Sequence Problem
3.5 Solutions of Assignment Problem by Hungarian Method
3.6 Variations in Assignment Problems
  3.6.1 Unbalanced Assignment Problem
  3.6.2 Maximization in Assignment Problem
3.7 Summing Up
3.8 Key Terms
3.9 Answers to 'Check Your Progress'
3.10 Questions and Exercises
3.11 References and Suggested Readings

UNIT 4: THEORY OF GAMES AND INVESTMENT ANALYSIS **131–172**

4.0 Introduction
4.1 Objectives
4.2 Game Theory: An Introduction
4.3 Saddle Point
  4.3.1 The Maximin-Minimax Principle
  4.3.2 Games Without Saddle Points
4.4 Mixed Strategies
4.5 Use of Dominance
4.6 Limitations of Game Theory in Competition
4.7 Investment Analysis
  4.7.1 Methods of Investment Analysis
  4.7.2 Break-Even Analysis
  4.7.3 Methods of Evaluation and Capital Investment
  4.7.4 Payback Period Method
  4.7.5 Average (Accounting) Rate of Return Method
  4.7.6 Time Adjusted or Discounted Cash Flow (DCF) Method
  4.7.7 Discounted Rate and Certainty Equivalent Approach
  4.7.8 Expected Monetary Value and Hillier and Hertz's Models
4.8 Summing Up
4.9 Key Terms
4.10 Answers to 'Check Your Progress'
4.11 Questions and Exercises
4.12 References and Suggested Readings

UNIT 5: SYSTEM ANALYSIS AND DESIGN **173–239**

5.0 Introduction
5.1 Objectives

5.2 Concept of System
5.2.1 Characteristics of a System
5.2.2 Types of Systems
5.2.3 Information System
5.3 System Development Life Cycle and its Phases
5.3.1 Design Development Methodologies
5.4 System Analysis
5.4.1 System Planning and Initial Investigation
5.4.2 Information Gathering
5.4.3 Structured Analysis
5.4.4 Data Flow Diagram (DFD)
5.4.5 Data Dictionary
5.4.6 Decision Trees
5.4.7 Decision Tables
5.4.8 Structured English
5.5 Process and Stages of System Design
5.5.1 Input/Output and File Design
5.5.2 Logical and Physical Design
5.5.3 File Organization
5.5.4 Database Design
5.5.5 Form Design
5.5.6 Design Methodologies
5.6 System Implementation
5.6.1 System Testing
5.6.2 Quality Assurance
5.6.3 Hardware/Software Selection
5.6.4 System Security
5.7 Summing Up
5.8 Key Terms
5.9 Answers to 'Check Your Progress'
5.10 Questions and Exercises
5.11 References and Suggested Readings

UNIT 6: INTRODUCTION TO E-COMMERCE 241–287

6.0 Introduction
6.1 Objectives
6.2 Electronic Commerce (E-Commerce): Fundamentals
6.2.1 Processes Involved in E-Commerce
6.2.2 Types/Categories of E-Commerce
6.2.3 Requirements of an E-Commerce System
6.2.4 Working of E-Commerce
6.2.5 Lure of E-Commerce
6.2.6 E-Business versus E-Commerce
6.3 Encryption
6.3.1 Components of Cryptography
6.3.2 Symmetric Key Encryption
6.3.3 Public Key Encryption
6.3.4 Comparison Between Symmetric and Public Key Encryption
6.4 Authentication
6.4.1 Authentication in E-Commerce
6.5 Digital Signatures
6.5.1 Attacks on Digital Signatures
6.5.2 Types of Digital Signature
6.6 Key Certification
6.6.1 Digital Certificates

6.7 Securing Channels of Communication
6.7.1 Secure Socket Layer (SSL)
6.7.2 Secure Electronic Transaction (SET)
6.7.3 SSL versus SET
6.8 Cryptography
6.8.1 Cryptographic Algorithms
6.9 Biometrics
6.9.1 Factors that Affect Biometric Systems
6.9.2 Design Issues in Biometric Systems
6.10 Future of E-Commerce
6.11 Summing Up
6.12 Key Terms
6.13 Answers to 'Check Your Progress'
6.14 Questions and Exercises
6.15 References and Suggested Readings

INTRODUCTION

NOTES

Operational Research, or simply OR, originated in the context of military operations, but today it is widely accepted as a powerful tool for planning and decision-making, especially in business and industry. The OR approach has provided a new tool for managing conventional management problems. In fact, operational research techniques do constitute a scientific methodology of analysing the problems of the business world. They provide an improved basis for taking management decisions. The practice of OR helps in tackling intricate and complex problems, such as that of resource allocation, product mix, inventory management, sequencing and scheduling, replacement and a host of similar problems of modern business and industry.

Operations research is an interdisciplinary branch of applied mathematics and formal science that uses mathematical methods, such as mathematical modelling, statistics and algorithms to arrive at optimal or near optimal solutions to complex problems. Basically, it is concerned with optimizing the maxima (profit, assembly line performance, bandwidth, etc) or minima (loss, risk, etc.) of some objective function. It also helps management achieve its goals using scientific methods. The field of operations research is closely related to Industrial Engineering and Industrial Engineers consider operations research techniques as their major toolset. Some of the primary tools used by operations researchers are statistics, optimization, probability theory, queuing theory, game theory, graph theory and decision analysis. Because of the computational nature of these fields, OR is linked to computer science and OR professionals use specific custom-written software for computation of data and decision-making.

The uniqueness of OR prompted industries is to use its formal tools, such as operations analysis, system analysis, management science, decision science, etc. Commercial industries, such as airlines, automobiles, communications, electronics, transportation, chemicals and mining use OR techniques to optimally utilize their limited resources and thereby maximize profits. Hence, OR is the application of the methods of science to complex problems arising in the direction and management of large systems of men, machines, materials and money in industry, business, government and defence. The distinctive approach is to develop a scientific model of the system, incorporating measurement of factors, such as chance and risk, with which to predict and compare the outcomes of alternative decision strategies and controls.

Operations research provides top-level administrators a quantitative basis for taking decisions which will help organizations to carry out their functions, such as planning, controlling and organizing, effectively. Decision-making is the key responsibility of managers and OR provides a scientific approach to them for solving problems. Decisions in an organization should be such that they can compete in the market. We can say that the OR and decision-making are interlinked. There are intangible factors also, such as human behaviour, which OR has to take into account when calculating for a solution.

Computers have brought about major changes, not only in all spheres of life but also in the way we communicate and work. Today, it is extremely difficult to imagine a world without computers. Since computers are used in every possible field today, it is important for us to understand their functioning in an effective way. Computerized systems are created to solve problems. A system is a collection of components that work together to realize some objective. Building such systems is not an easy process; it requires expert skills and capabilities to understand and follow a systematic procedure for designing

NOTES

and establishing any information system. The systems approach provides an organized way of dealing with a problem. Therefore, it becomes essential to study the subject, systems analysis and design. While analysing and designing a system, its flexibility, scalability, maintainability, performance and testability need to be considered. In the System Analysis and Design (SAD) process, System Development Life Cycle (SDLC) means a combination of various activities and deals mainly with the software development life cycle. In any system, the different components are connected with each other and are interdependent. Therefore, a system life cycle is an organizational process of developing and maintaining systems. It helps in establishing a system project plan because it gives a complete list of processes and subprocesses required for developing a system. System development life cycle defines combination of various activities. All procedures, requirements must be analysed and documented in the form of detailed data flow diagrams, data dictionary, logical data structures and miniature specifications.

You must be aware that organizations spend large amounts of money to understand the use of the Internet in business applications. The Internet, one of the fastest growing technologies in the world today, is used in different types of applications. The impact of the Internet in commerce can improve the efficiency and competitive power of organizations. Nowadays, mostly organizations want to present themselves on the most common technology of the society, i.e., the Internet. It has provided opportunities even for small organizations to attract global audiences for promoting their products. Companies perform their business transactions electronically on the Internet. The most common example of such type of transactions is selling and purchasing of goods. Emergence of the Internet in business is generally known as e-commerce, which is basically the process of purchasing and selling goods and services over electronic systems, such as the Internet.

The study of e-commerce includes obtaining knowledge about how business can be conducted through the Internet using e-commerce. Business in e-commerce is done at the e-marketplace, where goods and services are exchanged among different participants, such as buyers and sellers. Activities involved in e-commerce are performed on the basis of different models of e-commerce, such as Business to Business (B2B), Business to Consumer (B2C), Business to Government (B2G) and Consumer to Consumer (C2C). The different advantages of e-commerce are the access to global market, electronic catalogue facility, better and secure payment system, less competition, lower cost of running business and workflow automation. Different techniques, such as secure electronic payment system and electronic supply management are used to manage the transfer of money and goods in a safe and easy manner. The secure electronic payment system allows the use of smart cards, online cards and credit and debit cards to collect money from the customers purchasing from the Websites of the merchant or seller.

This book, *Operations Research and Computer in Business,* has been written in a simple and lucid language covering the fundamentals of operations research, system analysis and design, and e-commerce. The concepts are analysed in a logical format, beginning with an overview that helps the reader to easily understand the concept, followed by detailed explanations and solved examples. The book follows the Self Learning Material (SLM) format wherein each unit begins with an 'Introduction' to the topic followed by an outline of the 'Objectives'. The detailed content is then presented in a simple and structured manner interspersed with 'Check Your Progress' questions. A list of 'Key Terms', a 'Summing Up' and a set of 'Questions and Exercises' is also provided at the end of each unit for effective recapitulation.

UNIT 1 OPERATIONS RESEARCH: AN INTRODUCTION

NOTES

Structure

1.0 Introduction
1.1 Objectives
1.2 Operations Research: Meaning, Nature and Origin
1.2.1 Nature of Operations Research
1.3 Development of Operational Research
1.4 Operations Research in India
1.5 Operations Research as a Tool in Decision-Making
1.6 Operations Research and Management
1.6.1 Significance of Operations Research
1.6.2 Operations Research and Modern Business Management
1.7 Features and Methodology of Operations Research and Phases of Operations Research Study
1.7.1 Methodology of Operations Research
1.8 Models in Operations Research and Methods of Deriving the Solution
1.8.1 Advantages of a Model; 1.8.2 Classification of Models
1.9 Limitations of Operations Research
1.10 Summing Up
1.11 Key Terms
1.12 Answers to 'Check Your Progress'
1.13 Questions and Exercises
1.14 References and Suggested Readings

1.0 INTRODUCTION

In this unit, you will learn about Operations Research (OR), which is a very powerful tool for decision-making. The term Operations Research was coined by J.F. McCloskey and F.N. Trefethen in 1940 in Bowdsey in the United Kingdom. This innovative science was discovered during World War II for a specific military situation, when military management sought decisions based on the optimal consumption of limited military resources with the help of an organized and systematized scientific approach. This was termed as operations research or operational research. Thus, OR was known as an ability to win a war without really going into a battlefield or fighting it. It is a new field of scientific and managerial application.

You will learn the scope of OR and its applications. The different phases of applications include—implementing and formulating the problem, constructing a mathematical model, deriving the result from the model, testing and updating the model and controlling the final output or solution. You will learn about the requirements of a good model which must be capable of working on a new formulation without making any changes in its frame with minimum assumptions and with minimum variables and must not take extraordinary time to solve the problem.

You will understand how the various models of OR work and the techniques adopted. These techniques are: linear programming, waiting line or queuing theory, inventory control/planning, game theory, decision theory, network analysis, simulation,

NOTES

integrated production models, non-linear programming, dynamic programming, heuristic programming, integer programming, algorithmic programming, quadratic programming, parametric programming, probabilistic programming, search theory and theory of replacement. All these techniques involve higher mathematics. In real practice these techniques are used in combination to form more sophisticated and advanced programming models.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the definition and scope of operations research (OR)
- Define development of OR
- Analyse the nature of OR
- Know about OR in India
- Describe OR as a tool in decision-making
- Explain the relation between OR and management
- Know about the phases and applications of OR
- Understand the requirements of a good model
- Understand the working of various models and techniques of OR
- Derive solutions using OR

1.2 OPERATIONS RESEARCH: MEANING, NATURE AND ORIGIN

The term, operations research was first coined in 1940 by J.F. McCloskey and F.N. Trefethen in a small town Bowdsey in the United Kingdom. This new science came into existence in a military context. During World War II, military management called on scientists from various disciplines and organized them into teams to assist in solving strategic and tactical problems, relating to air and land defence. Their mission was to formulate specific proposals and plans for aiding the Military commands to arrive at decisions on optimal utilization of scarce military resources and attempts to implement these decisions effectively. This new approach to the systematic and scientific study of the operations of the system was called Operations Research (OR) or Operational Research. Hence, OR can be associated with 'an art of winning the war without actually fighting it'.

Definitions

Operations Research (OR) has been defined so far in various ways and it is perhaps still too young to be defined in some authoritative way. It is not possible to give uniformly acceptable definitions of OR. A few opinions about the definition of OR are given below. These have been changed according to the development of the subject.

> OR is a scientific method of providing executive departments with a quantitative basis for decisions regarding the operations under their control.
>
> *Morse and Kimball (1946)*

NOTES

OR is the scientific method of providing executive with an analytical and objective basis for decisions. *P.M.S. Blackett (1948)*

OR is a systematic method-oriented study of the basic structures, characteristics, functions and relationships of an organization to provide the executive with a sound, scientific and quantitative basis for decision-making.

E.L. Arnoff and M.J Netzorg

OR is a scientific approach to problem solving for executive management.

H.M. Wagner

OR is an aid for the executive in making his decisions by providing him with the quantitative information based on the scientific method of analysis.

C.Kittee

OR is the scientific knowledge through interdisciplinary team effort for the purpose of determining the best utilization of limited resources.

H.A. Taha

The various definitions given here bring out the following essential characteristics of operations research:

(*i*) System orientation

(*ii*) Use of interdisciplinary terms

(*iii*) Application of scientific methods

(*iv*) Uncovering new problems

(*v*) Quantitative solutions

(*vi*) Human factors

Scope of Operations Research

There is a great scope for economists, statisticians, administrators and the technicians working as a team to solve problems of defence by using the OR approach. Besides this, OR is useful in various other important fields like:

(*i*) Agriculture

(*ii*) Finance

(*iii*) Industry

(*iv*) Marketing

(*v*) Personnel Management

(*vi*) Production Management

(*vii*) Research and Development

Phases of Operations Research

The procedure to be followed in the study of OR generally involves the following major phases:

(*i*) Formulating the problem

(*ii*) Constructing a mathematical model

(*iii*) Deriving the solution from the model

(*iv*) Testing the model and its solution (updating the model)

(*v*) Controlling the solution

(*vi*) Implementation

NOTES

1.2.1 Nature of Operations Research

Looking to the basic features of the definitions concerning OR, we can state that, 'Operational Research can be considered as the application of scientific method by interdisciplinary teams to problems involving the control of organized (man-machine) systems to provide solutions, which best serve the purposes of the organization as a whole'.

Different characteristics constituting the nature of OR can be summed up as follows:

1. *Interdisciplinary Team Approach:* The Operations Research has the characteristics that it is done by a team of scientists drawn from various disciplines such as mathematics, statistics, economics, engineering, physics, etc. It is essentially an interdisciplinary team approach. Each member of the OR team is benefited from the viewpoint of the other so that a workable solution obtained through such collaborative study has a greater chance of acceptance by management.
2. *Systems Approach:* Operations Research emphasizes on the overall approach to the system. This characteristic of OR is often referred as system orientation. The orientation is based on the observation that in the organized systems the behaviour of any part ultimately has some effect on every other part. But all these effects are not significant and even not capable of detection. Therefore, the essence of system orientation lies in the systematic search for significant interactions in evaluating actions of any part of the organization. In OR an attempt is made to take account of all the significant effects and to evaluate them as a whole. OR thus considers the total system for getting the optimum decisions.
3. *Helpful in Improving the Quality of Solution:* Operations Research cannot give perfect answers or solutions to the problems. It merely gives bad answers to the problems which otherwise have worst answers. Thus, OR simply helps in improving the quality of the solution but does not result into a perfect solution.
4. *Scientific Method:* Operations Research involves scientific and systematic attack of complex problems to arrive at the optimum solution. In other words, Operations Research or OR uses techniques of scientific research. Thus OR comprehends both aspects, i.e., it includes both scientific research on the phenomena of operating systems and the associated engineering activities aimed at applying the results of research.
5. *Goal Oriented Optimum Solution:* Operations Research tries to optimize a well-defined function subject to given constraints and as such is concerned with the optimization theory.
6. *Use of Models:* Operations Research uses models built by quantitative measurement of the variables concerning a given problem and also derives a solution from the model using one or more of the diversified solution techniques. A solution may be extracted from a model either by conducting experiments on it or by mathematical analysis. The purpose is to help the management to determine its policy and actions scientifically.
7. *Require Willing Executives:* Operations Research does require the willingness on the part of the executive for experimentation to evaluate the costs and the consequences of the alternative solutions of the problem. It enables the decision-

maker to be objective in choosing an alternative from among many possible alternatives.

8. *Reduces Complexity:* Operations Research tries to reduce the complexity of business operations and does help the executive in correcting a troublesome function and to consider innovations which are too costly and complicated to experiment with the actual practice.

In view of the above, OR must be viewed as both a science and an art. As science, OR provides mathematical techniques and algorithms for solving appropriate decision problems. OR is an art because success in all the phases that precede and succeed the solution of a problem largely depends on the creativity and personal ability of the decision-making analysts.

NOTES

1.3 DEVELOPMENT OF OPERATIONAL RESEARCH

The subject of Operational Research (OR) was developed in military context during World War II, pioneered by the British scientists. At that time, the military management in England appointed a study group of scientists to deal with the strategic and tactical problems related to air and land defence of the country. The main reason for conducting the study was that they were having very limited military resources. It was, therefore, necessary to decide upon the most effective way of utilizing these resources. As the name implies, Operations Research was apparently invented because the team was dealing with research on military operations. The scientists studied the various problems and on the basis of quantitative study of operations suggested certain approaches which showed remarkable success. The encouraging results obtained by the British operations research teams consisting of personnel drawn from various fields like Mathematics, Physics, Biology, Psychology and other physical sciences, quickly motivated the United States military management to start similar activities. Successful innovations of the US teams included the development of new flight patterns, planning sea mining and effective utilization of electronic equipments. Similar OR teams also started functioning in Canada and France. These OR teams were usually assigned to the executive-in-charge of operations and as such their work came to be known as 'Operational Research' in the UK and by a variety of names in the United States an Operational Analysis, Operations Evaluation, Operations Research, Systems Analysis, Systems Evaluation and Systems Research. The name 'Operational Research' or 'Operations Research' or simply OR is most widely used now a days all over the world for the systematic and scientific study of the operations of the system. Till fifties, use of OR was mainly confined to military purposes.

After the end of the second world war, the success of military teams attracted the attention of industrial managers who were seeking solutions to their complex managerial problems. At the end of the war, expenditures on defence research were reduced in the UK and this led to the release of many operations research workers from the military at a time when industrial managers were confronted with the need to reconstruct most of Britain's manufacturing industries and plants that had been damaged in war. Executives in such industries sought assistance from the said operations research workers. But in the USA most of the war experienced operations research workers remained in military service as the defence research was increased

Check Your Progress

1. How was the concept of operations research started?
2. State one definition of OR.
3. Mention the essential characteristics of operations research.
4. Apart from defence, mention other areas where OR is applied.
5. What are major phases in the application of OR?

NOTES

and consequently operations research was expanded at the end of the war. It was only in the early 1950s, industry in the USA began to absorb the operations is research worker under the pressure for increased demands for greater productivity originated because of the outbreak of the Korean conflict and because of technological developments in industry. Thus, OR began to develop in industrial field in the United States since the year 1950. The Operations Research Society of America was formed in 1953 and in 1957 the International Federation of Operational Research Societies was established. Various journals relating to OR began to appear in different countries in the years that followed the mid-fifties. Courses and curricula in OR in different universities and other academic institutions began to proliferate in the United States. Other countries rapidly followed this and after the late fifties Operations Research was applied for solving business and industrial problems. Introduction of Electronic Data Processing (EDP) methods further enlarged the scope for application of OR techniques. With the help of a digited computer many complex problems can be studied on day-to-day basis. As a result, many industrial concerns are adopting OR as an integrated decision-making tool for their routine decision procedures.

1.4 OPERATIONS RESEARCH IN INDIA

Today, the impact of OR in Indian business and industry can be felt in many areas. A large number of management consulting firms are recently engaged in OR activities. Apart from military and business applications, OR activities include transportation systems, libraries, hospitals, city planning, financial institutions, etc. With increasing use of computers the operations research techniques have started playing a noticeable role in our country as well. The major Indian industries such as Delhi Cloth Mills, Indian Railways, Indian Airlines, Defence Organizations, Hindustan Liver, Tata Iron and Steel Company, Fertilizer Corporation of India and similar industries make use of operations research techniques for solving problems and making decisions.

Historically, Operations Research started developing in India after independence specially with the setting up of an Operation Research Unit at the Regional Research Laboratory at Hyderabad in 1949. Operations Research activities gained further impetus with the establishment of an Operations Research Unit in 1953 in the Indian Statistical Institute (ISI), Calcutta for applying Operations Research techniques in national planning and survey. Operational Research Society of India was formed in 1957 which joined International Federation of Operational Research Societies in 1960 by becoming its member. The said society helped the cause of the development of Operations Research activities in India in several ways and started publishing a journal of Operations Research entitled 'OPSEARCH' from 1963. Besides, the Indian Institute of Industrial Engineers has also promoted Operations Research in India and its journals viz., 'Industrial Engineering' and 'Management' are considered as important key journals relating to Operations Research in the country. Other important journals which deal with Operations Research in our country are the Journal of the National Productivity Council, Materials Management Journal of India and the Defence Science Journal. There are several institutions which train and produce people in the field of Operations Research to meet the need of OR practitioners in the country.

So far as the application of Operations Research in India is concerned it was Professor P.C. Mahalonobis of ISI, Calcutta who made the first important application. He formulated the Second Five Year Plan of our country with the help of OR technique

NOTES

to forecast the trends of demand, availability of resources and for scheduling the complex scheme necessary for developing our country's economy. It was estimated that India could become self-sufficient in food and solve her foreign exchange problems merely by reducing the wastage of food by 15%. Operational Research Commission made the use of OR techniques for planning the optimum size of the Caravelle fleet of Indian Airlines. Kirloskar company made use of assignments models for allocation of their salesmen to different areas so as to maximize their profit. Linear Programming (LP) models were also used by them to assemble various diesel engines at minimum possible cost. Various cotton textile leaders such as Binny, DCM, Calico, etc., are using linear programming techniques in cotton blending. Many other firms like Union Carbide, ICI, TELCO and Hindustan Liver, etc., are making use of OR techniques for solving many of their complex business problems. State Trading Corporation of India (STCI) has also set up a Management Sciences Group with the idea of promoting and developing the use of OR techniques in solving its management decision problems. Besides, many Universities and professional academic institutions are imparting training in OR in our country. The subject of OR has been included in the courses of such institutions. But in comparison with the western world the present state of OR in our country is much behind. Operations Research activities are very much limited and confined only to the big organized industries. Most popular practical application of Operations Research has been mainly that of Linear Programming. There is relative scarcity of well-trained operational researchers. The use of Operations Research is relatively a very costly affair. In spite of several limitations, our industrialists are gradually becoming conscious of the role of Operations Research techniques and in the coming years such techniques will have an increasingly important role to play in Indian business and industry.

1.5 OPERATIONS RESEARCH AS A TOOL IN DECISION-MAKING

Mathematical models have been constructed for OR problems and methods for solving the models are available in many cases. Such methods are usually termed as OR techniques. Some of the important OR techniques often used by decision-makers in modern times in business and industry are as under:

1. *Linear Programming:* This technique is used in finding a solution for optimizing a given objective, such as profit maximization or cost minimization under certain constraints. This technique is primarily concerned with the optimal allocation of limited resources for optimizing a given function. The name linear programming is given because of the fact that the model in such cases consists of linear equations indicating linear relationship between the different variables of the system. Linear programming technique solves product-mix and distribution problems of business and industry. It is a technique used to allocate scarce resources in an optimum manner in problems of scheduling, product-mix, and so on. Key factors under this technique include an objective function, choice among several alternatives, limits or constraints stated in symbols and variables assumed to be linear.
2. *Waiting Line or Queuing Theory:* Waiting line or queuing theory deals with mathematical study of queues. The queues are formed whenever the current demand for service exceeds the current capacity to provide that service. Waiting

NOTES

line technique concerns itself with the random arrival of customers at a service station where the facility is limited. Providing too much of capacity will mean idle time for servers and will lead to waste of 'money'. On the other hand, if the queue becomes long there will be a cost due to waiting of units in the queue. Waiting line theory, therefore, aims at minimizing the costs of both servicing and waiting. In other words, this technique is used to analyse the feasibility of adding facilities and to assess the amount and cost of waiting time. With its help we can find the optimal capacity to be installed which will lead to a sort of an economic balance between cost of service and cost of waiting.

3. *Inventory Control/Planning:* Inventory planning aims at optimizing inventory levels. Inventory may be defined as a useful idle resource which has economic value, e.g., raw materials, spare parts, finished products, etc. Inventory planning, in fact, answers the two questions, viz., how much to buy and when to buy. Under this technique the main emphasis is on minimizing costs associated with holding of inventories, procurement of inventories and the shortage of inventories.
4. *Game Theory:* Game theory is used to determine the optimum strategy in a competitive situation. Simplest possible competitive situation is that of two persons playing zero-sum game, i.e., a situation in which two persons are involved and one person wins exactly what the other loses. More complex competitive situations in real life can be imagined where game theory can be used to determine the optimum strategy.
5. *Decision Theory:* Decision theory concerns with making sound decisions under conditions of certainty, risk and uncertainty. As a matter of fact there are three different kinds of states under which decisions are made, viz., deterministic, stochastic and uncertainty and the decision theory explains how to select a suitable strategy to achieve some object or goal under each of these three states.
6. *Network Analysis:* Network analysis involves the determination of an optimum sequence of performing certain operations concerning some jobs in order to minimize overall time and/or cost. Programme Evaluation and Review Technique (PERT), Critical Path Method (CPM) and other network techniques such as Gantt Chart come under network analysis. Key concepts under this technique are network of events and activities, resource allocation, time and cost considerations, network paths and critical paths.
7. *Simulation:* Simulation is a technique of testing a model, which resembles a real life situation. This technique is used to imitate an operation prior to actual performance. Two methods of simulation are used—Monte Carlo method and the System simulation method. The former uses random numbers to solve problems which involve conditions of uncertainty and the mathematical formulation is impossible. In case of System simulation, there is a reproduction of the operating environment and the system allows for analysing the response from the environment to alternative management actions. This method draws samples from a real population instead of from a table of random numbers.
8. *Integrated Production Models:* This technique aims at minimizing cost with respect to work force, production and inventory. This technique is highly complex and is used only by big business and industrial units. This technique can be used only when sales and costs statistics for a considerable long period are available.

NOTES

9. *Some Other OR Techniques:* In addition, there are several other techniques such as non-linear programming, dynamic programming, search theory, the theory of replacement, etc. A brief mention of some of these is as follows:

   (i) *Non-linear programming:* A form of programming in which some or all of the variables are curvilinear. In other words, this means that either the objective function or constraints or both are not in linear form. In most of the practical situations, we encounter non-linear programming problems but for computation purpose we approximate them as linear programming problems. Even then there may remain some non-linear programming problems which may not be fully solved by presently known methods.

   (ii) *Dynamic programming:* It refers to a systematic search for optimal solutions to problems that involve many highly complex interrelations that are, moreover, sensitive to multistage effects such as successive time phases.

   (iii) *Heuristic programming:* It is also known as discovery method and refers to step-by-step search towards an optimum when a problem cannot be expressed in mathematical programming form. The search procedure examines successively a series of combinations that lead to stepwise improvements in the solution and the search stops when a near optimum has been found.

   (iv) *Integer programming:* It is a special form of linear programming in which the solution is required in terms of integral numbers (i.e., whole numbers) only.

   (v) *Algorithmic programming:* It is just the opposite of Heuristic programming. It may also be termed as similar to mathematical programming. This programming refers to a thorough and exhaustive mathematical approach to investigate all aspects of the given variables in order to obtain optimal solution.

   (vi) *Quadratic programming:* It refers to a modification of linear programming in which the objective equations appear in quadratic form, i.e., they contain squared terms.

   (vii) *Parametric programming:* It is the name given to linear programming when it is modified for the purpose of inclusion of several objective equations with varying degrees of priority. The sensitivity of the solution to these variations is then studied.

   (viii) *Probabilistic programming:* It is also known as stochastic programming and refers to linear programming that includes an evaluation of relative risks and uncertainties in various alternatives of choice for management decisions.

   (ix) *Search theory:* It concerns itself with search problems. A search problem is characterized by the need for designing a procedure to collect information on the basis of which one or more decisions are made. This theory is useful in places in which some events are known to occur but the exact location is not known. The first search model was developed during World War II to solve decision problems connected with air patrols and their search for submarines. Advertising agencies search for customers, personnel departments search for good executives are some of the examples of search theory's application in business.

NOTES

(x) *The Theory of Replacement:* It is concerned with the prediction of replacement costs and determination of the most economic replacement policy. There are two types of replacement models. First type of models deal in replacing equipments that deteriorate with time and the other type of models help in establishing replacement policy for those equipments which fail completely and instantaneously.

All these techniques are not simple but involve higher mathematics. The tendency today is to combine several of these techniques and form into more sophisticated and advanced programming models.

1.6 OPERATIONS RESEARCH AND MANAGEMENT

Operations Research plays significant role in industries and empowers management to make better decision. The following are the important applications of OR in business and management:

1.6.1 Significance of Operations Research

Operations Research has gained increasing importance since World War II in the technology of business and industry administration. It greatly helps in tackling the intricate and complex problems of modern business and industry. OR techniques are, in fact, examples of the use of scientific method of management. The significance of OR can be well understood under the following heads:

1. *OR provides a tool for scientific analysis.* OR provides the executives with a more precise description of the cause and effect relationship and risks underlying the business operations in measurable terms and this eliminates the conventional intuitive and subjective basis on which managements used to formulate their decisions decades ago. In fact, OR replaces the intuitive and subjective approach of decision-making by an analytical and objective approach. The use of OR has transformed the conventional techniques of operational and investment problems in business and industry. As such OR encourages and enforces disciplined thinking about organizational problems.
2. *OR provides solutions for various business problems.* OR techniques are being used in production, procurement, marketing, finance and other allied fields. It can be used to solve problems like how best can managers and executives allocate the available resources to various products so that in a given time the profits are maximum or the cost is minimum, possibility of an industrial enterprise to arrange the time and quantity of orders of its stocks such that the overall profit with given resources is maximum, competence of business managers to determine the number of men and machines to be employed and used in such a manner that neither remains idle and at the same time the customer or the public has not to wait unduly long for service, and similar other problems. Similarly we might have a complex of industries; steel, machine tools and others, all employed in the production of one item, say, steel. At any particular time we have a number of choices of allocating resources such as money, steel and tools for producing autos, building steel factories or tool factories. What

Check Your Progress

6. When was the subject of operational research developed?
7. Why is linear programming technique used?
8. Define waiting line theory.
9. Why is game theory used?
10. What does dynamic programming refers to?

NOTES

should be the policy which optimizes the total number of autos produced over a given period? OR techniques are capable of providing an answer in such a situation.

Planning decisions in business and industry are largely governed by the picture of anticipated demands. The potential long range profits of the business may vary in accordance with different possible demand patterns. The OR techniques serve to develop a scientific basis for coping with the uncertainties of future demands. Thus, in dealing with the problem of uncertainty over future sales and demands, OR can be used to generate 'a least risk' plan.

At times there may be a problem of finding an acceptable definition of long range company objectives. Management may be confronted with different viewpoints; some may stress the desirability of maximizing of net profit whereas others may focus attention primarily on the minimization of costs. OR techniques, specially that of mathematical programming such as linear programming can help resolve such dilemmas by permitting systematic evaluation of the best strategies for attaining different objectives. These techniques can also be used for estimating the worth of technical innovations as also of potential profits associated with the possible changes in rules and policies.

How much changes can there be in the data on which a planning formulation is based without undermining the soundness of the plan itself? How accurately must managements know cost coefficients, production performance figures and other factors before it can make planning decisions with confidence? Many of the basic data required for the development of long-range plans are uncertain. Such uncertainties though cannot be avoided but through various OR techniques the management can know how critical such uncertainties are and this in itself is a great help to business planners.

3. *OR enables proper deployment of resources.* OR renders valuable help in proper deployment of resources. For example, Programme Evaluation and Review Technique (PERT) enables us to determine the earliest and the latest times for each of the events and activities and thereby helps in identification of the critical path. All this helps in the deployment of resources from one activity to another to enable the project completion on time. This technique, thus provides for determining the probability of completing an event or project itself by a specified date.

4. *OR helps in minimizing waiting and servicing costs.* The waiting line or queuing theory helps management in minimizing the total waiting and servicing costs. This technique also analyses the feasibility of adding facilities and thereby helps businesspeople take correct and profitable decisions.

5. *OR enables management to decide when to buy and how much to buy.* The main objective of inventory planning is to achieve a balance between the cost of holding stocks and the benefits from stock-holding.

   Hence, the technique of inventory planning enables the management to decide when to buy and how much to buy.

6. *OR assists in choosing an optimum strategy.* Game theory is specially used to determine the optimum strategy in a competitive situation and enables

NOTES

businesspeople to maximize profits or minimize losses by adopting the optimum strategy.

7. *OR renders great help in optimum resource allocation.* Linear programming technique is used to allocate scarce resources in an optimum manner in problems of scheduling, product-mix and so on. This technique is popularly used by modern management in resource allocation and in ensuring optimal assignments.
8. *OR facilitates the process of decision-making.* Decision theory enables businessmen to select the best course of action when information is given in a probabilistic form. Through decision tree (a network showing the logical relationship between the different parts of a complex decision and the alternative courses of action in any phase of a decision situation) technique executive's judgement can systematically be brought into the analysis of the problems. Simulation is another important technique used to imitate an operation or process prior to actual performance. The significance or simulation lies in the fact that it enables in finding out the effect of alternative courses of action in a situation involving uncertainty where mathematical formulation is not possible. Even complex groups of variables can be handled through this technique.
9. *Through OR management can know the reactions of the integrated business systems.* The Integrated Production Models technique is used to minimize cost with respect to work force, production and inventory. This technique is quite complex and is usually used by companies having detailed information concerning their sales and costs statistics over a long period. Besides, various other OR techniques also help management people in taking decisions concerning various problems of business and industry. The techniques are designed to investigate how the integrated business system would react to variations in its component elements and/or external factors.
10. *OR techniques help a lot in the preparation of future (or would be) managers.* In fact, OR techniques substitute a means for improving the knowledge and skill of youngsters in the field of management.

1.6.2 Operations Research and Modern Business Management

From what has been stated above, we can say that operational research renders valuable service in the field of business management. It ensures improvement in the quality of managerial decisions in all functional areas of management. The role of OR in business management can be summed up as under:

OR techniques help the directing authority in optimum allocation of various limited resources, viz., men, machines, money, material, time, etc., to different competing opportunities on an objective basis for achieving effectively the goal of a business unit. They help the chief of executive in broadening management vision and perspectives in the choice of alternative strategies to the decision problems such as forecasting manpower, production capacities, capital requirements and plans for their acquisition.

OR is useful to the production management in (*i*) Selecting the building site for a plant, scheduling and controlling its development and designing its layout; (*ii*) Locating within the plant and controlling the movements of required production materials and finished goods inventories; (*iii*) Scheduling and sequencing production

by adequate preventive maintenance with optimum number of operatives by proper allocation of machines; and (*iv*) Calculating the optimum product-mix.

OR is useful to the personnel management to find out (*i*) Optimum manpower planning; (*ii*) The number of persons to be maintained on the permanent or full time roll; (*iii*) The number of persons to be kept in a work pool intended for meeting absenteeism; (*iv*) The optimum manner of sequencing and routing of personnel to a variety of jobs; and (*v*) In studying personnel recruiting procedures, accident rates and labour turnover.

OR techniques equally help the marketing management to determine (*i*) Where distribution points and warehousing should be located; their size, quantity to be stocked and the choice of customers; (*ii*) The optimum allocation of sales budget to direct selling and promotional expenses; (*iii*) The choice of different media of advertising and bidding strategies; and (*iv*) The consumer preferences relating to size, colour, packaging, etc., for various products as well as to outbid and outwit competitors.

OR is also very useful to the financial management in (*i*) Finding long range capital requirements as well as how to generate these requirements; (*ii*) Determining optimum replacement policies; (*iii*) Working out a profit plan for the firm; *(iv)* Developing capital investments plans; and (*v*) Estimating credit and investment risks.

In addition to all this, OR provides the business executives such an understanding of the business operations which gives them new insights and capability to determine better solutions for several decision-making problems with great speed, competence and confidence. When applied on the level of management where policies are formulated, OR assists the executives in an advisory capacity but on the operational level where production, personnel, purchasing, inventory and administrative decisions are made. It provides management with a means for handling and processing information. Thus, in brief, OR can be considered as scientific method of providing executive departments with a quantitative basis for taking decisions regarding operations under their control.

1.7 FEATURES AND METHODOLOGY OF OPERATIONS RESEARCH AND PHASES OF OPERATIONS RESEARCH STUDY

OR study generally involves three phases viz., the judgement phase, the research phase and the action phase. Of these three, the research phase is the longest and the largest, but the remaining two phases are very important since they provide the basis for and implementation of the research phase respectively.

The judgement phase includes *(i)* The determination of the problem; *(ii)* The establishment of the objectives and values related to the operation; and *(iii)* The determination of suitable measures of effectiveness.

The research phase utilizes *(i)* Observations and data collection for better understanding of the problem; *(ii)* Formulation of hypothesis and models; *(iii)* Observation and experimentation to test the hypothesis on the basis of additional

NOTES

data; and *(iv)* Predictions of various results from the hypothesis, generalization of the result and consideration of alternative methods.

The action phase in the OR consists of making recommendations for decision process. As such this phase deals with the implementation of the tested results of the model. This phase is executed primarily through the cooperation of the OR experts on the one hand and those who are responsible for operating the system on the other.

1.7.1 Methodology of Operations Research

In view of the above referred phases the methodology of OR generally invoives the following steps:

1. *Formulating the Problem:* The first step in an OR study is to formulate the problem in an appropriate form. Formulating a problem consists in identifying, defining and specifying the measures of the components of a decision model. This means that all quantifiable factors which are pertinent to the functioning of the system under consideration are defined in mathematical language as variables (factors which are not controllable), parameters or coefficient along with the constraints on the variables and the determination of suitable measures of effectiveness.
2. *Constructing the Model:* The second step consists in constructing the model by which we mean that appropriate mathematical expressions are formulated which describe interrelations of all variables and parameters. In addition, one or more equations or inequalities are required to express the fact that some or all of the controlled variables can only be manipulated within limits. Such equations or inequalities are termed as constraints or the restrictions. The model must also include an objective function which defines the measure of effectiveness of the system. The objective function and the constraints, together constitute a model of the problem that we want to solve. This model describes the technology and the economics of the system under consideration through a set of simultaneous equations and inequalities.
3. *Deriving the Solution:* Once the model is constructed the next step in an OR study is that of obtaining the solution to the model, i.e., finding the optimal values of the controlled variables—values that produce the best performance of the system for specified values of the uncontrolled variables. In other words, an optimum solution is determined on the basis of the various equations of the model satisfying the given constraints and inter-relations of the system and at the same time maximizing profit or minimizing cost or coming as close as possible to some other goal or criterion. How the solution can be derived depends on the nature of the model. In general, there are three methods available for the purpose viz., the analytical methods, the numerical methods and the simulation methods. Analytical methods involve expressions of the model by mathematical computations and the kind of mathematics required depends upon the nature of the model under consideration. This sort of mathematical analysis can be conducted only in some cases without any knowledge of the values of the variables but in others the values of the variables must be known concretely or numerically. In latter cases, we use the numerical methods which are concerned

NOTES

with iterative procedures through the use of numerical computations at each step. The algorithm (or the set of computational rules) is started with a trial or initial solution and continued with a set of rules for improving it towards optimality. The initial solution is then replaced by the improved one and the process is repeated until no further improvement is possible. But in those cases where the analytical as well as the numerical methods cannot be used for deriving the solution then we use simulation methods, i.e., we conduct experiments on the model in which we select values of the uncontrolled variables with the relative frequencies dictated by their probability distributions. The simulation methods involve the use of probability and sampling concepts and are generally used with the help of computers. Whichever method is used, our objective is to find an optimal or near-optimal solution, i.e., a solution which optimizes the measure of effectiveness in a model.

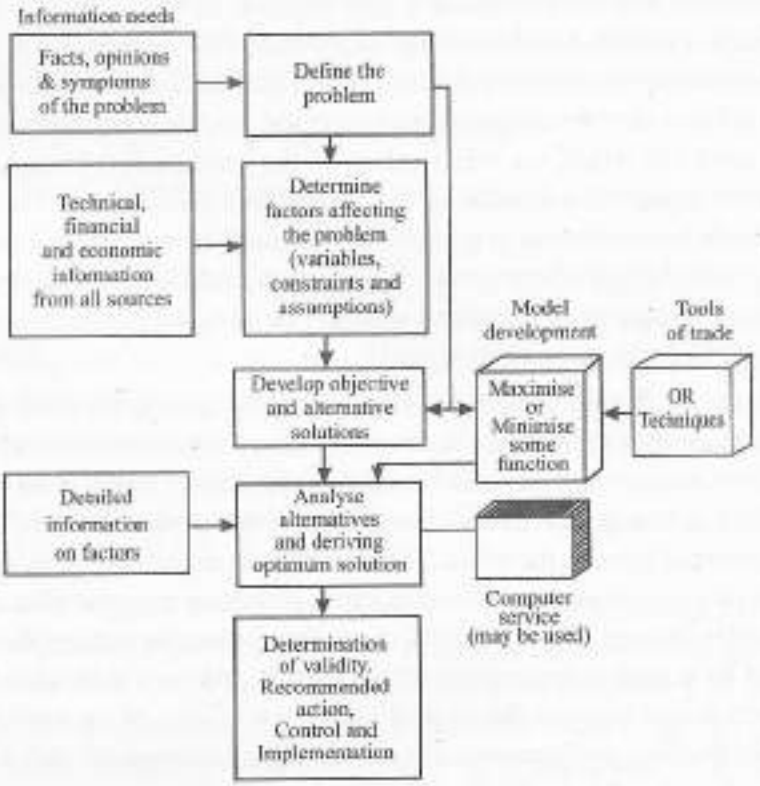
4. *Testing the Validity:* The solution values of the model, obtained as stated in step three above, are then tested against actual observations. In other words, effort is made to test the validity of the model used. A model is supposed to be valid if it can give a reliable prediction of the performance of the system represented through the model. If necessary, the model may be modified in the light of actual observations and the whole process is repeated till a satisfactory model is attained. The operational researcher quite often realizes that his model must be a good representation of the system and must correspond to reality which in turn requires this step of testing the validity of the model in an OR study. In effect, performance of the model must be compared with the policy or procedure that it is meant to replace.
5. *Controlling the Solution:* This step of an OR study establishes control over the solution by proper feedback of information on variables which might have deviated significantly. As such the significant changes in the system and its environment must be detected and the solution must accordingly be adjusted.
6. *Implementing the Results:* Implementing the results constitutes the last step of an OR study. The objective of OR is not merely to produce reports but to improve the performance of systems. The results of the research must be implemented if they are accepted by the decision-makers. It is through this step that the ultimate test and evaluation of the research is made and it is in this phase of the study the researcher has the greatest opportunity for learning.

Thus, the procedure for an OR study generally involves some major steps viz., formulating the problem, constructing the mathematical model to represent the system under study, deriving a solution from the model,.testing the model and the solution so derived, establishing controls over the solution and lastly putting the solution to work-implementation. Although the said phases and the steps are usually initiated in the order listed in an OR study but it should always be kept in mind that they are likely to overlap in time and to interact each phase usually continues until the study is completed.

NOTES

Flow Chart Showing OR Approach

OR approach can be well illustrated through the following flow chart:



1.8 MODELS IN OPERATIONS RESEARCH AND METHODS OF DERIVING THE SOLUTION

A model in OR is a simplified representation of an operation or as a process in which only the basic aspects or the most important features of a typical problem under investigation are considered. The objective of a model is to identify significant factors and interrelationships. The reliability of the solution obtained from a model depends on the validity of the model representing the real system.

A good model must possess the following characteristics:

(*i*) It should be capable of taking into account new formulation without having any changes in its frame.

(*ii*) Assumptions made in the model should be as small as possible.

(*iii*) Variables used in the model must be less in number ensuring that it is simple and coherent.

(*iv*) It should be open to parametric type of treatment.

(*v*) It should not take much time in its construction for any problem.

Check Your Progress

11. What is the significance of operations research?
12. How does operations research provides solution for various business problems?
13. How operations research is useful to the production management?
14. What does judgement phase includes?
15. What is the first step in OR study?

1.8.1 Advantages of a Model

There are certain significant advantages gained when using a model. These are:

(*i*) Problems under consideration become controllable through a model.

(*ii*) It provides a logical and systematic approach to the problem.

(*iii*) It provides the limitations and scope of an activity.

(iv) It helps in finding useful tools that eliminate duplication of methods applied to solve problems.

(v) It helps in finding solutions for research and improvements in a system.

(vi) It provides an economic description and explanation of either the operation, or the systems they represent.

NOTES

1.8.2 Classification of Models

The classification of models is a subjective problem. They may be distinguished as follows:

(i) Models by Degree of Abstraction
(ii) Models by Function
(iii) Models by Structure
(iv) Models by Nature of an Environment
(v) Models by the Extent of Generality

Models by Function

These models consist of (i) Descriptive models, (ii) Predictive models, and (iii) Normative or optimization models.

Descriptive and Predictive Models: These models describe and predict facts and relationships among the various activities of the problem. These models do not have an objective function as a part of the model to evaluate decision alternatives. In this model, it is possible to get information as to how one or more factors change as a result of changes in other factors.

Normative or Optimization Models: They are prescriptive in nature and develop objective decision-rule for optimum solutions.

Models by Structure

These models are represented by (i) Iconic or Physical models, (ii) Analog models, and (iii) Mathematic or Symbolic models.

Iconic or Physical Models: These are pictorial representations of real systems and have the appearance of the real thing. An iconic model is said to be scaled down or scaled up according to the dimensions of the model which may be smaller or greater than that of the real item, e.g., city maps, blue prints of houses, globe and so on. These models are easy to observe and describe but are difficult to manipulate and are not very useful for the purposes of prediction.

Analog Models: These are abstract than the iconic ones. There is no look alike correspondence between these models and real life items. The models in which one set of properties is used to represent another set of properties are called analog models. After the problem is solved, the solution is reinterpreted in terms of the original system. These models are less specific, less concrete but easier to manipulate than iconic models.

Mathematic or Symbolic Models: These are most abstract in nature in comparison to others. They employ a set of mathematical symbols to represent the components of the real system. These variables are related together by means of mathematical equations to describe the behaviour of the system. The solution of the problem is then obtained by applying well developed mathematical techniques to the model. The symbolic model is usually the easiest to manipulate experimentally and it is the most general and abstract. Its function is more explanatory than descriptive.

NOTES

Models by Nature of an Environment

These models can be classified into (*i*) Deterministic models, and (*ii*) Probabilistic or Stochastic models.

Deterministic Models: In these models, all parameters and functional relationships are assumed to be known with certainty when the decision is to be made. Linear programming and break even models are the examples of deterministic models.

Probabilistic or Stochastic Models: These models are those in which atleast one parameter or decision variable is a random variable. These models reflect to some extent the complexity of the real world and the uncertainty surrounding it.

Models by the Extent of Generality

These models can be categorized into (*i*) Specific models, and (*ii*) General models.

Specific Models: When a model presents a system at some specific time, it is known as a *specific model*. In these models, if the time factor is not considered them they are termed as *static models*. An inventory problem of determining economic order quantity for the next period assuming that the demand in the planning period would remain same as that of today is an example of static model. *Dynamic programming* may be considered as an example of *dynamic model*.

General Models: *Simulation* and *heuristic models* fall under the category of general models. These models are used to explore alternative strategies which have been overlooked previously.

1.9 LIMITATIONS OF OPERATIONS RESEARCH

OR though is a great aid to management as outlined above but still it cannot be a substitute for decision-making. The choice of a criterion as to what is actually best for a business enterprise is still that of an executive who has to fall back upon his experience and judgement. This is so because of the several limitations of OR. Important limitations are given below:

1. *The inherent limitations concerning mathematical expressions.* OR involves the use of mathematical models, equations and similar other mathematical expressions. Assumptions are always incorporated in the derivation of an equation or model and such an equation or model may be correctly used for the solution of the business problems when the underlying assumptions and variables in the model are present in the concerned problem. If this caution is not given due care then there always remains the possibility of wrong application of OR techniques. Quite often the operations researchers have been accused of having many solutions without being able to find problems that fit.
2. *High costs are involved in the use of OR techniques.* OR techniques usually prove very expensive. Services of specialized persons are invariably called for and along with it the use of computer and its maintenance is also considered while using OR techniques. Hence, only big concerns can think of using such techniques. Even in big business organizations we can expect that OR techniques will continue to be of limited use simply because they are not in many cases worth their cost. As opposed to this a typical manager, exercising intuition and judgement, may be able to make a decision very inexpensively. Thus, the use of OR is a costlier affair and this constitutes an important limitation of OR.

NOTES

3. *OR does not take into consideration the intangible factors, i.e., non-measurable human factors.* OR makes no allowance for intangible factors such as skill, attitude, vigour of the management in taking decisions but in many instances success or failure hinges upon the consideration of such non-measurable intangible factors. There cannot be any magic formula for getting an answer to management problems but it depends upon proper managerial attitudes and policies.
4. *OR is only a tool of analysis and not the complete decision-making process.* It should always be kept in mind that OR alone cannot make the final decision. It is just a tool and simply suggests best alternatives. In the final analysis many business decisions will involve human element. Thus, OR is at best a supplement to rather than a substitute for management; subjective judgement is likely to remain a principal approach to decision-making.
5. *Other limitations.* Among other limitations of OR, the following deserve mention:
   (*i*) *Bias.* The operational researchers must be unbiased. An attempt to shoehorn results into a confirmation of management's prior preferences can greatly increase the likelihood of failure.
   (*ii*) *Inadequate objective functions.* The use of a single objective function is often an insufficient basis for decisions. Laws, regulations, public relations, market strategies, etc., may all serve to overrule a choice arrived at in this way.
   (*iii*) *Internal resistance.* The implementation of an optimal decision may also confront internal obstacles such as trade unions or individual managers with strong preferences for other ways of doing the job.
   (*iv*) *Competence.* Competent OR analysis calls for the careful specification of alternatives, a full comprehension of the underlying mathematical relationships and a huge mass of data. Formulation of an industrial problem to an OR set programme is quite often a difficult task.
   (*v*) *Reliability of the prepared solution.* At times a non-linear relationship is changed to linear for fitting the problem to the LP pattern. This may disturb the solution.

1.10 SUMMING UP

- The term Operations Research (OR) was first coined in 1940 by J.F. McCloskey and F.N. Trefethen.
- OR is a scientific method of providing executive departments with a quantitative basis for decisions regarding the operations under their control.
- OR can be considered as being the application of scientific method by interdisciplinary teams to problems involving the control of organized (man-machine) systems to provide solutions which best serve the purpose of the organization as a whole.
- OR emphasizes on the overall approach to the system. This characteristic of OR is often referred as system oriented.
- OR involves scientific and systematic attack of complex problems to arrive at the optimum solution.

Check Your Progress

16. What is an OR model? What is its objective?
17. How are the models classified?
18. Define descriptive models.
19. Define deterministic models.
20. OR is only a tool of analysis and not the complete decision-making process. How?

NOTES

- Linear programming technique is used in finding a solution for optimizing a given objective, such as profit maximization or cost minimization under certain constraints.
- In OR, waiting line or queuing theory deals with mathematical study of queues which are formed whenever the current demand for service exceeds the current capacity to provide that service.
- In OR, decision theory concerns with making sound decisions under conditions of certainty, risk and uncertainty.
- OR techniques are being used in production, procurement, marketing, finance and other allied fields. Through OR, management can know the reactions of the integrated business systems. The Integrated Production Models technique is used to minimize cost with respect to work force, production and inventory.
- OR provides the business executives such an understanding of the business operations which gives them new insights and capability to determine better solutions for several decision-making problems with great speed, competence and confidence.
- OR study generally involves three phases, viz., the judjement phase, the research phase and the action phase.
- The procedure for an OR study generally involves some major steps viz., formulating the problem, constructing the mathematical model to represent the system under study, deriving a solution from the model, testing the model and the solution so derived, establishing controls over the solution and putting the solution to work implementation.
- OR is generally concerned with problems that are tactical rather than strategic in nature.
- A model in OR is a simplified representation of an operation or a process in which only the basic aspects or the most important features of a typical problem under investigation are considered.
- Models by function consist of descriptive models, predictive models and normative models.
- Models by structure are represented by iconic or physical models, analog models and mathematic or symbolic models.
- Models by nature of an environment can be classified into deterministic models and probabilistic models.
- Models by the extent of generality can be categorized into specific models and general models.

1.11 KEY TERMS

- **Operations research:** The application of scientific knowledge through interdisciplinary team effort for the purpose of determining the best utilization of limited resources
- **Model:** A simplified representation of an operation or a process that considers only the basic aspects or the most important features of a typical problem under investigation with an objective to identify significant factors and their interrelationships

NOTES

- **Descriptive model:** A type of model that describes facts and relationships among the various activities of the problem and collects information on factors that change as a result of changes in other factors
- **Normative or optimization models:** These models are prescriptive in nature and develop objective decision-rule for optimum solutions
- **Iconic or physical models:** Pictorial representations of real systems that have the appearance of the real thing.
- **Analog models:** These are abstract models and one set of properties is used to represent another set of properties
- **Mathematic or symbolic models:** These models are most abstract in nature and employ a set of mathematical symbols to represent the components of the real system

1.12 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The concept of Operations Research (OR) came into existence in a military context during World War II, when military management wanted to arrive at decisions on optimal utilization of scarce military resources with a new approach to the systematic and scientific study of the operations of the system.
2. OR is the scientific knowledge through interdisciplinary team effort for the purpose of determining the best utilization of limited resources.
3. Essential characteristics of operations research are:
   (*i*) System orientation
   (*ii*) Use of interdisciplinary terms
   (*iii*) Application of scientific methods
   (*iv*) Uncovering new problems
   (*v*) Quantitative solutions
   (*vi*) Human factors
4. Following are the areas where concept of OR is applied:
   (*i*) Agriculture
   (*ii*) Finance
   (*iii*) Industry
   (*iv*) Marketing
   (*v*) Personnel Management
   (*vi*) Production Management
   (*vii*) Research and Development
5. Major phases involved in the application of OR are:
   (*i*) Formulating the problem.
   (*ii*) Constructing a mathematical model
   (*iii*) Deriving the solution from the model
   (*iv*) Testing the model and its solution (updating the model)
   (*v*) Controlling the solution
   (*vi*) Implementation
6. The subject of Operational Research (OR) was developed in military context during World War II, pioneered by the British scientists.

NOTES

7. Linear programming technique is used in finding a solution for optimizing a given objective, such as profit maximization or cost minimization under certain constraints. This technique is primarily concerned with the optimal allocation of limited resources for optimizing a given function.
8. Waiting line or queuing theory deals with mathematical study of queues. The queues are formed whenever the current demand for service exceeds the current capacity to provide that service. Waiting line technique concerns itself with the random arrival of customers at a service station where the facility is limited.
9. Game theory is used to determine the optimum strategy in a competitive situation.
10. Dynamic programming refers to a systematic search for optimal solutions to problems that involve many highly complex interrelations that are, moreover, sensitive to multistage effects, such as successive time phases.
11. Operations Research has gained increasing importance since World War II in the technology of business and industry administration. It greatly helps in tackling the intricate and complex problems of modern business and industry.
12. OR techniques are being used in production, procurement, marketing, finance and other allied fields. It can be used to solve problems like how best can managers and executives allocate the available resources to various products so that in a given time the profits are maximum or the cost is minimum.
13. OR is useful to the production management in:
    (i) Selecting the building site for a plant, scheduling and controlling its development and designing its layout
    (ii) Locating within the plant and controlling the movements of required production materials and finished goods inventories
    (iii) Scheduling and sequencing production by adequate preventive maintenance with optimum number of operatives by proper allocation of machines
    (iv) Calculating the optimum product-mix.
14. The judgement phase includes:
    (i) A determination of the problem
    (ii) The establishment of the objectives and values related to the operation
    (iii) The determination of suitable measures of effectiveness.
15. The first step in an OR study is to formulate the problem in an appropriate form. Formulating a problem consists in identifying, defining and specifying the measures of the components of a decision model.
16. A model in OR is a simplified representation of an operation or as a process in which only the basic aspects or the most important features of a typical problem under investigation are considered. The objective of a model is to identify significant factors and interrelationships.
17. The classification of models is a subjective problem. They may be distinguished as follows:
    (i) Models by degree of abstraction
    (ii) Models by function

(iii) Models by structure

(iv) Models by nature of an environment

(v) Models by the extent of generality

NOTES

18. Descriptive models describe facts and relationships among the various activities of the problem. These models do not have an objective function as a part of the model to evaluate decision alternatives. In this model, it is possible to get information as to how one or more factors change as a result of changes in other factors.
19. In deterministic models, all parameters and functional relationships are assumed to be known with certainty when the decision is to be made. Linear programming and break even models are the examples of deterministic models.
20. It should always be kept in mind that OR alone cannot make the final decision. It is just a tool and simply suggests best alternatives. In the final analysis many business decisions will involve human element. Thus, OR is at best a supplement to rather than a substitute for management; subjective judgement is likely to remain a principal approach to decision-making.

1.13 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Where did the concept of operations research originate?
2. Name the fields where operations research can be used?
3. What do you mean by nature of operations research?
4. Which Indian companies use operations research?
5. What is inventory control? What is its importance?
6. What is decision theory?
7. What do you understand by network analysis?
8. What is non-linear programming?
9. What is integer programming?
10. Describe heuristic programming and algorithmic programming. When are these used?
11. What is search theory? In which situation is it applied?
12. What do you understand by theory of replacement? What is its significance?
13. How operations research enables proper deployment of resources?
14. How operations research techniques help in directing authority?
15. Define research phase.
16. What do you mean by testing the validity?
17. How competitive problems arise?
18. What is a model in operations research?
19. Write any one limitation of operations research.

NOTES

Long-Answer Questions

1. Explain the meaning and origin of operations research with the help of definitions and examples.
2. Discuss about the nature of operations research with the help of examples.
3. Write an essay on the development of operations research.
4. Discuss operations research in India.
5. How is operations research used as a tool in decision-making?
6. Explain the various types of operations research techniques.
7. How is operations research useful to management?
8. Explain the significance of operations research in management process.
9. Explain the role of operations research in modern business management.
10. Explain the various phases of operations research study.
11. Describe operations research approach with the help of a flow chart.
12. Explain the various categories under which operations research problems are classified.
13. How can you classify operations research models? Explain each type with the help of an example.
14. What should be the characteristics of a good operations research model? Explain.
15. What are the advantages of using a model? Explain and also give your own opinion.
16. Write about application of operations research in modern business environment. Explain with suitable examples.
17. Explain the limitations of operations research with the help of suitable examples.

1.14 REFERENCES AND SUGGESTED READINGS

Swaroop, Kanti, P.K. Gupta and Man Mohan. 2001. *Operations Research*. New Delhi: Sultan Chand & Sons.

Vohra, N. D. 2006. *Quantitative Techniques in Management*, 3rd edition. New Delhi: Tata McGraw-Hill Education.

Taha, Hamdy A. 1992. *Operations Research: An Introduction*. New York: Macmillan.

Gupta, P.K. and D.S. Hira. 2002. *Introduction to Operations Research*. New Delhi: S. Chand & Co.

Sharma, S. D. 2006. *Operations Research*. Uttar Pradesh: Kedar Nath Ram Nath & Co.

UNIT 2 LINEAR PROGRAMMING PROBLEM (LPP)

NOTES

Structure

2.0 Introduction
2.1 Objectives
2.2 Linear Programming: Meaning and Uses
  2.2.1 Meaning of Linear Programming
  2.2.2 Fields Where Linear Programming can be Used
2.3 Concepts, Notations and General Form of Linear Programming Model
  2.3.1 Basic Concepts and Notations
  2.3.2 General Form of the Linear Programming Model
2.4 Applications and Limitations of Linear Programming Problems
2.5 Formulation of Linear Programming Problem
  2.5.1 Graphic Solution
  2.5.2 General Formulation of Linear Programming Problem
  2.5.3 Matrix Form of Linear Programming Problem
2.6 Solution of Linear Programming Problem: Graphical Solution and Simplex Method
  2.6.1 Graphical Solution
  2.6.2 Some Important Definitions
  2.6.3 Canonical or Standard Forms of LPP
  2.6.4 Simplex Method
2.7 Artificial Variable Techniques
  2.7.1 The Big *M* Method
  2.7.2 The Two-Phase Simplex Method
2.8 Duality
  2.8.1 Duality in Linear Programming
2.9 Summing Up
2.10 Key Terms
2.11 Answers to 'Check Your Progress'
2.12 Questions and Exercises
2.13 References and Suggested Readings

2.0 INTRODUCTION

In this unit, you will learn the use of linear programming in decision-making. For a manufacturing process, a production manager has to take decisions as to what quantities and which process or processes are to be used so that the cost is minimum and profit is maximum. Currently, this method is used in solving a wide range of practical business problems. The word 'linear' means that the relationships are represented by straight lines. The word 'programming' means following a method for taking decisions systematically.

You will also learn a few basic terms like linearity, process and its level, criterion function, constraints, feasible solutions, optimum solution, etc. The term linearity implies straight line or proportional relationships among the relevant variables. Process means the combination of one or more inputs to produce a particular output. Criterion function is an objective function which is to be either maximized or minimized. Constraints are limitations under which one has to plan and decide. There are restrictions imposed upon decision variables. Feasible solutions are all those possible

NOTES

solutions considering given constraints. An optimum solution is considered the best among feasible solutions.

Finally, in this unit, you will learn about the canonical or standard form of LPP, artificial variable and duality. In the standard form, irrespective of the objective function, namely maximize or minimize, all the constraints are expressed as equations. Moreover, the Right Hand Side (RHS) of each constraint and all variables are non-negative. The simplex method and *M* method are the methods of solution by iterative procedure in a finite number of steps using matrix.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand what linear programming is
- Know the terms associated with a linear programming problem
- Learn how to formulate a linear programming problem
- Form a matrix of a linear programming problem
- Explain the applications and limitations of linear programming problems
- Solve a linear programming problem with two variables using the graphical method
- Describe linear programming problems in canonical form
- Solve linear programming problems using the simplex method
- Solve linear programming problems using the *M* method
- Explain the significance of artificial variable
- Describe duality and solve dual problems

2.2 LINEAR PROGRAMMING: MEANING AND USES

Decision-making has always been very important in the business and industrial world, particularly with regard to the problems concerning production of commodities. Which commodity/commodities to produce, in what quantities and by which process or processes, are the main questions before a production manager. English economist Alfred Marshall pointed out that the businessman always studies his production function and his input prices and substitutes one input for another till his costs become the minimum possible. All this sort of substitution, in the opinion of Marshall, is being done by businessman's trained instinct rather than with formal calculations. But now there does exist a method of formal calculations often termed as Linear Programming. This method was first formulated by a Russian mathematician L.V. Kantorovich, but it was developed later in 1947 by George B. Dantzig 'for the purpose of scheduling the complicated procurement activities of the United States Air Force'. Today, this method is being used in solving a wide range of practical business problems. The advent of electronic computers has further increased its applications to solve many other problems in industry. It is being considered as one of the most versatile management tools.

2.2.1 Meaning of Linear Programming

Linear Programming (LP) is a major innovation since World War II in the field of business decision-making, particularly under conditions of certainty. The word 'Linear'

NOTES

means that the relationships are represented by straight lines, i.e., the relationships are of the form $y = a + bx$ and the word 'Programming' means taking decisions systematically. Thus, LP is a decision-making technique under given constraints on the assumption that the relationships amongst the variables representing different phenomena happen to be linear. In fact, Dantzig originally called it 'programming of interdependent activities in a linear structure' but later shortened it to 'Linear Programming'. LP is generally used in solving maximization (sales or profit maximization) or minimization (cost minimization) problems subject to certain assumptions. Putting in a formal way, 'Linear Programming is the maximization (or minimization) of a linear function of variables subject to a constraint of linear inequalities.' Hence, LP is a mathematical technique designed to assist the organization in optimally allocating its available resources under conditions of certainty in problems of scheduling, product-mix and so on.

2.2.2 Fields Where Linear Programming can be Used

The problem for which LP provides a solution may be stated to maximize or minimize for some dependent variable which is a function of several independent variables when the independent variables are subject to various restrictions. The dependent variable is usually some economic objectives, such as profits, production, costs, work weeks, tonnage to be shipped, etc. More profits are generally preferred to less profits and lower costs are preferred to higher costs. Hence, it is appropriate to represent either maximization or minimization of the dependent variable as one of the firm's objective. LP is usually concerned with such objectives under given constraints with linearity assumptions. In fact, it is powerful to take in its stride a wide range of business applications. The applications of LP are numerous and are increasing every day. LP is extensively used in solving resource allocation problems. Production planning and scheduling, transportation, sales and advertising, financial planning, portfolio analysis, corporate planning, etc., are some of its most fertile application areas. More specifically, LP has been successfully applied in the following fields:

(*i*) *Agricultural applications:* LP can be applied in farm management problems as it relates to the allocation of resources such as acreage, labour, water supply or working capital in such a way that is maximizes net revenue.

(*ii*) *Contract awards:* Evaluation of tenders by recourse to LP guarantees that the awards are made in the cheapest way.

(*iii*) *Industrial applications:* Applications of LP in business and industry are of most diverse kind. Transportation problems concerning cost minimization can be solved by this technique. The technique can also be adopted in solving the problems of production (product-mix) and inventory control.

Thus, LP is the most widely used technique of decision-making in business and industry in modern times in various fields as stated above.

2.3 CONCEPTS, NOTATIONS AND GENERAL FORM OF LINEAR PROGRAMMING MODEL

The following are the concepts, notations and forms used in a linear programming model:

NOTES

2.3.1 Basic Concepts and Notations

There are certain basic concepts and notations to be first understood for easy adoption of the LP technique. A brief mention of such concepts is as follows:

1. *Linearity*: The term linearity implies straight line or proportional relationships among the relevant variables. Linearity in economic theory is known as constant returns which means that if the amount of input doubles, the corresponding output and profit are also doubled. Linearity assumption, thus, implies that if two machines and two workers can produce twice as much as one machine and one worker; four machines and four workers twice as much as two machines and two workers and so on.
2. *Process and its Level:* Process means the combination of particular inputs to produce a particular output. In a process, factors of production are used in fixed ratios, of course, depending upon technology and as such no substitution is possible with a process. There may be many processes open to a firm for producing a commodity and one process can be substituted for another. There is, thus, no interference of one process with another when two or more processes are used simultaneously. If a product can be produced in two different ways, then there are two different processes (or activities or decision variables) for the purpose of a linear programme.
3. *Criterion Function:* Criterion function is also known as objective function which states the determinants of the quantity either to be maximized or to be minimized. For example, revenue or profit is such a function when it is to be maximized or cost is such a function when the problem is to minimize it. An objective function should include all the possible activities with the revenue (profit) or cost coefficients per unit of production or acquisition. The goal may be either to maximize this function or to minimize this function. In symbolic form, let ZX denote the value of the objective function at the X level of the activities included in it. This is the total sum of individual activities produced at a specified level. The activities are denoted as $j = 1, 2, ..., n$. The revenue or cost coefficient of the jth activity is represented by C_j. Thus, $2X_1$, implies that X units of activity $j = 1$ yields a profit (or loss) of $C_1 = 2$.
4. *Constraints or Inequalities:* These are the limitations under which one has to plan and decide, i.e., restrictions imposed upon decision variables. For example, a certain machine requires one worker to be operated upon; another machine requires at least four workers (i.e., > 4); there are at most 20 machine hours (i.e., < 20) available; the weight of the product should be say 10 lbs and so on, are all examples of constraints or why are known as inequalities. Inequalities like $X > C$, (reads X is greater than C or $X < C$ (reads X is less than C) are termed as strict inequalities. The constraints may be in form of weak inequalities like $X \leq C$ (reads X is less than or equal to C) or $X \geq C$ (reads C is greater than or equal to C). Constraints may be in the form of strict equalities like $X = C$ (reads X is equal to C).

   Let b_i denote the quantity b of resource i available for use in various production processes. The coefficient attached to resource i is the quantity of resource i required for the production of one unit of product j.

5. *Feasible Solutions:* Feasible solutions are all those possible solutions which can be worked upon under given constraints. The region comprising of all feasible solutions is referred as *Feasible Region.*

6. *Optimum Solution:* Optimum solution is the best of the feasible solutions.

NOTES

2.3.2 General Form of the Linear Programming Model

Linear Programming problem mathematically can be stated as under:

Choose the quantities,

$$X_j \geq 0 \quad (j = 1,\ldots, n) \qquad \ldots(2.1)$$

This is also known as the non-negativity condition and in simple terms means that no X can be negative.

To maximize,

$$Z = \sum_{j=1}^{n} C_j X_j \qquad \ldots(2.2)$$

Subject to the constraints,

$$\sum_{j=1}^{n} a_{ij} X_j \leq b_i \qquad (i = 1,\ldots,m) \qquad \ldots(2.3)$$

The above is the usual structure of a linear programming model in the simplest possible form. This model can be interpreted as a profit maximization situation where n production activities are pursued at level X_j which have to be decided upon, subject to a limited amount of m resources being available. Each unit of the jth activity yields a return C and uses an amount a_{ij} of the ith resource. Z denotes the optimal value of the objective function for a given system.

Assumptions or the Conditions to be Fulfilled Underlying the LP Model

LP model is based on the assumptions of proportionality, additivity, certainty, continuity and finite choices.

Proportionality is assumed in the objective function and the constraint inequalities. In economic terminology this means that there are constant returns to scale, i.e., if one unit of a product contributes ₹ 5 toward profit, then 2 units will contribute ₹ 10, 4 units ₹ 20, and so on.

Certainty assumption means the prior knowledge of all the coefficients in the objective function, the coefficients of the constraints and the resource values. LP model operates only under conditions of certainty.

Additivity assumption means that the total of all the activities is given by the sum total of each activity conducted separately. For example, the total profit in the objective function is equal to the sum of the profit contributed by each of the products separately.

Continuity assumption means that the decision variables are continuous. Accordingly the combinations of output with fractional values, in case of product-mix problems, are possible and obtained frequently.

Finite choices assumption implies that finite number of choices are available to a decision-maker and the decision variables do not assume negative values.

NOTES

2.4 APPLICATIONS AND LIMITATIONS OF LINEAR PROGRAMMING PROBLEMS

The applications of linear programming problems are based on linear programming matrix coefficients and data transmission prior to solving the simplex algorithm. The problem can be formulated from the problem statement using linear programming techniques. The following are the objectives of linear programming:

- Identify the objective of the linear programming problem, i.e., which quantity is to be optimized. For example, maximize the profit.
- Identify the decision variables and constraints used in linear programming, for example, production quantities and production limitations are taken as decision variables and constraints.
- Identify the objective functions and constraints in terms of decision variables using information from the problem statement to determine the proper coefficients.
- Add implicit constraints, such as non-negative restrictions.
- Arrange the system of equations in a consistent form and place all the variables on the left side of the equations.

Applications of Linear Programming

Linear programming problems are associated with the efficient use of allocation of limited resources to meet desired objectives. A solution required to solve the linear programming problem is termed as optimal solution. The linear programming problems contain a very special subclass and depend on mathematical model or description. It is evaluated using relationships and are termed as straight-line or linear. The following are the applications of linear programming:

- Transportation problem
- Diet problem
- Matrix games
- Portfolio optimization
- Crew scheduling

Linear programming problem may be solved using a simplified version of the simplex technique called transportation method. Because of its major application in solving problems involving several product sources and several destinations of products, this type of problem is frequently called the transportation problem. It gets its name from its application to problems involving transporting products from several sources to several destinations. The formation is used to represent more general assignment and scheduling problems as well as transportation and distribution problems. The two common objectives of such problems are:

- To minimize the cost of shipping m units to n destinations.
- To maximize the profit of shipping m units to n destinations.

The goal of the diet problem is to find the cheapest combination of foods that will satisfy all the daily nutritional requirements of a person. The problem is formulated as a linear program where the objective is to minimize cost and meet constraints which require that nutritional needs be satisfied. The constraints are used to regulate

the number of calories and amounts of vitamins, minerals, fats, sodium and cholesterol in the diet.

Game method is used to turn a matrix game into a linear programming problem. It is based on the Min-Max theorem which suggests that each player determines the choice of strategies on the basis of a probability distribution over the player's list of strategies.

The portfolio optimization template calculates the optimal capital of investments that gives the highest return for the least risk. The unique design of the portfolio optimization technique helps in financial investments or business portfolios. The optimization analysis is applied to a portfolio of businesses to represent a desired and beneficial framework for driving capital allocation, investment and divestment decisions.

Crew scheduling is an important application of linear programming problem. It helps if any airline has a problem related to a large potential crew schedules variables. Crew scheduling models are a key to airline competitive cost advantage these days because crew costs are the second largest flying cost after fuel costs.

Limitations of Linear Programming Problems

Linear programming is applicable if constraints and objective functions are linear, but there are some limitations of this technique which are as follows:

- All the uncertain factors, such as weather conditions, growth rate of industry, etc., are not taken into consideration.
- Integer values are not taken as the solution, e.g., a value is required for fraction and the nearest integer is not taken for the optimal solution.
- Linear programming technique gives those practical-valued answers that are really not desirable with respect to linear programming problem.
- It deals with one single objective in real life problem which is more limited and the problems come with multi-objective.
- In linear programming, coefficients and parameters are assumed as constants but in realty they do not take place.
- Blending is a frequently encountered problem in linear programming. For example, if different commodities are purchased which have different characteristics and costs, then the problem helps to decide how much of each commodity would be purchased and blended within specified bound so that the total purchase cost is minimized.

2.5 FORMULATION OF LINEAR PROGRAMMING PROBLEM

A linear programming is a mathematical method for determining method to achieve the best outcome, i.e., maximum profit at lowest cost.

NOTES

Check Your Progress

1. What is linear programming?
2. What is meant by criterion function in linear programming?
3. Mention two areas where linear programming finds application.
4. What are the constraints in linear programming?

NOTES

2.5.1 Graphic Solution

The procedure for mathematical formulation of an LPP consists of the following steps:

Step 1: The decision variables of the problem are noted.

Step 2: The objective function to be optimized (maximized or minimized) as a linear function of the decision variables is formulated.

Step 3: The other conditions of the problem, such as resource limitation, market constraints, interrelations between variables, etc., are formulated as linear inequations or equations in terms of the decision variables.

Step 4: The non-negativity constraint from the considerations is added so that the negative values of the decision variables do not have any valid physical interpretation.

The objective function, the set of constraints and the non-negative constraint together form a linear programming problem.

2.5.2 General Formulation of Linear Programming Problem

The general formulation of the LPP can be stated as follows:

In order to find the values of n decision variables $X_1, X_2 \ldots X_n$ to maximize or minimize the objective function.

$$Z = C_1X_1 + C_2X_2 + \ldots + C_nX_n \quad \ldots (2.4)$$

$$\left.\begin{aligned} &a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n (\leq, =, \geq) b_1 \\ &a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n (\leq, =, \geq) b_2 \\ &\vdots \\ &a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{in}X_n (\leq, =, \geq) b_i \\ &\vdots \\ &a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n (\leq = \geq) b_m \end{aligned}\right\} \quad \ldots (2.5)$$

Here, the constraints can be inequality ≤ or ≥ or even in the form an equation (=) and finally satisfy the non-negative restrictions:

$$X_1 \geq 0, X_2 \geq 0 \ldots X_n \geq 0 \quad \ldots (2.6)$$

2.5.3 Matrix Form of Linear Programming Problem

The LPP can be expressed in the matrix form as follows:

Maximize or minimize $Z = CX \rightarrow$ Objective function

Subject to $AX (\leq, =, \geq) B \rightarrow$ Constant equation

$B > 0, X \geq 0 \rightarrow$ Non-negativity restrictions

Where, $X = (X_1, X_2 \cdots X_n)$

$C = (C_1, C_2 \cdots C_n)$

$$B = \begin{pmatrix} b_1 \\ b_2 \\ b_m \end{pmatrix} \quad A = \begin{pmatrix} a_{11}a_{12} \cdots a_{1n} \\ a_{21}a_{22} \cdots a_{2n} \\ \vdots \\ a_{m1}a_{m2} \cdots a_{mn} \end{pmatrix}$$

NOTES

Example 2.1: A manufacturer produces two types of models M_1 and M_2. Each model of the type M_1 requires 4 hours of grinding and 2 hours of polishing; whereas each model of the type M_2 requires 2 hours of grinding and 5 hours of polishing. The manufacturers have 2 grinders and 3 polishers. Each grinder works 40 hours a week and each polisher works for 60 hours a week. The profit on M_1 model is ₹ 3.00 and on model M_2 is ₹ 4.00. Whatever is produced in a week is sold in the market. How should the manufacturer allocate his production capacity to the two types of models, so that he may make the maximum profit in a week?

Solution:

Decision Variables: Let X_1 and X_2 be the number of units of M_1 and M_2.

Objective Function: Since the profit on both the models are given, we have to maximize the profit, viz.,

$\text{Max } Z = 3X_1 + 4X_2$

Constraints: There are two constraints: one for grinding and the other for polishing.

The number of hours available on each grinder for one week is 40 hours. There are 2 grinders. Hence, the manufacturer does not have more than $2 \times 40 = 80$ hours for grinding. M_1 requires 4 hours of grinding and M_2 requires 2 hours of grinding.

The grinding constraint is given by,

$4X_1 + 2X_2 \leq 80$

Since there are 3 polishers, the available time for polishing in a week is given by $3 \times 60 = 180$. M_1 requires 2 hours of polishing and M_2 requires 5 hours of polishing. Hence, we have $2X_1 + 5X_2 \leq 180$

Thus, we have,

$\text{Max } Z = 3X_1 + 4X_2$

Subject to, $4X_1 + 2X_2 \leq 80$

$2X_1 + 5X_2 \leq 180$

$X_1, X_2 \geq 0$

Example 2.2: A company manufactures two products A and B. These products are processed in the same machine. It takes 10 minutes to process one unit of product A and 2 minutes for each unit of product B and the machine operates for a maximum of 35 hours in a week. Product A requires 1 kg and B 0.5 kg of raw material per unit, the supply of which is 600 kg per week. The market constraint on product B is known to be 800 units every week. Product A costs ₹ 5 per unit and is sold at ₹ 10. Product B costs ₹ 6 per unit and can be sold in the market at a unit price of ₹ 8. Determine the number of units of A and B that should be manufactured per week to maximize the profit.

Solution:

Decision Variables: Let X_1 and X_2 be the number of products of A and B.

Objective Function: Cost of product A per unit is ₹ 5 and is sold at ₹ 10 per unit.

$\therefore$ Profit on one unit of product $A = 10 - 5 = 5$

$\therefore X_1$ units of product A, contributes a profit of ₹ $5X_1$ from one unit of product.

NOTES

Similarly, profit on one unit of $B = 8 - 6 = 2$

$\therefore X_2$ units of product B, contribute a profit of ₹ $2X_2$.

$\therefore$ The objective function is given by,

Max $Z = 5X_1 + 2X_2$

Constraints: Time requirement constraint is given by,

$10X_1 + 2X_2 \leq (35 \times 60)$

$10X_1 + 2X_2 \leq 2100$

Raw material constraint is given by,

$X_1 + 0.5X_2 \leq 600$

Market demand on product B is 800 units every week.

$\therefore X_2 \geq 800$

The complete LPP is,

Max $Z = 5X_1 + 2X_2$

Subject to, $10X_1 + 2X_2 \leq 2100$

$X_1 + 0.5X_2 \leq 600$

$X_2 \geq 800$

$X_1, X_2 \geq 0$

Example 2.3: A person requires 10, 12 and 12 units of chemicals A, B and C respectively for his garden. A liquid product contains 5, 2 and 1 units of A, B and C respectively per jar. A dry product contains 1, 2 and 4 units of A, B, C per carton. If the liquid product sells for ₹ 3 per jar and the dry product sells for ₹ 2 per carton, what should be the number of jar that needs to be purchased, in order to bring down the cost and meet the requirements?

Solution:

Decision Variables: Let X_1 and X_2 be the number of units of liquid and dry products.

Objective Function: Since the cost for the products are given, we have to minimize the cost.

Min $Z - 3X_1 + 2X_2$

Constraints: As there are 3 chemicals and their requirements are given, we have three constraints for these three chemicals.

$5X_1 + X_2 \geq 10$

$2X_1 + 2X_2 \geq 12$

$X_1 + 4X_2 \geq 12$

Hence, the complete LPP is,

Min $Z = 3X_1 + 2X_2$

NOTES

Subject to,

$$5X_1 + X_2 \geq 10$$
$$2X_1 + 2X_2 \geq 12$$
$$X_1 + 4X_2 \geq 12$$
$$X_1, X_2 \geq 0$$

Example 2.4: A paper mill produces two grades of paper, *X* and *Y*. Because of raw material restrictions, it cannot produce more than 400 tonnes of grade *X* and 300 tonnes of grade *Y* in a week. There are 160 production hours in a week. It requires 0.2 and 0.4 hours to produce a tonne of products *X* and *Y* respectively with corresponding profits of ₹ 200 and ₹ 500 per tonne. Formulate this as a LPP to maximize profit and find the optimum product mix.

Solution:

Decision Variables: Let X_1 and X_2 be the number of units of the two grades of paper, *X* and *Y*.

Objective Function: Since the profit for the two grades of paper *X* and *Y* are given, the objective function is to maximize the profit.

Max $Z = 200X_1 + 500X_2$

Constraints: There are 2 constraints one with reference to raw material, and the other with reference to production hours.

Max $Z = 200X_1 + 500X_2$

Subject to,

$$X_1 \leq 400$$
$$X_2 \leq 300$$
$$0.2X_1 + 0.4X_2 \leq 160$$

Non-negative restriction $X_1, X_2 \geq 0$

Example 2.5: A company manufactures two products *A* and *B*. Each unit of *B* takes twice as long to produce as one unit of *A* and if the company were to produce only *A* it would have time to produce 2000 units per day. The availability of the raw material is enough to produce 1500 units per day of both *A* and *B* together. Product *B* requiring a special ingredient, only 600 units of it can be made per day. If *A* fetches a profit of ₹ 2 per unit and *B* a profit of ₹ 4 per unit, find the optimum product mix by graphical method.

Solution: Let X_1 and X_2 be the number of units of the products *A* and *B* respectively.

The profit after selling these two products is given by the objective function,

Max $Z = 2X_1 + 4X_2$

Since the company can produce at the most 2000 units of the product in a day and product *B* requires twice as much time as that of product *A*, production restriction is given by,

$$X_1 + 2X_2 \leq 2000$$

Since the raw material is sufficient to produce 1500 units per day of both *A* and *B*, we have $X_1 + X_2 \leq 1500$.

NOTES

There are special ingredients for the product B we have $X_2 \leq 600$.

Also, since the company cannot produce negative quantities $X_1 \geq 0$ and $X_2 \geq 0$.

Hence, the problem can be finally put in the form:

Find X_1 and X_2 such that the profits, $Z = 2X_1 + 4X_2$ is maximum.

Subject to,
$$X_1 + 2X_2 \leq 2000$$
$$X_1 + X_2 \leq 1500$$
$$X_2 \leq 600$$
$$X_1, X_2 \geq 0$$

Example 2.6: A firm manufacturers 3 products A, B and C. The profits are ₹ 3, ₹ 2 and ₹ 4, respectively. The firm has 2 machines and the following is the required processing time in minutes for each machine on each product.

		Product		
		A	*B*	*C*
Machines	*C*	4	3	5
	D	3	2	4

Machine C and D have 2000 and 2500 machine minutes, respectively. The firm must manufacture 100 units of A, 200 units of B and 50 units of C, but not more than 150 units of A. Set up an LP problem to maximize the profit.

Solution: Let X_1, X_2, X_3 be the number of units of the product A, B, C respectively.

Since the profits are ₹ 3, ₹ 2 and ₹ 4 respectively, the total profit gained by the firm after selling these three products is given by,

$$Z = 3X_1 + 2X_2 + 4X_3$$

The total number of minutes required in producing these three products at machine C is given by $4X_1 + 3X_2 + 5X_3$ and at machine D is given by $3X_1 + 2X_2 + 4X_3$.

The restrictions on the machine C and D are given by 2000 minutes and 2500 minutes.

$$4X_1 + 3X_2 + 5X_3 \leq 2000$$
$$3X_1 + 2X_2 + 4X_3 \leq 2500$$

Also, since the firm manufactures 100 units of A, 200 units of B and 50 units of C, but not more than 150 units of A, the further restriction becomes,

$$100 \leq X_1 \leq 150$$
$$200 \leq X_2 \geq 0$$
$$50 \leq X_3 \geq 0$$

Hence, the allocation problem of the firm can be finally put in the form:

Find the value of X_1, X_2, X_3 so as to maximize,

$$Z = 3X_1 + 2X_2 + 4X_3$$

NOTES

Example 2.7: Solve the following LPP by graphical method.

Minimize, $Z = 20X_1 + 10X_2$

Subject to, $X_1 + 2X_2 \leq 40$

$3X_1 + X_2 \geq 30$

$4X_1 + 3X_2 \geq 60$

$X_1, X_2 \geq 0$

Solution: Replace all the inequalities of the constraints by equation,

$X_1 + 2X_2 = 40$ If $X_1 = 0 \Rightarrow X_2 = 20$

If $X_2 = 0 \Rightarrow X_1 = 40$

$\therefore X_1 + 2X_2 = 40$ passes through (0, 20) (40, 0)

$3X_1 + X_2 = 30$ passes through (0, 30) (10, 0)

$4X_1 + 3X_2 = 60$ passes through (0, 20) (15, 0)

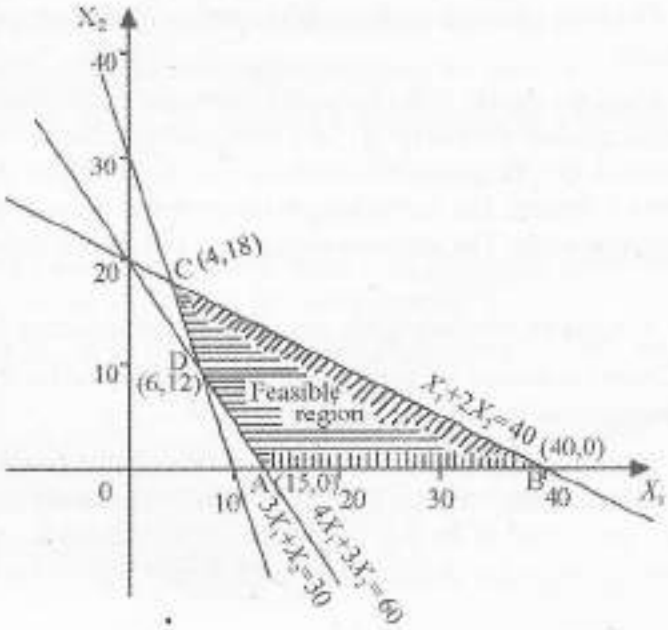
Plot each equation on the graph.

The feasible region is *ABCD*.

C and *D* are points of intersection of lines.

$X_1 + 2X_2 = 40$, $3X_1 + X_2 = 30$

And, $4X_1 + 3X_2 = 60$



On solving, we get *C* (4, 18) and *D* (6, 12)

Corner Points	*Value of* $Z = 20X_1 + 10X_2$
A (15, 0)	300
B (40, 0)	800
C (4, 18)	260
D (6, 12)	240 (Minimum value)

∴ The minimum value of Z occurs at D (6, 12). Hence, the optimal solution is $X_1 = 6, X_2 = 12$.

Example 2.8: Find the maximum value of $Z = 5X_1 + 7X_2$

Subject to the constraints,

$X_1 + X_2 \leq 4$

$3X_1 + 8X_2 \leq 24$

$10X_1 + 7X_2 \leq 35$

$X_1, X_2 > 0$

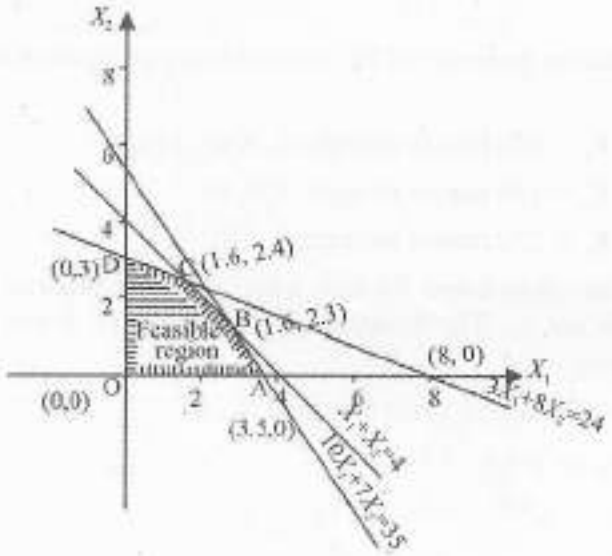
Solution: Replace all the inequalities of the constraints by forming equations.

$X_1 + X_2 = 4$ passes through (0, 4) (4, 0)

$3X_1 + 8X_2 = 24$ passes through (0, 3) (8, 0)

$10X_1 + 7X_2 = 35$ passes through (0, 5) (3.5, 0)

Plot these lines in the graph and mark the region below the line as the inequality of the constraint is ≤ and is also lying in the first quadrant.



The feasible region is *OABCD*.

B and *C* are points of instruction of lines,

$X_1 + X_2 = 4$, $10X_1 + 7X_2 = 35$

And, $3X_1 + 8X_2 = 24$

On solving we get,

B (1.6, 2.3)

C (1.6, 2.4)

Corner Points	*Value of* $Z = 5X_1 + 7X_2$
O (0, 0)	0

NOTES

NOTES

A (3.5, 0)	17.5
B (1.6, 2.3)	25.1
C (1.6, 2.4)	24.8 (Maximum value)
D (0, 3)	21

$\therefore$ The maximum value of Z occurs at C (1.6, 2.4) and the optimal solution is $X_1 = 1.6, X_2 = 2.4$.

Example 2.9: A company makes 2 types of hats. Each hat A needs twice as much labour time as the second hat B. If the company is able to produce only hat B, then it can make about 500 hats per day. The market limits daily sales of the hat A and hat B to 150 and 250 hats. The profits on hat A and hat B are ₹ 8 and ₹ 5, respectively. Solve graphically to get the optimal solution.

Solution: Let X_1 and X_2 be the number of units of type A and type B hats, respectively.

Maximize $Z = 8X_1 + 5X_2$

Subject to, $2X_1 + 2X_2 \leq 500$

$X_1 \geq 150$

$X_2 \geq 250$

$X_1, X_2 \geq 0$

First rewrite the inequality of the constraint into an equation and plot the lines in the graph.

$2X_1 + X_2 = 500$ passes through (0, 500) (250, 0)

$X_1 = 150$ passes through (150, 0)

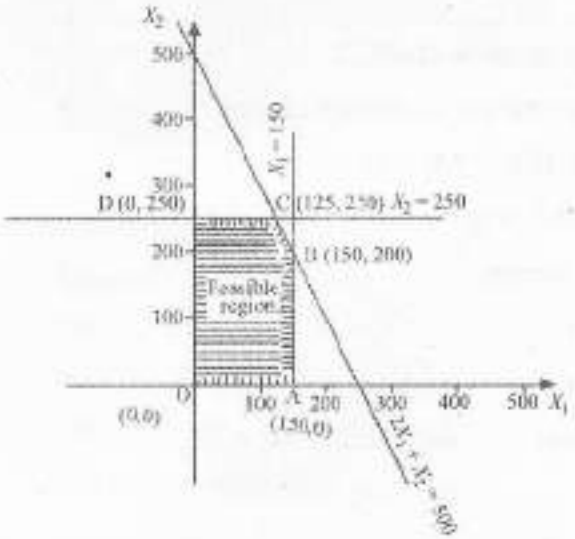
$X_2 = 250$ passes through (0, 250)

We mark the region below the lines lying in the first quadrant as the inequality of the constraints are $\leq$. The feasible region is $OABCD$. B and C are points of intersection of lines:

$2X_1 + X_2 = 500$, where $X_1 = 150$ and $X_2 = 250$

On solving, we get $B = (150, 200)$

$C = (125, 250)$



NOTES

Corner Points	*Value of* $Z = 8X_1 + 5X_2$
O (0, 0)	0
A (150, 0)	1200
B (150, 200)	2200
C (125, 250)	2250 (Maximum $Z = 2250$)
D (0, 250)	1250

The maximum value of Z is attained at C (125, 250)

$\therefore$ The optimal solution is $X_1 = 125$, $X_2 = 250$

i.e., The company should produce 125 hats of type A and 250 hats of type B in order to get the maximum profit of ₹ 2250.

Example 2.10: By graphical method solve the following LPP.

Maximize $Z = 3X_1 + 4X_2$

Subject to, $5X_1 + 4X_2 \leq 200$

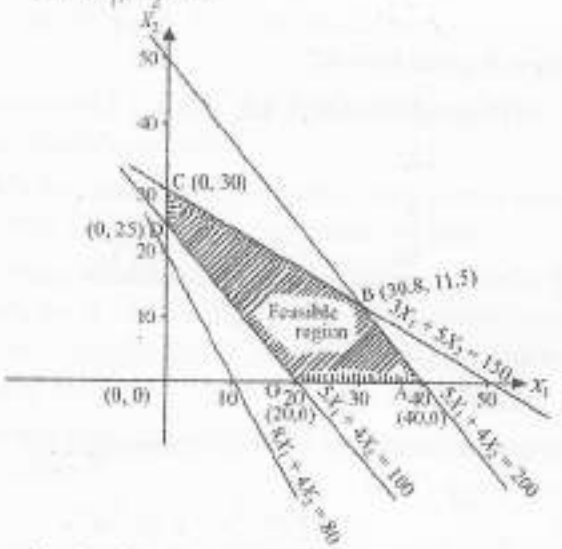
$3X_1 + 5X_2 \leq 150$

$5X_1 + 4X_2 \geq 100$

$8X_1 + 4X_2 \geq 80$

and $X_1, X_2 \geq 0$

Solution:



Feasible region is given by $OABCD$.

Corner Points	*Value of* $Z = 3X_1 + 4X_2$
O (20, 0)	60
A (40, 0)	120
B (30.8, 11.5)	138.4 (Maximum value)
C (0, 30)	120
D (0, 25)	100

$\therefore$ The maximum value of Z is attained at B (30.8, 11.5)

$\therefore$ The optimal solution is $X_1 = 30.8$, $X_2 = 11.5$

NOTES

Example 2.11: Use graphical method to solve the following LPP.

Maximize, $Z = 6X_1 + 4X_2$

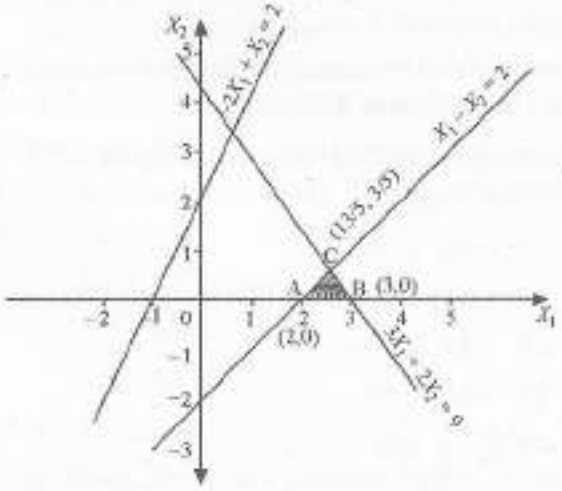
Subject to, $-2X_1 + X_2 \leq 2$

$X_1 - X_2 \leq 2$

$3X_1 + 2X_2 \leq 9$

$X_1, X_2 \geq 0$

Solution:



The feasible region is given by *ABC*.

Corner Points	*Value of* $Z = 6X_1 + 4X_2$
A (2, 0)	12
B (3,0)	18
C (13/5, 3/5)	$\frac{90}{5} = 18$ (Maximum value)

The maximum value of Z is attained at C (13/5, 3/5)

$\therefore$ The optimal solution is $X_1 = 13/5, X_2 = 3/5$

Example 2.12: Use graphical method to solve the following LPP.

Minimize $Z = 3X_1 + 2X_2$

Subject to, $5X_1 + X_2 \geq 10$

$X_1 + X_2 \geq 6$

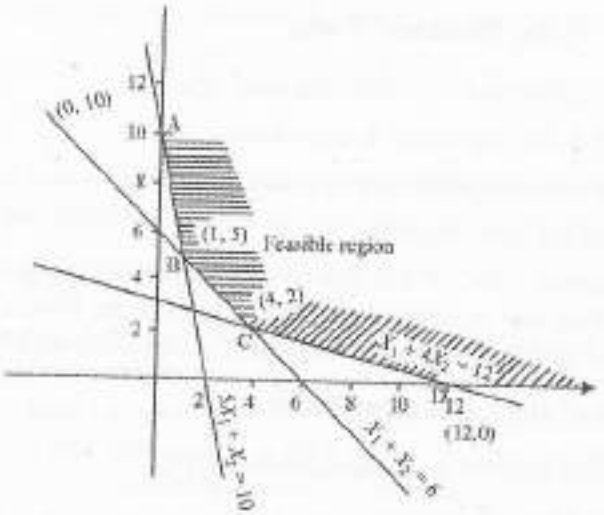
$X_1 + 4X_2 \geq 12$

$X_1, X_2 \geq 0$

Solution:

Corner Points	*Value of* $Z = 3X_1 + 2X_2$
A (0, 10)	20
B (1, 5)	13 (Minimum value)
C (4, 2)	16
D (12, 0)	36

NOTES



Since the minimum value is attained at B (1,5), the optimum solution is,

$X_1 = 1, X_2 = 5$

Note: In this problem, if the objective function is maximization then the solution is unbounded, as maximum value of Z occurs at infinity.

2.6.2 Some Important Definitions

1. A set of values $X_1, X_2, \ldots X_n$ which satisfies the constraints of the LPP is called its solution.
2. Any solution to a LPP which satisfies the non-negativity restrictions of the LPP is called its feasible solution.
3. Any feasible solution which optimizes (minimizes or maximizes) the objective function of the LPP is called its optimum solution.
4. Given a system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution. Such m variables are called basic variables and the remaining variables are called non-basic variables.
5. A basic feasible solution is a basic solution which also satisfies all basic variables are non-negative.

   Basic feasible solutions are of two types:

   (*i*) *Non-Degenerate:* A non-degenerate basic feasible solution is the basic feasible solution which has exactly m positive X_i ($i = 1, 2, \ldots, m$), i.e., none of the basic variables is zero.

   (*ii*) *Degenerate:* A basic feasible solution is said to be degenerate if one or more basic variables are zero.
6. If the value of the objective function Z can be increased or decreased indefinitely, such solutions are called unbounded solutions.

2.6.3 Canonical or Standard Forms of LPP

The general LPP can be put in either canonical or standard forms.

In the standard form, irrespective of the objective function, namely maximize or minimize, all the constraints are expressed as equations. Moreover, RHS of each constraint and all variables are non-negative.

Check Your Progress

5. What do you understand by solution in the context of linear programming problem?
6. What is a 'basic solution' of an LPP?
7. What is meant by basic and non-basic variables?
8. What do you understand by basic feasible solution?

NOTES

Characteristics of the Standard Form

(*i*) The objective function is of maximization type.

(*ii*) All constraints are expressed as equations.

(*iii*) Right hand side of each constraint is non-negative.

(*iv*) All variables are non-negative.

In the canonical form, if the objective function is of maximization, all the constraints other than non-negativity conditions are '≤' type. If the objective function is of minimization, all the constraints other than non-negative conditions are '≥' type.

Characteristics of the Canonical Form

(*i*) The objective function is of maximization type.

(*ii*) All constraints are of '≤' type.

(*iii*) All variables X_i are non-negative.

Notes:

1. Minimization of a function Z is equivalent to maximization of the negative expression of this function, i.e., Min Z = –Max (–Z).
2. An inequality in one direction can be converted into an inequality in the opposite direction by multiplying both sides by (–1).
3. Suppose we have the constraint equation,

   $$a_{11}X_1+a_{12}X_2+\ldots\ldots+a_{1n}X_n=b_1$$

   This equation can be replaced by two weak inequalities in opposite directions.

   $$a_{11}X_1+a_{12}X_2+\ldots\ldots+a_{1n}X_n\le b_1$$

   $$a_{11}X_1+a_{12}X_2+\ldots\ldots+a_{1n}X_n\ge b_1$$

4. If a variable is unrestricted in sign, then it can be expressed as a difference of two non-negative variables, i.e., X_1 is unrestricted in sign, then $X_1=X_1'-X_1''$, where X_1, X_1', X_1'' are ≥ 0.
5. In standard form, all the constraints are expressed in equation, which is possible by introducing some additional variables called slack variables and surplus variables so that a system of simultaneous linear equations is obtained. The necessary transformation will be made to ensure that $b_i \ge 0$.

Definition

(*i*) If the constraints of a general LPP be,

$$\sum_{j=1}^{n} a_{ij}\, X_j \le b_i \ (i=1, 2, \ldots, m),$$

then the non-negative variables S_i, which are introduced to convert the inequalities (≤) to the equalities $\sum_{j=1}^{n} a_{ij}\, X_j + S_i = b_i \ (i=1, 2, \ldots, m)$, are called slack variables.

Slack variables are also defined as the non-negative variables which are added in the LHS of the constraint to convert the inequality '≤' into an equation.

(*ii*) If the constraints of a general LPP be,

$$\sum_{j=1}^{n} a_{ij}\, X_j \ge b_i \ (i=1, 2, \ldots, m),$$

NOTES

then, the non-negative variables S_i which are introduced to convert the inequalities '≥' to the equalities $\sum_{j=1}^{n} a_{ij} X_j - S_i = b_i$ $(i = 1, 2, ..., m)$ are called surplus variables.

Surplus variables are defined as the non-negative variables which are removed from the LHS of the constraint to convert the inequality '≥' into an equation.

2.6.4 Simplex Method

Simplex method is an iterative procedure for solving LPP in a finite number of steps. This method provides an algorithm which consists of moving from one vertex of the region of feasible solution to another in such a manner that the value of the objective function at the succeeding vertex is less or more as the case may be that at the previous vertex. This procedure is repeated and since the number of vertices is finite, the method leads to an optimal vertex in a finite number of steps or indicates the existence of unbounded solution.

Definition

(*i*) Let X_B be a basic feasible solution to the LPP.

Max $Z = C_X$

Subject to $A_X = b$ and $X \geq 0$, such that it satisfies $X_B = B^{-1}b$,

Where B is the basic matrix formed by the column of basic variables.

The vector $C_B = (C_{B1}, C_{B2} \ldots C_{Bm})$, where C_{Bi} are components of C associated with the basic variables is called the cost vector associated with the basic feasible solution X_B.

(*ii*) Let X_B be a basic feasible solution to the LPP.

Max $Z = C_X$, where $A_X = b$ and $X \geq 0$.

Let C_B be the cost vector corresponding to X_B. For each column vector a_j in A_1, which is not a column vector of B, let

$$a_j = \sum_{i=1}^{m} a_i b_j$$

Then the number $Z_j = \sum_{i=1}^{m} C_{Bi} a_{ij}$ is called the evaluation corresponding to a_j and the number $(Z_j - C_j)$ is called the net evaluation corresponding to j.

Simplex Algorithm

For the solution of any LPP by simplex algorithm, the existence of an initial basic feasible solution is always assumed. The steps for the computation of an optimum solution are as follows:

Step 1: Check whether the objective function of the given LPP is to be maximized or minimized. If it is to be minimized then we convert it into a problem of maximization by,

Min Z = –Max $(-Z)$

NOTES

Step 2: Check whether all b_i $(i = 1, 2, \ldots, m)$ are positive. If any one of b_i is negative, then multiply the inequation of the constraint by -1 so as to get all b_i to be positive.

Step 3: Express the problem in the standard form by introducing slack/surplus variables to convert the inequality constraints into equations.

Step 4: Obtain an initial basic feasible solution to the problem in the form $X_B = B^{-1}b$ and put it in the first column of the simplex table. Form the initial simplex table shown as follows:

		C_j	C_1	C_2	C_3	..	..	0 0 0
C_B	S_B	X_B	X_1	X_2	X_3	X_4	..... X_n	S_1, S_2 ..... S_m
C_{B1}	S_1	b_1	a_{11}	a_{12}	a_{13}	a_{14}	..... a_{1n}	1 0 0
C_{B2}	S_2	b_2	a_{21}	a_{22}	a_{23}	a_{24}	..... a_{2n}	1 0 0

Step 5: Compute the net evaluations $Z_j - C_j$ by using the relation:

$Z_j - C_j = C_B (a_j - C_j)$

Examine the sign of $Z_j - C_j$:

(*i*) If all $Z_j - C_j \geq 0$, then the initial basic feasible solution X_B is an optimum basic feasible solution.

(*ii*) If at least one $Z_j - C_j > 0$, then proceed to the next step as the solution is not optimal.

Step 6: To find the entering variable, i.e., key column.

If there are more than one negative $Z_j - C_j$ choose the most negative of them. Let it be $Z_r - C_r$ for some $j = r$. This gives the entering variable X_r and is indicated by an arrow at the bottom of the rth column. If there are more than one variable having the same most negative $Z_j - C_j$, then any one of the variable can be selected arbitrarily as the entering variable.

(*i*) If all $X_{ir} \leq 0$ $(i = 1, 2, \ldots, m)$ then there is an unbounded solution to the given problem.

(*ii*) If at least one $X_{ir} > 0$ $(i = 1, 2, \ldots, m)$, then the corresponding vector X_r enters the basis.

Step 7: To find the leaving variable or key row:

Compute the ratio $(X_{Bi}/X_{kr}, X_{ir}>0)$

If the minimum of these ratios be X_{Bi}/X_{kr}, then choose the variable X_k to leave the basis called the key row and the element at the intersection of the key row and the key column is called the key element.

Step 8: Form a new basis by dropping the leaving variable and introducing the entering variable along with the associated value under C_B column. The leaving element is converted to unity by dividing the key equation by the key element and all other elements in its column to zero by using the formula:

New element = Old element

$$- \left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}} \right]$$

Step 9: Repeat the procedure of Step (5) until either an optimum solution is obtained or there is an indication of unbounded solution.

Example 2.13: Use simplex method to solve the following LPP.

Maximize, $Z = 3X_1 + 2X_2$

Subject to, $X_1 + X_2 \le 4$

$X_1 - X_2 \le 2$

$X_1, X_2 \ge 0$

Solution: By introducing the slack variables S_1 and S_2 convert the problem into standard form.

Max, $Z = 3X_1 + 2X_2 + 0S_1 + 0S_2$

Subject to, $X_1 + X_2 + S_1 = 4$

$X_1 - X_2 + S_2 = 2$

$X_1, X_2, S_1, S_2 \ge 0$

$$\begin{bmatrix} X_1 & X_2 & S_1 & S_2 \\ 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

An initial basic feasible solution is given by,

$X_B = B^{-1}b$,

Where, $B = I_2, X_B = (S_1, S_2)$

i.e., $(S_1, S_2) = I_2\,(4, 2) = (4, 2)$

Initial Simplex Table

$Z_j = C_B\, a_j$

$$Z_1 - C_1 = C_B a_1 - C_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (1\ 1) - 3 = -3$$

$$Z_2 - C_2 = C_B a_2 - C_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (1\ 1) - 2 = -2$$

$$Z_3 - C_3 = C_B a_3 - C_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (1\ 0) - 0 = -0$$

$$Z_4 - C_4 = C_B a_4 - C_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} (0\ 1) - 0 = -0$$

NOTES

NOTES

		C_j	3	2	0	0	
C_B	B	X_B	X_1	X_2	S_1	S_2	Min $\frac{X_B}{X_1}$
0	S_1	4	1	1	1	0	4/1 = 4
←0	S_2	2	①	−1	0	1	2/1 = 2
	Z_j	0	0	0	0	0	
	$Z_j - C_j$		−3↑	−2	0	0	

Since, there are some $Z_j - C_j = 0$, the current basic feasible solution is not optimum.

Since, $Z_1 - C_1 = -3$ is the most negative, the corresponding non-basic variable X_1 enters the basis.

The column corresponding to this X_1 is called the key column.

$$\text{Ratio} = \text{Min}\left\{\frac{X_{Bi}}{X_{ir}}, X_{ir} > 0\right\}$$

$$= \text{Min}\left\{\frac{4}{1}, \frac{2}{1}\right\}, \text{ which corresponds to } S_2$$

∴ The leaving variable is the basic variable S_2. This row is called the key row. Convert the leading element X_{21} to units and all other elements in its column n, i.e., (X_1) to zero by using the formula:

New element = Old element –

$$\left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}}\right]$$

To apply this formula, first we find the ratio, namely

$$\frac{\text{The element to be zero}}{\text{Key element}} = \frac{1}{1} = 1$$

Apply this ratio for the number of elements that are converted in the key row. Multiply this ratio by key row element shown as follows:

1×2

1×1

1×-1

1×0

1×1

Now, subtract this element from the old element. The element to be converted into zero is called the old element row. Finally, we have

$4 - 1 \times 2 = 2$

$1 - 1 \times 1 = 0$

$1 - 1 \times -1 = 2$

$1-1\times 0=1$

$0-1\times 1=-1$

$\therefore$ The improved basic feasible solution is given in the following simplex table.

First Iteration

		C_j	3	2	0	0	
C_B	B	X_B	X_1	X_2	S_1	S_2	Min $\frac{X_B}{X_2}$ ←
←0	S_1	2	0	(2)	1	−1	2/2 = 1
3	X_1	2	1	−1	0	1	–
	Z_j	6	3	−3	0	0	
	$Z_j - C_j$		0	−5↑	0	0	

Since, $Z_2 - C_2$ is the most negative, X_2 enters the basis.

To find Min $\left(\frac{X_B}{X_{i2}}, X_{i2} > 0\right)$

Min $\left[\frac{2}{2}, \frac{2}{-1}\right] = 1$

($\therefore$ Negative or zero values are not considered)

This gives the outgoing variables. Convert the leaving element into one. This is done by dividing all the elements in the key row by 2. The remaining elements are converted to zero by using the following formula.

Here, $-\frac{1}{2}$ is the common ratio. Put this ratio 5 times and multiply each ratio by the key row element.

$-\frac{1}{2}\times 2$

$-\frac{1}{2}\times 0$

$-\frac{1}{2}\times 2$

$-1/2\times 1$

$-1/2\times -1$

Subtract this from the old element. All the row elements which are converted into zero are called the old elements.

$2-\left(-\frac{1}{2}\times 2\right)=3$

$1-(-1/2\times 0)=1$

$-1-(-1/2\times 2)=0$

$0-(-1/2\times 1)=1/2$

$1-(-1/2\times -1)=1/2$

NOTES

NOTES

Second Iteration

		C_j	3	2	0	0
C_B	B	X_B	X_1	X_2	S_1	S_2
2	X_2	1	0	1	1/2	−1/2
3	X_1	3	1	0	1/2	1/2
	Z_j	11	3	2	5/2	1/2
	$Z_j - C_j$		0	0	5/2	1/2

Since all $Z_j - C_j \geq 0$, the solution is optimum. The optimal solution is Max $Z = 11$, $X_1 = 3$, and $X_2 = 1$.

Example 2.14: Solve the LPP

$$\text{Max } Z = 3x_1 + 2x_2$$

Subject to,

$$4x_1 + 3x_2 \leq 12$$
$$4x_1 + x_2 \leq 8$$
$$4x_1 - x_2 \leq 8$$
$$x_1, x_2 \geq 0$$

Solution: Convert the inequality of the constraint into an equation by adding slack variables S_1, S_2, S_3.

$$\text{Max } Z = 3x_1 + 2x_2 + 0S_1 + 0S_2 + 0S_3$$

Subject to,

$$4x_1 + 3x_2 + S_1 = 12$$
$$4x_1 + x_2 + S_2 = 8$$
$$4x_1 - x_2 + S_3 = 8$$
$$x_1, x_2, S_1, S_2, S_3 \geq 0$$

$$\begin{bmatrix} x_1 & x_2 & S_1 & S_2 & S_3 \\ 4 & 3 & 1 & 0 & 0 \\ 4 & 1 & 0 & 1 & 0 \\ 4 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 8 \\ 8 \end{bmatrix}.$$

Initial Table

		C_j	3	2	0	0	0	
C_B	*Basis*	x_B	x_1	x_2	S_1	S_2	S_3	*Min* $\frac{x_B}{x_1}$
0	S_1	12	4	3	1	0	0	12/4 = 3
0	S_2	8	4	1	0	1	0	8/4 = 2
←0	S_3	8	(4)	−1	0	0	1	8/4 = 2
	Z_j	0	0	0	0	0	0	
	$Z_j - C_j$		−3	−2	0	0	0	

↑

NOTES

$\therefore Z_1 - C_1$ is most negative, x_1 enters the basis. And the $\min\left(\frac{x_B}{x_{il}}, x_{il} > 0\right) = \min$ $(3, 2, 2) = 2$ gives S_3 as the leaving variable.

Convert the leading element into 1, by dividing key row element by 4 and the remaining elements into 0.

Initial Simplex Table

	C_j	3	2	0	0	0	
C_B	Basis	x_B	x_1	x_2	S_1	S_2	S_3	$Min \frac{x_B}{x_2}$
0	S_1	4	0	4	1	0	−1	4/4 = 1
←0	S_2	0	0	②	0	1	−1	0/2 = 0
3	x_1	2	1	−1/4	0	0	¼	—
	Z_j	⑥	3	−3/4	0	0	¾	
	$Z_j - C_j$		0	−11/4	0	0	¾	

↑

$$8 - \left(\frac{4}{4} \times 8\right) = 0 \qquad 12 - \left(\frac{4}{4} \times 8\right) = 4$$

$$4 - \left(\frac{4}{4} \times 4\right) = 0 \qquad 4 - \left(\frac{4}{4} \times 4\right) = 0$$

$$1 - \left(\frac{4}{4} \times -1\right) = 2 \qquad 3 - \left(\frac{4}{4} \times -1\right) = 4$$

$$0 - \left(\frac{4}{4} \times 0\right) = 0 \qquad 1 - \left(\frac{4}{4} \times 0\right) = 1$$

$$1 - \left(\frac{4}{4} \times 0\right) = 1 \qquad 0 - \left(\frac{4}{4} \times 0\right) = 0$$

$$0 - \left(\frac{4}{4} \times 1\right) = -1 \qquad 0 - \left(\frac{4}{4} \times 1\right) = -1$$

Since, $Z_2 - C_2 = -\frac{11}{4}$ is the most negative, x_2 enters the basis.

To find the outgoing variable, find $\text{Min}\left(\frac{x_B}{x_{i2}}, x_{i2} > 0\right)$

$$\text{Min}\left(\frac{4}{4}, \frac{0}{2}, -\right) = 0$$

First Iteration

Therefore, S_2 leaves the basis. Convert the leading element into 1 by dividing the key row elements by 2 and make the remaining elements in that column as zero using the formula.

$$\text{New element} = \text{Old element} - \left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}}\right]$$

NOTES

		C_j	3	2	0	0	0	
C_B	Basis	x_B	x_1	x_2	S_1	S_2	S_3	Min $\frac{x_B}{S_3}$
←0	S_1	4	0	0	1	−2	(1)	4/1 = 4
2	x_2	0	0	1	0	1/2	−1/2	—
3	x_1	2	1	0	0	1/8	1/8	$\frac{2}{1/8}$ = 16
	Z_j	6	3	2	0	11/8	−5/8	
	$Z_j - C_j$		0	0	0	11/8	−5/8	

↑

Second Iteration

Since $Z_5 - C_5 = -5/8$ is most negative, S_3 enters the basis and

$$\text{Min}\left(\frac{x_B}{S_{13}}, S_{13}\right) = \text{Min}\left(\frac{4}{1}, 16\right) = 4.$$

Therefore, S_1 leaves the basis. Convert the leading element into one and remaining elements as zero.

Third Iteration

		C_j	3	2	0	0	0
C_B	Basis	x_B	x_1	x_2	S_1	S_2	S_3
0	S_3	4	0	0	1	−2	1
2	x_2	2	0	1	1/2	−1/2	0
3	x_1	3/2	1	0	−1/8	3/8	0
	Z_j	17/2	3	2	5/8	1/8	0
	$Z_j - C_j$		0	0	5/8	1/8	0

Since all $Z_j - C_j \geq 0$, the solution is optimum and it is given by $x_1 = 3/2$, $x_2 = 2$ and Max $Z = 17/2$.

Example 2.15: Using simplex method solve the LPP.

$$\text{Max } Z = x_1 + x_2 + 3x_3$$

Subject to,

$$3x_1 + 2x_2 + x_3 \leq 3$$
$$2x_1 + x_2 + 2x_3 \leq 2$$
$$x_1, x_2, x_3 \geq 0$$

Solution: Rewrite the inequality of the constraints into an equation by adding slack variables.

$$\text{Max } Z = x_1 + x_2 + 3x_3 + 0S_1 + 0S_2$$

Subject to,

$$3x_1 + 2x_2 + x_3 + S_1 \leq 3$$
$$2x_1 + x_2 + 2x_3 + S_2 \leq 2$$

Initial basic feasible solution is,

$$x_1 = x_2 = x_3 = 0$$
$$S_1 = 3,\ S_2 = 2 \text{ and } Z = 0$$

NOTES

$$\begin{bmatrix} x_1 & x_2 & x_3 & S_1 & S_2 \\ 3 & 2 & 1 & 1 & 0 \\ 2 & 1 & 2 & 0 & 1 \\ 1 & 1 & 3 & 0 & 0 \end{bmatrix}$$

		C_j	1	1	3	0	0	
C_B	Basis	x_B	x_1	x_2	x_3	S_1	S_2	Min $\frac{x_B}{x_3}$
0	S_1	3	3	2	1	1	0	3/1 = 3
←0	S_2	2	2	1	②	0	1	2/2 = 1
	Z_j	0	0	0	0	0	0	
	$Z_j - C_j$		−1	−1	−3	0	0	

↑

Since $Z_3 - C_3 = -3$ is the most negative, the variable x_3 enters the basis. The column corresponding to x_3 is called the *key column*.

To determine the key row or leaving variable, find $\text{Min}\left(\frac{x_B}{x_{i3}}, x_{i3} > 0\right)$ $\text{Min}\left(\frac{3}{1}, \frac{2}{2}\right) = 1$

Therefore, the leaving variable is the basic variable S_2, the row is called the *key row* and the intersection element 2 is called the *key element*.

Convert this element into one by dividing each element in the key row by 2 and the remaining elements in that key column as zero using the formula

$$\text{New element} = \text{Old element} - \left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}}\right]$$

First Iteration

		C_j	1	1	3	0	0
C_B	Basis	x_B	x_1	x_2	x_3	S_1	S_2
0	S_1	2	2	3/2	0	1	−1/2
3	x_3	1	1	1/2	1	0	1/2
	Z_j	3	3	3/2	3	0	3/2
	$Z_j - C_j$		2	1/2	0	0	3/2

Since all $Z_j - C_j \geq 0$, the solution is optimum and it is given by $x_1 = 0$, $x_2 = 0$, $x_3 = 1$, Max $Z = 3$.

Example 2.16: Use simplex method to solve the LPP.

$$\text{Min } Z = x_2 - 3x_3 + 2x_5$$

Subject to,

$$3x_2 - x_3 + 2x_5 \leq 7$$

$$-2x_2 + 4x_3 \leq 12$$

NOTES

$$-4x_2 + 3x_3 + 8x_5 \le 10$$

$$x_2, x_3, x_5 \ge 0$$

Solution: Since the given objective function is of minimization, we shall convert it into maximization using Min Z = –Max(–Z) = –Max Z^*

$$\text{Max } Z^* = -x_2 + 3x_3 - 2x_5$$

Subject to,

$$3x_2 - x_3 + 2x_5 \le 7$$

$$-2x_2 + 4x_3 \le 12$$

$$-4x_2 + 3x_3 + 8x_5 \le 10$$

We rewrite the inequality of the constraints into an equation by adding slack variables S_1, S_2, S_3 and the standard form of LPP becomes.

$$\text{Max } Z = -x_2 + 3x_3 - 2x_5 + 0S_1 + 0S_2 + 0S_3$$

Subject to,

$$3x_2 - x_3 + 2x_5 + S_1 = 7$$

$$-2x_2 + 4x_3 + S_2 = 12$$

$$-4x_2 + 3x_3 + 8x_5 + S_3 = 10$$

$$x_2, x_3, x_5, S_1, S_2, S_3 \ge 0$$

∴ The initial basic feasible solution is given by $S_1 = 7$, $S_2 = 12$, $S_3 = 10$. ($x_2 = x_3 = x_5 = 0$)

Initial Table

		C_j	–1	3	–2	0	0	0	
C_B	*Basis*	x_B	x_2	x_3	x_5	S_1	S_2	S_3	*Min* $\frac{x_B}{x_3}$
0	S_1	7	3	–1	2	1	0	0	—
←0	S_2	12	–2	④	0	0	1	0	12/4 = 3
0	S_3	10	–4	3	8	0	0	1	10/3 = 3.33
	Z_j	0	0	0	0	0	0		
	$Z_j - C_j$		1	–3	2	0	0	0	

↑

Since $Z_2 - C_2 = -3 < 0$, the solution is not optimum.

The incoming variable is x_3 (key column) and the outgoing variable (key row) is given by,

$$\text{Min}\left(\frac{x_B}{x_{i3}}, x_{i3} > 0\right) = \text{Min}\left(-, \frac{12}{4}, \frac{10}{3}\right) = 3.$$

Hence, S_2 leaves the basis.

NOTES

First Iteration

		C_j	−1	3	−2	0	0	0	
C_B	B	x_B	x_2	x_3	x_5	S_1	S_2	S_3	$Min \frac{x_B}{x_2}$
←0	S_1	10	(5/2)	0	2	1	1/4	0	$\frac{10}{5/2} = 4$
3	x_3	3	−1/2	1	0	0	1/4	0	—
0	S_3	1	5/2	0	8	0	−3/4	1	2/5
	Z_j	9	−3/2	3	0	0	3/4	0	
	$Z_j - C_j$		−1/2	0	2	0	3/4	0	

↑

Since $Z_1 - C_1 < 0$, the solution is not optimum. Improve the solution by allowing the variable x_2 to enter into the basis and the variable S_1 to leave the basis.

Second Iteration

		C_j	−1	3	−2	0	0	0
C_B	B	x_B	x_2	x_3	x_5	S_1	S_2	S_3
−1	x_2	4	1	0	4/5	2/5	1/10	0
3	x_3	5	0	1	2/5	1/5	3/10	0
0	S_3	11	0	0	10	1	−1/2	1
	Zj	11	−1	3	2/5	1/5	8/10	0
	$Z_j - C_j$		0	0	12/5	1/5	8/10	0

Since all $Z_j - C_j \geq 0$, the solution is optimum.

∴ The optimal solution is given by Max $Z^* = 11$

$$x_2 = 4, x_3 = 5, x_5 = 0$$

∴ Min $Z = -\text{Max}(-Z) = -11$

∴ Min $Z = -11, x_2 = 4, x_3 = 5, x_5 = 0.$

Example 2.17: Solve the following LPP using simplex method.

Max $Z = 15x_1 + 6x_2 + 9x_3 + 2x_4$

Subject to,

$$2x_1 + x_2 + 5x_3 + 6x_4 \leq 20$$
$$3x_1 + x_2 + 3x_3 + 25x_4 \leq 24$$
$$7x_1 + x_4 \leq 70$$
$$x_1, x_2, x_3, x_4 \geq 0$$

Solution: Rewriting the inequality of the constraint into an equation by adding slack variables S_1, S_2 and S_3, the standard form of LPP becomes.

Max $Z = 15x_1 + 6x_2 + 9x_3 + 2x_4 + 0S_1 + 0S_2 + 0S_3$

Subject to,

$$2x_1 + x_2 + 5x_3 + 6x_4 + S_1 = 20$$
$$3x_1 + x_2 + 3x_3 + 25x_4 + S_2 = 24$$
$$7x_1 + x_4 + S_3 = 70$$
$$x_1, x_2, x_3, x_4, S_1, S_2, S_3 \geq 0$$

NOTES

The initial basic feasible solution is $S_1 = 20$, $S_2 = 24$, $S_3 = 70$ ($x_1 = x_2 = x_3 = x_4 = 0$ non-basic)

The initial simplex table is given by:

		C_j	15	6	9	2	0	0	0	
C_B	Basis	x_B	x_1	x_2	x_3	x_4	S_1	S_2	S_3	$Min \frac{x_B}{x_1}$
0	S_1	20	2	1	5	6	1	0	0	20/2 = 10
←0	S_2	24	(3)	1	3	25	0	1	0	24/3 = 8←
0	S_3	70	7	0	0	1	0	0	1	70/7 = 10
	Z_j	0	0	0	0	0	0	0	0	
	$Z_j - C_j$		−15	−6	−9	−2	0	0	0	

↑

∴ As some of $Z_j - C_j \leq 0$ the current basic feasible solution is not optimum. $Z_1 - C_1 = -15$ is the most negative value and hence x_1 enters the basis and the variable S_2 leaves the basis.

First Iteration

		C_j	15	6	9	2	0	0	0	
C_B	Basis	x_B	x_1	x_2	x_3	x_4	S_1	S_2	S_3	$Min \frac{x_B}{x_2}$
←0	S_1	4	0	(1/3)	3	−32/3	1	−2/3	0	$\frac{4}{1/3} = 12$
15	x_1	8	1	1/3	1	25/3	0	1/3	0	$\frac{8}{1/3} = 24$
0	S_3	14	0	−7/3	−7	−172/3	0	−7/3	1	—
	Z_j	120	15	5	15	125	0	5	0	
	$Z_j - C_j$		0	−1	6	123	0	5	0	

↑

Since $Z_2 - C_2 = -1 < 0$ the solution is not optimal therefore, x_2 enters the basis and the basic variable S_1 leaves the basis.

Second Iteration

		C_j	−5	6	9	−2	0	0	0
C_B	B	x_B	x_1	x_2	x_3	x_4	S_1	S_2	S_3
6	x_2	12	0	1	9	−32	3	−2	0
15	x_1	4	1	0	−2	57/3	−1	1	0
0	S_3	42	0	0	14	−132	7	−7	1
	Z_j	132	15	6	24	93	3	3	0
	$Z_j - C_j$		0	0	15	91	3	3	0

Since all $Z_j - C_j \geq 0$, the solution is optimal and is given by,

Max Z = 132, $x_1 = 4$, $x_2 = 12$, $x_3 = 0$, $x_4 = 0$.

Check Your Progress

9. When is an objective function minimized? When is it maximized?
10. What is meant by a feasible solution?
11. What is a feasible region?
12. What is an optimal solution?
13. What are non-degenerate and degenerate type basic feasible solutions?
14. What is the simplex method?
15. How is a leaving element converted to unity in a simplex algorithm?
16. What is the role of the slack variable?

2.7 ARTIFICIAL VARIABLE TECHNIQUES

This section will consider the LPP in which constraints may also have the ≥ and = signs after ensuring that all $b_i \geq 0$. In such cases, the basis matrix cannot be obtained

NOTES

as an identify matrix in the starting simplex table, therefore we introduce a new type of variable called the *artificial variable*. These variables are fictitious and cannot have any physical meaning. The artificial variable technique is merely a device to get the starting basic feasible solution so that the simplex procedure may be adopted as usual until the optimal solution is obtained. To solve such LPP, there are following two methods:

(*i*) The Big *M* Method or the Method of Penalties

(*ii*) The Two-Phase Simplex Method

2.7.1 The Big *M* Method

The following steps are involved in solving an LPP using the Big *M* method.

Step 1: Express the problem in the standard form.

Step 2: Add non-negative artificial variables to the left side of each of the equations corresponding to constraints of the type ≥ or =. However, the addition of these artificial variables causes a violation of the corresponding constraints. Therefore, we would like to get rid of these variables and not allow them to appear in the final solution. This is achieved by assigning a very large penalty ($-M$ for maximization and M for minimization) in the objective function.

Step 3: Solve the modified LPP by the simplex method, until one of the following three cases arises.

1. If no artificial variable appears in the basis and the optimality conditions are satisfied, then the current solution is an optimal basic feasible solution.
2. If at least one artificial variable appears in the basis at the zero level and the optimality condition is satisfied, then the current solution is an optimal basic feasible solution (though a degenerated solution).
3. If at least one artificial variable appears in the basis at the positive level and the optimality condition is satisfied, then the original problem has no feasible solution. The solution satisfies the constraints but does not optimize the objective function since it contains a very large penalty M, and is called a pseudo-optimal solution.

Note: While applying the simplex method, whenever an artificial variable happens to leave the basis, we drop that artificial variable and omit all the entries corresponding to its column from the simplex table.

Example 2.18: Use the penalty method to

Maximize $Z = 3x_1 + 2x_2$

Subject to the constraints,

$$2x_1 + x_2 \leq 2$$

$$3x_1 + 4x_2 \geq 12$$

$$x_1, x_2 \geq 0$$

Solution: By introducing the slack variable $S_1 \geq 0$, the surplus variable $S_2 \geq 0$ and the artificial variable $A_1 \geq 0$, the given LPP can be reformulated as:

Maximize $Z = 3x_1 + 2x_2 + 0S_1 + 0S_2 - MA_1$,

Subject to $2x_1 + x_2 + S_1 = 2$

$$3x_1 + 4x_2 - S_2 + A_1 = 12$$

The starting feasible solution is $S_1 = 2$, $A_1 = 12$.

Initial Table

NOTES

		C_j	3	2	0	0	$-M$	
C_B	B	x_B	x_1	x_2	S_1	S_2	A_1	$Min\ x_B/x_2$
← 0	S_1	2	2	①	1	0	0	2/1 = 2
$-M$	A_1	12	3	4	0	−1	1	12/4 = 3
	Z_j	$-12M$	$-3M$	$-4M$	0	M	$-M$	
	$Z_j - C_j$	–	$-3M-3$	$-4M-2$ ↑	0	M	0	

Since some of $Z_j - C_j \leq 0$, the current feasible solution is not optimum. Choose the most negative $Z_j - C_j = -4M-2$.

∴ the x_2 variable enters the basis and the basic variable S_1 leaves the basis.

		C_j	3	2	0	0	$-M$
C_B	B	x_B	x_1	x_2	S_1	S_2	A_1
2	x_1	2	2	1	1	0	0
$-M$	A_1	4	−5	0	−4	−1	1
	Z_j	$4-4M$	$4+5M$	2	$2+4M$	M	$-M$
	$Z_j - C_j$		$5M+1$	0	$4M+2$	M	0

Since all $Z_j - C_j \geq 0$ and an artificial variable appears in the basis at the positive level, the given LPP does not possess any feasible solution. But the LPP possesses a pseudo-optimal solution.

Example 2.19: Solve the following LPP.

Minimize $Z = 4x_1 + x_2$

Subject to $3x_1 + x_2 = 3$

$$4x_1 + 3x_2 \geq 6$$

$$x_1 + 2x_2 \leq 3$$

$$x_1, x_2 \geq 0$$

Solution: Since the objective function is minimization, we covert it into maximization using,

Min Z = –Max $(-z)$

Maximize $Z = -4x_1 - x_2$

Subject to $3x_1 + x_2 = 3$

NOTES

$$4x_1 + 3x_2 \geq 6$$

$$x_1 + 2x_2 \leq 3$$

$$x_1, x_2 \geq 0$$

Convert the given LPP into the standard form by adding the artificial variables A_1, A_2, the surplus variable S_1 and the slack variable S_2 to get the initial basic feasible solution.

Maximize $Z = -4x_1 - x_2 + 0S_1 + 0S_2 - MA_1 - MA_2$

Subject to $3x_1 + x_2 + A_1 = 3$

$$4x_1 + 3x_2 - S_1 + A_2 = 6$$

$$x_1 + 2x_2 + S_2 = 3$$

The starting feasible solution is $A_1 = 3, A_2 = 6, S_2 = 3$.

Initial Solution

		C_j	–4	–1	$-M$	0	$-M$	0	
C_B	B	x_B	x_1	x_2	A_1	S_1	A_2	S_2	$Min \frac{x_B}{x_2}$
$-M$	A_1	3	3	1	1	0	0	0	3/3 = 1
$-M$	A_2	6	4	3	0	–1	1	0	6/4 = 1.5
← 0	S_2	3	1	(2)	0	0	0	1	3/1 = 3
	Z_j	$-9M$	$-7M$	$-4M$	$-M$	M	$-M$	0	
	$Z_j - C_j$		$-7M + 4$	$-4M + 1$↑	0	M	0	0	

Since some of $Z_j - C_j \geq 0$, the current feasible solution is not optimum. As $Z_1 - C_1$ is the most negative, x_1 enters the basis and the basic variable A_2 leaves the basis.

First Iteration

		C_j	–4	–1	$-M$	0	$-M$	0	
C_B	B	x_B	x_1	x_2	A_1	S_1	A_2	S_2	$Min \frac{x_B}{x_1}$
$-M$	A_1	3/2	5/2	0	1	0	0	–1/2	3/5
← $-M$	A_2	3/2	(5/2)	0	0	–1	1	–3/2	3/5
–1	x_2	3/2	3/2	1	0	0	0	1.2	3
	Z_j	$-3M-3/2$	$-5M-1/2$	–1	$-M$	$+M$	$-M$	$2M-1/2$	
	$Z_j - C_j$		$-5M-7/2$↑	0	0	M	0	$2M-1/2$	

Since $Z_1 - C_1$ is negative, the current feasible solution is not optimum. Therefore, the x_1 variable enters the basis and the artificial variable A_2 leaves the basis.

Second Iteration

		C_j	−4	−1	−M	0	0	
C_B	B	x_B	x_1	x_2	A_1	S_1	S_2	$Min\frac{x_B}{x_1}$
−M	A_1	0	0	0	1	(1)	1	0
← −M	x_1	$\frac{3}{5}$	1	0	0	$-\frac{2}{5}$	$-\frac{3}{5}$	–
−1	x_2	$\frac{6}{5}$	0	1	0	$-\frac{1}{5}$	$\frac{4}{5}$	–
	Z_j	$-\frac{18}{5}$	−4	−1	−M	$-M+\frac{9}{5}$		
	$Z_j - C_j$		0	0	0	$-M+\frac{9}{5}\uparrow$		

Since $Z_4 - C_4$ is the most negative, S_1 enters the basis and the artificial variable A_1 leaves the basis.

Third Iteration

		C_j	−4	−1	0	0	
C_B	B	x_B	x_1	x_2	S_1	S_2	$Min\frac{x_B}{x_2}$
← 0	S_1	0	0	0	1	(1)	0
−4	x_1	$\frac{3}{5}$	1	0	0	$-\frac{1}{5}$	–
−1	x_2	6/5	0	1	0	1	6/5
	Z_j	$\frac{18}{5}$	−4	−1	0	−1/5	
	$Z_j - C_j$		0	0	0	−1/5↑	

Since $Z_4 - C_1$ is the most negative, S_2 enters the basis and S_1 leaves the basis.

Fourth Iteration

		C_j	−4	−1	0	0
C_B	B	x_B	x_1	x_2	S_1	S_2
0	S_2	0	0	0	1	1
−4	x_1	3/5	1	0	$\frac{1}{5}$	0
−1	x_2	6/5	0	1	−1	0
	Z_j	−18/5	−4	−1	1/5	0
	$Z_j - C_j$		0	0	1/5	0

Since all $Z_j - C_j \geq 0$, the solution is optimum and is given by $x_1 = 3/5$, $x_2 = 6/5$ and Max Z = −18/5.

∴ Min Z = −Max (−Z) = 18/5.

Example 2.20: Solve by Big M method.

Maximize $Z = x_1 + 2x_2 + 3x_3 - x_4$

Subject to, $x_1 + 2x_2 + 3x_3 = 15$

$2x_1 + x_2 + 5x_3 = 20$

$x_1 + 2x_2 + x_3 + x_4 = 10$

Solution: Since the constraints are equations, introduce artificial variables $A_1, A_2 \geq 0$. The reformulated problem is given as follows.

Maximize $Z = x_1 + 2x_2 + 3x_3 - x_4 - MA_1 - MA_2$

Subject to, $x_1 + 2x_2 + 3x_3 + A_1 = 15$

$2x_1 + x_2 + 5x_3 + A_2 = 20$

$x_1 + 2x_2 + x_3 + x_4 = 10$

Initial solution is given by $A_1 = 15$, $A_2 = 20$ and $x_4 = 10$.

	C_j	1	2	3	−1	$-M$	$-M$	
C_B	B	x_B	x_1	x_2	x_3	x_4	A_1	A_2	$Min \frac{x_B}{x_3}$
$-M$	A_1	15	1	2	3	0	1	0	15/3 = 5
← $-M$	A_2	20	2	1	(5)	0	0	1	20/5 = 4 ←
−1	x_4	10	1	2	1	1	0	0	10/1 = 10
	Z_j	$-35M$	$-3M$	$-3M$	$-8M$	−1	$-M$	$-M$	
		−10	− 1	− 2	− 1				
	$Z_j - C_j$		$-3M - 2$	$-3M - 4$	$-8M - 4$	0	0	0	

↑

Since $Z_3 - C_3$ is most negative, x_3 enters the basis and the basic variable A_2 leaves the basis.

First Iteration

	C_j	1	2	3	−1	$-M$	
C_B	B	x_B	x_1	x_2	x_3	x_4	A_1	$Min \frac{x_B}{x_2}$
← $-M$	A_1	3	−1/5	(7/5)	0	0	1	$\frac{3}{7/5} = \frac{15}{7}$
3	x_3	4	2/5	1/5	1	0	0	$\frac{4}{1/5} = \frac{20}{1}$
−1	x_4	6	3/5	9/5	0	1	0	$\frac{6}{9/5} = \frac{30}{9}$
	Z_j	$-3M + 6$	$\frac{1}{5}M + \frac{3}{5}$	$-\frac{7}{5}M - \frac{6}{5}$	3	−1	$-M$	
	$Z_j - C_j$		$\frac{1}{5}M - \frac{2}{5}$	$-\frac{7}{5}M - \frac{16}{5}$	0	0	0	

↑

Since $Z_2 - C_2$ is most negative, x_2 enters the basis and the basic variable A_1 leaves the basis.

NOTES

Second Iteration

		C_j	1	2	3	−1	
C_B	B	x_B	x_1	x_2	x_3	x_4	$Min \frac{x_B}{x_1}$
2	x_2	15/7	$-\frac{1}{7}$	1	0	0	—
3	x_3	25/7	3/7	0	1	0	25/3
← −1	x_4	15/7	$\left(\frac{6}{7}\right)$	0	0	1	15/6
	Z_j	$\frac{90}{7}$	$\frac{1}{7}$	2	3	−1	
	$Z_j - C_j$		−6/7	0	0	0	

↑

Since $Z_1 - C_1 = -6/7$ is negative, the current feasible solution is not optimum. Therefore, x_1 enters the basis and the basic variable x_4 leaves the basis.

Third Iteration

		C_j	1	2	3	−1
C_B	B	x_B	x_1	x_2	x_3	x_4
2	x_2	15/6	0	1	0	1/6
3	x_3	15/6	0	0	1	3/6
3	x_1	15/6	1	0	0	7/6
	Z_j	15	1	2	3	3
	$Z_j - C_j$		0	0	0	4

Since all $Z_j - C_j \geq 0$, the solution is optimum and is given by $x_1 = x_2 = x_3 = 15/6 = 5/2$, and Max $Z = 15$.

Example 2.21: Use penalty method to solve the following LPP:

Minimize $Z = 5x + 3y$

Subject to,

$$2x + 4y \leq 12$$
$$2x + 2y = 10$$
$$5x + 2y \geq 10$$
$$x, y \geq 0$$

Solution: First we convert the objective function from minimization to maximization using

Min $Z = -\text{Max}(-Z)$.

Rewrite the given LPP into standard form by adding slack variable $S_1 \geq 0$, surplus variable $S_2 \geq 0$ and the artificial variables $A_1, A_2 \geq 0$.

Maximize $Z = -5x - 3y + 0S_1 + 0S_2 - MA_1 - MA_2$

Subject to,

$$2x + 4y + S_1 = 12$$
$$2x + 2y + A_1 = 10$$

$$5x + 2y - S_2 + A_2 = 10$$
$$x, y, S_1, S_2, A_1, A_2 \geq 0$$

Initial feasible solution is given by $S_1 = 12$, $A_1 = 10$ and $A_2 = 10$

Since all $Z_j - C_j \geq 0$ (see Table below) and no artificial variable is in the basis, the solution is optimum and is given by:

$x = 4, y = 1$, Max $Z = -23$

Min $Z = -$Max $(-Z) = 23$

NOTES

Table

		C_j	-5	-3	0	$-M$	0	$-M$	
C_B	B	x_B	x	y	S_1	A_1	S_2	A_2	$Min \frac{x_B}{x}$
0	S_1	12	2	4	1	0	0	0	12/2 = 6
$-M$	A_1	10	2	2	0	1	0	0	10/2 = 5
$\leftarrow -M$	A_2	10	(5)	2	0	0	-1	1	10/5 = 2
	Z_j	$10M$	$-7M$	$-4M$	0	$-M$	M	$-M$	
	$Z_j - C_j$		$-7M + 5\uparrow$	$-4M + 3$	0	0	M	0	$Min \frac{x_B}{y}$
$\leftarrow 0$	S_1	8	0	(16/5)	1	0	2/5		$\frac{8 \times 5}{16} = 5/2$
$-M$	A_1	6	0	6/5	0	1	2/5	—	30/6 = 5
-5	x	2	1	2/5	0	0	$-1/5$	—	10/2 = 5

$\uparrow$

		C_j	-5	-3	0	$-M$	0	$-M$	
C_B	B	x_B	x	y	S_1	A_1	S_2	A_2	$Min \frac{x_B}{x}$
	Z_j	$-6M - 10$	-5	$-6/5\,M - 2$	0	$-M$	$-2/5\,M + 1$	—	
	$Z_j - C_j$		0	$-6/5\,M + 1\uparrow$	0	0	$-\frac{2}{5}M + 1$		$Min \frac{x_B}{S_2}$
-3	y	5/2	0	1	5/16	0	1/8	—	20
$\leftarrow -M$	A_1	3	0	0	$-3/8$	1	$-1/4$	—	12
-5	x	1	1	0	$-1/8$	0	$-1/4$	—	—
	Z_j	$-3M - 25/2$	-5	-3	$-3/8M$ $-5/16$	$-M$ 0	$* + 7/8$ $-1/4M$	—	
	$Z_j - C_j$		0	0	$3/8M$ $-5/16$	0	$7/8$ $-M/4\uparrow$	—	
-3	y	1	0	1	1/2	—	0	—	
0	S_2	12	0	0	$-3/2$	—	1	—	
-5	x	4	1	0	$-1/2$	—	0	—	
	Z_j	-23	-5	-3	1	—	0	—	
	$Z_j - C_j$		0	0	1	—	0	—	

NOTES

2.7.2 The Two-Phase Simplex Method

The two-phase simplex method is another method to solve a given LPP involving some artificial variables. The solution is obtained in two phases.

Phase I

In this phase, we construct an auxiliary LPP leading to a final simplex table containing a basic feasible solution to the original problem.

Step 1: Assign a cost –1 to each artificial variable and a cost 0 to all other variables and get a new objective function $Z^* = -A_1 - A_2 - A_3 ...$ where A_i are artificial variables.

Step 2: Write down the auxiliary LPP in which the new objective function is to be maximized, subject to the given set of constraints.

Step 3: Solve the auxiliary LPP by the simplex method until either of the following three cases arise:

(*i*) Max $Z^* < 0$ and at least one artificial variable appears in the optimum basis at the positive level.

(*ii*) Max $Z^* = 0$ and at least one artificial variable appears in the optimum basis at the zero level.

(*iii*) Max $Z^* = 0$ and no artificial variable appears in the optimum basis.

In case (i), the given LPP does not possess any feasible solution, whereas in cases (ii) and (iii), we go to phase II.

Phase II

Use the optimum basic feasible solution of phase I as a starting solution for the original LPP. Assign the actual costs to the variable in the objective function and a zero cost to every artificial variable in the basis at the zero level. Delete the artificial variable column that is eliminated from the basis in phase I from the table. Apply the simplex method to the modified simplex table obtained at the end of phase I till an optimum basic feasible solution is obtained or till there is an indication of an unbounded solution.

Since all $Z_j - C_j \geq 0$, an optimum basic feasible solution to the auxiliary LPP has been attained. But since max Z^* is negative and the artificial variable A_1 appears in the basis at the positive level, the original problem does not possess any feasible solution.

Example 2.22: Use the two-phase simple method to solve:

Maximize $Z = 5x_1 - 4x_2 + 3x_3$

Subject to $2x_1 + x_2 - 6x_3 = 20$

$6x_1 + 5x_2 + 10x_3 \leq 76$

$8x_1 - 3x_2 + 6x_3 \leq 50$

$x_1, x_2, x_3 \geq 0$

Solution: Introducing the slack variables $S_1, S_2 \geq 0$ and an artificial variable $A_1 \geq 0$ in the constraints of the given LPP, the problem is reformulated in the standard form.

The initial basic feasible solution is given by $A_1 = 20$, $S_1 = 76$ and $S_2 = 50$.

Phase I

Assigning a cost –1 to the artificial variable A_1 and a cost 0 to the other variables, the objective function of the auxiliary LPP becomes,

NOTES

Maximize $Z^* = 0x_1 + 0x_2 + 0x_3 + 0S_1 + 0S_2 - 1A_1$

Subject to $2x_1 + x_2 - 6x_3 + A_1 = 20$

$6x_1 + 5x_2 + 10x_3 + S_1 = 76$

$8x_1 - 3x_2 + 6x_3 + S_2 = 50$

$x_1, x_2, x_3, S_1, S_2, A_1 \geq 0$

	C_j		0	0	0	−1	0	0	
	C_B	B	x_B	x_1	x_2	x_3	A_1	S_1	S_2	$Min \frac{x_B}{x_1}$
-	1	A_1	20	2	1	−6	1	0	0	20/2 = 10
	0	S_1	76	6	5	10	0	1	0	76/6 = 12.66
←	0	S_2	50	8	−3	6	0	0	1	50/8 = 6.25
		Z_j	−20	−2	−1	6	−1	0	0	
		$Z_j - C_j$		−2↑	−1	6	0	0	0	

	C_j		0	0	0	−1	0	0	
	C_B	B	x_B	x_1	x_2	x_3	A_1	S_1	S_2	$Min \frac{x_B}{x_2}$
←	−1	A_1	15/2	0	7/4	−15/2	1	0	−1/4	30/7
	0	S_1	77/2	0	29/4	11/2	0	1	−3/4	154/29
	0	x_1	25/4	1	−3/8	3/4	0	0	1/8	—
		Z_j	−15/2	0	−7/4	15/2	−1	0	1/4	
		$Z_j - C_j$		0	−7/4↑	15/2	0	0	1/4	
	0	x_2	30/7	0	1	−30/7	4/7	0	−1/7	
	0	S_1	52/7	0	1	256/7	−29/7	1	2/7	
	0	x_1	55/7	1	0	−6/7	3/4	0	1/14	
		Z_j	0	0	0	0	0	0	0	
		$Z_j - C_j$		0	0	0	0	1	0	

Since all $Z_j = C_j \geq 0$, an optimum solution to the auxiliary LPP has been obtained. Also, Max $Z^* = 0$ with no artificial variable in the basis. We now go to phase II.

Phase II

Consider the final simplex table of phase I. Consider the actual cost associated with the original variables. Delete the artificial variable A_1 column from the table as it is eliminated in phase I.

	C_j		5	−4	3	0	0
C_B	B	x_B	x_1	x_2	x_3	S_1	S_2
−4	x_2	30/7	0	1	−30/7	0	−1/7
0	S_1	52/7	0	0	256/7	1	2/7
5	x_1	55/7	1	0	−6/7	0	1/14
	Z_j	155/7	5	−4	90/7	0	13/4
	$Z_j - C_j$	0	0	0	69/7	0	13/14

Since all $Z_j - C_j \geq 0$, an optimum basic feasible solution has been reached. Hence, an optimum feasible solution to the given LPP is $x_1 = 55/7$, $x_2 = 30/7$, $x_3 = 0$ and Max $Z = 155/7$.

NOTES

2.8 DUALITY

For every given linear programming problem, there is another intimately related linear programming problem referred to as its dual. The duality theorem states that 'for every maximization (or minimization) problem in linear programming, there is a unique similar problem of minimization (or maximization) involving the same date which describes the original problem'. The original problem is referred to as the 'primal'. The 'dual' of a dual problem is the primal. Thus the primal and dual problems are replicas of each other. Further, the maximum feasible value of the primal objective function equals to the minimum feasible value of the dual objective function. This means that the solutions of the primal and the dual problems are related which infact yields several advantages.

The transformation of a given primal problem into a dual problem involves the following considerations:

(1) If the objective of the primal is maximization, the objective of the dual is minimization.

(2) The primal has m-constraints while its dual has m-unknowns.

(3) The primal has n-unknowns while its dual has n-constraints.

(4) The n-coefficients of the objective function of primal (C_j) become the n-constant terms (b_j) of its dual.

(5) The m-constant terms of the primal (b_i) become the m-constant terms of the objective function (C_j) of its dual.

(6) The coefficients of the variables of the primal are transformed in their position in the dual. This means that the first column of the coefficients in the primal becomes the first row in the dual, the second column becomes the second row and so on.

(7) The n-variables (X_n) of the primal are replaced by the m new variables (Y_m) of its dual. This change affects the system of restrictions as well as the objective function.

(8) The sign of the inequalities in the set of restrictions of the primal ($\leq$) is reversed in the set of restrictions in its dual ($\geq$). In other words, if the inequalities in the primal are of the type $\leq$, then, they are of $\geq$ type in the dual.

(9) The sign of the inequalities restricting the variable ($\geq X_j$) to non-negative values in the primal is equal to the inequality sign of the new variable ($\geq Y_j$)of its dual.

(10) For writing the dual of the given maximization problem, we should first ensure that all the constraint inequalities are of the $\leq$ type and for writing the dual of the given minimization problem, the constraint inequalities should be of the $\geq$ type. We can see the application of these considerations with the help of given examples.

2.8.1 Duality in Linear Programming

Every LPP (called the primal) is associated with another LPP (called its dual). Either of the problem can be considered as primal with the other as dual.

The importance of the duality concept is due to two main reasons:

(*i*) If the primal contains a large number of constraints and a smaller number of variables, the labour of computation can be considerably reduced by converting it in to the dual problem and then solving it. (*ii*) The interpretation of the dual variables

from the cost or economic point of view, proves extremely useful in making future decisions in the activities being programmed.

NOTES

Formulation of Dual Problems

For formulating a dual problem, we first write the problem in the canonical form. The following changes are used in formulating the dual problem:

(1) Change the objective function of maximization in the primal into minimization in the dual, and vice versa.

(2) The variables in the primal should be equal to the constraints in the dual and vice versa.

(3) The cost coefficients $C_1, C_2 \ldots C_n$ in the objective function of the primal should be the RHS constant of the constraints in the dual and vice versa.

(4) In forming the constraints for the dual, we consider the transpose of the body matrix of the primal problem.

(5) The variables in both the problems are non-negative.

(6) If the variable in the primal is unrestricted in sign, then the corresponding constraint in the dual will be an equation, and vice versa.

Definition of the Dual Problem

A dual problem refers to a linear program in which the objective function is a linear combination of m values that are the limits in the m constraints of the primal problem.

Let the primal problem be:

$$\text{Max } Z = C_1x_1 + C_2x_2 + \ldots + C_nx_n$$

Subject to,

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1$$
$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2$$
$$\vdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m$$
$$x_1, x_2 \cdots x_n \ge 0$$

Dual: The dual problem is defined as,

$$\text{Min } Z' = b_1w_1 + b_2w_2 + \ldots + b_mw_m$$

Subject to

$$a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m \ge C_1$$
$$a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m \ge C_2$$
$$\vdots$$
$$a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_n \ge C_n$$
$$w_1, w_2 \cdots w_m \ge 0$$

where $w_1, w_2, w_3 \ldots w_m$ are called dual variables.

Example 2.23: Write the dual of the primal LP problem given as follows:

$$\text{Max } Z = x_1 + 2x_2 + x_3$$

Subject to,

$$2x_1 + x_2 - x_3 \le 2$$
$$-2x_1 + x_2 - 5x_3 \ge -6$$

NOTES

$$4x_1 + x_2 + x_3 \leq 6$$
$$x_1, x_2, x_3 \geq 0$$

Solution: Since the problem is not in the canonical form, we interchange the inequality of the second constraint.

Max $Z = x_1 + 2x_2 + x_3$

Subject to, $2x_1 + x_2 - x_3 \leq 2$
$2x_1 - x_2 + 5x_3 \leq 6$
$4x_1 + x_2 + x_3 \leq 6$
and $x_1, x_2, x_3 \geq 0$

Dual: Let w_1, w_2, w_3 be the dual variables.

Min $Z' = 2w_1 + 6w_2 + 6w_3$

Subject to, $2w_1 + 2w_2 + 4w_3 \geq 1$
$+w_1 - w_2 + w_3 \geq 2$
$-w_1 + 5w_2 + w_3 \geq 1$
$w_1, w_2, w_3 \geq 0$

Example 2.24: Write the dual of the following LPP.

Min $Z = 2x_2 + 5x_3$

Subject to, $x_1 + x_2 \geq 2$
$2x_1 + x_2 + 6x_3 \leq 6$
$x_1 - x_2 + 3x_3 = 4$
$x_1, x_2, x_3 \geq 0$

Solution: Since the given primal problem is not in the canonical form, we interchange the inequality of the constraint. Also, the third constraint is an equation. This equation can be converted into two inequations.

Min $Z = 0x_1 + 2x_2 + 5x_3$

Subject to, $x_1 + x_2 + 0x_3 \geq 2$
$-2x_1 - x_2 - 6x_3 \geq 6$
$,x_1 - x_2 + 3x_3 \leq 4$
$x_1 - x_2 + 3x_3 \geq 4$
$x_1, x_2, x_3 \geq 0$

Again, on rearranging the constraint, we have,

Min $Z = 0x_1 + 2x_2 + 5x_3$

Subject to,
$x_1 + x_2 + 0x_3 \geq 2$
$-2x_1 - x_2 - 6x_3 \geq -6$

$$x_1 - x_2 + 3x_3 \geq 4$$
$$-x_1 + x_2 - 3x_3 \geq -4$$
$$x_1, x_2, x_3 \geq 0$$

NOTES

Dual: Since there are four constraints in the primal, we have four dual variables, namely w_1, w_2, w_3', w_3''.

Max $Z' = 2w_1 - 6w_2 + 4w_3' - 4w_3''$

Subject to
$$w_1 - 2w_2 + w_3' - w_3'' \leq 0$$
$$w_1 - w_2 - w_3' + w_3'' \leq 2$$
$$0w_1 - 6w_2 + 3w_3' - 3w_3'' \leq 5$$
$$w_1, w_2, w_3', w_3'' \geq 0$$

Let $w_3 = w_3' - w_3''$

Max $Z' = 2w_1 - 6w_2 + 4(w_3' - w_3'')$

Subject to
$$w_1 - 2w_2 + (w_3' - w_3'') \leq 0$$
$$w_1 - w_2 - (w_3' - w_3'') \leq 2$$

Finally, we have, $0w_1 - 6w_2 + 3(w_3' - w_3'') \leq 5$

Max $Z' = 2w_1 - 6w_2 + 4w_3$

Subject to
$$w_1 - 2w_2 + w_3 \leq 0$$
$$w_1 - w_2 - w_3 \leq 2$$
$$0w_1 - 6w_2 + 3w_3 \leq 5$$
$$w_1, w_2 \geq 0, w_3 \text{ is unrestricted.}$$

Example 2.25: Find the dual of the LPP given as follows:

Max $Z = 3x_1 - x_2 + x_3$

Subject to
$$4x_1 - x_2 \leq 8$$
$$8x_1 + x_2 + 3x_3 \geq 12$$
$$5x_1 - 6x_3 \leq 13$$
$$x_1, x_2, x_3 \geq 0$$

Solution: Since the problem is not in the canonical form, we interchange the inequality of the second constraint.

Max $Z = 3x_1 - x_2 + x_3$

Subject to
$$4x_1 - x_2 + 0x_3 \leq 8$$
$$-8x_1 - x_2 - 3x_3 \leq -12$$
$$5x_1 + 0x_2 - 6x_3 \leq 13$$
$$x_1, x_2, x_3 \geq 0$$

NOTES

Max $Z = Cx$

Subject to $Ax \le B$

$x \ge 0$

$$C = (3 \; -11) \times \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \; b = \begin{pmatrix} 8 \\ -12 \\ 13 \end{pmatrix}$$

$$A = \begin{pmatrix} 4 & -1 & 0 \\ -8 & -1 & -3 \\ 5 & 0 & -6 \end{pmatrix}$$

Dual: Let w_1, w_2, w_3 be the dual variables. The dual problem is:

Min $Z' = b^T W$

Subject to $A^T W \ge C^T$ and $W \ge 0$

i.e., Min $Z' = (8 \quad -12 \quad 13) \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$

Subject to $\begin{pmatrix} 4 & -8 & 5 \\ -1 & -1 & 0 \\ 0 & -3 & 6 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \ge \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

Min $Z_1 = 8w_1 - 12w_2 + 13w_3$

Subject to $4w_1 - 8w_2 + 5w_3 \ge 3$

$-w_1 - w_2 + 0w_3 \ge -1$

$0w_1 - 3w_2 + 6w_3 \ge 1$

$w_1, w_2, w_3 \ge 0$

Example 2.26: Give the dual of the problem given as follows:

Max $Z = x + 2y$

Subject to $2x + 3y \ge 4$

$3x + 4y = 5$

$x \ge 0$ and y unrestricted.

Solution: Since the variable y is unrestricted, it can be expressed as $y = y' - y''$, $y', y'' \ge 0$. On reformulating the given problem, we have,

Max $Z = x + 2(y' - y'')$

Subject to $-2x - 3(y' - y'') \le -4$

$3x + 4(y' - y'') \le 5$

$$3x + 4(y' - y'') \geq 5$$
$$x, y', y'' \geq 0$$

Since the problem is not in the canonical form, we rearrange the constraints.

NOTES

Max $Z = x + 2y' - 2y''$

Subject to
$$-2x - 3y' + 3y'' \leq -4$$
$$3x + 4y' - 4y'' \leq 5$$
$$-3x - 4y' + 4y'' \leq -5$$

Dual: Since there are three variables and three constraints, in the dual we have three variables namely, w_1, w_2', w_2''.

Min $Z' = -4w_1 + 5w_2' - 5w_2''$

Subject to
$$-2w_1 - 3w_2' - 3w_2'' \geq 1$$
$$-3w_1 + 4w_2' - 4w_2'' \geq 2$$
$$3w_1 - 4w_2' + 4w_2'' \geq -2$$
$$w_1, w_2', w_2'' \geq 0$$

Let $w_2 = w_2' - w_2''$, so that the dual variable w_2 is unrestricted in sign. Hence, the dual is:

Min $Z' = -4w_1 + 5(w_2' - w_2'')$

Subject to
$$-2w_1 + 3(w_2' - w_2'') \geq 1$$
$$-3w_1 + 4(w_2' - w'') \geq 2$$
$$3w_1 + 4(w_2' + w'') \geq -2$$

i.e., Min $Z' = -4w_1 + 5w_2$

Subject to
$$-2w_1 + 3w_2 \geq 1$$
$$-3w_1 + 4w_2 \geq 2$$
$$3w_1 - 4w_2 \geq -2$$

$w_1 \geq 0$ and w_2 is unrestricted.

i.e., Min $Z' = -4w_1 + 5w_2$

Subject to
$$-2w_1 + 3w_2 \geq 1$$
$$-3w_1 + 4w_2 \geq 2$$
$$-3w_1 + 4w_2 \geq 2$$

i.e., Min $Z' = -4w_1 + 5w_2$

Subject to $-2w_1 + 3w_2 \geq 1$

$-3w_1 + 4w_2 = 2$, $w_1 \geq 0$ and w_2 is unrestricted.

Example 2.27: Write the dual of the following primal LPP.

Min $Z = 4x_1 + 5x_2 - 3x_3$

NOTES

Subject to $x_1 + x_2 + x_3 = 22$

$3x_1 + 5x_2 - 2x_3 \leq 65$

$x_1 + 7x_2 + 4x_3 \geq 120$

$x_1 + x_2 \geq 0$ and x_3 is unrestricted.

Solution: Since the variable x_3 is unrestricted, $x_3 = x_3' - x_3''$. Also, bring the problem into canonical form by rearranging the constraints.

$\text{Min } Z = 4x_1 + 5x_2 - 3(x_3' - x_3'')$

Subject to $x_1 + x_2 + (x_3' - x_3'') \leq 22$

$x_1 + x_2 + x_3' - x_3'' \geq 22$

$-3x_1 - 5x_2 + 2(x_3' - x_3'') \geq -65$

$x_1 + 7x_2 + 4(x_3' - x_3'') \geq 120$

$x_1, x_2, x_3' - x_3'' \geq 0$

$\text{Min } Z = 4x_1 + 5x_2 - 3x_3' + 3x_3''$

Subject to $x_1 + x_2 + x_3' - x_3'' \geq 22$

$-x_1 - x_2 - x_3' + x_3'' \geq -22$

$-3x_1 - 5x_2 + 2x_3' - 2x_3'' \geq -65$

$x_1 + 7x_2 + 4x_3' - 4x_3'' \geq 120$

$x_1, x_2, x_3' - x_3'' \geq 0$

Dual: Since there are four constraints in the primal problem, in the dual there are four variables, namely w_1', w_1'', w_2, w_3 so that the dual is given by:

$\text{Max } Z' = 22(w_1' - w_1'') - 65w_2 + 120w_3$

Subject to $w_1' - w_1'' - 3w_2 + w_3 \leq 4$

$w_1' - w_1'' - 5w_2 + 7w_3 \leq 5$

$w_1' - w_1'' + 2w_2 + 4w_3 \leq -3$

$-w_1' + w_1'' - 2w_2 - 4w_3 \leq 3$

$w_1', w_1'', w_2, w_3 \geq 0$

Let $w_1 = w_1' - w_1''$, i.e., the variable w_1 is unrestricted.

i.e., $\text{Max } Z' = 22(w_1' - w_1'') - 65w_2 + 120w_3$

Subject to $w_1' - w_1'' - 3w_2 + w_3 \leq 4$

$w_1' - w_1'' - 5w_2 + 7w_3 \leq 5$

$-(w_1' - w_1'') - 2w_2 - 4w_3 \geq 3$

$-(w_1' - w_1'') - 2w_2 - 4w_3 \leq 3$

NOTES

i.e., $\text{Max } Z' = 22w_1 - 65w_2 + 120w_3$

Subject to
$$w_1 - 3w_2 + w_3 \le 4$$
$$w_1 - 5w_2 + 7w_3 \le 5$$
$$-w_1 - 2w_2 - 4w_3 \ge 3$$
$$-w_1 - 2w_2 - 4w_3 \le 3$$

Thus, we have,

$\text{Min } Z' = 22w_1 - 65w_2 + 120w_3$

Subject to
$$w_1 - 3w_2 + 4w_3 \le 4$$
$$w_1 - 5w_2 + 7w_2 \le 5$$
$$-w_1 - 2w_2 - 4w_3 = 3$$

$w_2, w_3 \ge 0$ and w_1 is unrestricted.

Important Results in Duality

1. The dual of the dual is primal.
2. If one is a maximization problem, then the other is a minimization one.
3. The necessary and sufficient condition for any LPP and its dual to have an optimal solution is that both must have a feasible solution.
4. The fundamental duality theorem states that if either the primal or the dual problem has a definite optimal solution, then the other problem also has a definite optimal solution and the maximum values of the objective function in both the problems are the same, i.e., Max Z = Min Z. The solution of the other problem can be read from the $Z_j - C_j$ row below the columns of slack and surplus variables.
5. The existence theorem states that if either problem has an unbounded solution, then the other problem has no feasible solution.
6. Complementary slackness theorem: According to this,
   (*i*) If a primal variable is positive, then the corresponding dual constraint is an equation at the optimum, and vice versa.
   (*ii*) If a primal constraint is a strict inequality, then the corresponding dual variable is zero at the optimum, and vice versa.

Economic Interpretation of Dual Variables

If we interpret our primal LP problem as a classical 'Resource Allocation' problem, then its dual can be interpreted as a 'Resource Valuation' problem. Thus the values of the optimal dual variables have fascinating economic interpretation. The primal problem is used to describe a production problem in which the objective function represents the gain obtained from the production of goods, while the constraints characterize bounds on the production amounts due to the presence of limited resources. The available quantity of each resource is then measured.

Economic Interpretation: Consider the following primal problem:

Maximize $C_1x_1 + \ldots + C_nx_n$

Subject to, All $x_i \ge 0$

$$a_{11}x_1 + \ldots + a_{1n}x_n \leq b_1$$
$$a_{m1}x_1 + \ldots + a_{mn}x_n \leq b_m$$

Where,

NOTES

n = Economic activities

m = Resources

C_j = Revenue per unit of activity j

b_i = Maximum availability of resource i

a_{ij} = Consumption of resource i per unit of activity j

If $(x_1, \ldots, x_n)$ is optimal for the primal and $(y_1, \ldots, y_m)$ is optimal for the dual, then we can state that:

$$C_1x_1 + \ldots + C_nx_n = b_1y_1 + \ldots + b_my_m$$

Here,

Left hand side = Maximal revenue

Right hand side = Resource i (Availability of resource i) × Revenue per unit of resource i.

In other words we can say that the value of y_i at optimal is dual price of resource i.

2.9 SUMMING UP

- Linear Programming (LP) is a major innovation since World War II in the field of business decision-making, particularly under conditions of certainty.
- The word 'Linear' means that the relationships are represented by straight lines, i.e., the relationships are of the form $y = a + bx$ and the word 'Programming' means taking decisions systematically.
- LP is generally used in solving maximization (sales or profit maximization) or minimization (cost minimization) problems subject to certain assumptions.
- LP is extensively used in solving resource allocation problems.
- The term linearity implies straight line or proportional relationships among the relevant variables.
- Process means the combination of particular inputs to produce a particular output.
- Criterion function is also known as objective function which states the determinants of the quantity either to be maximized or to be minimized.
- Feasible solutions are all those possible solutions which can be worked upon under given constraints. The region comprising of all feasible solutions is referred as feasible region.
- Optimum solution is the best of the feasible solutions. Z denotes the optimal value of the objective function for a given system.
- LP model is based on the assumptions of proportionality, additivity, certainty, continuity and finite choices.
- The applications of linear programming problems are based on linear programming matrix coefficients and data transmission prior to solving the simplex algorithm.

Check Your Progress

17. What is artificial variable?
18. What is duality?
19. What is primal?
20. Why the primal problem is used?

NOTES

- The portfolio optimization template calculates the optimal capital of investments that gives the highest return for the least risk.
- The objective function, the set of constraints and the non-negative constraints together form a linear programming problem.
- Simple linear programming problem with two decision variables can be easily solved by graphical method.
- A set of values $X_1, X_2, ..., X_n$ which satisfies the constraints of the LPP is called its solution.
- Given a system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution. Such m variables are called basic variables and the remaining variables are called non-basic variables.
- A basic feasible solution is a basic solution which also satisfies all basic variables which are non-negative.
- If the value of the objective function Z can be increased or decreased indefinitely, such solutions are called unbounded solutions.
- Simplex method is an iterative procedure for solving LPP in a finite number of steps.
- For the solution of any LPP by simplex algorithm, the existence of an initial basic feasible solution is always assumed.
- For every given linear programming problem there is another intimately related linear programming problem referred to as its dual.
- The duality theorem states that 'for every maximization (or minimization) problem in linear programming there is a unique similar problem of minimization (or maximization) involving the same date which describes the original problem'. The original problem is referred to as the 'primal'. The 'dual' of a dual problem is the primal. Thus the primal and dual problems are replicas of each other.

2.10 KEY TERMS

- **Linear programming:** A decision-making technique under a set of given constraints and is based on the assumption that the relationships amongst the variables representing different phenomena are linear
- **Decision variables:** Variables that form objective function and on which the cost or profit depends
- **Linearity:** It is the line straight or proportional relationships among the relevant variables and in economic theory it is known as constant return
- **Criterion function:** An objective function which states the determinants of the quantity to be either maximized or minimized
- **Constraints:** Limitations under which planning is decided or restrictions imposed on decision variables
- **Feasible solution:** Any solution to a LPP which satisfies the non-negativity restrictions of the LPP
- **Feasible region:** The region comprising all feasible solutions

NOTES

- **Optimal solution:** Any feasible solution which optimizes (minimizes or maximizes) the objective function of the LPP
- **Certainty:** Assumption that includes prior knowledge of all the coefficients in the objective function, the coefficients of the constraints and the resource values. LP model operates only under conditions of certainty
- **Finite choices:** An assumption that implies that finite numbers of choices are available to a decision-maker and the decision variables do not assume negative values
- **Basic solution:** In a given system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution
- **Basic feasible solution:** A basic solution which also satisfies the condition in which all basic variables are non-negative
- **Slack variables:** If the constraints of a general LPP be given as $\Sigma a_{ij} X_j \leq b_i$ ($i = 1, 2, ..., m$; $j = 1, 2, ..., n$), then the non-negative variables S_i is introduced to convert the inequalities '$\leq$' to the equalities are called slack variables
- **Surplus variables:** If the constraints of a general LPP be $\Sigma a_{ij} X_j \geq b_i$ ($i = 1, 2, ..., m$; $j = 1, 2, ..., n$), then non-negative variables S_i introduced to convert the inequalities '$\geq$' to the equalities are called surplus variables

2.11 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Linear programming is a decision-making technique under a set of given constraints and is based on the assumption that the relationships amongst the variables representing different phenomena are linear.
2. Criterion function is objective function which states the determinants of the quantity, to be either maximized or minimized.
3. Linear programming finds application in agricultural and various industrial problems.
4. Constraints are limitations under which planning is decided, these are restrictions imposed on decision variables.
5. Solution of a linear programming is a set of values $X_1, X_2, ..., X_n$, satisfying the constraints of the LPP.
6. In a given system of m linear equations with n variables ($m < n$), any solution which is obtained by solving m variables keeping the remaining $n - m$ variables zero is called a basic solution.
7. In a given system of m linear equations with n variables ($m < n$), where m variables are solved keeping remaining $n - m$ variables zero, m variables are called basic variables and the remaining variables are called non-basic variables.
8. Basic feasible solution is a basic solution which satisfies the condition in which all basic variables are non-negative.
9. An objective function is maximized when it is a profit function. It is minimized when it is a cost function.
10. Feasible solution of a LPP is a solution that satisfies the non-negativity restrictions of the LPP.

NOTES

11. Feasible region is the region comprising all feasible solutions.
12. Optimal solution of a LPP is a feasible solution which optimizes (minimizes or maximizes) the objective function of the LPP.
13. Non-degenerate and degenerate solutions are the basic feasible solutions. In a problem which has exactly m positive variables, X_i ($i = 1, 2, ..., m$), i.e., none of the basic variables is zero, then it is called non-degenerate type and if one or more basic variables are zero then such basic feasible solution is said to be degenerate type.
14. Simplex method is an iterative procedure for solving LPP in a finite number of steps. This method provides an algorithm which consists of moving from one vertex of the region of feasible solution to another in such a manner that the value of the objective function at the succeeding vertex is less or more as the case may be that at the previous vertex.
15. The leaving element is converted to unity by dividing the key equation by the key element and all other elements in its column to zero by using the formula:

    New element

$$= \text{Old element} - \left[\frac{\text{Product of elements in key row and key column}}{\text{Key element}}\right]$$

16. By introducing slack variable, the problem is converted into standard form.
17. When the basis matrix cannot be obtained as an identify matrix in the starting simplex table, therefore we introduce a new type of variable called the *artificial variable*. These variables are fictitious and cannot have any physical meaning.
18. For every given linear programming problem there is another intimately related linear programming problem referred to as its dual. The duality theorem states that 'for every maximization (or minimization) problem in linear programming there is a unique similar problem of minimization (or maximization) involving the same date which describes the original problem'.
19. The original problem is referred to as the 'primal'. The 'dual' of a dual problem is the primal. Thus the primal and dual problems are replicas of each other.
20. The primal problem is used to describe a production problem in which the objective function represents the gain obtained from the production of goods, while the constraints characterize bounds on the production amounts due to the presence of limited resources. The available quantity of each resource is then measured.

2.12 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is meant by proportionality in linear programming?
2. What do you understand by certainty in linear programming?
3. What is meant by continuity in linear programming?
4. What are finite choices in the context of linear programming?
5. What are the basic constituents of an LP model?

NOTES

6. What is the canonical form of a LPP?
7. What are characteristics of the canonical form?
8. What are slack variables? Where are they used? Explain in brief.
9. What do you understand by surplus variables?
10. What is the simplex method?
11. Does every LPP solution have an optimal solution? Explain.
12. What is dual?
13. How will you transform a given primal problem into a dual problem?

Long-Answer Questions

1. A company manufactures 3 products *A*, *B* and C. The profits are: ₹ 3, ₹ 2 and ₹ 4 respectively. The company has two machines and given below is the required processing time in minutes for each machine on each product.

	Products		
Machines	***A***	***B***	***C***
I	4	3	5
II	2	2	4

Machines I and II have 2000 and 2500 minutes respectively. The company must manufacturers 100 *A*'s 200 *B*'s and 50 *C*'s but no more than 150 *A*'s. Find the number of units of each product to be manufactured by the company to maximize the profit. Formulate the above as a LP Model.

2. A company produces two types of leather belts *A* and *B*. *A* is of superior quality and *B* is of inferior quality. The respective profits are ₹ 10 and ₹ 5 per belt. The supply of raw material is sufficient for making 850 belts per day. For belt *A*, a special type of buckle is required and 500 are available per day. There are 700 buckles available for belt *B* per day. Belt *A* needs twice as much time as that required for belt *B* and the company can produce 500 belts if all of them were of the type *A*. Formulate a LP Model for the given problem.
3. The standard weight of a special purpose brick is 5 kg and it contains two ingredients B_1 and B_2, where B_1 costs ₹ 5 per kg and B_2 costs ₹ 8 per kg. Strength considerations dictate that the brick contains not more than 4 kg of B_1 and a minimum of 2 kg of B_2 since the demand for the product is likely to be related to the price of the brick. Formulate the given problem as a LP Model.
4. Egg contains 6 units of vitamin *A* per gram and 7 units of vitamin *B* per gram and 12 units of vitamin *B* per gram and costs 20 paise per gram. The daily minimum requirement of vitamin *A* and vitamin *B* are 100 units and 120 units respectively. Find the optimal product mix.
5. A company desires to devote the excess capacity of the three machines lathe, shaping machine and milling machine to make three products *A*, *B* and *C*. The available time per month in these machinery are tabulated below:

Machine	***Lathe***	***Shaping***	***Milling***
Available Time/Month	200 hrs	100 hrs	180 hrs

NOTES

The time taken to produce each unit of the products A, B and C on the machines is displayed in the table below.

	Lathe	*Shaping*	*Milling*
Product A hrs	6	2	4
Product B hrs	2	2	–
Product C hrs	3	–	3

The profit per product would be ₹ 20, ₹ 16 and ₹ 12 respectively on product A, B and C.

Formulate a LPP to find the optimum product-mix.

6. Solve the following by graphical method:

(*i*) Max $Z = X_1 - 3X_2$

Subject to, $X_1 + X_2 \le 300$

$X_1 - 2X_2 \le 200$

$2X_1 + X_2 \le 100$

$X_2 \le 200$

$X_1, X_2 \ge 0$

(*ii*) Max $Z = 5X + 8Y$

Subject to, $3X + 2Y \le 36$

$X + 2Y \le 20$

$3X + 4Y \le 42$

$X, Y \ge 0$

7. Solve graphically the following LPP:

Max $Z = 20X_1 + 10X_2$

Subject to, $X_1 + 2X_2 \le 40$

$3X_1 + X_2 \ge 30$

$4X_1 + 3X_2 \ge 60$

and $X_1, X_2 \ge 0$

8. Using simplex method, find non-negative values of X_1, X_2 and X_3 when

(*i*) Max $Z = X_1 + 4X_2 + 5X_3$

Subject to the constraints,

$3X_1 + 6X_2 + 3X_3 \le 22$

$X_1 + 2X_2 + 3X_3 \le 14$ and

$3X_1 + 2X_2 \le 14$

(*ii*) Max $Z = 10X_1 + 6X_2$

Subject to, $X_1 + X_2 \le 2$

$2X_1 + X_2 \le 4$

$3X_1 + 8X_2 \le 12$

$X_1, X_2 \ge 0$

(*iii*) Max $Z = X_1 + 2X_2 + X_3$

Subject to, $2X_1 + X_2 - X_3 \ge -2$

$-2X_1 + X_2 - 5X_3 \le 6$

NOTES

$$4X_1 + X_2 + X_3 \leq 6$$
$$X_1, X_2, X_3 \geq 0$$

9. Write the dual of the following LP problem:

Maximize, $Z = 5X_1 + 6X_3$

Subject to, $X_1 + 2X_2 = 5$

$$-X_1 + 5X_2 \geq 3$$
$$4X_1 + 7X_2 \leq 8$$

and X_1 unrestricted in sign

$$X_2 \geq 0$$

10. Write the primal of the following dual:

Minimize, $ZY = 10Y_1 + 8Y_2$

Subject to, $Y_1 + 2Y_2 \geq 5$

$$2Y_1 - Y_2 \geq 12$$
$$Y_1 + 3Y_2 \geq 4$$

and $Y_1 \geq 0; \geq Y_2$ unrestricted in sign.

2.13 REFERENCES AND SUGGESTED READINGS

Swaroop, Kanti, P.K. Gupta and Man Mohan. 2001. *Operations Research*. New Delhi: Sultan Chand & Sons.

Vohra, N. D. 2006. *Quantitative Techniques in Management*, 3rd edition. New Delhi: Tata McGraw-Hill Education.

Taha, Hamdy A. 1992. *Operations Research: An Introduction*. New York: Macmillan.

Gupta, P.K. and D.S. Hira. 2002. *Introduction to Operations Research*. New Delhi: S. Chand & Co.

Sharma, S. D. 2006. *Operations Research*. Uttar Pradesh: Kedar Nath Ram Nath & Co.

UNIT 3 TRANSPORTATION AND ASSIGNMENT PROBLEMS

NOTES

Structure

3.0 Introduction
3.1 Objectives
3.2 Transportation Problem: Different Initial Allocation Methods
  3.2.1 Basics of Transportation Problem
  3.2.2 Applications of Transportation Problem
  3.2.3 Solutions of Transportation Problem
3.3 Optimality and MODI Method of Solving Transportation Problems
3.4 Assignment Problem
  3.4.1 Basics of Assignment Problem
  3.4.2 Job Sequence Problem
3.5 Solutions of Assignment Problem by Hungarian Method
3.6 Variations in Assignment Problems
  3.6.1 Unbalanced Assignment Problem
  3.6.2 Maximization in Assignment Problem
3.7 Summing Up
3.8 Key Terms
3.9 Answers to 'Check Your Progress'
3.10 Questions and Exercises
3.11 References and Suggested Readings

3.0 INTRODUCTION

In this unit, you will learn about the transportation problem, which is a subclass of a Linear Programming Problem (LPP). Transportation problems deal with the objective of transporting various quantities of a single homogeneous commodity initially stored at various origins to different destinations in a way that keeps transportation cost at a minimum. You will learn about applications of the transportation problem and solution and rules to solve such problems. The solution of any transportation problem is obtained in two stages, initial solution and optimal solution. There are three methods of obtaining an initial solution which include North West Corner Rule, Least Cost Method and Vogel's Approximation Method (VAM). VAM is preferred since the solution obtained this way is very close to the optimal solution. The optimal solution of any transportation problem is a feasible solution that minimizes the total cost. An optimal solution is the second stage of a solution obtained by improving the initial solution. The MODI or MOdified DIstribution method is used to obtain optimal solutions and optimality tests.

In this unit, you will also learn the basic concepts of an assignment problem, which can be presented in the form of $n \times n$ cost matrix of real numbers. For solving an assignment problem, the Hungarian method is used. You will learn about maximization in assignment problems and the job sequence problem. To solve a job sequence problem, the assignment problem has to be balanced. An assignment problem is balanced if the cost matrix is a square matrix, otherwise it is termed as unbalanced. To make an unbalanced assignment problem balanced, dummy rows or dummy columns are added with all entries as zeroes.

NOTES

3.1 OBJECTIVES

After going through this unit, you will be able to:

- Know about the transportation problem
- Explain the three methods of finding an initial solution
- Understand special cases in transportation
- Explain restricted transportation problem
- Use the MODI method for finding optimal solution
- Define assignment problem
- Understand how to maximize an assignment problem
- Solve a job sequence problem
- Solve an assignment problem using the Hungarian method
- Describe variations in assignment problem

3.2 TRANSPORTATION PROBLEM: DIFFERENT INITIAL ALLOCATION METHODS

3.2.1 Basics of Transportation Problem

The Transportation Problem (TP) is one of the subclasses of LPP (Linear Programming Problem) in which the objective is to transport various quantities of a single homogeneous commodity that are initially stored at various origins to different destinations in such a way that the transportation cost is minimum. To achieve this objective we must know the amount and location of available supplies and the quantities demanded. In addition we must know the costs that result from transporting one unit of commodity from various origins to various destinations.

Elementary Transportation Problem

Consider a transportation problem with m origins (rows) and n destinations (columns). Let C_{ij} be the cost of transporting one unit of the product from the ith origin to jth destination, a_i the quantity of commodity available at origin i, b_j the quantity of commodity needed at destination j. X_{ij} is the quantity transported from ith origin to jth destination. This transportation problem can be stated in the following tabular form.

		Destinations				
		1	2	3	... n	Capacity
Origins	1	C_{11} X_{11}	C_{12} X_{12}	C_{13} X_{13}	... ... C_{1n} X_{1n}	a_1
	2	C_{21} X_{21}	C_{22} X_{22}	C_{23} X_{23}	... ... C_{2n} X_{2n}	a_2
	3	C_{31} X_{31}	C_{32} X_{32}	C_{33}	... ... C_{3n} X_{3n}	a_3
	m	C_{m1} X_{m1}	C_{m2} X_{m2}	C_{m3} X_{m3}	... ... C_{mn} X_{mn}	a_m
	Demand	b_1	b_2	b_3	... b_n	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

The linear programming model representing the transportation problem is given by,

$$\text{Minimize } Z = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}X_{ij}$$

Subject to the constraints,

$$\sum_{j=1}^{n} X_{ij} = a_i \qquad i = 1, 2, ..., n \text{ (Row sum)}$$

$$\sum_{i=1}^{m} X_{ij} = b_j \qquad j = 1, 2, ..., n \text{ (Column sum)}$$

$$X_{ij} \geq 0 \quad \text{For all } i \text{ and } j$$

The given transportation problem is said to be balanced if,

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$$

i.e., the total supply is equal to the total demand.

Definitions

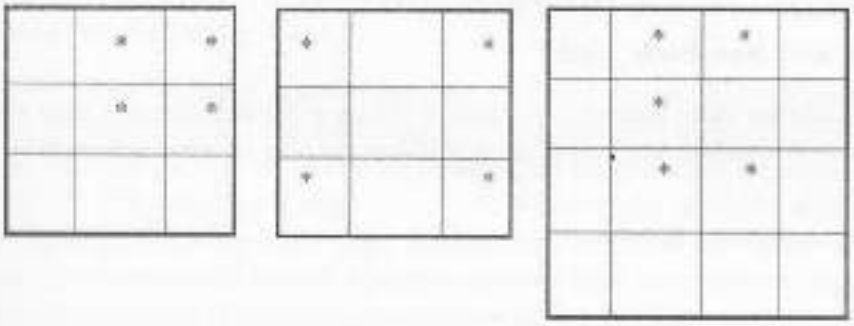
Feasible Solution: Any set of non-negative allocations ($X_{ij}>0$) which satisfies the row and column sum is called a feasible solution.

Basic Feasible Solution: A feasible solution is called a basic feasible solution if the number of non-negative allocations is equal to $m + n - 1$, where m is the number of rows and n the number of columns in a transportation table.

Non-Degenerate Basic Feasible Solution: Any feasible solution to a transportation problem containing m origins and n destinations is said to be non-degenerate, if it contains $m + n - 1$ occupied cells and each allocation is in independent positions.

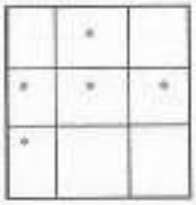
The allocations are said to be in independent positions if it is impossible to form a closed path. Closed path means by allowing horizontal and vertical lines and when all the corner cells are occupied.

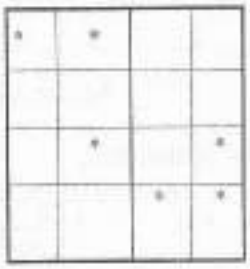
The allocations in the following tables are not in independent positions.



The allocations in the following tables are in independent positions.

NOTES





Degenerate Basic Feasible Solution: If a basic feasible solution contains less than $m + n - 1$ non-negative allocations it is said to be degenerate.

Transportation Algorithm

This algorithm can be used for minimizing the transportation cost for goods from O origins to D destinations and there may be $O*D$ number of direct routes from O origins to D destinations. Problem is balanced when sum of supplies at O sources is equal to sum of demands at D destinations. If it is not so, then this problem is not balanced. There may be two such situations. Supply may be lesser than demand and in that case it is balanced by adding dummy supply node. If demand is lesser than supply then dummy demand node is added to make it a balanced problem. Thus, before starting to use this algorithm, problem should be made balanced, if it is not balanced.

Data is presented in tabular form. As a convention, origins are put on left side of the table with quantity to be supplied listed towards right side and demands are put on top with quantity of demand towards the bottom side. Unit cost of transportation is put at the top of every cell within a small box. Zero unit cost shows unshipped unit column in case supply is in excess of the demand. Similarly, a unit cost either penalty or zero shows shortage row in supplies that are lesser that demand.

The algorithm has two phases. In phase I this makes allocation for supplies on demands by making utilizing an approach of minimum unit cost for generating a feasible solution. This feasible solution may not be optimal. Optimization is done in second phase and in this phase checking is done for optimality conditions and improvement is done for reducing the cost if optimality conditions are not satisfied. This second phase adopts iterative steps and stops only when optimality conditions are satisfied. Once done, no further steps are required.

Basic and Non-Basic Cells

Basic cells are those that indicate positive values and non-basic cells have zero value for flow. According to transportation problem number of basic cells will be exactly $m + n - 1$.

Algorithm follows as below:

Step 0

Initialization: Before starting to solve the problem, it should be balanced. If not then make it balanced by 'unshipped supply' column in case demand is less than supply or by adding 'shortage' row in case supply is less than the supply. Put zero for unit costs in the column for unshipped supply. Put either penalty costs or zero in a row that shows shortage.

NOTES

Phase I: To Find Initial 'Feasible Solution'

Step 1: Locate cell with minimum cost having positive supply as well as demand, then make allocation in that cell having residuals as minimum.

Step 2: Reduce residual supply/demand as per allocations made above (Step 1). Do this till all demands are met and then proceed to Phase II. If all demands are not met go to Step 1.

Phase II: Optimal Solution

Carry out check for Optimality.

Step 3: For rows and columns find dual values, u_i for rows and v_j for columns. For this, set u_1 (first dual value) to 0 followed by solution of triangular dual equations one by one. These **dual equations** are to be applied for basic cells only as below as,

$$C_{ij} = u_i + v_j$$

Where, C_{ij} denotes unit cost, as given for that cell and either v_j or u_i is already known. Here, v_j denotes dual value of column j and u_i denotes that for row i. These are taken as already known. Other dual value is computed from equation,

$$C_{ij} = u_i + v_j$$

Thus, dual values of every cell can be computed by setting the first to zero and appropriate order is used for dual equations for basic cells.

Step 4: Optimality conditions are expressed as **reduced costs** in case of all non-basic cells as given below:

$$\Delta_{ij} = C_{ij} - (u_i + v_j)$$

Reduced cost (Δ_{ij}) in case of non-basic cells represent net change in unit cost resulting due to movement of cell ij to solve and adjusting around a cycle that has been created in this way for basic cells to find current solution. This is shown in Step 5. Hence if one Δ_{ij} is positive, then by use of this cell total transportation will increase, but when Δ_{ij} is negative for a cell, this will cause reduction in total transportation. If all reduced costs (Δ_{ij}) are positive, i.e., $\Delta_{ij} > 0$, ***optimality conditions*** are satisfied and no further improvement is possible and algorithm terminates at this point. But if at least one Δ_{ij} is negative, optimality conditions are not satisfied and there is possibility of reduction in costs. If optimality conditions are not satisfied algorithm continues as given in Step 5.

Adjustment for Reducing Cost

Step 5: Select a cell 'ij' that is most negative for Δ_{ij}. This becomes entering variable. Put (+) sign for identifying it. To maintain constraints of balance in supply and demand, locate basic cells for ith row as well as jth column that compensates for increase in value in cell 'ij'. Put negative (–) sign in these cells. This process should be continued to get one cycle in which (+) and (–) are marked. Such a cycle will unique and if 'dead ends' are encountered in such a process, make a back track and one amongst other alternatives are tried. A cycle has rows/column having non-basic cells for holding compensating (+) or (–) sign. This may require trial and error for to finding it.

Step 6: After determining every basic cell within this cycle, adjustment is obtained as minimum value in basic cells that are negative. This is known as adjustment amount and let it be called '**aa**'. Add this to every cell value that is positive and marked with

NOTES

(+) sign, and subsequently deduct this from cells having negative (–) sign and then drop those from the basis that becomes zero. In case two or more cells become zero from such adjustment, drop only one of these that have greatest C_{ij} value. This is necessary for maintaining basic cells having $m + n - 1$ number for computing dual values. Reduction in cost that is associated with such a change is found as product of reduced cost Δ_{ij} for incoming cell multiplied by cell value previously held by outgoing cell.

After finding this new solution, move to the third Step 3 for checking conditions of optimality. If optimality condition is satisfied, algorithm terminates.

3.2.2 Applications of Transportation Problem

The Travelling Salesman Problem

Assume that a salesman has to visit n cities. He wishes to start from a particular city, visits each city once and then returns to his starting point. His objective is to select the sequence in which the cities are visited in such a way that his total travelling time is minimized.

To visit 2 cities A and B, there is no choice. To visit 3 cities we have 2! possible routes. For 4 cities we have 3! possible routes. In general to visit n cities there are $(n - 1)!$ possible routes.

Mathematical Formulation

Let C_{ij} be the distance or time or cost of going from city i to city j. The decision variable X_{ij} be 1 if the salesman travels from city i to city j and otherwise 0.

The objective is to minimize the travelling time.

$$Z = \sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij} X_{ij}$$

Subject to the constraints,

$$\sum_{j=1}^{n} X_{ij} = 1, i = 2, ..., n$$

$$\sum_{i=1}^{n} X_{ij} = 1, j = 2, ..., n$$

Subject to the additional constraint that X_{ij} is so chosen that no city is visited twice before all the cities are completely visited.

In particular, going directly from i to i is not permitted. Which means $C_{ij} = \infty$, when $i = j$.

In travelling salesman problem we cannot choose the element along the diagonal and this can be avoided by filling the diagonal with infinitely large elements.

The travelling salesman problem is very similar to the assignment problem except that in the former case, there is an additional restriction that X_{ij} is so chosen that no city is visited twice before the tour of all the cities is completed.

Treat the problem as an assignment problem and solve it using the same procedures. If the optimal solution of the assignment problem satisfies the additional

NOTES

constraint, then it is also an optimal solution of the given travelling salesman problem. If the solution to the assignment problem does not satisfy the additional restriction then after solving the problem by assignment technique we use the method of enumeration.

Example 3.1: A travelling salesman has to visit 5 cities. He wishes to start from a particular city, visit each city once and then return to his starting point. Cost of going from one city to another is shown below. You are required to find the least cost route.

To City

From City	*A*	*B*	*C*	*D*	*E*
A	∞	4	10	14	2
B	12	∞	6	10	4
C	16	14	∞	8	14
D	24	8	12	∞	10
E	2	6	4	16	∞

Solution: First we solve this problem as an assignment problem.

Subtract the minimum element in each row from all the elements in its row.

	A	*B*	*C*	*D*	*E*
A	∞	2	8	12	0
B	8	∞	2	6	0
C	8	6	∞	0	6
D	16	0	4	∞	2
E	0	4	2	14	∞

Subtract the minimum element in each column from all the elements in its column.

	A	*B*	*C*	*D*	*E*
A	∞	2	8	12	0
B	8	∞	2	6	0
C	8	6	∞	0	6
D	16	0	4	∞	2
E	0	4	2	14	∞

We have the first modified matrix. Draw minimum number of lines to cover all zeros.

	A	*B*	*C*	*D*	*E*
A	∞	2	8	12	0
B	8	∞	2	6	0
C	8	6	∞	0	6
D	16	0	4	∞	2
E	0	4	2	14	∞

Here $N = 4 < n = 5$. Subtract the smallest uncovered element from all the uncovered elements and add to the element which is in the point of intersection of lines. Hence, we get the second modified matrix.

	A	*B*	*C*	*D*	*E*
A	∞	0	6	12	0
B	8	2	0	6	0
C	8	6	∞	0	6
D	18	0	4	∞	4
E	0	2	0	14	∞

$N = 5 = n = 5$ = Order of matrix. We make assignment.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & \not{8} & 6 & 12 & \textcircled{0} \\ 8 & \infty & \textcircled{0} & 6 & \not{4} \\ 8 & 4 & \infty & \textcircled{0} & 6 \\ 18 & \textcircled{0} & 4 & \infty & 4 \\ \textcircled{0} & 2 & \not{8} & 14 & \infty \end{bmatrix} \end{array}\begin{array}{l} A \to E \\ B \to C \\ C \to D \\ D \to B \\ E \to A \\ A \to E \to A \end{array}$$

As the salesman should go from A to E and then come back to A without covering B, C, D which is contradicting the fact that no city is visited twice before all the cities are visitied.

Hence, we obtain the next best solution by bringing the next minimum non-zero element namely 4.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & \textcircled{0} & 6 & 12 & 0 \\ 8 & \infty & \textcircled{0} & 6 & \not{4} \\ 8 & 4 & \infty & \textcircled{0} & 6 \\ 18 & \not{8} & 4 & \infty & \textcircled{4} \\ \textcircled{0} & 2 & \not{8} & 14 & \infty \end{bmatrix} \end{array}$$

$A \to B, B \to C, C \to D, D \to E, E \to A$

Since all the cities have been visited and no city is visited twice before completing the tour of all the cities, we have an optimal solution to the travelling salesman.

The least cost route is $A \to B \to C \to D \to E \to A$.

Total Cost = 4 + 6 + 8 + 10 + 2 = ₹ 30.

Example 3.2: A machine operator processes five types of items on his machine each week and must choose a sequence for them. The set-up cost per change depends on the items presently on the machine and the set-up to be made according to the following table.

To Item

$$\textbf{From Item}\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 4 & 7 & 3 & 4 \\ 4 & \infty & 6 & 3 & 4 \\ 7 & 6 & \infty & 7 & 5 \\ 3 & 3 & 7 & \infty & 7 \\ 4 & 4 & 5 & 7 & \infty \end{bmatrix} \end{array}$$

If he processes each type of item once and only once in each week, how should he sequence the items on his machine in order to minimize the total set-up cost?

Solution: Reduce the cost matrix and make assignments in rows and columns having single row.

Modify the matrix by subtracting the least element from all the elements in its row and also in its column.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 1 & 4 & 0 & 1 \\ 1 & \infty & 3 & 0 & 1 \\ 2 & 1 & \infty & 2 & 0 \\ 0 & 0 & 4 & \infty & 4 \\ 0 & 0 & 1 & 3 & \infty \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 1 & 4 & 0 & 1 \\ 1 & \infty & 3 & 0 & 1 \\ 2 & 1 & \infty & 2 & 0 \\ 0 & 0 & 4 & \infty & 4 \\ 0 & 0 & 1 & 3 & \infty \end{bmatrix} \end{array}$$

Here, $N = 4 < n = 5$, i.e., $N < n$.

Subtract the smallest uncovered element from all the uncovered elements and add to the element which is at the point of intersection of lines and get the reduced second modified matrix.

$$\begin{array}{c} \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & 0 & 3 & 0 & 1 \\ 0 & \infty & 2 & 0 & 1 \\ 1 & 0 & \infty & 1 & 0 \\ 0 & 0 & 4 & \infty & 5 \\ 0 & 0 & 1 & 4 & \infty \end{bmatrix} \end{array}$$

Here, $N = 5 = n = 5$ = Order of matrix. We make the assignment.

Assignment

We make a solution by considering the next smallest non-zero element by considering 1.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} \infty & \not{0} & 3 & \not{0} & \textcircled{1} \\ \not{0} & \infty & 2 & \textcircled{0} & 1 \\ 1 & \textcircled{0} & \infty & 1 & \not{0} \\ \textcircled{0} & \not{0} & 4 & \infty & 5 \\ \not{0} & \not{0} & \textcircled{1} & 4 & \infty \end{bmatrix} \end{array}\begin{array}{l} \\ A \to E \\ B \to D \\ C \to B \\ D \to A \\ E \to C \end{array}$$

$A \to E, E \to C, C \to B, B \to D, D \to A$

We get the solution $A \to E \to C \to B \to D \to A$.

This schedule provides the required solution in which each item is not processed once in a week.

i.e., $A \to E \to C \to B \to D \to A$.

The total set-up cost comes to ₹ 21.

3.2.3 Solutions of Transportation Problem

Optimal solution is a feasible solution (not necessarily basic) which minimizes the total cost.

The solution of a Transportation Problem (TP) can be obtained in two stages, namely initial solution and optimum solution.

Initial solution can be obtained by using any one of the three methods, viz.

(*i*) North West Corner Rule (NWCR)

(*ii*) Least Cost Method or Matrix Minima Method

(*iii*) Vogel's Approximation Method (VAM)

VAM is preferred over the other two methods, since the initial basic feasible solution obtained by this method is either optimal or very close to the optimal solution.

NOTES

The cells in the transportation table can be classified as occupied cells and unoccupied cells. The allocated cells in the transportation table is called *occupied cells* and *empty cells* in a transportation table is called *unoccupied cells*.

The improved solution of the initial basic feasible solution is called optimal solution which is the second stage of solution, that can be obtained by MODI (MOdified DIstribution Method).

North West Corner Rule

Step 1: Starting with the cell at the upper left corner (North West) of the transportation matrix we allocate as much as possible so that either the capacity of the first row is exhausted or the destination requirement of the first column is satisfied, i.e., $X_{11} = \text{Min}\ (a_1, b_1)$.

Step 2: If $b_1 > a_1$, we move down vertically to the second row and make the second allocation of magnitude $X_{22} = \text{Min}(a_2, b_1 - X_{11})$ in the cell (2, 1).

If $b_1 < a_1$, move right horizontally to the second column and make the second allocation of magnitude $X_{12} = \text{Min}\ (a_1, X_{11} - b_1)$ in the cell (1, 2).

If $b_1 = a_1$, there is a tie for the second allocation. We make the second allocations of magnitude,

$$X_{12} = \text{Min}\ (a_1 - a_1, b_1) = 0 \text{ in the cell } (1,2)$$

or, $$X_{21} = \text{Min}\ (a_2, b_1 - b_1) = 0 \text{ in the cell } (2,1)$$

Step 3: Repeat Steps 1 and 2 moving down towards the lower right corner of the transportation table until all the rim requirements are satisfied.

Example 3.3: Obtain the initial basic feasible solution of a transportation problem whose cost and rim requirement table is as follows:

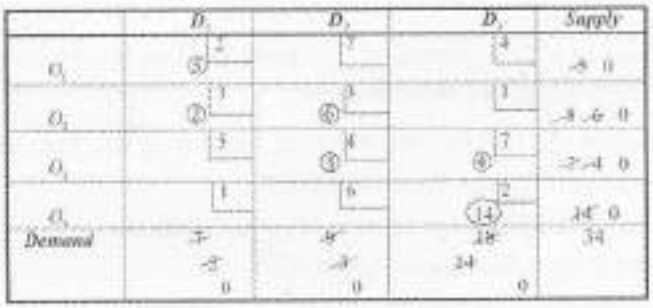
Origin/Destination	D_1	D_2	D_3	*Supply*
O_1	2	7	4	5
O_2	3	3	1	8
O_3	5	4	7	7
O_4	1	6	2	14
Demand	7	9	18	34

Solution: Since $\Sigma a_i = 34 = \Sigma b_j$, there exists a feasible solution to the transportation problem. We obtain initial feasible solution as follows.

The first allocation is made in the cell (1, 1) the magnitude being,

$X_{11} = \text{Min}\ (5, 7) = 5.$

The second allocation is made in the cell (2, 1) and the magnitude of the allocation is given by $X_{21} = \text{Min}\ (8, 7 - 5) = 2.$



The third allocation is made in the cell (2, 2) the magnitude $X_{22} = \text{Min}\ (8 - 2, 9) = 6.$

The magnitude of the fourth allocation is made in the cell (3, 2) given by X_{32} = Min (7, 9 – 6) = 3.

The fifth allocation is made in the cell (3, 3) with magnitude X_{33} = Min (7 – 3, 14) = 4.

The final allocation is made in the cell (4, 3) with magnitude X_{43} = Min (14, 18 – 4) = 14.

Hence, we get the initial basic feasible solution to the given TP and is given by,

$X_{11} = 5; X_{23} = 2; X_{22} = 6; X_{32} = 3; X_{33} = 4; X_{43} = 14$

Total Cost $= 2 \times 5 + 3 \times 2 + 3 \times 6 + 3 \times 4 + 4 \times 7 + 2 \times 14$

$= 10 + 6 + 18 + 12 + 28 + 28 = ₹ 102$

Least Cost or Matrix Minima Method

Step 1: Determine the smallest cost in the cost matrix of the transportation table. Let it be C_{ij}. Allocate X_{ij} = Min (a_i, b_j) in the cell (i, j).

Step 2: If $X_{ij} = a_i$ cross off the ith row of the transportation table and decrease b_j by a_i. Then go to Step 3.

If $X_{ij} = b_j$ cross off the jth column of the transportation table and decrease a_i by b_j. Go to Step 3.

If $X_{ij} = a_i = b_j$ cross off either the ith row or the jth column but not both.

Step 3: Repeat Steps 1 and 2 for the resulting reduced transportation table until all the rim requirements are satisfied. Whenever the minimum cost is not unique make an arbitrary choice among the minima.

Example 3.4: Obtain an initial feasible solution to the following TP using matrix minima method.

	D_1	D_2	D_3	D_4	*Supply*
O_1	1	2	3	4	6
O_2	4	3	2	0	8
O_3	0	2	2	1	10
Demand	4	6	8	6	24

Solution: Since $\Sigma a_i = \Sigma b_j = 24$, there exists a feasible solution to the TP using the steps in the least cost method, the first allocation is made in the cell (3, 1) the magnitude being $X_{31} = 4$. This satisfies the demand at the destination D_1 and we delete this column from the table as it is exhausted.

	D_1	D_2	D_3	D_4	*Supply*
O_1	1	2 ⑥	3	4	6 0
O_2	4	3	2 ②	0 ⑥	8 2
O_3	0 ④	2	2 ⑧	1	10 6
Demand	4 0	6 0	8 2 0	6 0	24

The second allocation is made in the cell (2, 4) with magnitude X_{24} = Min (6, 8) = 6. Since it satisfies the demand at the destination D_4, it is deleted from the table. From the reduced table, the third allocation is made in the cell (3, 3) with magnitude X_{33} = Min (8, 6) = 6. The next allocation is made in the cell (2, 3) with magnitude X_{23} of Min (2, 2) = 2. Finally, the allocation is made in the cell (1, 2) with magnitude X_{12} = Min (6, 6) = 6. Now, all the requirements have been satisfied and hence, the initial feasible solution is obtained.

NOTES

The solution is given by,

$X_{12} = 6; X_{23} = 2; X_{24} = 6; X_{31} = 4; X_{33} = 6$

Since the total number of occupied cells = $5 < m + n - 1$

We get a degenerate solution.

Total Cost = $6 \times 2 + 2 \times 2 + 6 \times 0 + 4 \times 0 + 6 \times 2$

$= 12 + 4 + 12 = ₹ 28.$

Vogel's Approximation Method (VAM)

Vogel Approximation Method (VAM) is used to find the feasible solution for transportation of goods where the solution is either optimal or near to the optimal solution. Typically, this method is used to reduce the transportation costs by interpreting using a mathematical table the transportation costs from one place to another. In the table, the column represents the demand centres while the row represents the supply points. The following are the general steps used in VAM:

Step 1: Identify the minimum and next minimum numbers in a column and repeat the same for the row.

Step 2: The above step is repeated for all other columns and rows.

Step 3: Now, subtract the two numbers identified for each column and each row such that the difference is positive.

Step 4: Identify the maximum difference among all the rows and also among all the columns.

Step 5: Assign all the demand units for that minimum number in that column which has got the maximum difference (repeat the same for the row).

Step 6: Remove that column and row completely and repeat the above process until all the demand units are filled up completely.

Vogel's Approximation Method (VAM) also takes costs into account in allocation.

The steps involved in VAM method for finding the initial solution are as follows.

Step 1: Find the penalty cost, namely the difference between the smallest and next smallest costs in each row and column.

Step 2: Among the penalties as found in Step (1) choose the maximum penalty. If this maximum penalty is more than one (i.e., if there is a tie) choose any one arbitrarily.

Step 3: In the selected row or column as by Step (2) find out the cell having the least cost. Allocate to this cell as much as possible depending on the capacity and requirements.

Step 4: Delete the row or column which is fully exhausted. Again, compute the column and row penalties for the reduced transportation table and then go to Step (2). Repeat the procedure until all the rim requirements are satisfied.

Note: If the column is exhausted, then there is a change in row penalty and vice versa.

Example 3.5: Find the initial basic feasible solution for the following transportation problem using VAM.

		Destination				
		D_1	D_2	D_3	D_4	*Supply*
Origin	O_1	11	13	17	14	250
	O_2	16	18	14	10	300
	O_3	21	24	13	10	400
	Demand	200	225	275	250	950

Solution: Since $\Sigma a_i = \Sigma b_j = 950$, the problem is balanced and there exists a feasible solution to the problem.

First, we find the row and column penalty P_i as the difference between the least and the next least cost. The maximum penalty is 5. Choose the first column arbitrarily. In this column, choose the cell having the least cost name (1, 1). Allocate to this cell with minimum magnitude (i.e., Min (250, 200) = 200.) This exhausts the first column. Delete this column. Since a column is deleted, then there is a change in row penalty P_{II} and column penalty P_{II} remains the same. Continuing in this manner, we get the remaining allocations as given in the following table below.

I Allocation

	D_1	D_2	D_3	D_4	Supply	P_i
O_1	11 (200)	13	17	14	50 250	2
O_2	16	18	14	10	300	4
O_3	21	24	13	10	400	3
Demand	200 0	225	275	250		
P_i	5↑	5	(1)	0		

II Allocation

	D_1	D_3	D_4	Supply	P_{II}
O_1	13 (50)	17	14	50	(1)
O_2	18	14	10	300	4
O_3	24	13	10	400	3
Demand	225 175	275	250		
P_{II}	5↑	(1)	0		

III Allocation

	D_2	D_3	D_4	Supply	P_{III}
O_2	18 (175)	14	10	300 125	4
O_3	24	13	10	400	3
Demand	175 0	275	250		
P_{III}	6↑	1	0		

IV Allocation

	D_1	D_4	Supply	P_{IV}
O_2	14	10 (125)	125 0	4 ←
O_3	13	10	400	3
Demand	275	250 125		
P_{IV}	1	0		

V Allocation

	D_1	D_4	Supply	P_V
O_3	13 (275)	10	400 125	3
Demand	275 0	125		
P_V	13↑	10		

VI Allocation

	D_4	Supply	P_{VI}
O_3	10 (125)	125 0	10 ←
Demand	125 0		
P_{VI}	10		

Finally, we arrive at the initial basic feasible solution which is shown in the following table.

	D_1	D_2	D_3	D_4	Supply
O_1	11 (200)	13 (50)	17	14	250
O_2	16	18 (175)	14	10 (125)	300
O_3	21	24	13 (275)	10 (125)	400
Demand	200	225	275	250	

There are 6 positive independent allocations which equals to $m + n - 1 = 3 + 4 - 1$. This ensures that the solution is a non-degenerate basic feasible solution.

NOTES

Check Your Progress

1. What is a transportation problem?
2. What must one know to achieve the objectives of a transportation problem?
3. How are transportation problems presented?
4. Give the mathematical presentation of a transportation problem.
5. Define feasible, basic and non-degenerate solutions of a transportation problem.
6. How the solution of a TP can be obtained?
7. Name the methods used to obtain initial solution.

NOTES

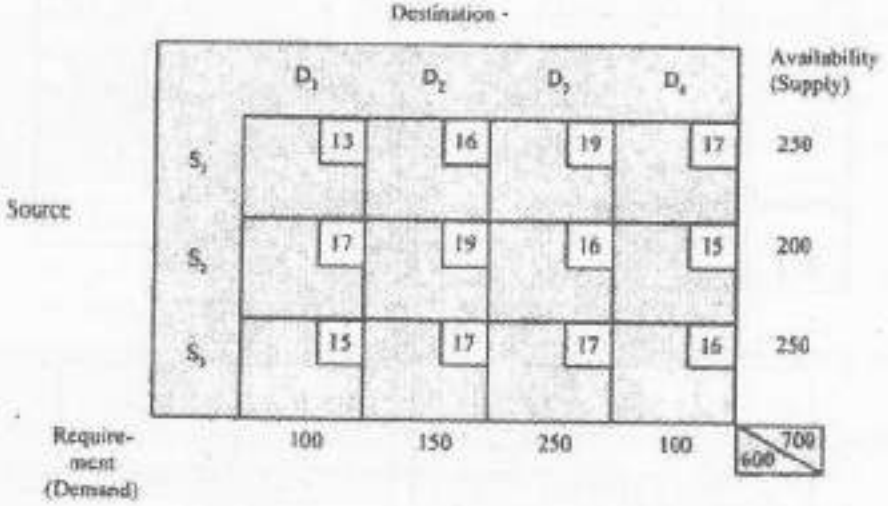
∴ Transportation Cost,

$= 11 \times 200 + 13 \times 50 + 18 \times 175 + 10 \times 125 + 13 \times 275 + 10 \times 125$

$= ₹ 12{,}075.$

Examples to Find Solution of Transportation Problem Using Vogel's Approximation Method or VAM

Example 3.6: An organization has four destinations (D_1, D_2, D_3 and D_4) and three sources (S_1, S_2 and S_3) for supply of goods. The transportation cost per unit is given below. The total availability is 700 units which exceeds the cumulative demand of 600 units. Find the optimal transportation scheme for this condition using the Vogel's Approximation Method or VAM.



Solution: The solution is obtained as follows.

Step 1: First check for balance of supply and demand

Supply = 250 + 200 + 250 = 700 units

Demand = 100 + 150 + 250 + 100 = 600 units

Decision Rule

(*i*) If Supply = Demand then go to next step.

(*ii*) If Supply > Demand then add a 'Dummy Destination' with zero transportation cost.

(*iii*) If Supply < Demand then add a 'Dummy Source' with zero transportation cost.

In the given problem, **Supply > Demand.**

Hence, we add a 'Dummy Destination' say D_5 with zero transportation cost and balance demand which is difference in supply and demand (= 100 units). The initial transportation matrix is now formulated with transportation cost in each route. Each cell of the transportation matrix represents a possible route. In the following table, dummy column is introduced for balancing the supply and demand.

NOTES

Source \ Destination	D_1	D_2	D_3	D_4	D_5	Supply
S_1	13	16	19	17	0	250
S_2	17	19	16	15	0	200
S_3	15	17	17	16	0	250
Demand	100	150	250	100	100	700 / 700

Step 2: (*i*) Decide the nature of problem, i.e., minimization of transportation cost..

(*ii*) Make initial assignment using the Vogel's approximation method.

(i) Select the lowest transportation cost route in the initial matrix. For example, it is route S_1D_5, S_2D_5 and S_3D_5 in the given problem with zero transportation cost. Allocate the minimum of remaining balance of supply (in last column) and demand (in last row).

In this method, we calculate the difference between the two least-cost routes for each row and column. The difference is called as penalty cost for not using the last-cost route. Following table shows the first calculation of 'Penalty' cost in VAM.

	D_1	D_2	D_3	D_4	D_5	Supply
S_1	13	16	19	17	0	250
S_2	17	19	16	15	0	200
S_3	15	17	17	16	0	250 150
Demand	100	150	250	100	(100)	

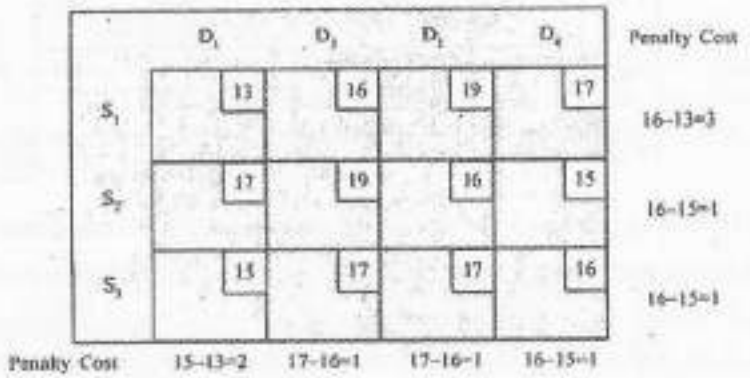
Highest of all calculated penalty costs is for S_3 and S_2. Therefore, allocation is to be made in row of source S_3. The route or cell which should be selected should be the lowest cost of this row, i.e., the route S_3D_5. Hence, first allocation in Vogel's method is as follows.

	D_1	D_2	D_3	D_4	Penalty Cost
S_1	13	16	19	17	16–13=3
S_2	17	19	16	15	16–15=1
S_3	15	17	17	16	16–15=1
Penalty Cost	15–13=2	17–16=1	17–16=1	16–15=1	

NOTES

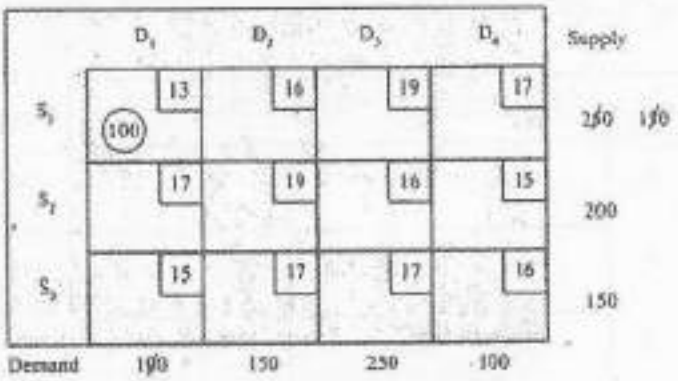
With the first allocation, destination D_5 is used. Leave out this column and develop the remaining matrix for calculating the penalty cost. We obtain the following matrix.

Now for this, source S_1 has highest penalty cost. For this row, the least cost route is S_1D_1. Hence, next assignment is due in this route:



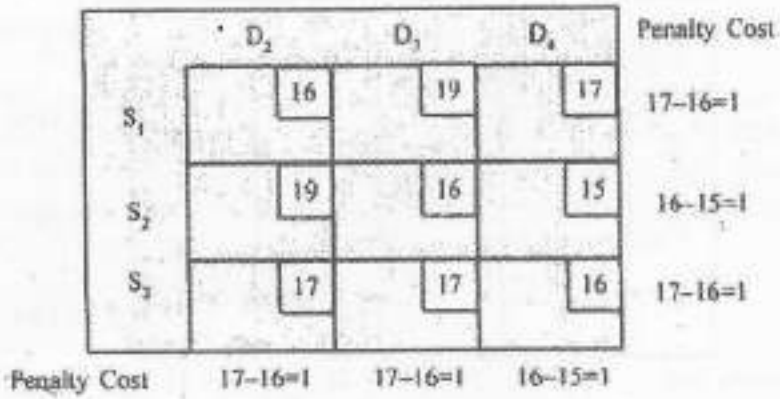
Second Calculation of Penalty Cost in VAM

Second allocation in Vogel's method is obtained as follows:



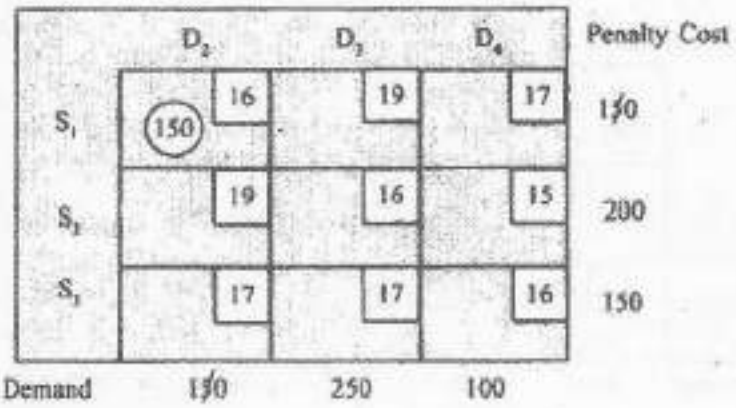
After second allocation, since destination D_1 is used, leave this column and proceed for calculation of next penalty cost. Allocation is done in route S_1D_2. Since there is tie between all routes, break the tie by arbitrarily selecting any route, for example S_1D_2 in this case.

Third Calculation of Penalty Cost in VAM

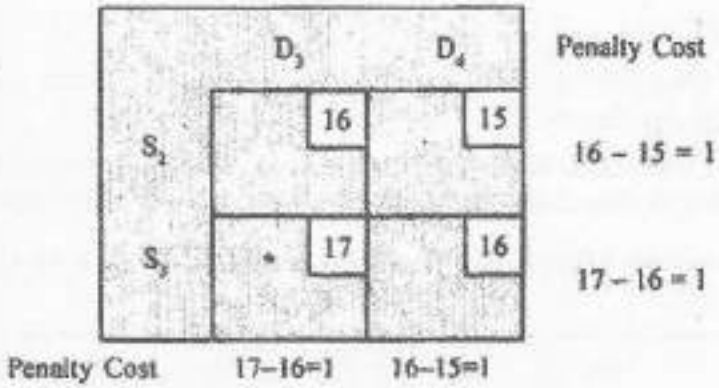


NOTES

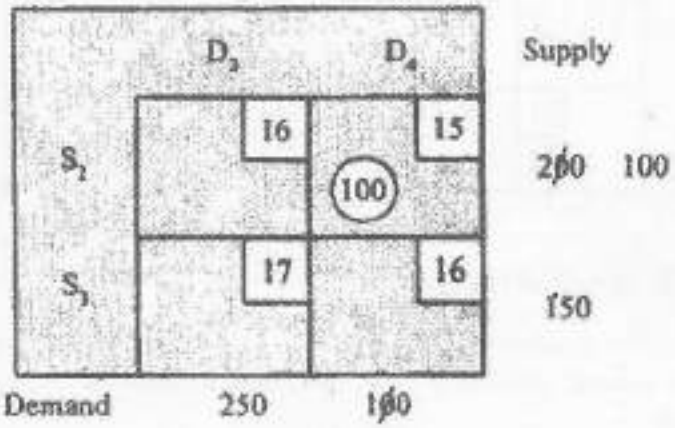
Third Allocation in VAM



Fourth Calculation of Penalty Cost in VAM



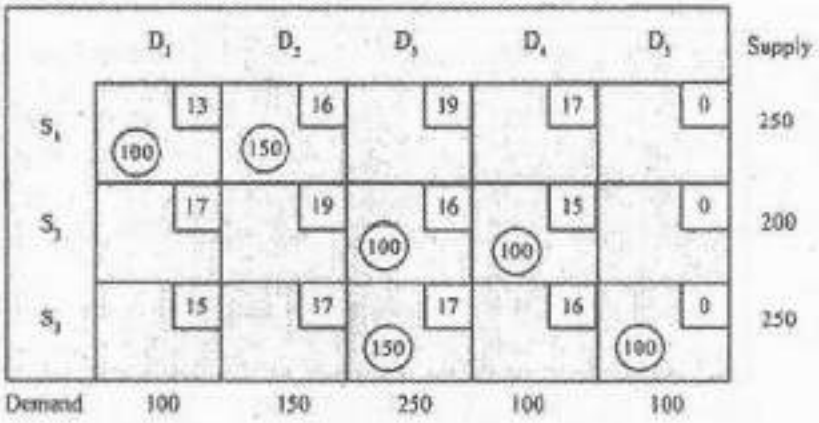
Fourth Allocation in VAM



With the fourth allocation, column D_4 is used. In the only left column D_3, the allocations of 100 units and 150 units are done in route S_2D_3 and S_4D_3, respectively. Thus, we obtain the following allocations using the Vogel's Approximation Method or VAM.

NOTES

Final Allocation Through Vogel's Method



The initial cost for this allocation is:

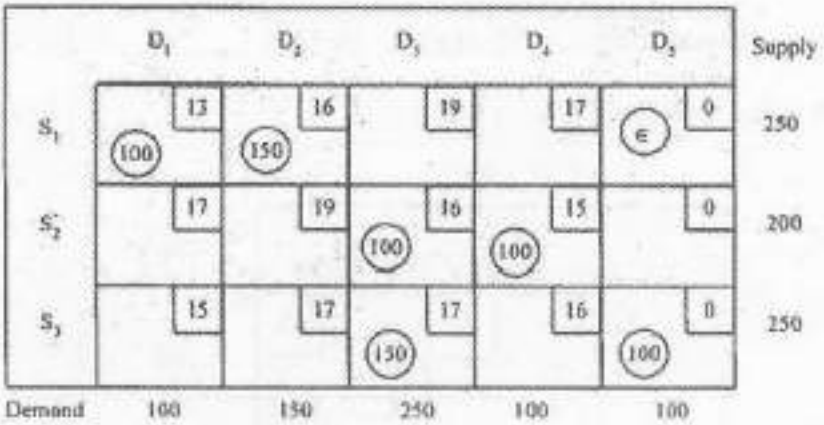
$(13 \times 100 + 16 \times 150 + 16 \times 100 + 15 \times 100 + 17 \times 150 + 0 \times 100)$ or equal to Rs **9350**

Step 3: Verify for degeneracy, $(m + n - 1) = 7$.

Number of filled cell = 6, which is one less than $(m + n + 1)$. Hence, go to Step 4 for removing the degeneracy.

Step 4: Now we allocate in the least cost unfilled cell. This cell is route S_1D_5 or S_2D_5. Select route S_1D_5. We obtain the following matrix after removing degeneracy.

Final Allocation After Removing Degeneracy in Vogel's Method



Optimization of Initial Assignment

The initial feasible assignment done by using Vogel's approximation method does not guarantee optimal solution. Hence, next step is to check the optimality of the initial solution.

Step 5: Check the optimality of the initial solution. For this, calculate the opportunity cost of un-occupied routes.

NOTES

First, we start with any row (or column). Select row 1, i.e., source S_1. For this row define row value, $u_1 = 0$. Now consider all filled routes of this row. For these routes, calculate column values v using following equation:

$$u_i + v_j = C_{ij} \text{ (For any filled route)}$$

Where u_i = Row value

v_j = Column value

C_{ij} = Unit cost of assigned route

Once first set of column values is known, say in this case v_j is known, locate other routes of filled cells in these columns. Calculate next of u_i or v_j values using the above equation. In this method, for all rows and columns, u_i and v_j values are determined for a non-degenerate initial solution.

Step 6: Check the optimality.

Calculate the opportunity of non-allocated or unfilled routes. For this, use the following equation:

$$\textbf{Opportunity unassigned route} = u_i + v_j - C_{ij}$$

Where, u_i = Row value

v_j = Column value

C_{ij} = Unit cost of unassigned route

If the opportunity cost is negative for all unassigned routes, the initial solution is optimal. If in case any of the opportunity costs is positive, then go to next step.

	D_1	D_2	D_3	D_4	D_5	Supply	u_i
S_1	13 (100)	16 (50)	19	17	8 (100)	250	0
S_2	17	19	16 (100)	15 (100)	0	200	0
S_3	15	17 (100)	17 (150)	16	0	250	1
Demand	100	150	250	100	100		
v_j	13	16	16	15	0		

Step 7: Make a loop of horizontal and vertical lines which joins some filled routes with the unfilled route, which has a positive opportunity cost. Note that all the corner points of the loop are either filled cells or positive opportunity cost unassigned cells.

For this, we start with row, S_1 and take $u_1 = 0$. Now S_1D_1, S_1D_2, and S_1D_5 are filled cells. Hence, for filled cells ($v_j = C_{ij} - u_i$).

$v_1 = 13 - 0 = 13$

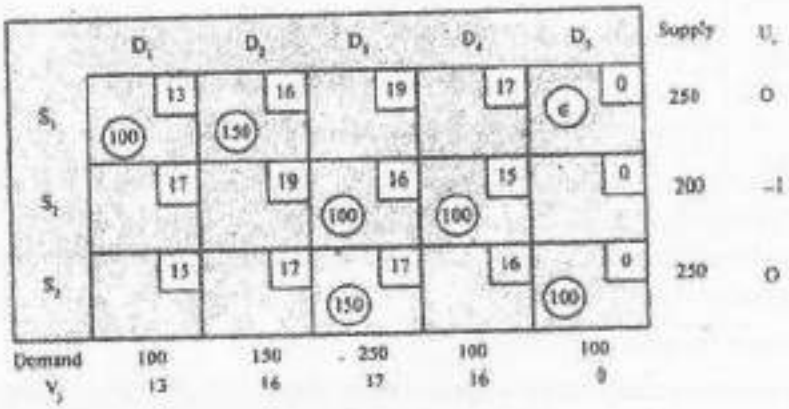
$v_2 = 16 - 0 = 16$

$v_5 = 0 - 0 = 0$

NOTES

The optimality of Vogel's method's initial solution is as follows.

Calculation of u_i and v_j for Vogel Approximation Method's Initial Solutions



Opportunity cost of above assignment using VAM is as follows:

Unassigned Route	Opportunity Cost ($u_i + v_j - C_{ij}$)
S_1D_3	0 + 17 – 19 = –2
S_1D_4	0 + 16 – 17 = –1
S_2D_1	–1 + 13 – 17 = –5
S_2D_2	–1 + 16 – 19 = –4
S_2D_5	–1 + 0 – 0 = –1
S_3D_1	0 + 13 – 15 = –2
S_3D_2	0 + 16 – 17 = –1
S_3D_4	0 + 16 – 16 = 0

Since all opportunity costs are negative or zero, the initial assignment of Vogel's solution is optimal with total cost of Rs. 9350.

Example 3.7: Distances between factory and its warehouses and demand at each warehouse are given in the following table. Calculate the values of penalty to all the rows and columns for the reduced transportation problem and repeat the same procedure till the entire requirement has been met. Solve this problem using Vogel's Approximation Method or VAM.

Table: Transportation Table

Factory/Warehouse	W1	W2	W3	Supply
F1	16	22	14	200
F2	18	14	18	150
F3	8	14	16	100
Demand	**175**	**125**	**150**	

Solution: The solution is obtained as follows.

Step 1: Compute the penalty for each row and column of the transportation problems. The penalty for the first row is, (16 - 14) = 2. Similarly the values of penalty for the second and the third row are 4 and 6.

	W_1	W_2	W_3	Supply	Penalty
F_1	16	22	14	200	2
F_2	18	14	18	150	4
F_3	8	14	16	100	6
Demand	175	125	150		
	8	**0**	**2**		

Similarly, the values of penalty for the first, second and the third columns are 8, 0 and 2 respectively.

Step 2: Identify the row or column with the largest penalty value. In this case, the first column with a penalty value is 8.

Step 3: The cell with the least cost is chosen and the possible number of goods is assigned to that cell. Therefore, assign 100 to the cell (F_3, W_1).

Step 4: If the remaining row supply or column demand is zero, remove that row/ column.

Now, the transportation problem can be reduced as illustrated in the following table:

	W_1	W_2	W_3	Supply	Penalty
F_1	16	22	14	200	2
F_2	18	14	18	150	4
Demand	**75**	**125**	**150**		
Penalty	**2**	**8**	**4**		

Step 5: The process is repeated for the reduced transportation problem till the entire supply at the factories is assigned to satisfy the demand at different warehouses.

Now, the W_2 column has the highest penalty, i.e., 8. Therefore, assign 125 units to the cell (F_2, W_2) since the cell has the least cost in the W_2 column.

Then the transportation problem can further be reduced as illustrated in the following table:

	W_1	W_3	Supply	Penalty
F_1	16	14	200	2
F_2	18	18	25	0
Demand	**75**	**150**		
Penalty	**2**	**4**		

NOTES

Now, the W_3 column has the highest penalty, i.e., 4. Next assign 150 units to the cell (F_1, W_3) since the cell has the least cost. Then remove the W_3 column and the remaining units are assigned to the cells (F_1, W_1) and (F_2, W_1). Thus, 50 units are assigned to the cell (F_1, W_1) and 25 units to the cell (F_2, W_1).

Since the number of cells occupied is 5, i.e., (3+3−1), the solution obtained is a feasible solution. Thus, the cost obtained using VAM is:

$(50 \times 16) + (25 \times 18) + (100 \times 8) + (125 \times 14) + (150 \times 14) =$ Rs.5,900.

3.3 OPTIMALITY AND MODI METHOD OF SOLVING TRANSPORTATION PROBLEMS

Optimality Test

Once the initial basic feasible solution has been computed, the next step in the problem is to determine whether the solution obtained is optimum or not.

Optimality test can be conducted to any initial basic feasible solution of a TP provided such allocations has exactly $m + n - 1$ non-negative allocations, where m is the number of origins and n is the number of destinations. Also, these allocations must be in independent positions.

To perform this optimality test, we shall discuss the MOdified DIstribution (MODI) method. The various steps involved in the MODI method for performing optimality test are as follows:

MODI Method

Step 1: Find the initial basic feasible solution of a TP by using any one of the three methods.

Step 2: Find out a set of numbers u_i and v_j for each row and column satisfying $u_i + v_j = C_{ij}$ for each occupied cell. To start with, we assign a number '0' to any row or column having maximum number of allocations. If this maximum number of allocations is more than one, choose any one arbitrarily.

Step 3: For each empty (unoccupied) cell, we find the sum u_i and v_j written in the bottom left corner of that cell.

Step 4: Find out for each empty cell the net evaluation value $\Delta_{ij} = C_{ij} - (u_i + v_j)$ and which is written at the bottom right corner of that cell. This step gives the optimality conclusion,

(*i*) If all $\Delta_{ij} > 0$ (i.e., all the net evaluation value) the solution is optimum and a unique solution exists.

(*ii*) If $\Delta_{ij} \geq 0$ then the solution is optimum, but an alternate solution exists.

(*iii*) If at least one $\Delta_{ij} < 0$, the solution is not optimum. In this case, we go to the next step to improve the total transportation cost.

Step 5: Select the empty cell having the most negative value of Δ_{ij}. From this cell we draw a closed path by drawing horizontal and vertical lines with the corner cells occupied. Assign sign '+' and '−' alternately and find the minimum allocation from the cell having negative sign. This allocation should be added to the allocation having positive sign and subtracted from the allocation having negative sign.

NOTES

Step 6: The previous step yields a better solution by making one (or more) occupied cell as empty and one empty cell as occupied. For this new set of basic feasible allocations repeat from the Step (2) till an optimum basic feasible solution is obtained.

Example 3.8: Solve the following transportation problem.

		Destination				
		P	*Q*	*R*	*S*	*Supply*
Source	*A*	21	16	25	13	11
	B	17	18	14	23	13
	C	32	17	18	41	19
	Demand	6	10	12	15	43

Origin\Dest	*P*	*Q*	*R*	*S*	*Supply*	P_1	P_{II}	P_{III}	P_{IV}	P_V	P_{VI}
A	21	16	25	13 ⑪	11	3	–	–	–	–	–
B	17 ⑥	18	14 ⑤	23 ④	13	3	4	4	4 ←	–	–
C	32	17 ⑩	18 ⑨	41	19	1	1	1	1	1	17
Demand	6	10	12	15	43						
P_1	4	1	4	10↑							
P_{II}	15	1	4	18↑							
P_{III}	15↑	1	4	–							
P_{IV}	–	1	4	–							
P_V	–	17	18↑	–							
P_{VI}	–	17↑	–	–							

Solution: We first find the initial basic feasible solution by using VAM. Since $\Sigma a_i = \Sigma b_j$ the given TP is a balanced one. Therefore, there exists a feasible solution.

Finally, we have the initial basic feasible solution as given in the following table:

		Destination			
		P	*Q*	*R*	*S*
Source	*A*	21	16	25	13 ⑪
	B	17 ⑥	18	14 ⑤	23 ④
	C	32	17 ⑩	18 ⑨	41

From this table, we see that the number of non-negative independent allocations is $6 = m + n - 1 = 3 + 4 - 1$.

Hence, the solution is non-degenerate basic feasible.

∴ The Initial Transportation Cost

$= 11 \times 13 + 3 \times 14 + 4 \times 23 + 6 \times 17 + 17 \times 10 + 18 \times 9 = ₹\ 711.$

To Find the Optimal Solution

We apply the MODI method in order to determine the optimum solution. We determine a set of numbers u_i and v_j for each row and column, with $u_i + v_j = C_{ij}$ for each occupied cell. To start with we give $u_2 = 0$ as the second row has the maximum number of allocation.

Now, we find the sum u_i and v_j for each empty cell and enter at the bottom left corner of that cell.

$C_{21} = u_2 + v_1 = 17 = 0 + v_1 = 17 \Rightarrow v_1 = 17$

$C_{23} = u_2 + v_3 = 14 = 0 + v_3 = 14 \Rightarrow v_3 = 14$

NOTES

$C_{24} = u_2 + v_4 = 23 = 0 + v_4 = 23 \Rightarrow v_4 = 23$

$C_{14} = u_1 + v_4 = 13 = u_1 + 23 = 13 \Rightarrow u_1 = 10$

$C_{33} = u_3 + v_3 = 18 = u_3 + 14 = 18 \Rightarrow u_3 = 4$

Next, we find the net evaluations $\Delta_{ij} = C_{ij} - (u_i + v_j)$ for each unoccupied cell and enter at the bottom right corner of that cell.

Initial Table

		C_j	4	3	0	0	$-M$	0	
C_B	B	X_B	W_1	W_2	S_1	S_2	A_1	S_3	Min X_B/w_2
0	S_1	1	1	1	1	0	0	0	1
$-M$	A_1	1	0	1	0	-1	1	0	1
←0	S_3	1	-1	②	0	0	0	1	$\frac{1}{2}$
	Z_j	$-M$	0	$-M$	0	M	$-M$	0	
	$Z_j - C_j$		-4	$-M-3$	0	M	0	0	
				↑					Min $\frac{X_B}{W_1}$
←0	S_1	1/2	(3/2)	0	1	0	0	$-1/2$	1/3
$-M$	A_1	1/2	1/2	0	0	-1	1	$-1/2$	1
3	W_1	1/2	$-1/2$	1	0	0	0	1/2	–
	Z_j	$-\frac{1}{2}M - 3/2$	$\frac{1}{2}M - 3/2$	3	0	M	$-M$	$\frac{1}{2}M + 3/2$	
	$Z_j - C_j$		$-\frac{M}{2} + \frac{5}{2}$ ↑	0	0	M	0	$\frac{M+3}{2}$	

Since all $\Delta_{ij} > 0$, the solution is optimal and unique. The optimum solution is given by,

$X_{14} = 11; X_{21} = 6; X_{23} = 3; X_{24} = 4; X_{32} = 10; X_{33} = 9$

The Minimum Transportation Cost,

$= 11 \times 13 + 17 \times 6 + 3 \times 14 + 4 \times 23 + 10 \times 17 + 9 \times 18 = ₹\ 711.$

Degeneracy in Transportation Problem

In a TP, if the number of non-negative independent allocations is less than $m + n - 1$, where m is the number of origins (rows) and n is the number of destinations (columns) there exists a degeneracy. This may occur either at the initial stage or at the subsequent iteration.

To resolve this degeneracy, we adopt the following steps.

Step 1: Among the empty cell, we choose an empty cell having the least cost which is of an independent position. If this cell is more than one, choose any one arbitrarily.

Step 2: To the cell as chosen in Step (1) we allocate a small positive quantity $\in > 0$.

The cells containing '$\in$' are treated like other occupied cells and degeneracy is removed by adding one (more) accordingly. For this modified solution, we adopt the steps involved in MODI method till an optimum solution is obtained.

Example 3.9: Solve the transportation problem for minimization.

		Destinations			
		1	2	3	Capacity
Sources	1	2	2	3	10
	2	4	1	2	15
	3	1	3	1	40
	Demand	20	15	30	65

Solution: Since $\Sigma a_i = \Sigma b_j$ the problem is a balanced TP. Hence, there exists a feasible solution. We find the initial solution by North West Corner rule as given here.

	1	2	3	Capacity
1	(10) 2	2	3	10
2	(10) 4	(5) 1	2	15
3	1	(10) 3	(30) 1	40
Demand	20	15	30	

Since, the number of occupied cells = 5 = $m + n - 1$ and all the allocations are independent we get an initial basic feasible solution.

The Initial Transportation Cost

$= 10 \times 2 + 4 \times 10 + 5 \times 1 + 10 \times 3 + 1 \times 30 = 20 + 40 + 5 + 30 + 30 = ₹ 125.$

To Find the Optimal Solution (MODI Method)

We use the previous table to apply the MODI method. We find out a set of numbers u_i and v_j for which $u_i + v_j = C_{ij}$ only for occupied cell. To start with as the maximum number of allocations is 2 in more than one row and column, we choose arbitrarily column 1 and assign a number 0 to this column, i.e., $v_1 = 0$. The remaining numbers can be obtained as follows.

$C_{11} = u_1 + v_1 = 2 = u_1 + 0 = 2 \Rightarrow u_1 = 2$

$C_{21} = u_2 + v_1 = 4 \Rightarrow u_2 = 4 - 0 = 4$

$C_{22} = u_2 + v_2 = 1 \Rightarrow v_2 = 1 - u_2 = 1 - 4 = -3$

$C_{32} = u_3 + v_2 = 3 = u_3 = 3 - v_2 = 3 - (-3) = 6$

$C_{33} = u_3 + v_1 = 1 = v_3 = 1 - u_3 = 1 - 6 = -5$

Initial Table

	1	2	3	u_i
1	(10) 2	−1 2 3	−3 6	$2 = u_1$
2	(10) − 4	(5) + 1	−1 2 3	$4 = u_2$
3	6 + 1 −5	(10) − 3	(30) 1	$6 = u_3$
v_j	$v_1 = 0$	$v_2 = -3$	$v_3 = -5$	

NOTES

We find the sum of u_i and v_j for each empty cell and write at the bottom left corner of that cell. Find the net evaluations $\Delta_{ij} = C_{ij} - (u_i + v_j)$ for each empty cell and enter at the bottom right corner of the cell.

The solution is not optimum as the cell (3, 1) has a negative Δ_{ij} value. We improve the allocation by making this cell namely (3, 1) as an allocated cell. We draw a closed path from this cell and assign sign '+' and '–' alternately. From the cell having '–' sign we find the minimum allocation given by Min (10, 10) = 10. Hence, we get two occupied cells.

(2, 1) (3, 2) becomes empty and the cell (3, 1) is occupied and resulting in a degenerate solution.

Number of allocated cells = $4 < m + n - 1 = 5$.

We get degeneracy. To resolve we add the empty cell (1, 2) and allocate $\in > 0$. This cell namely (1, 2) is added as it satisfies the two steps for resolving the degeneracy. We assign a number 0 to the first row, namely $u_1 = 0$ we get the remaining numbers as follows.

$$C_{11} = u_1 + v_1 = 2 \Rightarrow v_1 = 2 - u_1 = 2 - 0 = 2$$
$$C_{12} = u_1 + v_2 = 2 \Rightarrow v_2 = 2 - u_1 = 2 - 0 = 2$$
$$C_{31} = u_3 + v_1 = 1 \Rightarrow u_3 = 1 - v_1 = 1 - 2 = -1$$
$$C_{33} = u_3 + v_3 = 1 \Rightarrow v_3 = 1 - u_3 = 1 - (-1) = 2$$
$$C_{22} = u_2 + v_2 = 1 \Rightarrow u_2 = 1 - v_2 = 1 - 2 = -1$$

Next, we find the sum of u_i and v_j for the empty cell and enter at the bottom left corner of that cell and also the net evaluation $\Delta_{ij} = C_{ij} - (u_i + v_j)$ for each empty cell and enter at the bottom right corner of the cell.

Iteration Table

	1	2	3	u_i
1	2 ⑩	2 ∈	3 2 1	0
2	4 1 3	1 ⑤	2 1 1	–1
3	1 ⑩	3 1 2	1 ⑳	–1
v_j	2	2	2	

The modified solution is given in the following table. This solution is also optimal and unique as it satisfies the optimality condition that all $\Delta_{ij} > 0$.

	1	2	3	Capacity
1	2 (10)	2 (∈)	3	10
2	4	1 (15)	2	15
3	1 10	3	1 30	40
Demand	(20)	15	(30)	65

$X_{11} = 10; X_{22} = 15; X_{33} = 30; X_{12} = \in; X_{31} = 10$

Total Cost $= 10 \times 2 + 2 \times \in + 15 \times 1 + 10 \times 1 + 30 \times 1$

$= 75 + 2 \in = ₹ 75.$

Unbalanced Transportation

The given TP is said to be unbalanced if $\Sigma a_i \neq \Sigma b_j$, i.e., if the total supply is not equal to the total demand.

There are following two possible cases:

Case (i): $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$

If the total supply is less than the total demand, a dummy source (row) is included in the cost matrix with zero cost; the excess demand is entered as a rim requirement for this dummy source (origin). Hence, the unbalanced transportation problem can be converted into a balanced TP.

Case (ii): $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$

Therefore, the total supply is greater than the total demand. In this case, the unbalanced TP can be converted into a balanced TP by adding a dummy destination (column) with zero cost. The excess supply is entered as a rim requirement for the dummy destination.

Example 3.10: Solve the transportation problem when the unit transportation costs, demands and supplies are as given below:

		Destination				
		D_1	D_2	D_3	D_4	Supply
Origins	O_1	6	1	9	3	70
	O_2	11	5	2	8	55
	O_3	10	12	4	7	70
	Demand	85	35	50	45	

Solution: Since the total demand $\Sigma b_j = 215$ is greater than the total supply $\Sigma a_i = 195$, the problem is an unbalanced TP.

We convert this into a balanced TP by introducing a dummy origin O_4 with cost zero and giving supply equal to 215–195=20 units. Hence, we have the converted problem as follows:

NOTES

Origins	Destination	D_1	D_2	D_3	D_4	Supply
	O_1	6	1	9	3	70
	O_2	11	5	2	8	55
	O_3	10	12	4	7	70
	O_4	0	0	0	0	20
	Demand	85	35	50	45	215

As this problem is balanced, there exists a feasible solution to this problem. Using VAM we get the initial solution.

	D_1	D_2	D_3	D_4	Supply	P_1	P_{II}	P_{III}	P_{IV}	P_V	P_{VI}	P_{VII}
O_1	6 (65)	1 (5)	9	3	70	2	2	2	–	–	–	–
O_2	11	5 (30)	2 (25)	8	55	3	3	3	3	6 ←	–	–
O_3	10	12	4 (25)	7 (45)	70	3	3	3	3	3	3	4 ←
O_4	0 (20)	0	0	0	20	0	–	–	–	–	–	–
Demand	85	35	50	45								
P_1	6↑	1	2	3								
P_{II}	4↑	4	2	3								
P_{III}	–	4↑	2	4								
P_{IV}	–	7↑	2	1								
P_V	–	–	2	1								
P_{VI}	–	–	4	7↑								
P_{VII}	–	–	4	–								

The initial solution to the problem is given by,

	D_1	D_2	D_3	D_4
O_1	6 (65)	1 (5)	9	3
O_2	11	5 (30)	2 (25)	8
O_3	10	12	4 (25)	7 (45)
O_4	0 (20)	0	0	0

There are 7 independent non-negative allocations equal to $m + n - 1$. Hence, the solution is a non-degenerate one.

The Total Transportation Cost,

$= 6 \times 65 + 5 \times 1 + 5 \times 30 + 2 \times 25 + 4 \times 25 + 7 \times 45 + 20 \times 0$

$= ₹ 1010.$

To Find the Optimal Solution

We apply the steps in the MODI method to the previous table.

	D_1	D_2	D_3	D_4	u_i
O_1	[6] (65) −	[1] + (5)	[9] 0 9	[3] 3 0	0
O_2	[11] 10 1	[5] (30)	[2] + (25)	[8] 5 3	4
O_3	[10] 12 −2	12 7 5	[4] −(25)	[7] (45)	6
O_4	[0] (20)	[0] −5 5	[0] −4 4	[0] −5 5	
v_j	6	1	−2	1	

Since all $\Delta_{ij} \geq 0$ the solution is not optimum and to solve we introduce the cell (3,1) as this cell has the most negative value of Δ_{ij}. We modify the solution by adding and subtracting the minimum allocation given by Min (65, 30, 25). While doing this, the occupied cell (3, 3) becomes empty.

First Iteration Table

	D_1	D_2	D_3	D_4	u_i
O_1	[6] (40)	[1] (30)	[9] −2 11	[3] 3 0	6
O_2	[11] 10 1	[5] (5)	[2] (50)	[8] 7 1	10
O_3	[10] (25)	[12] 5 7	[4] 2	[7] (45)	10
O_4	[0] (20)	[0] −5 5	[0] −8 8	[0] −3 3	0
v_j	0	−5	−8	−3	

As the number of independent allocations are equal to $m+n-1$, we check the optimality.

Since, all $\Delta_{ij} \geq 0$, the solution is optimal and an alternate solution exists as $\Delta_{14} = 0$. Therefore, the optimum allocation is given by,

$X_{11} = 40$; $X_{12} = 30$; $X_{22} = 5$; $X_{23} = 50$; $X_{31} = 25$; $X_{34} = 45$; $X_{41} = 20$

The Optimum Transportation Cost

$= 6 \times 40 + 1 \times 30 + 5 \times 5 + 2 \times 50 + 10 \times 25 + 7 \times 45 + 0 \times 20$

$= ₹\ 960$.

Example 3.11: A product is produced by 4 factories F_1, F_2, F_3 and F_4. Their unit production cost are ₹ 2, 3, 1 and 5 respectively. Production capacity of the factories are 50, 70, 40 and 50 units respectively. The product is supplied to 4 stores S_1, S_2, S_3 and S_4 and the requirements of which are 25, 35, 105 and 20, respectively.

Find the transportation plan such that the total production and transportation cost is minimum. Unit costs of transportation are given as follow:

NOTES

NOTES

Factory	Stores				
		S_1	S_2	S_3	S_4
	F_1	2	4	6	11
	F_2	10	8	7	5
	F_3	13	3	9	12
	F_4	4	6	8	3

Solution: We form the transportation table which consists of both production and transportation costs.

	S_1	S_2	S_3	S_4	*Capacity*
F_1	4	6	8	13	50
F_2	13	11	10	8	70
F_3	14	4	10	13	30
F_4	9	11	13	8	50
Demand	25	35	105	20	

Total capacity = 200 units

Total demand = 185 units

Therefore, $\sum a_i > \sum b_j$. Hence, the problem is unbalanced. We convert it into a balanced one by adding a dummy store S_5 with cost 0 and the excess supply is given as the rim requirement to this store, namely (200–185) units.

	S_1	S_2	S_3	S_4	S_5	*Supply*
F_1	4 (25)	6 (5)	8 (20)	13	0	50
F_2	13	11	10 (50)	8 (20)	0	70
F_3	14	4 (30)	10	13	0	30
F_4	9	11	13 (35)	8	0 (15)	50
Demand	25	35	105	20	15	200

The initial basic feasible solution is obtained by least cost method. We get the solution containing 8 non-negative independent allocations equal to $m + n - 1$. So the solution is a non-degenerate solution.

The Total Transportation Cost,

$= 4 \times 25 + 6 \times 5 + 8 \times 20 + 10 \times 50 + 8 \times 20 + 4 \times 30 + 13 \times 35 + 0 \times 15$

$= ₹ 1525.$

To Find the Optimal Solution

We apply MODI method to the previous table as it has $m + n - 1$ independent non-negative allocations.

Initial Table

	S_1		S_2		S_3		S_4		S_5		u_i
F_1	(25)	4	(5)	6	(20)	8	6	13 7	−5	0 5	0
F_2	6	13 7	8	11 3	(50)	10	(20)	8 –	−3	0 3	2
F_3	2	14 12	(30)	4	6	10 4	4	3 9	−7	0 7	−2
F_4	9	9 0	11	11 0	(35)	13 –	11	8 −3		0 (15)	5
v_j		4		6		8		6		−5	

The solution is not optimum as the cell (4, 4) is having a negative net evaluation value, i.e., $\Delta_{4,4} = -3$. We draw a closed path from this cell and have a modified allocation by adding and subtracting the allocation Min (35, 20) = 20. This modified allocation is given in the table.

First Iteration Table

	S_1		S_2		S_3		S_4		S_5		u_i
F_1	(25)	4	(5)	6	(20)	8	3	3 0	−5	0 5	0
F_2	6	13 7	8	11 3	(70)	10	(5)	8 −3	−3	0 3	2
F_3	2	14 12	(30)	4		10	1	13 12	−7	0 7	−2
F_4	9	9 0	11	11 0	(15)	13	20	8	(15)	0	5
v_j		4		6		8		3		−5	

Since all the values of $\Delta_{ij} \geq 0$ the solution is optimum, but an alternate solution exists.

The optimum solution or the transportation plan is given by,

$X_{11} = 25$ units; $X_{12} = 30$ units

$X_{12} = 5$ units; $X_{43} = 15$ units

$X_{33} = 20$ units; $X_{44} = 20$ units

$X_{23} = 70$ units; $X_{45} = 15$ units

This is the surplus capacity that are not transported which are manufactured in factory F_4. The optimum production with transportation cost is as follows:

$-4 \times 25 + 6 \times 5 + 8 \times 20 + 10 \times 70 + 4 \times 30 + 8 \times 20 + 13 \times 15 + 0 \times 15$

= ₹ 1465.

NOTES

Check Your Progress

8. How is an optimality test conducted?
9. When is a solution said to be optimal transportation problem.
10. State the necessary and sufficient conditions for the existence of a feasible solution to a transportation problem.
11. What is the purpose of the MODI method?

NOTES

3.4 ASSIGNMENT PROBLEM

The assignment problem is used to find the best possible assignment for the given situations.

3.4.1 Basics of Assignment Problem

The assignment problem is one of the fundamental combinatorial optimization problems. It helps to find a maximum weight identical in nature in a weighted bipartite graph. The assignment problem is also termed as a special case of transportation problem.

Suppose there are n jobs to be performed and n persons are available for doing these jobs. Assume that each person can do each job at a time, though with varying degrees of efficiency. Let C_{ij} be the cost if the ith person is assigned to the jth job. The solution to the problem is to find an assignment (which job should be assigned to which person on one-one basis) so that the total cost of performing all jobs is minimum. Problems of this kind are known as assignment problems.

The assignment problem can be stated in the form of $n \times n$ cost matrix $[C_{ij}]$ of real numbers as given in the following table:

		Jobs				
		1	2	3... j...		n
Persons	1	C_{11}	C_{12}	C_{13}...	C_{1j}...	C_{1n}
	2	C_{21}	C_{22}	C_{23}...	C_{2j}...	C_{2n}
	3	C_{31}	C_{32}	C_{33}...	C_{3j}...	C_{3n}
	i	C_{i1}	C_{i2}	C_{i3}...	C_{ij}...	C_{in}
	n	C_{n1}	C_{n2}	C_{n3}...	C_{nj}...	C_{nn}

Mathematical Formulation of the Assignment Problem

Mathematically, the assignment problem can be stated as follows:

Minimize, $Z = \sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}x_{ij}$ $i = 1, 2, \ldots, n$

$j = 1, 2, \ldots, n$

Subject to the restrictions:

$$x_{ij} = \begin{cases} 1 \text{ if the } i\text{th person is assigned } j\text{th job} \\ 0 \text{ if not} \end{cases}$$

$\sum_{j=1}^{n} x_{ij} = 1$ (one job is done by the ith person)

and $\sum_{i=1}^{n} x_{ij} = 1$ (only one person should be assigned the jth job)

Where x_{ij} denotes that the jth job is to be assigned to the ith person.

3.4.2 Job Sequence Problem

NOTES

Job sequencing is basically the planning of the jobs in sequential manner and is an essential part of any work. Without proper planning and scheduling one can not achieve the desired output and profit. For sequencing a job, generally the two techniques are used termed as Priority Rules and Johnson's Rules. Priority rules give the guidelines for properly sequencing the job, where as Johnson's rule is used to minimize the completion time for a set of jobs to be done on two different machines. Using these rules one can assign jobs and maximize product and profit.

Basic Characteristics of Job Sequencing

1. Only one single job should be scheduled for a machine at a time.
2. Do not stop the process in between before completion.
3. New processing can be started after the completion of the previous processing.
4. Any job is scheduled for processing as per the order and due date requirements.
5. If the jobs are transferred from one machine to another due to some reason, then the time involved in transferring the jobs is considered negligible.

Priority Rules: These rules are used to get specific guidelines for job sequencing. The rules do not consider job setup cost and time while analysing processing times. In it job processing time and due dates are given importance because the due dates are fixed to give delivery in time to the customers. The rules are very useful for process-focussed amenities, for example health clinics, print shops and manufacturing industries. Hence, priority rules minimize the time for completing a job, sequences the jobs in the organization, checks if any job is late and maximizes resource utilization. The most popular priority rules are as follows:

- **First Come First Serve (FCFS):** The job to be processed first is the job that turned up first in the organization.
- **Earliest Due Date (EDD):** The job to be processed first is the job that has earliest due date.
- **Shortest Processing Time (SPT):** The job to be processed first and completed is the job that is shortest in nature; in other words the job can be processed in short time.
- **Longest Processing Time (LPT):** The job to be processed first is the job that is very important or of high priority though it can take longer processing time.
- **Critical Ratio (CR):** The job to be processed first is analysed on the basis of critical ratio, which is an index number calculated from time remaining until due date divided by the remaining work time.

Johnson's Rule: This rule is applied to minimize the completion time for a set of jobs that are to be processed on two different machines or at two consecutive work stations. The main objectives of the rules are,

- To minimize the processing time while sequencing a set of jobs on two different machines or work stations.
- To minimize the complete idle time on the processing machines.
- To minimize the flow time of the job, i.e., from the start of the first job until the completion of the last job.

NOTES

Necessary Conditions for Johnson's Rules: The necessary conditions to efficient complete the processing of the jobs are as follows:

- Knowledge about job time for each job at the specific work station.
- Job time must not depend on sequencing of jobs.
- All the jobs to follow the predefined work sequence.
- Avoid job priority.

Four Steps Johnson's Rule: The following are the important Four Steps in Johnson rule:

Step 1: List all the jobs and the processing time of each machine to which thes jobs are scheduled.

Step 2: Choose the job which has the shortest processing time. If the shorte time has been scheduled on the first machine or work station then the job is selecte first for processing. In case the shortest time is scheduled on the second machine work station then the job is processed at the end.

Step 3: After scheduling the job for processing go to Step 4.

Step 4: Repeat Step 2 again to schedule the processing of remaining jobs and fill the sequence columns towards the centre till all the jobs are scheduled.

The following example will help you to understand how the sequences in scheduled.

For example, there are five jobs to be done at a factory and each job must b processed through two work stations at two different machines, drill machine an lathe machine. Using Johnson's rule we can schedule the sequence of jobs.

The time (in hours) for processing each job is given in the following table:

Jobs	*Work Station 1 (Drill)*	*Work Station 2 (Lathe)*
A	5	2
B	3	6
C	9	4
D	12	8
E	8	14

Using the Steps of Johnson's rule, the job processing sequences are scheduled a follows:

Step 1: In the given table, the job with the shortest processing time is job **A**, in wor station 2 (with a time of 2 hours). Because it is at the second work station, schedule A last.

				A

Step 2: Next shortest time is of job **B** (with a time of 3 hours). Because it is at the firs work station, schedule it at first priority and eliminate it from the list.

B				A

Step 3: The next shortest time is of job **C** (with a time of 4 hours), but it is at th second work station. Therefore, place it at last before **A**.

B			C	A

NOTES

Step 4: There is a tie between job **D** (with a time of 8 hours at work station 2) and job E (with a time of 8 hours at work station 1) for the shortest remaining job. Because job E is at the first work station, so place it first after job **B**. Then place job **D** in the last sequencing position. You will get the job sequence schedule as follows:

B	E	D	C	A

The final sequential times at both the work stations will be:

Jobs	*B*	*E*	*D*	*C*	*A*
Work Station 1 (Drill)	3	8	12	9	5
Work Station 2 (Lathe)	6	14	8	4	2

3.5 SOLUTIONS OF ASSIGNMENT PROBLEM BY HUNGARIAN METHOD

The solution of an assignment problem can be arrived at by using the Hungarian method. The steps involved in this method are as follows:

Step 1: Prepare a cost matrix. If the cost matrix is not a square matrix then add a dummy row (column) with zero cost element.

Step 2: Subtract the minimum element in each row from all the elements of the respective rows.

Step 3: Further modify the resulting matrix by subtracting the minimum element of each column from all the elements of the respective columns. Thus, obtain the modified matrix.

Step 4: Then, draw minimum number of horizontal and vertical lines to cover all zeroes in the resulting matrix. Let the minimum number of lines be N. Now, there are two possible cases.

Case (i) If $N = n$, where n is the order of matrix an optimal assignment can be made. So make the assignment to get the required solution.

Case (ii) If $N < n$, then proceed to Step 5.

Step 5: Determine the smallest uncovered element in the matrix (element not covered by N lines). Subtract this minimum element from all uncovered elements and add the same element at the intersection of horizontal and vertical lines. Thus, the second modified matrix is obtained.

Step 6: Repeat Steps (3) and (4) until we get the Case (*i*) of Step 4.

Step 7: To make zero assignment examine the rows successively until a row with exactly single zero is found. Circle (O) this zero to make the assignment. Then mark a cross (×) over all zeroes if lying in the column of the circled zero showing that they can not be considered for future assignment. Continue in this manner until all the zeroes have been examined. Repeat the same procedure for the columns also.

Step 8: Repeat the Step 6 successively until one of the following situation arises:

Case (i) If no unmarked zero is left, then the process ends.

Case (ii) If there lies more than one of the unmarked zero in any column or row, then circle one of the unmarked zeroes arbitrarily and mark a cross in the cells of remaining zeroes in its row or column. Repeat the process until no unmarked zero is left in the matrix.

Check Your Progress

12. What is an assignment problem?
13. How is an assignment problem presented to find the minimum cost?

NOTES

Step 9: Thus, exactly one marked circled zero in each row and each column of the matrix is obtained. The assignment corresponding to these marked circled zeroes will give the optimal assignment.

Example 3.12: Using the following cost matrix and determine (*i*) Optimal job assignment and (*ii*) The cost of assignments.

		Jobs				
		1	2	3	4	5
Mechanics	*A*	10	3	3	2	8
	B	9	7	8	2	7
	C	7	5	6	2	4
	D	3	5	8	2	4
	E	9	10	9	6	10

Solution: Select the smallest element in each row and subtract this smallest element from all the other elements in its row.

	1	2	3	4	5
A	8	1	1	0	6
B	7	5	6	0	5
C	5	3	4	0	2
D	1	3	6	0	2
E	3	4	3	0	4

Subtract the minimum element from each column and subtract this element from all the elements in its column. With this we get the first modified matrix.

	1	2	3	4	5
A	7	0	0	0	4
B	6	4	5	0	3
C	4	2	3	0	0
D	0	2	5	0	0
E	2	3	2	0	2

In this modified matrix, we draw the minimum number of lines to cover all zeroes (horizontal and vertical).

	1	2	3	4	5
A	~~7~~	~~0~~	~~0~~	~~0~~	~~4~~
B	6	4	5	0	3
C	4	2	3	0	0
D	0	2	5	0	0
E	2	3	2	0	2

Number of lines drawn to cover all zeroes is $N = 4$.

The order of matrix is $n = 5$.

Hence, $N < n$, less than order of matrix.

NOTES

Now, we get the second modified matrix by subtracting the smallest uncovered element from the remaining uncovered elements and adding it to the element at the point of intersection of lines.

	1	2	3	4	5
A	9	0	8	2	6
B	6	2	3	0	3
C	4	8	1	8	0
D	0	8	3	8	8
E	2	1	0	8	2

Number of lines drawn to cover all zeroes is $N = 5$

The order of matrix is $n = 5$

Hence, $N = n$, i.e., order of matrix. Now, we determine the optimum assignment.

Assignment

	1	2	3	4	5
A	9	0	8	2	6
B	6	2	3	0	3
C	4	8	1	8	0
D	0	8	3	8	8
E	2	1	0	8	2

The first row contains more than one zero. So, proceed to the second row. It has exactly one zero. The corresponding cell is $(B, 4)$. Circle this zero thus making an assignment. Mark (×) for all other zeroes in its column to show that they cannot be used for making other assignments. Now, row 5 has a single zero in the cell $(E, 3)$. Make an assignment in this cell and cross the 2nd zero in the 3rd column.

Now, row 1 has a single zero in the column 2, i.e., in the cell $(A, 2)$.

Make an assignment in this cell and cross the other zeroes in the 2nd column. This leads to a single zero in the column 1 of the cell $(D, 1)$. Make an assignment in this cell and cross the other zeroes in the 4th row. Finally, we have a single zero left in the 3rd row making an assignment in the cell $(C, 5)$. Thus, we have the following assignment.

Job	*Mechanic*	*Cost*
1	D	3
2	A	3
3	E	9
4	B	2
5	C	4
		₹ 21

Therefore, $1 \rightarrow D$, $2 \rightarrow A$, $3 \rightarrow E$, $4 \rightarrow B$, $5 \rightarrow C$ with minimum cost ₹ 21.

NOTES

3.6 VARIATIONS IN ASSIGNMENT PROBLEMS

The following situations are considered as variations in assignment problem:

3.6.1 Unbalanced Assignment Problem

Any assignment problem is said to be unbalanced if the cost matrix is not a square matrix, i.e., the number of rows and columns are not equal. To make it balanced, we add a dummy row or dummy column with all the entries as zero.

Example 3.13: A company has 4 machines to do 3 jobs. Each job can be assigned to only one machine. The cost of each job on each machine is given below. Determine the job assignments that will minimize the total cost.

		Machines			
		W	*X*	*Y*	*Z*
Jobs	*A*	18	24	28	32
	B	8	13	17	18
	C	10	15	19	22

Solution: Since the cost matrix is not a square matrix, we add a dummy row *D* with all the elements 0.

	W	*X*	*Y*	*Z*
A	18	24	28	32
B	8	13	17	18
C	10	15	19	22
D	0	0	0	0

Subtract the minimum element in each row from all the elements in its row.

	W	*X*	*Y*	*Z*
A	0	6	10	14
B	0	5	9	10
C	0	5	9	12
D	0	0	0	0

Since each column has a minimum element 0, we draw minimum number of lines to cover all zeros.

	W	*X*	*Y*	*Z*
A	0	6	10	14
B	0	5	9	10
C	0	5	9	12
D	0	0	0	0

∴ The number of lines drawn to cover all zeroes is N = 2 < n = 4, which is less than the order of matrix, hence we form a second modified matrix.

	W	*X*	*Y*	*Z*
A	0		5	9
B	0	0	4	5
C	0	0	4	7
D	5	0	0	0

Here, $N - 3 < n - 4$, which is less than the order of matrix.

Again we subtract the smallest uncovered element from all the uncovered elements and add to the element at the point of intersection

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Here, $N = 4 = n = 4$ = Order of matrix. Hence, we make an assignment.

Assignment

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 \\ 9 & 4 & 0 & 0 \end{bmatrix} \end{array}\begin{array}{l} A \to W \\ B \to X \\ C \to Y \\ D \to Z \end{array}$$

or

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 \\ 9 & 4 & 0 & 0 \end{bmatrix} \end{array}\begin{array}{l} A \to W \\ B \to Y \\ C \to X \\ D \to Z \end{array}$$

Since D is a dummy job, machine Z is assigned no job.

Therefore, optimum cost = 18 + 13 + 19 = ₹ 50.

3.6.2 Maximization in Assignment Problem

In this, the objective is to maximize the profit. To solve this, we first convert the given profit matrix into the loss matrix by subtracting all the elements from the highest element. For this converted loss matrix we apply the steps in Hungarian method to get the optimum assignment.

Example 3.14: The owner of a small machine shop has four mechanics available to assign jobs for the day. Five jobs are offered with expected profit for each mechanic on each jobs, which are as follows:

Jobs

$$\begin{array}{c} \\ \textit{Mechanics} \\ \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 62 & 78 & 50 & 111 & 82 \\ 71 & 84 & 61 & 73 & 59 \\ 87 & 92 & 111 & 71 & 81 \\ 48 & 64 & 87 & 77 & 80 \end{bmatrix} \end{array}\begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \end{array}$$

By using the assignment method, find the assignment of mechanics to the job that will result in maximum profit. Which job should be declined ?

Solution: The given profit matrix is not a square matrix as the number of jobs is not equal to the number of mechanics. Hence, we introduce a dummy mechanic 5 with all the elements 0.

Jobs

$$\begin{array}{c} \\ \textit{Mechanics} \\ \\ \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 62 & 78 & 50 & 111 & 82 \\ 71 & 84 & 61 & 73 & 59 \\ 87 & 92 & 111 & 71 & 81 \\ 48 & 64 & 87 & 77 & 80 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}\begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

NOTES

Now we convert this profit matrix into loss matrix by subtracting all the elements from the highest element 111.

Loss Matrix

	A	B	C	D	E
1	49	33	61	0	29
2	40	27	50	38	52
3	24	19	0	40	30
4	63	47	24	34	31
5	111	111	111	111	111

We subtract the smallest element from all the elements in the respective rows.

	A	B	C	D	E
1	49	33	61	0	29
2	13	0	23	11	25
3	24	19	0	40	30
4	39	23	0	10	7
5	0	0	0	0	0

Since each column has minimum element as zero, we draw minimum number of lines to cover all zeros.

	A	B	C	D	E
1	49	33	61	0	29
2	13	0	23	11	25
3	24	19	0	40	30
4	39	23	0	10	7
5	0	0	0	0	0

Here the number of lines drawn to cover all zeroes is $N = 4$, which is less than the order of matrix.

We form the second modified matrix by subtracting the smallest uncovered element from the remaining uncovered elements and adding to the element that is at the point of intersection of lines.

	A	B	C	D	E
1	49	40	68	0	29
2	6	0	23	4	18
3	17	19	0	33	23
4	32	23	0	3	0
5	0	7	7	0	0

Here, $N = 5 = n = 5$ = Order of matrix.

We make the assignment.

	A	B	C	D	E
1	49	40	68	0	29
2	6	0	23	4	18
3	17	19	0	33	23
4	32	23	0	3	0
5	0	7	7	0	0

The optimum assignment is:

Jobs	*Mechanics*
A	5
B	2
C	3
D	1
E	4

Since the 5th mechanic is a dummy, job *A* assigned to the 5th mechanic is declined.

The maximum profit is given by, 84 + 111 + 111 + 80 = ₹ 386.

3.7 SUMMING UP

- The transportation problem is one of the subclasses of LPP (Linear Programming Problem) in which the objective is to transport various quantities of a single homogeneous commodity that are initially stored at various origins to different destinations in such a way that the transportation cost is minimum.
- Any set of non-negative allocations ($X_{ij} > 0$) which satisfies the row and column sum is called a feasible solution.
- A feasible solution is called a basic feasible solution if the number of non-negative allocations is equal to $m + n - 1$, where m is the number of rows and n is the number of columns in a transportation table.
- Optimal solution is a feasible solution (not necessarily basic) which minimizes the total cost.
- Initial solution can be obtained by using any one of the three methods, viz. North West Corner Rule (NWCR), Least Cost Method or Matrix Minima Method and Vogel's Approximation Method (VAM).
- VAM is preferred over the other two methods since the initial basic feasible solution obtained by this method is either optimal or very close to the optimal solution.
- The cells in the transportation table can be classified as occupied cells and unoccupied cells. The allocated cells in the transportation table are called occupied cells and empty cells in a transportation table are called unoccupied cells.
- The improved solution of the initial basic feasible solution is called optimal solution which is the second stage of solution that can be obtained by MODI (MOdified DIstribution Method).
- Maximization case in transportation problem is done by maximizing the total profit for which the profit matrix is given.
- Once the initial basic feasible solution has been computed, the next step in the problem is to determine whether the solution obtained is optimum or not.
- The given TP is said to be unbalanced if the total supply is not equal to the total demand.
- If the total supply is less than the total demand, a dummy source (row) is included in the cost matrix with zero cost; the excess demand is entered as a

NOTES

Check Your Progress

14. Name the method used for solving assignment problems.
15. How many steps are there in the Hungarian method?
16. How is the presence of an alternate optimal solution established?
17. How do you convert the maximization assignment problem into a minimization one?
18. How will you solve an assignment problem if profit is to be maximized?
19. What is an unbalanced assignment? How can it be balanced?

NOTES

rim requirement for this dummy source (origin). Hence, the unbalanced transportation problem can be converted into a balanced TP.

- The assignment problem is one of the fundamental combinatorial optimization problems. It helps to find a maximum weight identical in nature to a weighted bipartite graph. The assignment problem is also termed as a special case of transportation problem.
- Job sequencing is basically the planning of the jobs in sequential manner and is an essential part of any work.
- Priority rules give the guidelines for properly sequencing the job whereas Johnson's rule is used to minimize the completion time for a set of jobs to be done on two different machines. Using these rules one can assign jobs and maximize product and profit.
- Any assignment problem is said to be unbalanced if the cost matrix is not a square matrix, i.e., the number of rows and columns are not equal. To make it balanced we add a dummy row or dummy column with all the entries as zero.

3.8 KEY TERMS

- **Transportation problem:** A problem for transportation of various quantities of a single homogeneous commodity, initially stored at various origins to different destinations at the minimum cost
- **Feasible solution:** A set of non-negative allocations where some quantity is transferred from an origin i to a destination j, ($X_{ij} > 0$) and satisfies the row and column sum is a feasible solution
- **Basic feasible solution:** A feasible solution where number of non-negative allocations is equal to $m + n - 1$, where m is the number of rows and n is the number of columns in a transportation table
- **Non-degenerate basic feasible solution:** Any feasible solution to a transportation problem containing m origins and n destinations is known as non-degenerate, if it contains $m + n - 1$ occupied cells and each allocation is in independent positions
- **Degenerate basic feasible solution:** A basic feasible solution that contains less than $m + n - 1$ non-negative allocations
- **Assignment problem:** A problem of finding an assignment of jobs among persons so that total cost of performing all jobs is at minimum
- **Cost matrix:** An assignment problem can be stated in form of $n \times n$ matrix $[C_{ij}]$ with i rows and j columns of real numbers. Here, C_{ij} is the cost when ith person is assigned jth job
- **Unbalanced assignment:** An assignment is unbalanced if the cost matrix is not a square matrix

3.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Transportation problem deals with transportation of various quantities of a single homogeneous commodity initially stored at various origins to different destinations at the minimum cost.
2. One must know: (*i*) Amount and location of available supplies, (*ii*) Quantity demanded at destination and (*iii*) Costs of transporting one unit of commodity from various origins to various destinations.

NOTES

3. Transportation problems are presented in tabular form with m rows (as origins) and n columns (as destinations). Every cell of this table has cost as C_{ij} of transporting one unit of product from ith origin to jth destination and X_{ij} which is the quantity transported from ith origin to jth destination.
4. If there are m origins and n destinations and C_{ij} be the cost of transporting one unit of product from ith origin to jth destination and X_{ij} be the quantity transported from ith origin to jth destination and Z is the cost given then it can be mathematically represented as:

   $Z = \sum_{i=1}^{n} \sum_{j=1}^{m} C_{ij} X_{ij}$ where $1 <= i <= n$ and $1 <= j <= m$.

5. Feasible solution: Any set of non-negative allocations ($X_{ij} > 0$) which satisfies the row and column sum is called a feasible solution.

   Basic feasible solution: A feasible solution is called a basic feasible solution if the number of non-negative allocations is equal to $m + n - 1$, where m is the number of rows n the number of columns in a transportation table.

   Non-degenerate basic feasible solution: Any feasible solution to a transportation problem containing m origins and n destinations is said to be non-degenerate, if it contains $m + n - 1$ occupied cells and each allocation is in independent positions.

6. The solution of a Transportation Problem (TP) can be obtained in two stages, namely initial solution and optimum solution.
7. Initial solution can be obtained by using any one of the following three methods:
   (*i*) North West Corner Rule (NWCR)
   (*ii*) Least Cost Method or Matrix Minima Method
   (*iii*) Vogel's Approximation Method (VAM)
8. Optimality test can be conducted to any initial basic feasible solution of a TP provided such allocations has exactly $m + n - 1$ non-negative allocations, where m is the number of origins and n is the number of destinations. Also, these allocations must be in independent positions.
9. The basic feasible solution to a transportation problem is said to be optimal if it minimizes the total transportation cost.
10. The necessary and sufficient condition for the existence of a feasible solution is a solution that satisfies all conditions of supply and demand.
11. The purpose of MODI method is to get the optimal solution of a transportation problem.
12. It is a problem of finding an assignment of jobs among persons so that total cost of performing all jobs is at minimum.
13. An assignment problem can be stated in form of $n \times n$ matrix $[C_{ij}]$ with i rows and j columns of real numbers. Here, C_{ij} is the cost when ith person is assigned jth job.
14. Hungarian method is used to solve assignment problems.
15. There are nine steps in the method of solution by Hungarian method.
16. If the final cost matrix contains more than a required number of zeroes at independent positions, then it indicates the presence of an alternate optimal solution.

NOTES

17. The maximization assignment problem can be converted into minimization assignment problem by subtracting all the elements in the given profit matrix from the highest element in that matrix.
18. The given profit matrix can be converted into a loss matrix or minimization type by subtracting all the elements from the highest element of the given matrix. For this minimization problem apply steps of the Hungarian method to get an optimal assignment.
19. If the cost matrix of the assignment problem is not a square matrix, the problem is said to be unbalanced. To make it balanced, extra row(s) or column(s) are added with zero elements.

3.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is transportation problem?
2. What is transportation algorithm?
3. List the merits and limitations of the North West Corner rule.
4. Vogel's Approximation Method results in the most economical initial basic feasible solution. How?
5. How will you identify if a transportation problem has an alternate optimal solution or not?
6. What is the coefficient of X_{ij} of constraints in a transportation problem?
7. What is an initial basic feasible solution?
8. When does degeneracy occur in an $m \times n$ transportation problem?
9. What is a cost matrix?
10. Write the first three steps of the Hungarian procedure for solving an assignment problem.
11. While solving a problem using the Hungarian method at what stage are you able to make an assignment? Explain in brief.
12. What do you mean by maximization in an assignment problem?

Long-Answer Questions

1. Give the mathematical formulation of a transportation problem.
2. Write an algorithm to solve a transportation problem.
3. Define the terms feasible solution, basic solution, non-degenerate solution and, optimal solution in a transportation problem.
4. Explain the following briefly with examples:
   (*i*) North West Corner Rule
   (*ii*) Least Cost Method
   (*iii*) Vogel's Approximation Method
5. Explain degeneracy in a transportation problem. Describe a method to resolve it.
6. What do you mean by an unbalanced transportation problem? Explain the process of converting an unbalanced transportation problem into a balanced one.

7. What do you understand by transportation model?

8. Obtain the initial solution for the following transportation problem using (*i*) North West Corner Rule, (*ii*) Least Cost Method and (*iii*) VAM

		Destination			
		A	*B*	*C*	*Supply*
Source	1	2	7	4	5
	2	3	3	1	8
	3	5	4	7	7
	4	1	6	2	14
	Demand	7	9	18	34

9. Solve the following transportation problem where the cell entries denote the unit transportation costs.

		Destination				
		A	*B*	*C*	*D*	*Supply*
Origin	*P*	5	4	2	6	20
	Q	8	3	5	7	30
	R	5	9	4	6	50
	Demand	10	40	20	30	100

10. Solve the following transportation problem.

		Destination			
		1	**2**	**3**	*Capacity*
Source	1	2	2	3	10
	2	4	1	2	15
	3	1	3	1	40
	Demand	20	15	30	

11. Find the minimum transportation cost.

		Warehouse				
		D_1	D_2	D_3	D_4	*Supply*
Factory	F_1	19	30	50	10	7
	F_2	70	30	40	60	9
	F_3	40	8	70	30	18
	Demand	5	8	7	14	

12. Solve the following transportation problem using Vogel's Approximation Method.

		Warehouse						
		A	*B*	*C*	*D*	*E*	*F*	*Available*
Factory	1	9	12	9	6	9	10	5
	2	7	3	7	7	5	5	6
	3	6	5	9	11	3	11	2
	4	6	8	11	2	2	10	9
	Requirement	4	4	6	2	4	2	

NOTES

NOTES

13. Solve the following transportation problem to maximize the profit.

		Destination				
		A	*B*	*C*	*D*	*Supply*
Source	1	40	25	22	33	100
	2	44	35	30	30	30
	3	38	38	28	30	70
	Demand	40	20	60	30	

14. Write the steps for finding a solution using the Hungarian method.
15. Solve the following assignment problems:

(*i*)

Jobs \ *Men*	*A*	*B*	*C*	*D*
I	1	4	6	3
II	9	7	10	9
III	4	5	11	7
IV	8	7	8	5

(*ii*)

Jobs \ *Men*	*A*	*B*	*C*	*D*
I	10	25	15	20
II	15	30	5	15
III	35	20	12	24
IV	17	25	24	20

(*iii*)

Tasks \ *Men*	*A*	*B*	*C*	*D*	*E*
I	1	3	2	8	8
II	2	4	3	1	5
III	5	6	3	4	6
IV	3	1	4	2	2
V	1	5	6	5	4

16. There are five jobs to be assigned one each to 5 machines. The associated cost matrix is as follows.

Jobs \ *Machines*	1	2	3	4	5
A	11	17	8	16	20
B	9	7	12	6	15
C	13	16	15	12	16
D	21	24	17	28	26
E	14	10	12	11	15

How should the job be assigned to various machines?

17. A company is faced with the problem of assigning 4 machines to 6 different jobs (one machine to one job only). The profits are estimated as follows.

NOTES

		Machines			
		A	B	C	D
Jobs	1	3	6	2	6
	2	7	1	4	4
	3	3	8	5	8
	4	6	4	3	7
	5	5	2	4	3
	6	5	7	6	4

Solve the problem to maximize the profit.

3.11 REFERENCES AND SUGGESTED READINGS

Swaroop, Kanti, P.K. Gupta and Man Mohan. 2001. *Operations Research*. New Delhi: Sultan Chand & Sons.

Vohra, N. D. 2006. *Quantitative Techniques in Management*, 3rd edition. New Delhi: Tata McGraw-Hill Education.

Taha, Hamdy A. 1992. *Operations Research: An Introduction*. New York: Macmillan.

Gupta, P.K. and D.S. Hira. 2002. *Introduction to Operations Research*. New Delhi: S. Chand & Co.

Sharma, S. D. 2006. *Operations Research*. Uttar Pradesh: Kedar Nath Ram Nath & Co.

UNIT 4 THEORY OF GAMES AND INVESTMENT ANALYSIS

NOTES

Structure

4.0 Introduction
4.1 Objectives
4.2 Game Theory: An Introduction
4.3 Saddle Point
  4.3.1 The Maximin-Minimax Principle
  4.3.2 Games Without Saddle Points
4.4 Mixed Strategies
4.5 Use of Dominance
4.6 Limitations of Game Theory in Competition
4.7 Investment Analysis
  4.7.1 Methods of Investment Analysis
  4.7.2 Break-Even Analysis
  4.7.3 Methods of Evaluation and Capital Investment
  4.7.4 Payback Period Method
  4.7.5 Average (Accounting) Rate of Return Method
  4.7.6 Time Adjusted or Discounted Cash Flow (DCF) Method
  4.7.7 Discounted Rate and Certainty Equivalent Approach
  4.7.8 Expected Monetary Value and Hillier and Hertz's Models
4.8 Summing Up
4.9 Key Terms
4.10 Answers to 'Check Your Progress'
4.11 Questions and Exercises
4.12 References and Suggested Readings

4.0 INTRODUCTION

In this unit, you will learn about the concept of game theory. Competition is a watchword of modern life. We say that a competitive situation exists if two or more individuals are making decisions in situation that involves conflicting interests and in which the outcome is controlled by the decisions of all parties concerned. Game theory applies to those competitive situations which are technically known as 'competitive games' or simply 'games'. You will also learn about the saddle point, the maximin and minimax principle as well as the games that do not require saddle point. The concept of a dominated strategy is very useful for reducing the size of the payoff table. A mixed strategy is an assignment of a probability of each pure strategy. Investment is the employment of funds on assets with the aim of earning income or capital appreciation. Investments have two attributes–time and risk. Investment is basically putting aside and employing money in financial instruments in the present, with the expectation of a positive rate of return in future. In brevity, an investment is a sacrifice in the present in the expectation of a future gain. An investor needs to be updated with the investment schemes, the process and policies.

NOTES

4.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the concept of game theory
- Analyse the various types of games
- Understand the value of a game
- Explain saddle point and its significance
- Discuss the maximin and the minimax principle
- Know the games with and without saddle points
- Understand mixed strategies
- Discuss the various methods of investment analysis
- Understand the significance of break-even analysis
- Calculate payback period and average (accounting) rate of return
- Explain what Discounted Cash Flow (DCF) method is

4.2 GAME THEORY: AN INTRODUCTION

Competition is a watchword of modern life. We say that a competitive situation exists if two or more individuals are making decisions in situation that involves conflicting interests and in which the outcome is controlled by the decisions of all parties concerned. We assme that in a competitive situation, each participant acts in a rational manner and tries to resolve the conflict of interests in his favour. It is in this context that game theory has developed. Professor John von Neumann and Oscar Morgenstern published their book entitled 'The Theory of Games and Economic Behaviour' wherein they provided a new approach to many problems involving conflict situations—an approach now widely used in Economics, Business Administration, Sociology, Psychology and Political Science as well as in Military Training. Fundamentally, the theory of games attempts to provide an answer to the question: what may be considered a rational course of action for an individual confronted with a situation whose outcome depends not only upon his own actions but also upon the actions of others, who in turn, are faced with a similar problem of choosing a rational course of action? In fact, the theory of games is simply the logic of rational decisions.

Meaning of the Term 'Game'

The term 'game' represents a conflict between two or more parties. Game theory is really the 'science of conflict.' Game theory is not concerned with finding an optimum or winning strategy for a particular conflict situation but it provides general rules concerning the logic that underlies strategic behaviour of all types.

Game theory applies to those competitive situations which are technically known as 'competitive games' or simply 'games'. Situations, in order to be termed as games must possess the following properties:

1. The number of competitors is finite.
2. There is a conflict of interests between the participants.
3. Each of the participants has available to him a finite test of possible courses of action, i.e., has several choices as to his appropriate actions; this list being not necessarily the same for each competitor.

NOTES

4. The rules governing these choices are specified and known to all players; a play of the game results when each of the players chooses a single course of action from the list of courses available to him.
5. The outcome of the game is affected by choices made by all of the players; the choices are to be made simultaneously so that no competitor knows his opponent's choice until he is already committed to his own.
6. The outcome for all specific sets of choices by all of the players is known in advance and numerically defined. The outcome of a play consists of the particular set of courses of action undertaken by the competitors. Each outcome determines a set of payments (+ ve, –ve or zero), one to each competitor.

Standard Conventions in Game Theory

In order to eliminate the necessity for written descriptions of the 'payoffs' (as we have shown in the above two examples), a standard set of conventions has been established in game theory. It is the usual practice to omit a description like 'player X wins two points' and replace it with integer 2. The positive algebraic sign which is assumed to accompany this number indicates that it is player X who benefits from this payoff. Similarly, instead of saying 'player Y wins three points' one simply indicates this with the value –3, the minus sign indicating that it is player Y who benefits from this particular payoff.

Another standard convention which is usually followed is that player X has choices between the rows and player Y has choices between the columns.

Keeping these two conventions in view we can write the above stated two illustrations of games as shown:

ILLUSTRATION 1

	Y	
X	2	3
	–3	–1

ILLUSTRATION 2

	Y	
X	3	–4
	–2	1

A further convention in game theory is known as matrix notation. In this case, games are represented in the form of a matrix (a rectangular array of numbers deriving from matrix algebra). When games are expressed in this fashion, the resulting matrix is commonly known as *payoff matrix*. The above stated two illustrations can be put in the form of payoff matrices as follows:

ILLUSTRATION 1

$$\begin{array}{c} \\ X \end{array}\begin{array}{c} Y \\ \begin{bmatrix} 2 & 3 \\ -3 & -1 \end{bmatrix} \end{array}$$

ILLUSTRATION 2

$$\begin{array}{c} \\ X \end{array}\begin{array}{c} Y \\ \begin{bmatrix} 3 & -4 \\ -2 & 1 \end{bmatrix} \end{array}$$

The term *'strategy'* is often talked about in game theory. It refers to the total pattern of choices employed by any player. It may be defined as a complete set of plans of action specifying precisely what the player will do under every possible future contingency that might occur during the play of the game. For example, if player X chooses to play his first row half of the time and his second row half of the time, his strategy for the game is 1/2, 1/2. Thus, strategy of a player is the decision rule he uses for making the choice from his list of courses of action. *Strategy could be a pure or a mixed one.* When a player plays one row all of the time (or one column all of the time in the case of player Y) he is said to be playing a pure strategy. In a mixed

NOTES

strategy, player *X* will plan each of his rows a certain portion of the time and player Y will play each of his columns a certain part of the time. In business we have a close analogy when for example, a manager follows a certain course of action, say *A*, until an alternate course of action, say *B*, appears to be more profitable. If action *A* appears more attractive again, the manager switches back to it.

Types of Games

Games can be of several types. Important ones are as follows:

1. *Two-Person Games and n-Person Games:* In two-person games the players may have many possible choices open to them for each play of the game but the number of players remains only two. But games can as well involve many people as active participants, each with his own set of choices for each play of the game. In case of three players, it can be named as three-person game. Thus, in case of more than two persons, the game is generally named as *n*-person game.
2. *Zero-Sum and Non-Zero-Sum Game:* A zero-sum game is one in which the sum of the payments to all the competitors is zero for every possible outcome of the game. In other words, in such a game the sum of the points won equals the sum of the points lost, i.e., one player wins at the expense of the other (others). Two-person matrix game is always zero-sum game since one player loses what the other wins. But in a non-zero-sum game the sum of the payoffs from any play of the game may be either positive or negative but not zero. A well known example of a non-zero-sum game is the case of two competing firms each with a choice regarding its advertising campaign. If we assume that the market is quite small so that sales remain at nearly the same level whether the firms engage in advertising or not. If neither firm advertises, they both save the cost of advertising and thus each receives a positive benefit or payoff (representing a non-zero-sum outcome). In the event both firms spend equal amounts of money to advertise, neither firm's sales will increase given equal quality of the advertising campaigns and thus both firms will have incurred the cost of advertising without any benefits. Since both firms will lose money, this is clearly a non-zero-sum outcome.
3. *Games of Perfect Information and Games of Imperfect Information:* Whatever strategy is adopted by either player, if the same can also be discovered by his competitor, then such games are known as games of perfect information. In case of games of imperfect information neither player knows the entire situation and must be guided in part by his guess as to what the real situation is.
4. *Games with Finite, i.e., Limited Number of Moves (or plays) and Games with Unlimited Number of Moves:* Games with finite number of moves are those where the number of moves is limited to a fixed magnitude before play begins, but if the game could continue over an extended period of time and no limit is put on the number of moves then it is referred as a game with unlimited number of moves.
5. *Constant-Sum Game:* In some situation zero-sum game is correctly referred to as constant-sum game. Take, for example, a competitive struggle for share of the market by two firms (a duopoly). In such a case, every percentage point gained by one of the firms is necessarily lost by the other. This situation is ordinarily called a two person, zero-sum game. But strictly speaking this is a constant-sum rather than a zero-sum game because the sum of the two market

shares is the fixed number 100 (per cent), not zero. It should however, be remembered that there is no significant analytic difference between the constant-sum and the zero-sum games. This is because there is no change in strategic possibilities from a given game to another game in which some constant amount that cannot be changed by players is added to the original payoffs.

NOTES

6. *2 × 2 Two-Person Games and 2 × m and m × 2 Games:* Two person zero-sum games with only two choices open to each player are denoted as 2 × 2 two-person games but a game in which one of the players has more than two choices of rows or columns and in which the other player has exactly two choices is referred to as $m \times 2$ or $2 \times m$ game respectively.
7. *3 × 3 and Larger Games:* Two-person zero-sum games can be of size 3 × 3 and larger. If each of the two players has 3 choices then the game is of 3 × 3 type and if the choices open to any of the player or to both players are more than three, the game is referred to that of a larger size. 3 × 3 and larger games quite often present complications and in such situations linear programming as a solution method may allow us to solve for the optimum strategies.
8. *Negotiable (or Cooperative) and Non-Negotiable (or Non-Cooperative) Games:* Negotiation amongst the players is possible in *n*-person and non-zero-sum games but the same is not necessarily required. On this basis we can divide the analysis of such games into two parts—games in which the participants can negotiate and the games in which negotiation is not permitted. The former type of games are known as negotiable games and the latter type of games are known as non-negotiable games.

4.3 SADDLE POINT

The saddle point in a payoff matrix is one which is the smallest value in its row and the largest value in its column. The saddle point is also known as equilibrium point in the theory of games. An element of a matrix that is simultaneously minimum of the row in which it occurs and the maximum of the column in which it occurs is a saddle point of the matrix game. In a game having a saddle point optimum strategy for player *X* is always to play the row containing a saddle point and for the player *Y* to play the column that contains a saddle point. Saddle point also gives the valus of such a game. Saddle point in a payoff matrix concerning a game may be there and may not be there. If there is a saddle point we can easily find out the optimum strategies and the value of the game by what is known as the solution by saddle point without doing much calculations. But when saddle point is not there we have to use algebraic methods for working out the solutions concerning game problems.

4.3.1 The Maximin-Minimax Principle

This principle is used for the selection of optimal strategies by two players. Consider two players *A* and *B*. *A* is a player who wishes to maximize his gains, while player *B* wishes to minimize his losses. Since *A* would like to maximize his minimum gain, we obtain for player *A*, the value called *maximin value* and the corresponding strategy is called the *maximin strategy*.

On the other hand, since player *B* wishes to minimize his losses, a value called the *minimax value*, which is the minimum of the maximum losses is found. The corresponding strategy is called the *minimax strategy*. When these two are equal (maximin value = minimax value), the corresponding strategies are called *optimal*

NOTES

strategies and the game is said to have a *saddle point*. The value of the game is given by the saddle point.

The selection of maximin and minimax strategies by A and B is based upon the so called maximin-minimax principle, which guarantees the best of the worst results.

Saddle Point: A saddle point is a position in the payoff matrix where the maximum of row minima coincides with the minimum of column maxima. The payoff at the saddle point is called the *value* of the game.

We shall denote the maximin value by $\underline{\gamma}$, the minimax value of the game by $\overline{\gamma}$ and the value of the game by γ.

Notes:

1. A game is said to be fair if,
   Maximin value = Minimax value = 0, i.e., if $\overline{\gamma} = \underline{\gamma} = 0$
2. A game is said to be strictly determinable if,
   Maximin value = Minimax value $\neq 0$. $\underline{\gamma} = \gamma = \overline{\gamma}$.

Example 4.1: Solve the game whose payoff matrix is given by,

Player B

Player A	B_1	B_2	B_3
A_1	1	3	1
A_2	0	−4	−3
A_3	1	5	−1

Solution:

Player B

Player A	B_1	B_2	B_3	Row minima
A_1	1	3	1	1
A_2	0	−4	−3	−4
A_3	1	5	−1	−1
Column maxima	1	5	1	

Maxi (minimum) = Max (1, − 4, − 1) = 1

Mini (maximum) = Min (1, 5, 1) = 1

i.e., Maximin value $\underline{\gamma} = 1$ = Minimax value $\overline{\gamma}$

∴ Saddle point exists. The value of the game is the saddle point, which is 1. The optimal strategy is the position of the saddle point and is given by (A_1, B_1).

Example 4.2: For what value of λ is the game with the following matrix strictly determinable?

Player B

Player A	B_1	B_2	B_3
A_1	λ	6	2
A_2	−1	λ	−7
A_3	−2	4	λ

NOTES

Solution: Ignoring the value of λ, the payoff matrix is given by,

		Player B			
		B_1	B_2	B_3	Row minima
Player A	A_1	λ	6	2	2
	A_2	−1	λ	−7	−7
	A_3	−2	4	λ	−2
Column maxima		−1	6	2	

The game is strictly determinable if,

$$\underline{\gamma} = \gamma = \overline{\gamma}. \text{ Hence, } \underline{\gamma} = 2, \quad \overline{\gamma} = -1$$

$$\Rightarrow \qquad -1 \leq \lambda \leq 2.$$

Example 4.3: Determine which of the following two-person zero-sum games are strictly determinable and fair. Give the optimum strategy for each player in the case of strictly determinable games.

(*i*)

		Player B	
		B_1	B_2
Player A	A_1	−5	2
	A_2	−7	−4

(*ii*)

		Player B	
		B_1	B_2
Player A	A_1	1	1
	A_2	4	−3

Solution:

(*i*)

		Player B		
		B_1	B_2	Row minima
Player A	A_1	−5	2	−5
	A_2	−7	−4	−7
Column maxima		−5	2	

Maxi (minimum) = $\underline{\gamma}$ = Max (−5, −7) = −5

Mini (maximum) = $\overline{\gamma}$ = Min (−5, 2) = −5

Since $\underline{\gamma} = \overline{\gamma} = -5 \neq 0$, the game is strictly determinable. There exists a saddle point = − 5. Hence, the value of the game is −5. The optimal strategy is the position of the saddle point given by (A_1, B_1).

(*ii*)

		Player B		
		B_1	B_2	Row minimum
Player A	A_1	1	1	1
	A_2	4	−3	−3
Column maximum		4	1	

Maxi (minimum) = $\underline{\gamma}$ = Max (1, − 3) = 1.

Mini (maximum) = $\overline{\gamma}$ = Min (4, 1) = 1.

Since $\underline{\gamma} = \overline{\gamma} = 1 \neq 0$, the game is strictly determinable. Value of the game is 1. The optimal strategy is (A_1, B_2).

Example 4.4: Solve the game whose payoff matrix is given below.

$$\begin{bmatrix} -2 & 0 & 0 & 5 & 3 \\ 3 & 2 & 1 & 2 & 2 \\ -4 & -3 & 0 & -2 & 6 \\ 5 & 3 & -4 & 2 & -6 \end{bmatrix}$$

Solution:

Player B

Player A	B_1	B_2	B_3	B_4	B_5	Row minima
A_1	−2	0	0	5	3	−2
A_2	3	2	1	2	2	1
A_3	−4	−3	0	−2	6	−4
A_4	5	3	−4	2	−6	−6
Column maxima	5	3	1	5	6	

Maxi (minimum) = $\underline{\gamma}$ = Max (−2, 1, −4, −6) = 1.

Mini (maximum) = $\overline{\gamma}$ = Min (5, 3, 1, 5, 6) = 1.

Since, $\underline{\gamma} = \overline{\gamma} = 1$, there exists a saddle point. Value of the game is 1. The position of the saddle point is the optimal strategy and is given by $[A_2, B_3]$.

4.3.2 Games Without Saddle Points

A game without saddle point can be solved by various solution methods.

2 × 2 Games Without Saddle Point

Consider a 2 × 2 two-person zero-sum game without any saddle point, having the payoff matrix for player A as,

$$\begin{array}{c} \\ A_1 \\ A_2 \end{array} \begin{array}{c} \begin{array}{cc} B_1 & B_2 \end{array} \\ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \end{array}$$

The optimum mixed strategies,

$$S_A = \begin{bmatrix} A_1 & A_2 \\ p_1 & p_2 \end{bmatrix} \text{ and } S_B \begin{bmatrix} B_1 & B_2 \\ q_1 & q_2 \end{bmatrix}$$

where,

$$p_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}, p_1 + p_2 = 1 \Rightarrow p_2 = 1 - p_1$$

$$q_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}, q_1 + q_2 = 1 \Rightarrow q_2 = 1 - q_1$$

The value of the game $(\gamma) = \dfrac{a_{11}a_{22} - a_{12}a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}$

Example 4.5: Solve the following game and determine its value.

$$\begin{array}{c} \\ A \end{array} \begin{array}{c} \begin{array}{cc} B & \end{array} \\ \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \end{array}$$

NOTES

Solution: It is clear that the payoff matrix does not possess any saddle point. The players will use mixed strategies. The optimum mixed strategy for the players are,

$$S_A = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix}, p_1 + p_2 = 1$$

$$S_B = \begin{pmatrix} B_1 & B_2 \\ q_1 & q_2 \end{pmatrix}, q_1 + q_2 = 1$$

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} = \frac{4 - (-4)}{4 + 4 - (-4 - 4)} = \frac{8}{16} = \frac{1}{2}$$

$$p_2 = 1 - p_1 \Rightarrow p_2 = 1 - \frac{1}{2} = \frac{1}{2}$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} = \frac{4 - (-4)}{4 + 4 - (-4 - 4)} = \frac{8}{16} = \frac{1}{2}$$

$$q_2 = 1 - q_1 = 1 - \frac{1}{2} = \frac{1}{2}$$

The optimum strategy is, $S_A = \left(\frac{1}{2}, \frac{1}{2}\right)$; $S_B = \left(\frac{1}{2}, \frac{1}{2}\right)$

The value of the game is, $\gamma = \frac{a_{22}a_{11} - a_{12}a_{21}}{(a_{22} + a_{11}) - (a_{12} + a_{21})}$

$$= \frac{(4 \times 4) - [-4 \times (-4)]}{(4 + 4) - [-4 + (-4)]} = 0.$$

Example 4.6: In a game of matching coins with two players, suppose A wins one unit of value when there are two heads, wins nothing when there are two tails and losses 1/2 unit of value when there are one head and one tail. Determine the payoff matrix, the best strategies for each player and the value of the game to A.

Solution: The payoff matrix for the player A is given by,

$$\text{Player } A \quad \begin{matrix} & \text{Player } B \\ & \begin{matrix} H & T \end{matrix} \\ \begin{matrix} H \\ T \end{matrix} & \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 0 \end{bmatrix} \end{matrix}$$

Let this be,

$$\begin{matrix} & \begin{matrix} B_1 & B_2 \end{matrix} \\ \begin{matrix} A_1 \\ A_2 \end{matrix} & \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \end{matrix}.$$

The optimum mixed strategies,

$$S_A = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix}, p_1 + p_2 = 1$$

$$S_B = \begin{pmatrix} B_1 & B_2 \\ q_1 & q_2 \end{pmatrix}, q_1 + q_2 = 1.$$

NOTES

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} = \frac{0 - \left(-\frac{1}{2}\right)}{1 + 0 - \left(-\frac{1}{2} - \frac{1}{2}\right)}$$

$$= \frac{\frac{1}{2}}{2} = \frac{1}{4}$$

$$p_2 = 1 - p_1 = 1 - \frac{1}{4} = \frac{3}{4}$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} = \frac{0 - \left(-\frac{1}{2}\right)}{1 + 0 - \left(-\frac{1}{2} - \frac{1}{2}\right)} = \frac{1}{4}$$

$$q_2 = 1 - q_1 = 1 - \frac{1}{4} = \frac{3}{4}$$

$$\text{Value of the game} = \frac{1 \times 0 - \left(-\frac{1}{2}\right)\left(-\frac{1}{2}\right)}{1 + 0 - \left(-\frac{1}{2} - \frac{1}{2}\right)} = \frac{\frac{-1}{4}}{2}$$

$$\gamma = -\frac{1}{8}$$

∴ The optimum mixed strategy is given by,

$$S_A = \left(\frac{1}{4}, \frac{3}{4}\right); S_B = \left(\frac{1}{4}, \frac{3}{4}\right)$$

and the value of game is –1/8.

Example 4.7: Solve the following payoff matrix. Also determine the optimal strategies and value of the game.

$$\begin{matrix} & B \\ A & \begin{bmatrix} 5 & 1 \\ 3 & 4 \end{bmatrix} \end{matrix}$$

Solution:

$$\begin{matrix} & B \\ A & \begin{bmatrix} 5 & 1 \\ 3 & 4 \end{bmatrix} \end{matrix}$$

Let this be,

$$\begin{matrix} & B_1 & B_2 \\ A_1 & \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ A_2 & \end{matrix}$$

The optimum mixed strategies,

$$S_A = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix} \text{ and } S_B = \begin{pmatrix} B_1 & B_2 \\ q_1 & q_2 \end{pmatrix}$$

Where,

$$p_1 = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} = \frac{4 - 3}{(5 + 4) - (1 + 3)} = \frac{1}{5}$$

NOTES

$$p_2 = 1 - p_1 \Rightarrow p_2 = 1 - \frac{1}{5} = \frac{4}{5}$$

$$q_1 = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})} = \frac{4 - 1}{(5 + 4) - (1 + 3)} = \frac{3}{5}$$

$$q_2 = 1 - q_1 \Rightarrow q_2 = 1 - \frac{3}{5} = \frac{2}{5}$$

Value of game, $$\gamma = \frac{(5 \times 4) - (1 \times 3)}{(5 + 4) - (1 + 3)} = \frac{17}{5}$$

∴ The optimum mixed strategies,

$$S_A = \left(\frac{1}{5}, \frac{4}{5}\right); S_B = \left(\frac{3}{5}, \frac{2}{5}\right)$$

$$\text{Value of game} = \frac{17}{5}$$

4.4 MIXED STRATEGIES

In game theory, the strategy of a player in a game is a complete plan of action for any situation that may occur. This determines the complete behaviour of player and the player's strategy determines the action that the player will take at any stage of the game.

A strategy profile is also sometimes termed as strategy combination. It is a set of strategies for each player that specifies all actions in a game. It must include one and only one strategy for every player. Sometimes the strategy concept is by mistake confused with that of a move. A move is an action taken by a player at some point during the play of a game whereas a strategy is a complete algorithm for playing the game because it guides a player what to do for every possible situation throughout the game. A player's strategy set describes what strategies are available for playing the game. Strategies are of two types, pure and mixed. A pure strategy provides a complete definition of how a player will play a game. A mixed strategy is an assignment of a probability to each pure strategy. This allows for a player to randomly select a pure strategy. Since probabilities are continuous, there are infinitely many mixed strategies available to a player, even if their strategy set is finite. Certainly, a pure strategy can be considered as a degenerate case of a mixed strategy in which that specific pure strategy is selected with probability 1 and every other strategy with probability 0.

A totally mixed strategy is a mixed strategy in which the player assigns a strictly positive probability to every pure strategy. The totally mixed strategies are important for equilibrium refinement.

Consider the payoff matrix table of pure coordination game (see Table 4.1). Here one player chooses the row and the other chooses a column. The row player receives the first payoff, the column player the second. If row opts to play A with probability 1, i.e., play A for sure then the player is said to be playing a pure strategy. If column opts to flip a coin and play A if the coin lands heads and B if the coin lands tails then the player is said to be playing a mixed strategy and not a pure strategy.

Check Your Progress

1. Define the term game. What is the function of game theory?
2. In what areas is the game theory applied?
3. What are the properties that a game must possess? Name any two.
4. Define the term saddle point.
5. What is the use of the maximin-minimax principle?

NOTES

Table 4.1 Pure Coordination Game

	A	*B*
A	1,1	0,0
B	0,0	1,1

Example 4.8: Find the optimum strategies and the value of the game from the following payoff matrix concerning two-person game:

$$\begin{array}{c} \\ \text{Player } X \end{array} \begin{array}{c} \text{Player } Y \\ \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} \end{array}$$

Solution: In the given game, there is no saddle point because there is no one value which is smallest value in its row and largest in its column. Therefore the players will resort to what is known as mixed strategy, i.e., player X will play each of his rows a certain portion of time and player Y will play each of his columns a certain part of the time. The question then is to determine what proportion of the time a player should spend on his respective rows and columns. This can be done by the use of algebraic method stated as follows.

Let Q equal the proportion of time player X spends playing the first row, then $1–Q$ must equal the time he spends playing his second row (because one equals the time available for play). Similarly, suppose player Y spends time R in playing first column and $1–R$ proportion of time he spends playing the second column. All this can be stated as under:

$$\begin{array}{c} \\ \text{Player } X \end{array} \begin{array}{c} \begin{array}{cc} \text{Player } Y & \\ R & 1-R \end{array} \\ \begin{array}{c} Q \\ 1-Q \end{array} \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} \end{array}$$

Now, we must find out the values of Q and R. Let us analyse the situation from X's view point. He would like to devise a strategy that will maximize his winning (or minimize his losses) irrespective of what his opponent Y does. For this X would like to divide his play between his rows in such a manner that his expected winnings or losses when Y plays the first column will equal his expected winnings or losses when Y plays the second column. Expected winnings indicate the sum, overtime, of the payoffs multiplied by the probabilities that these payoffs will obtain. This can be calculated as shown below:

X's EXPECTED WINNINGS

	When Y Plays Column One	*When Y Plays Column Two*
X Plays Row One: Q of the Time	(1) (Q)	(4) (Q)
X Plays Row Two: $1–Q$ of the Time.	$5(1 - Q)$	$3(1 - Q)$
X's Total Expected Winnings	$Q + 5(1 - Q)$	$4Q + 3(1 - Q)$

Equating the expected winnings of X when Y plays column one with when Y plays column two we can find the value of Q as follows:

NOTES

$$Q + 5(1 - Q) = 4Q + 3(1 - Q)$$

or $$Q + 5 - 5Q = 4Q + 3 - 3Q$$

or $$-5Q = -2$$

or $$Q = 2/5$$

$\therefore$ $$1 - Q = 3/5$$

This means that player X should play his first row 2/5 of the time and his second row 3/5 of the time if he wants to maximize his expected winnings from the game.

On the similar basis expected losses of Y can be worked out as under:

Y's EXPECTED LOSSES

	Y Plays Column One R of the Time	*Y Plays Column Two 1—R of the Time*	*Y's Total Expected Losses*
When X Plays Row One	$1(R)$	$4(1 - R)$	$R + 4(1 - R)$
When X Plays Row Two	$5(R)$	$3(1 - R)$	$5R + 3(1 - R)$

Equating the expected losses of Y when X plays row one with when X plays row two, we can find the value of R as follows:

$$R + 4(1 - R) = 5R + 3(1 - R)$$

$$R + 4 - 4R = 5R + 3 - 3R$$

or $$-5R = -1$$

or $$R = 1/5$$

$\therefore$ $$1 - R = 4/5$$

This means that player Y should play his first column 1/5 of the time and his second column 4/5 of the time if he wants to minimize his expected losses in the game.

Now we can illustrate the original game with the appropriate strategies for each of the player as follows:

$$\begin{array}{ccc} & & \textit{Player } Y \\ & & \begin{array}{cc} 1/5 & 4/5 \end{array} \\ \textit{Player } X & \begin{array}{c} 2/5 \\ 3/5 \end{array} & \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} \end{array}$$

Alternative Method or Short Cut Method for Finding the Above Strategies

Original game

$$\begin{array}{cc} & Y \\ X & \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} \end{array}$$

Step 1: Subtract the smaller payoff in each row from the larger one and the smaller payoff in each column from the larger one.

$$\begin{array}{cccc} & Y & & \\ X & \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} & \begin{array}{c} 3 \\ 2 \end{array} & \begin{array}{l} \text{(i.e., } 4 - 1 = 3) \\ \text{(i.e., } 5 - 3 = 2) \end{array} \\ & \begin{array}{cc} 4 & 1 \end{array} & & \end{array}$$

(i.e. $5 - 1 = 4$) (i.e. $4 - 3 = 1$)

NOTES

Step 2: Interchange each of these pairs of subtracted numbers found in Step 1 above

$$\begin{array}{c} \\ X \end{array}\begin{array}{c} Y \\ \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} \\ 1 \quad 4 \end{array}\begin{array}{c} \\ \begin{matrix} 2 \\ 3 \end{matrix} \\ \\ \end{array}$$

Step 3: Put each of the interchanged numbers over the sum of the pair of numbers.

$$\begin{array}{c} \\ X \end{array}\begin{array}{c} Y \\ \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} \\ 1/(1+4) \quad 4/(1+4) \end{array}\begin{array}{c} \\ \begin{matrix} 2/(2+3) \\ 3/(2+3) \end{matrix} \\ \\ \end{array}$$

Step 4: Simplify the fraction to obtain the proper proportions or the required strategies.

$$\begin{array}{c} \\ X \end{array}\begin{array}{c} Y \\ \begin{bmatrix} 1 & 4 \\ 5 & 3 \end{bmatrix} \\ 1/5 \quad 4/5 \end{array}\begin{array}{c} \\ \begin{matrix} 2/5 \\ 3/5 \end{matrix} \\ \\ \end{array}$$

Now we determine the value of the game. Looking at the game from X's point of view we can argue as follows:

(*i*) During the 1/5 of the time Y plays column one, X wins 1 point 2/5 of the time (when X plays row one) and 5 points 3/5 of the time (when X plays row second).

(*ii*) During the 4/5 of the time Y plays column two, X wins 4 points 2/5 of the time (when X plays row one) and 3 points 3/5 of the time (when X plays row second).

Thus total expected winnings of player X are the sum of the above two statements as under:

$$\frac{1}{5}\left\{(1)\left(\frac{2}{5}\right)+(5)\left(\frac{3}{5}\right)\right\}+\frac{4}{5}\left\{(4)\left(\frac{2}{5}\right)+(3)\left(\frac{3}{5}\right)\right\}$$

$$=\left(\frac{1}{5}\right)\left(\frac{17}{5}\right)+\left(\frac{4}{5}\right)\left(\frac{17}{5}\right)$$

$$=\frac{17}{25}+\frac{68}{25}$$

$$=\frac{85}{25}$$

$$=\frac{17}{5}$$

Thus the value of the game is 17/5 which means that player X can expect to win an average payoff of 17/5 points for each play of the game if he adopts the strategy we have determined as stated above. If the value of the game determined above had a negative sign, it would simply signify that Y was the winner. The same result we can also get by looking at the game from Y's point of view doing similar calculations.

4.5 USE OF DOMINANCE

NOTES

The concept of dominance is very useful for reducing the size of the game. Applying this concept, we can convert a bigger game into a smaller game. Hence, we must examine the possibility of reducing the game size using the principle of dominance.

Rule for Dominance

(*i*) If all the elements in a column are greater than or equal to the corresponding elements in another column, then that column is dominated.

(*ii*) Similarly, if all the elements in a row are less than or equal to the corresponding elements in another row, then that row is dominated.

Dominated rows or columns may be deleted which reduces the size of the game. Always look for dominance when solving a game.

Example 4.9: Determine the optimum strategies and the value of the game from the following payoff matrix concerning a two person 4 × 2 game:

$$\begin{matrix} & Y \\ X & \begin{bmatrix} -6 & -2 \\ -3 & -4 \\ 2 & -9 \\ -7 & -1 \end{bmatrix} \end{matrix}$$

Solution: There is no saddle point in the given game. This game also cannot be solved by dominance because there is no row which is always preferred by X irrespective of what Y might do.

Player X can actually think of this 4 × 2 game as being six sub-games each of size 2 × 2 (because he can choose not to play any two of his rows if he so desires). These six sub-games can be enlisted as follows:

$$\begin{matrix} & & Y \\ & & 2/5 & 3/5 \\ X & 1/5 & -6 & -2 \\ & 4/5 & -3 & -4 \end{matrix}$$

← *Sub-game No.* (*i*)
Strategies have been noted in this matrix payoff.

Value of the game is $\left\{-\frac{6}{5}+\left(-\frac{12}{5}\right)\right\}$

$$= -\frac{18}{5} = -3.60$$

$$\begin{matrix} & & Y \\ & & 7/15 & 8/15 \\ X & 11/15 & -6 & -2 \\ & 4/15 & 2 & -9 \end{matrix}$$

← *Sub-game No.* (*ii*)
Strategies have been noted in this matrix payoff.

Value of the game is $\left(\frac{11}{15}\times -6\right)+\left(\frac{4}{15}\times 2\right)$

$$=\left(-\frac{66}{15}\right)+\left(\frac{8}{15}\right)=-\frac{58}{15}=-3.87$$

$$\begin{matrix} & Y \\ X & \begin{bmatrix} \textcircled{6} & -2 \\ -7 & -1 \end{bmatrix} \end{matrix}$$

← *Sub-game No. (iii)*
There is saddle point here.
Hence value to this sub-game is –6

$$\begin{matrix} & Y \\ X & \begin{bmatrix} -3 & \textcircled{-4} \\ 2 & -9 \end{bmatrix} \end{matrix}$$

← *Sub-game No. (iv)*
There is saddle point here.
Hence value to this sub-game is –4

$$\begin{matrix} & & Y & \\ & & 3/7 & 4/7 \\ X & 6/7 & \begin{bmatrix} -3 & -4 \\ -7 & -1 \end{bmatrix} \\ & 1/7 & \end{matrix}$$

← *Sub-game No. (v)*
Strategies have been noted in this matrix payoff
Value of the game is $\left(-\frac{18}{7}\right)+\left(-\frac{7}{7}\right)$

$$=-\frac{25}{7}=-3.57$$

$$\begin{matrix} & & Y & \\ & & 8/17 & 9/17 \\ X & 6/17 & \begin{bmatrix} 2 & -9 \\ -7 & -1 \end{bmatrix} \\ & 11/17 & \end{matrix}$$

← *Sub-game No. (vi)*
Strategies have been noted in this matrix payoff
Value of the game is

$$\left(\frac{12}{17}\right)+\left(-\frac{77}{17}\right)=-\frac{65}{17}=-3.82$$

If we look at the values of all these sub-games we find that, the values being negative, Y will win and X will lose in all the cases. Player X who is making the choice will naturally prefer the sub-game with the smallest negative value and this is sub-game No. (*v*) in the example. Thus X will play a two-row mixed strategy between the second and fourth rows of the original game and he will expect to lose an average of 3.57 points per play of the game which will be his minimum possible loss. Being an intelligent opponent player Y will also adopt a strategy consisting of the same number of columns. In brief, sub-game No. (*v*)th's strategies will be adopted by players X and Y with a game value equal to –3.57.

4.6 LIMITATIONS OF GAME THEORY IN COMPETITION

Game theory has been used in business and industry to develop bidding tactics, pricing policies, advertising strategies and timing of the introduction of new models into the market. The theory provides the basis for rational decisions and thus improves the decision making process. In spite of all this the theory of games suffers from certain limitations and this marks a question about the relevancy of game theory in practical life. The important limitations are:

(1) Businessmen do not have all the knowledge required by the theory of games. They do not even know all of the alternative strategies available to themselves and the strategies open to their rivals.

(2) There is a great deal of uncertainty that cannot conveniently be inserted into game theory models. The outcome of a game may not be known with certainty in many cases. It may, for example, depend upon the toss of a coin or upon

cutting a deck of cards. As such we usually restrict to games with known outcomes. This, therefore, is an important limitation.

(3) The minimax strategy implies that the businessman minimizes the chance of the maximum loss. It is thus a very conservative rule. It is a much less accurate description of the dynamic businessman who is constantly in quest of profit.

(4) The techniques of solving games involving mixed strategies particularly in case of a larger payoff matrix is very complicated and this lessens the significance of this sort of analysis. There are many situations concerning real games wherein the game theory is as yet quite helpless to resolve.

(5) Though the game theory points out some of the important aspects of competitive situation but still it is not a model that offers a complete solution particularly when the game does not happen to be a strictly adversary game. For example, in an oligopoly situation a reduction in price of a commodity may result in the expansion of the entire quantity demanded and not merely shift demand units from one firm, say *A* to another firm say *B*. The game theory will not be able to provide a complete solution to this sort of oligopoly game problem.

NOTES

4.7 INVESTMENT ANALYSIS

Investment analysis is the study of the likely return from a proposed investment with the objective of evaluating the amount an investor may pay for it. There are various methods of investment analysis, including cash on cash return, payback period, internal rate of return and net present value. Each provides some measure of the estimated return on an investment based on various assumptions and investment horizons.

4.7.1 Methods of Investment Analysis

Investment is the employment of funds on assets with the aim of earning income or capital appreciation. Investments have two attributes—time and risk. Present consumption is sacrificed to get a return in the future. The sacrifice that has to be made is certain but the return in the future may be uncertain. This attribute of investment indicates the risk factor. The risk is undertaken with a view to reap some return from the investment. For a layman, investment means some monetary commitment. A person's commitment to buy a flat or a house for his personal use may be an investment from his point of view. This cannot be considered as an actual investment as it involves sacrifice but does not yield any financial return.

Financial investment is the allocation of money to assets that are expected to yield some gain over a period of time. It is an exchange of financial claims, such as stocks and bonds, for money. They are expected to yield returns and experience capital growth over the years.

Check Your Progress

6. What is totally mixed strategy?
7. What should be looked up to solve a game?
8. State the rules of dominance.
9. Name some of the fields where game theory is applied.
10. Give any three limitations of game theory.

Objectives of Investment

The main objectives of investment are increasing the rate of return and reducing the risk. Other objectives like safety, liquidity and hedge against inflation can be considered as subsidiary objectives.

Return: Investors always expect a good rate of return from their investments. Rate of return could be defined as the total income the investor receives during the holding period stated as a percentage of the purchasing price at the beginning of the holding period.

$$\text{Return} = \frac{\text{End Period Value} - \text{Beginning Period Value} + \text{Dividend}}{\text{Beginning Period Value}} \times 100$$

NOTES

Rate of return is stated semi-annually or annually to help comparison among the different investment alternatives. If it is a stock, the investor gets the dividend as well as the capital appreciation as returns. Market return of the stock indicates the price appreciation for the particular stock. If a particular share is purchased in 1998 at ₹50, disposed at ₹60 in 1999 and the dividend yield is ₹5, then the return would be calculated as follows:

$$\text{Return} = \frac{\text{Capital Appreciation \& Dividend}}{\text{Purchase Price}} \times 100$$

$$\text{Return} = \frac{10+5}{50} \times 100 = 30\%$$

Risk: Risk of holding securities is related with the probability of actual return becoming less than the expected return. The word risk is synonymous with the phrase variability of return. Investments' risk is just as important as measuring its expected rate of return because minimizing risk and maximizing the rate of return are interrelated objectives in the investment management. Investments with rate of returns that vary from period to period is riskier than those investments with rates of return that do not vary much. Every investor likes to reduce the risk of his investment by proper combination of different securities.

Liquidity: Marketability of the investment provides liquidity to the investment. The liquidity depends upon the marketing and trading facility. If a portion of the investment could be converted into cash without much loss of time, it would help the investor meet the emergencies. Stocks are liquid only if they command good market by providing adequate return through dividends and capital appreciation.

Hedge Against Inflation: Since there is inflation in almost all the economy, the rate of return should ensure a cover against the inflation. The return rate should be higher than the rate of inflation, otherwise the investor will have loss in real terms. Growth stocks would appreciate in their values overtime and provide a protection against inflation. The return thus earned should assure the safety of the principal amount, regular flow of income and be a hedge against inflation.

Safety: The selected investment avenue should be under the legal and regulatory framework. If it is not under the legal framework, it is difficult to represent the grievances, if any. Approval of the law itself adds a flavour of safety. Even though approved by law, the safety of the principal differs from one mode of investment to another. Investments done with the government assure more safety than with the private party. From the safety point of view, investments can be ranked as follows: bank deposits, government bonds, UTI units, non-convertible debentures, convertible debentures, equity shares and deposits with the non-banking financial companies.

Process of Investment

The process of investment involves a series of activities leading to the purchase of securities or other investment alternatives. The investment process can be divided into five stages: (*i*) Framing of investment policy (*ii*) Investment analysis (*iii*) Valuation (*iv*) Portfolio construction (*v*) Portfolio evaluation. Figure 4.1 explains the stages and factors connected thereof.

NOTES

Fig. 4.1 Investment Process

Types of Investment

Negotiable Securities

These are the securities which may be freely bought, sold and transferred.

Variable Income Securities

Equity Shares: The equity shares attract the interest of many. In the early 1990s, the stock market was the best and safest place for the common individual to invest. Since 1996 the share market prices have been low. This made the retail investors to turn away from the stock market.

The stock market classifies shares into growth shares, income shares, defensive shares, cyclical shares and speculative shares.

(i) *Growth Shares:* The stocks that have higher rate of growth than the industrial growth rate in profitability are referred to as 'growth shares'.

(ii) *Income Shares:* The stocks belong to companies that have comparatively stable operations and limited growth opportunities. Bank shares and some of the fasts moving consumer good stocks, such as Cadburys, Nestle and Hindustan Lever may be termed as income shares.

(iii) *Defensive Shares:* Defensive stocks are relatively unaffected by the market movements.

(iv) *Cyclical Shares:* The business cycle affects the cyclical shares. The upward and downward movements of the business cycle affect the business prospects of certain companies and their stock prices. Such shares provide low to moderate current yield. Capital gain may be highly variable.

(v) *Speculative Shares:* Shares that have lot of speculative trading in them are referred to as 'speculative shares'. During the bull and bear phases of the market, this type of shares attract the attention of the traders.

The stocks, which fall under one category in one period may switch over to another category in another period. The classification should not be considered rigid. For example, growth shares may be speculative shares.

Fixed Income Securities

Preference Shares: Preference shares are no longer regarded as inferior to the equity capital. Corporates like Siemens has placed ₹150 crore worth of preference shares. High tax paying companies or investors prefer to subscribe to the preference shares, and investors with a low tax burden would prefer to go in for debt instruments. The

NOTES

conversion options provided in the preference shares also make it attractive. Th biggest advantage is the tax-exempt status of the dividend of the preference share.

Debentures: Corporate debentures are an option available to the investors who ar willing to sacrifice liquidity for higher return. Manufacturing companies like Gujara Industries Power and Tata Steel have issued debentures. If the debentures are no actively traded in the debt segment of the capital market, the investors may have t hold the instrument till maturity. If the instruments were actively traded in the secondar market, it would have perhaps changed hands at a considerable premium, thereb lowering the yield on par with the present interest rate. These reasons contribut towards high coupon rates on debentures.

Bonds: Bonds are similar to the debentures but they are issued by the public secto undertakings. The value of the bond in the market depends upon the interest rate an the maturity. The coupon rate is the nominal interest rate offered on the bonds. Th coupon rate is contractual involving the terms and conditions of the issuance of th debt security. Being contractual, it cannot be changed during the tenure of th instrument. The investors are not affected by lowering of the bank rates. When th bank rates are lowered, actually, the value of the bonds, which carry interest rate above the bank rate would appreciate. IDBI and ICICI have issued various bonds t suit the needs of the investors. Some of them are deep discount bond, educatio benefit bond, retirement benefit bond and index bond.

IVPs and KVPs: These are saving certificates issued by the post office with th name Indira Vikas Patra (IVP) and Kisan Vikas Patra (KVP). IVPs are available i face values of ₹500, 1,000 and 5,000. KVPs are in the denominations of ₹ 1,000, 5,000 and 10,000. The capital is doubled in 5.5 years with the return of 13.47 pe cent. IVPs are like bearer bonds, transferable by hand delivery, and therefore ar attractive to the persons who prefer cash transactions. No income tax concession i available for this type of investment.

Government Securities: The securities issued by the Central Government, stat government and quasi government agencies are known as government securities o 'gilt edged securities'. As government guaranteed security is a claim on th government, it is a secured financial instrument, which guarantees the income an the capital. The rate of interest on these securities is relatively lower because of thei high liquidity and safety.

Money Market Securities: Money market securities have very short-term maturity, say less than a year. There are three common money market instruments as follows:

(*i*) *Treasury bills:* A treasury bill is basically an instrument of short-term borrowing by the Government of India.

(*ii*) *Commercial Papers*: Commercial Paper (CP) is a short-term negotiable instrument with fixed CP maturity period. It is an unsecured promissory note issued by the company either directly or through bank/merchant banks. The maturity period of commercial paper was originally three (minimum) to six (maximum) months from the date of issue.

(*iii*) *Certificate of Deposit:* The Certificate of Deposit (CD) is a marketable receipt of funds deposited in a bank for a fixed period at a specified rate of interest. They are bearer documents and readily negotiable. The denominations of the CD and the interest rate on them are high. It is mainly preferred by institutional investors and companies rather than the individuals.

NOTES

Non-Negotiable Securities

Deposits

Deposits earn a fixed rate of return. Even though bank deposits resemble fixed income securities, they are not negotiable instruments. Some of the deposits have been dealt subsequently.

(a) ***Bank Deposits:*** It is the simple investment avenue open for an investor. He has to open an account and deposit the money. Traditionally, the banks offered current account, savings account and fixed deposit account. Current account does not offer any interest rate.

The deposits in the banks are considered to be safe because of the RBI regulation. The risk averse investors prefer the bank deposits.

Post Office Deposits: Like the banks, post office also offers fixed deposit facility and monthly income scheme. 'Post office monthly income scheme' is a popular scheme for the retired. An interest rate of 13 per cent is paid monthly. The term of the scheme is 6 years, at the end of which a bonus of 10 per cent is paid. The annual yield to maturity works out to be 15.01 per cent per annum. After three years, premature closure is allowed without any penalty. If the closure is after one year, a penalty of 5 per cent is charged.

NBFC Deposits: In recent years, there has been a significant increase in the importance of Non-Banking Financial Companies (NBFC) in the process of financial intermediation. The NBFC comes under the purview of the RBI. The Amendment to the RBI Act in January 1997, made registration compulsory for the NBFCs.

Tax Sheltered Savings Scheme

Tax sheltered savings schemes are of great importance to the investors in the tax-paying category. The tax sheltered savings schemes offer tax relief to those who participate in their schemes according to the income tax laws. The important tax sheltered savings schemes are as follows:

Public Provident Fund (PPF) Scheme: PPF earns an interest rate of 12 per cent per year, which is exempted from the income tax under Section 88. Individuals and Hindu Undivided Families (HUF) can participate in this scheme.

National Savings Scheme (NSS): This scheme helps in deferring the tax payment. Individuals and HUF are eligible to open NSS account in the designated post office.

National Savings Certificate (6 yrs.): This scheme is offered by the post office and the maturity period is 6 years. These certificates come in the denominations of ₹500, 1,000, 5,000 and 10,000. The contribution and the interest for the first 5 years are covered by Section 88. The interest is cumulative at the rate of 12 per cent per annum and payable biannually is covered by Section 80L. No withdrawals are permitted. There is no deduction at maturity.

Life Insurance

Life insurance is a contract for payment of a sum of money to the person assured (or to the person entitled to receive the same) on the happening of event insured against. Usually the contract provides for the payment of an amount on the date of maturity or at specified dates at periodic intervals or if unfortunate death occurs. Among other

NOTES

things, the contracts also provide for the payment of premium periodically to the corporation by the policy holders. Life insurance eliminates risk.

4.7.2 Break-Even Analysis

The cost concepts that are relevant to business operations and decisions can be grouped on the basis of their nature and purpose.

1. ***Fixed and Variable Costs: Fixed costs*** are those that are fixed in volume for a certain quantity of output. Fixed cost does not vary with variation in the output between zero and a certain level of output. In other words, costs that do not vary or are fixed for a certain level of output are known as *fixed costs*. The fixed costs include (*i*) Costs of managerial and administrative staff, (*ii*) Depreciation of machinery, building and other fixed assets, (*iii*) Maintenance of land, etc. The concept of fixed cost is associated with the short-run.

   Variable costs are those which vary with the variation in the total output. Variable costs include cost of raw material, running cost of fixed capital, such as fuel, repairs, routine maintenance expenditure, direct labour charges associated with the level of output, and the costs of all other inputs that vary with output.

2. ***Total, Average and Marginal Costs: Total Cost (TC)*** is the total actual cost incurred on the production of goods and service. It refers to the total outlays of money expenditure, both explicit and implicit, on the resources used to produce a given level of output. It includes both fixed and variable costs. The total cost for a given output is given by the cost function.

   Average Cost (AC) is of statistical nature—it is not actual cost. It is obtained by dividing the Total Cost (TC) by the total output (Q), i.e.,

$$AC = \frac{TC}{Q}$$

   Marginal Cost (MC) is defined as the addition to the total cost on account of producing one additional unit of the product. Or, marginal cost is the cost of the marginal unit produced. Marginal cost is calculated as $TC_n - TC_{n-1}$ where n is the number of units produced. Using cost function, MC can be defined as

$$MC = \frac{\partial TC}{\partial Q}$$

   Total, average and marginal cost concepts are used in the economic analysis of firm's production activities.

In traditional theory of firm, the basic objective of the firm is to maximize profit. Maximum profit does not necessarily coincide with the minimum cost, as far as the traditional theory of firm is concerned. Besides, profit is maximum at a specific level of output which is difficult to know before hand. Even if it is known, it cannot be achieved at the outset of production. In real life, firms begin their activity even at a loss, in anticipation of profit in the future. However, the firms can plan their production better if they know the level of production where cost and revenue break-even, i.e., the profitable and non-profitable range of production. Break-even analysis or what is also known as *profit contribution analysis* is an important analytical technique used to study the relationship between the total costs, total revenue and total profits and losses over the whole range of stipulated output. The break-even analysis is a technique of having a preview of profit prospects and a tool of profit planning. It integrates the

NOTES

ost and revenue estimates to ascertain the profits and losses associated with different evels of output.

The relationship between cost and output and between price and output may e linear or non-linear in nature.

Linear Cost and Revenue Functions

To illustrate the break-even analysis under linear cost and revenue conditions, let us assume linear cost and linear revenue functions are given as follows.

$$\text{Cost function: } TC = 100 + 10Q \quad \ldots(4.1)$$

$$\text{Revenue function: } TR = 15Q \quad \ldots(4.2)$$

The cost function given in Equation (4.1) implies that the firm's total fixed cost is given at ₹100 and its variable cost varies at a constant rate of ₹10 per unit in response to increase in output. The revenue function given in Equation (4.2) implies that the price for the firm's product is given in the market at ₹15 per unit of sale.

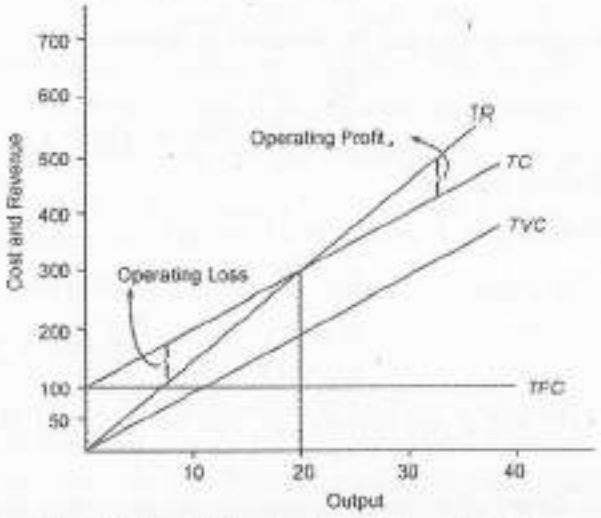


Fig. 4.2 Break-Even Analysis: Linear Functions

What the firm needs to carry out the break-even analysis of its business operations is to make a chart of its Total Fixed Cost (*TFC*), Total Variable Cost (*TVC*), Total Cost (*TC*) and the Total Revenue (*TR*), and graph them to find the break-even point. The process of break-even analysis is illustrated graphically in Figure 4.2. The line *TFC* shows the total fixed cost at ₹100 for a certain level of output, and the line *TVC* shows the variable cost rising with a slope $\Delta TVC/\Delta Q = 10/1 = 10$. The line *TC* has been obtained by plotting the *TC* function. It can also be obtained by a vertical summation of *TFC* and *TVC* at various levels of output. The line *TR* shows the Total Revenue (*TR*) obtained as $Q \times P$. The *TR* and *TC* lines intersect at point *B*, where output is equal to 20 units. The point *B* shows that at $Q = 20$, firm's total cost equals its total revenue. That is, at $Q = 20$, *TC* breaks-even with *TR*. Point *B* is, therefore, the break-even point and $Q = 20$ is the break-even output. Below this level of output, *TC* exceeds *TR*. The vertical difference between *TC* and *TR*, (i.e., *TC–TR*) is known as operating loss. Beyond $Q = 20$, $TR > TC$, and *TR–TC* is known as operating profit. It may thus be inferred that a firm producing a commodity under cost and revenue conditions given in Equation 4.1 and Equation 4.2, respectively, must produce at least 20 units to make its total cost and total revenue break-even.

NOTES

The break-even output can also be calculated algebraically. We know that at break-even point, $TR = TC$

That is, in terms of TR and TC functions,

$$15Q = 100 + 10Q$$

$$5Q = 100$$

$$Q = 20$$

Thus, 20 is the break-even output. Given the TR and TC functions, production beyond 20 units will yield increasing profits, at least in the short-run.

Algebra of Break-Even Analysis: The break-even analysis can also be presented algebraically. At break-even volume,

$$TR = TC$$

and that $TR = (P \times Q)$ and $TC = TFC + TVC$. In break-even analysis TVC is defined as $TVC = AVC \times Q$. Thus,

$$TC = TFC + AVC \times Q$$

Now, break-even quantity (Q_B) can be obtained as follows.

$$TR = TC$$

$$Q_B \cdot P = TFC + AVC \cdot Q_B \quad \ldots(4.3)$$

where Q_B = break-even volume.

Rearranging Equation (4.3), we get

$$Q_B \cdot P - AVC \cdot Q_B = TFC$$

$$Q_B(P - AVC) = TFC$$

$$Q_B = (TFC/P) - AVC \quad \ldots(4.4)$$

If firm's TFC, AVC and P are known, Q_B can be obtained straightaway from Equation (4.4).

Limitations: The theory of break-even analysis, as presented above, is applicable only if cost and revenue functions are linear. Under the condition of linear cost and revenue functions, TC and TR are straight lines and they intersect at only one point (as shown in Figure 4.2.) dividing the whole range of output into two parts—profitable and non-profitable. It may give the impression that the whole output beyond the break-even level is profitable. In real life, however, conditions are different due to changing price and cost levels. In reality, the cost and revenue functions may be non-linear. Non-linearity arises because AVC and price vary with variation in the output. As a result, the Total Cost (TC) may increase at increasing rates while the Total Revenue (TR) increases at decreasing rates. Therefore, at some stage of output, TC may exceed TR. Thus, there might be two break-even points (as shown in Figure 4.3) instead of one. This limits the profitable range of output and determines the lower and upper limits of profitable output. The analyst should, therefore, pre-test and verify the validity of cost and revenue functions rather than assuming straightaway the linearity conditions.

Use of Break-Even Analysis

(*i*) Sales volume can be determined to earn a given amount of return on capital.

(*ii*) Profit can be forecast if estimates of revenue and cost are available.

(*iii*) Effect of change in the volume of sales, sale price, cost of production, can be appraised.

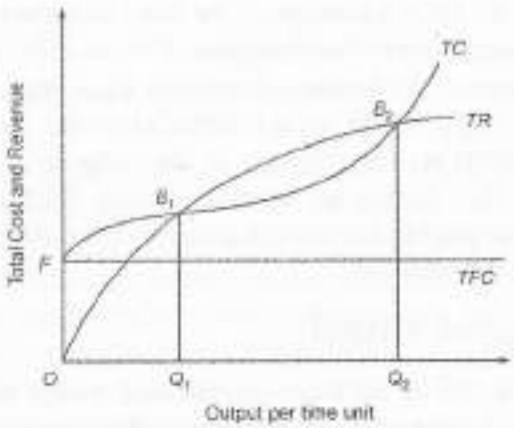


Fig. 4.3 Break-Even Analysis: Non-Linear Functions

(iv) Choice of products can be made from the alternatives available. Product-mix can also be determined.

(v) Impact of increase or decrease in fixed and variable costs can be highlighted.

(vi) Effect of high fixed costs and low variable costs to the total cost can be studied.

(vii) Valid inter-firm comparisons of profitability can be made.

(viii) Cash break-even chart helps proper planning of cash requirements.

(ix) Break-even analysis emphasizes the importance of capacity utilization for achieving economies.

(x) Further help is provided by margin of safety and angle of incidence.

Limitations of Break-Even Analysis

We have discussed above that the break-even analysis is based on linear assumptions. The linearity assumption can be removed by pre-testing the cost and revenue functions and by using, if necessary, the non-linearity conditions. Nevertheless, the break-even analysis as such has certain other limitations.

First, the break-even analysis can be applied only to a single product system. Under the condition of multiple products and joint operations, the break-even analysis can be applied only if product-wise cost can be ascertained which is, of course, extremely difficult.

Second, break-even analysis cannot be applied usefully where cost and price data cannot be ascertained beforehand and where historical data are not relevant for estimating future costs and prices. Despite these limitations, the break-even analysis may serve a useful purpose in production planning if relevant data can be easily obtained.

4.7.3 Methods of Evaluation and Capital Investment

As regards the use of evaluation methods, most Indian companies use payback criterion. In addition to payback and/or other methods, companies also use Internal Rate of Return (IRR) and Net Present Value (NPV) methods. A few companies use Accounting Rate of Return (ARR) method. IRR is the second most popular technique in India.

The major reason for payback to be more popular than the DCF (Discounted Cash Flow) techniques is the executives' lack of familiarity with DCF techniques. Other factors are lack of technical people and sometimes unwillingness of top

management to use the DCF techniques. One large manufacturing and marketing organization, for example, thinks that conditions of its business are such that the DCF techniques are not needed. By business conditions the company perhaps means its marketing nature, and its products being in seller's markets. Another company feels that replacement projects are very frequent in the company, and therefore, it is not necessary to use the DCF techniques for such projects. Both these companies have fallacious approaches towards investment analysis. They should subject all capital expenditures to formal evaluation.

4.7.4 Payback Period Method

The PayBack (PB) is one of the most popular and widely recognized traditional methods of evaluating investment proposals. **Payback** is the number of years required to recover the original cash outlay invested in a project. If the project generates constant annual cash inflows, the payback period can be computed by dividing cash outlay by the annual cash inflow. That is:

$$\text{Payback} = \frac{\text{Initial Investment}}{\text{Annual Cash Inflow}} = \frac{C_0}{C}$$

Payback (Constant Cash Flows)

Assume that a project requires an outlay of ₹50,000 and yields annual cash inflow of ₹12,500 for 7 years. The payback period for the project is:

$$PB = \frac{₹\ 50{,}000}{₹\ 12{,}000} = 4 \text{ years}$$

Unequal Cash Flows: In case of unequal cash inflows, the payback period can be found out by adding up the cash inflows until the total is equal to the initial cash outlay. Consider the following example.

Payback (Uneven Cash Flows)

Suppose that a project requires a cash outlay of ₹20,000, and generates cash inflows of ₹ 8,000; ₹7,000; ₹4,000; and ₹3,000 during the next 4 years. What is the project's payback? When we add up the cash inflows, we find that in the first three years ₹19,000 of the original outlay is recovered. In the fourth year cash inflow generated is ₹3,000 and only ₹1,000 of the original outlay remains to be recovered. Assuming that the cash inflows occur evenly during the year, the time required to recover ₹1,000 will be (₹1,000/₹3,000) × 12 months = 4 months. Thus, the payback period is 3 years and 4 months.

Acceptance Rule

Many firms use the payback period as an investment evaluation criterion and a method of ranking projects. They compare the project's payback with a predetermined, standard payback. The project would be accepted if its payback period is less than the maximum or **standard payback** period set by management. As a ranking method, it gives highest ranking to the project, which has the shortest payback period and lowest ranking to the project with highest payback period. Thus, if the firm has to choose between two mutually exclusive projects, the project with shorter payback period will be selected.

NOTES

Evaluation of Payback

Payback is a popular investment criterion in practice. It is considered to have certain virtues.

- ***Simplicity:*** The most significant merit of payback is that it is simple to understand and easy to calculate. The business executives consider the simplicity of method as a virtue. This is evident from their heavy reliance on it for appraising investment proposals in practice.
- ***Cost Effective:*** Payback method costs less than most of the sophisticated techniques that require a lot of the analysts' time and the use of computers.
- ***Short-Term Effects:*** A company can have more favourable short-run effects on earnings per share by setting up a shorter standard payback period. It should, however, be remembered that this may not be a wise long-term policy as the company may have to sacrifice its future growth for current earnings.
- ***Risk Shield:*** The risk of the project can be tackled by having a shorter standard payback period as it may ensure guarantee against loss. A company has to invest in many projects where the cash inflows and life expectancies are highly uncertain. Under such circumstances, payback may become important, not so much as a measure of profitability but as a means of establishing an upper bound on the acceptable degree of risk.
- ***Liquidity:*** The emphasis in payback is on the early recovery of the investment. Thus, it gives an insight into the liquidity of the project. The funds so released can be put to other uses.

In spite of its simplicity and the so-called virtues, the payback may not be a desirable investment criterion since it suffers from a number of serious limitations:

- ***Cash Flows After Payback:*** Payback fails to take account of the cash inflows earned after the payback period. For example, consider the following projects *X* and *Y*:

	Cash Flows (₹)					
Project	C_0	C_1	C_2	C_3	*Payback*	*NPV*
X	– 4,000	0	4,000	2,000	2 years	+ 806
Y	– 4,000	2,000	2.000	0	3 years	– 530

As per the payback rule, both the projects are equally desirable since both return the investment outlay in two years. If we assume an opportunity cost of 10 per cent, Project *X* yields a positive net present value of ₹806 and Project *Y* yields a negative net present value of ₹530. As per the NPV rule, Project *X* should be accepted and Project *Y* rejected. Payback rule gave wrong results because it failed to consider ₹2,000 cash flow in third year for Project *X*.

- ***Cash Flows Ignored:*** Payback is not an appropriate method of measuring the profitability of an investment project as it does not consider all cash inflows yielded by the project. Considering Project *X* again, payback rule did not take into account its entire series of cash flows.
- ***Cash Flow Patterns:*** Payback fails to consider the pattern of cash inflows, i.e., magnitude and timing of cash inflows. In other words, it gives equal weights to returns of equal amounts even though they occur in different time periods. For example, compare the following projects *C* and *D* where

NOTES

they involve equal cash outlay and yield equal total cash inflows over equal time periods:

	Cash Flows (₹)					
Project	C_0	C_1	C_2	C_3	*Payback*	*NPV*
C	−5,000	3,000	2,000	2,000	2 years	+ 881
D	−5,000	2,000	3,000	2,000	2 years	+ 798

Using payback period, both projects are equally desirable. But Project *C should be* preferable as larger cash inflows' come earlier in its life. This is indicated by the NPV rule; project *C* has higher NPV (₹881) than Project *D* (₹798) at 10 per cent opportunity cost. It should be thus clear that payback is not a measure of profitability. As such, it is dangerous to use it as a decision criterion.

- ***Administrative Difficulties:*** A firm may face difficulties in determining the maximum acceptable payback period. There is no rational basis for setting a maximum payback period. It is generally a subjective decision.
- ***Inconsistent with Shareholder Value:*** Payback is not consistent with the objective of maximizing the market value of the firm's shares. Share values do not depend on payback periods of investment projects.

Let us re-emphasize that the payback is not a valid method for evaluating the acceptability of the investment projects. It can, however, be used along with the NPV rule as a first step in roughly screening the projects. In practice, the use of DCF techniques has been increasing but payback continues to remain a popular and primary method of investment evaluation.

Payback Reciprocal and the Rate of Return

Payback is considered theoretically useful in a few situations. One significant argument in favour of payback is that its reciprocal is a good approximation of the rate of return under certain conditions.

The payback period is defined as follows:

$$\text{Payback} = \frac{\text{Initial Investment}}{\text{Annual Cash Inflow (Annuity)}} = \frac{C_0}{C}$$

The formula for the present value of an annuity is given by the following equation:

$$C_0 = C\left[\frac{1-\frac{1}{(1+r)^n}}{r}\right] = \frac{C}{r} - \frac{C}{r}\left[\frac{1}{(1+r)^n}\right]$$

Multiplying both sides by r, we get

$$rC_0 = C - C\left[\frac{1}{(1+r)^n}\right]$$

Solving for r, we find

$$r = \frac{C}{C_0} - \frac{C}{C_0}\left[\frac{1}{(1+r)^n}\right]$$

where C_0 is the initial investment, C is annual cash inflow, r is rate of return and n is the life of investment.

The first right-hand term is the reciprocal of the payback period. The second right-hand term is payback reciprocal multiplied by $1/(1 + r)^n$. If n is very large or extends to infinity, the second term becomes insignificant (almost equal to zero), and we are left with the term C/C_0. Thus, IRR is equal to the reciprocal of payback.

The reciprocal of payback will be a close approximation of the internal rate of return if the following two conditions are satisfied:

- The life of the project is large or at least twice the payback period.
- The project generates equal annual cash inflows.

The payback reciprocal is a useful technique to quickly estimate the true rate of return. But its major limitation is that every investment project does not satisfy the conditions on which this method is based. When the useful life of the project is not at least twice the payback period, the payback reciprocal will always exceed the rate of return. Similarly, it cannot be used as an approximation of the rate of return if the project yields uneven cash inflows.

4.7.5 Average (Accounting) Rate of Return Method

The **Accounting Rate of Return** (ARR), also known as the **Return On Investment** (ROI), uses accounting information, as revealed by financial statements, to measure the profitability of an investment. The accounting rate of return is the ratio of the average after tax profit divided by the average investment. The average investment would be equal to half of the original investment if it were depreciated constantly. Alternatively, it can be found out by dividing the total of the investment's book values after depreciation by the life of the project. The accounting rate of return, thus, is an average rate and can be determined by the following equation:

$$\text{ARR} = \frac{\text{Average Income}}{\text{Average Investment}}$$

Average income should be defined in terms of earnings after taxes without an adjustment for interest viz. EBIT $(1 - T)$ or net operating profit after tax. Thus

$$\text{ARR} = \frac{\left[\sum_{t=1}^{n} \text{EBIT}_t (1-T)\right]/n}{(I_0 + I_n)/2}$$

where EBIT is earnings before interest and taxes, T tax rate, I_0 book value of investment in the beginning, I_n book value of investment at the end of n number of years.

Accounting Rate of Return

A project will cost ₹40,000. Its stream of Earnings Before Depreciation, Interest and Taxes (EBDIT) during first year through five years is expected to be ₹10,000, ₹12,000, ₹14,000, ₹16,000 and ₹20,000. Assume a 50 per cent tax rate and depreciation on straight-line basis. Project's ARR is computed in Table 4.2.

$$\text{Accounting Rate of Return} = \frac{3,200}{20,000} \times 100 = 16 \text{ per cent}$$

A variation of the ARR method is to divide average earnings after taxes by the original cost of the project instead of the average cost. Thus, using this version, the ARR would be: ₹3,200 , ₹ 40,000 ₹ 100 = 8 per cent. This version of the ARR method is less consistent as earnings are averaged but investment is not.

Table 4.2 Calculation of Accounting Rate of Return

	(₹)					
Period	*1*	*2*	*3*	*4*	*5*	*Average*
Earnings before depreciation, interest and taxes (EBDIT)	10,000	12,000	14,000	16,000	20,000	14,400
Depreciation	8,000	8,000	8,000	8,000	8,000	8,000
Earnings before interest and taxes (EBIT)	2,000	4,000	6,000	8,000	12,000	6,400
Taxes at 50%	1,000	2,000	3,000	4,000	6,000	3,200
Earnings before interest and after taxes [EBIT (1– *T*)]	1,000	2,000	3,000	4,000	6,000	3,200
Book value of investment:						
Beginning	40,000	32,000	24,000	16,000	8,000	
Ending	32,000	24,000	16,000	8,000	—	
Average	36,000	28,000	20,000	12,000	4,000	20,000

Acceptance Rule

As an accept-or-reject criterion, this method will accept all those projects whose ARR is higher than the minimum rate established by the management and reject those projects which have ARR less than the minimum rate. This method would rank a project as number one if it has highest ARR and lowest rank would be assigned to the project with lowest ARR.

Evaluation of ARR Method

The ARR method may claim some merits:

- ***Simplicity:*** The ARR method is simple to understand and use. It does not involve complicated computations.
- ***Accounting Data:*** The ARR can be readily calculated from the accounting data; unlike in the NPV and IRR methods, no adjustments are required to arrive at cash flows of the project.
- ***Accounting Profitability:*** The ARR rule incorporates the entire stream of income in calculating the project's profitability.

The ARR is a method commonly understood by accountants, and frequently used as a performance measure. As a decision criterion, however, it has serious shortcomings.

- ***Cash Flows Ignored:*** The ARR method uses accounting profits, not cash flows, in appraising the projects. Accounting profits are based on arbitrary assumptions and choices and also include non-cash items. It is, therefore, inappropriate to rely on them for measuring the acceptability of the investment projects.
- ***Time Value Ignored:*** The averaging of income ignores the time value of money. In fact, this procedure gives more weightage to the distant receipts.
- ***Arbitrary Cut-Off:*** The firm employing the ARR rule uses an arbitrary cut-off yardstick. Generally, the yardstick is the firm's current return on its

assets (book-value). Because of this, the growth companies earning very high rates on their existing assets may reject profitable projects (i.e., with positive NPVs) and the less profitable companies may accept bad projects (i.e., with negative NPVs).

The ARR method continues to be used as a performance evaluation and control measure in practice. But its use as an investment criterion is certainly undesirable. It may lead to unprofitable allocation of capital.

4.7.6 Time Adjusted or Discounted Cash Flow (DCF) Method

The various methods for time adjustment and discounted cash flow are as follows:

1. **Net Present Value (NPV):** The difference between PV of cash flows and PV of cash outflows is equal to NPV; the firm's opportunity cost of capital being the discount rate.

$$NPV = \left[\frac{C_1}{(1+k)} + \frac{C_2}{(1+k)^2} + \frac{C_3}{(1+k)^3} + \cdots + \frac{C_n}{(1+k)^n}\right] - C_0$$

$$NPV = \sum_{t=1}^{n} \frac{C_t}{(1+k)^t} - C_0$$

Acceptance Rule

- Accept if NPV > 0 (i.e., NPV is positive)
- Reject if NPV < 0 (i.e., NPV is negative)
- Project may be accepted if NPV = 0

Merits	*Demerits*
• Considers all cash flows	• Requires estimates of cash flows which is a tedious task
• True measure of profitability	• Requires computation of the opportunity cost of capital which poses practical difficulties
• Based on the concept of the	• Sensitive to discount rates time value of money

- Satisfies the value-additivity principle (i.e., NPV's of two or more projects can be added)
- Consistent with the Share-holders' Wealth Maximization (SWM) principle.

2. **Internal Rate of Return (IRR):** The discount rate which equates the present value of an investment's cash inflows and outflows is its internal rate of return.

$$\left[\frac{C_1}{(1+r)} + \frac{C_2}{(1+r)^2} + \frac{C_3}{(1+r)^3} + \cdots + \frac{C_n}{(1+r)^n}\right] = C_0$$

$$NPV = \sum_{t=1}^{n} \frac{C_t}{(1+r)^t} - C_0 = 0$$

Acceptance Rule

- Accept if IRR > k
- Reject if IRR < k
- Project may be accepted if IRR = k

NOTES

Merits	*Demerits*
• Considers all cash flows	• Requires estimates of cash flows which is a tedious task
• True measure of profitability	• Does not hold the value additivity principle (i.e., IR ₹of two or more projects do not add)
• Based on the concept of the time value of money	• At times fails to indicate correct choice between mutually exclusive projects
• Generally, consistent with wealth maximization principle	• At times yields multiple rates • Relatively difficult to compute

3. **Profitability Index (PI):** The ratio of the present value of the cash flows to the initial outlay is profitability index or benefit-cost ratio:

$$\text{PI} = \frac{\text{PV of Annual Cash Flows}}{\text{Initial Investment}}$$

$$PI = \frac{\sum_{t=1}^{n} \frac{C_t}{(1+k)^t}}{C_0}$$

Acceptance Rule

- Accept if PI > 1.0
- Reject if PI < 1.0
- Project may be accepted if PI = 1.0

Merits	*Demerits*
• Considers all cash flows	• Requires estimates of the cash flows which is a tedious task
• Recognizes the time value of money	• At times fails to indicate correct choice between mutually exclusive projects
• Relative measure of profitability	
• Generally consistent with the wealth maximization principle	

Net Present Value (NPV) Method

The Net Present Value (NPV) method is the classic economic method of evaluating the investment proposals. It is a DCF technique that explicitly recognizes the time value of money. It correctly postulates that cash flows arising at different time periods differ in value and are comparable only when their equivalents—present values—are found out. The following steps are involved in the calculation of NPV:

- Cash flows of the investment project should be forecasted based on realistic assumptions.
- Appropriate discount rate should be identified to discount the forecasted cash flows. The appropriate discount rate is the project's opportunity cost of capital, which is equal to the required rate of return expected by investors on investments of equivalent risk.
- Present value of cash flows should be calculated using the opportunity cost of capital as the discount rate.

NOTES

- Net present value should be found out by subtracting present value of cash outflows from present value of cash inflows. The project should be accepted if NPV is positive (i.e., NPV > 0).

Let us consider an example.

Calculating Net Present Value

Assume that Project *X* costs ₹2,500 now and is expected to generate year-end cash inflows of ₹900, ₹ 800, ₹700, ₹600 and ₹500 in years 1 through 5. The opportunity cost of the capital may be assumed to be 10 per cent.

The net present value for Project *X* can be calculated by referring to the present value table (Table C at the end of the book). The calculations are as follows:

$$NPV = \left[\frac{₹\ 900}{(1+0.10)} + \frac{₹\ 800}{(1+0.10)^2} + \frac{₹\ 700}{(1+0.10)^3} + \frac{₹\ 600}{(1+0.10)^4} + \frac{₹\ 500}{(1+0.10)^5}\right] - ₹\ 2{,}500$$

$$NPV = [₹\ 900(PVF_{1,\,0.10}) + ₹\ 800(PVF_{2,\,0.10}) + ₹\ 700(PVF_{3,\,0.10}) + ₹\ 600(PVF_{4,\,0.10}) + ₹\ 500(PVF_{5,\,0.10})] - ₹\ 2{,}500$$

$$NPV = [₹\ 900 \times 0.909 + ₹\ 800 \times 0.826 + ₹\ 700 \times 0.751 + ₹\ 600 \times 0.683 + ₹\ 500 \times 0.620] - ₹\ 2{,}500$$

$$NPV = ₹\ 2{,}725 - ₹\ 2{,}500 = + ₹\ 225$$

Project *X*'s present value of cash inflows (₹2,725) is greater than that of cash outflow (₹2,500). Thus, it generates a positive net present value (NPV = +₹225). Project *X* adds to the wealth of owners; therefore, it should be accepted.

The formula for the net present value can be written as follows:

$$NPV = \left[\frac{C_1}{(1+k)} + \frac{C_2}{(1+k)^2} + \frac{C_3}{(1+k)^3} + \cdots + \frac{C_n}{(1+k)^n}\right] - C_0$$

$$NPV = \sum_{t=1}^{n} \frac{C_t}{(1+k)^t} - C_0$$

where C_1, C_2... represent net cash inflows in year 1, 2..., k is the opportunity cost of capital, C_0 is the initial cost of the investment and n is the expected life of the investment. It should be noted that the cost of capital, k, is assumed to be known and is constant.

Acceptance Rule

It should be clear that the acceptance rule using the NPV method is to accept the investment project if its net present value is positive (NPV > 0) and to reject it if the net present value is negative (NPV < 0). Positive NPV contributes to the net wealth of the shareholders, which should result in the increased price of a firm's share. The positive net present value will result only if the project generates cash inflows at a rate higher than the opportunity cost of capital. A project with zero NPV (NPV = 0) may be accepted. A zero NPV implies that project generates cash flows at a rate just equal to the opportunity cost of capital. The NPV acceptance rules are:

- Accept the project when NPV is positive — NPV > 0
- Reject the project when NPV is negative — NPV < 0
- May accept the project when NPV is zero — NPV = 0

The NPV method can be used to select between mutually exclusive projects; the one with the higher NPV should be selected. Using the NPV method, projects

NOTES

would be ranked in order of net present values; that is, first rank will be given to the project with highest positive net present value and so on.

Level Cash Flows

An easy procedure can be followed to calculate the IRR for a project that produces level or equal cash flows each period. To illustrate, let us assume that an investment would cost ₹20,000 and provide annual cash inflow of ₹5,430 for 6 years. If the opportunity cost of capital is 10 per cent, what is the investment's NPV? The ₹5,430 is an annuity for 6 years. The NPV can be found as follows:

$$NPV = -₹\,20{,}000 + ₹\,5{,}430(PVFA_{6,0.10})$$

$$= -₹\,20{,}000 + ₹\,5{,}430 \times 4.355 = ₹\,3{,}648$$

How much is the project's IRR?

The IRR of the investment can be found out as follows:

$$NPV = -₹\,20{,}000 + ₹\,5{,}430(PVFA_{6,r}) = 0$$

$$₹\,20{,}000 = ₹\,5{,}430(PVFA_{6,r})$$

$$PVFA_{6,r} = \frac{₹\,20{,}000}{₹\,5{,}430} = 3.683$$

The rate, which gives a PVFA of 3.683 for 6 years, is the project's internal rate of return. Looking up PVFA in Table D (given at the end of the book) across the 6-year row, we find it approximately under the 16 per cent column. Thus, 16 per cent is the project's IRR that equates the present value of the initial cash outlay (₹20,000) with the constant annual cash inflows (₹5,430 per year) for 6 years.

Discounted Payback Method

One of the serious objections to the payback method is that it does not discount the cash flows for calculating the payback period. We can discount cash flows and then calculate the payback. The **discounted payback period** is the number of periods taken in recovering the investment outlay on the present value basis. The discounted payback period still fails to consider the cash flows occurring after the payback period.

Let us consider an example. Projects P and Q involve the same outlay of ₹4,000 each. The opportunity cost of capital may be assumed as 10 per cent. The cash flows of the projects and their discounted payback periods are shown in Table 4.3.

Table 4.3 Discounted Payback Illustrated

	Cash Flows (₹)					Simple PB	Discounted PB	NPV at 10%
	C_0	C_1	C_2	C_3	C_4			
P	−4,000	3,000	1,000	1,000	1,000	2 yrs	—	—
PV of cash flows	−4,000	2,727	826	751	683		2.6 yrs	987
Q	−4,000	0	4,000	1,000	2,000	2 yrs	—	—
PV of cash flows	−4,000	0	3,304	751	1,366		2.9 yrs	1,421

The projects are indicated of same desirability by the simple payback period. When cash flows are discounted to calculate the discounted payback period, Project P recovers the investment outlay faster than Project Q, and therefore, it would be preferred over Project Q. Discounted payback period for a project will be always

igher than simple payback period because its calculation is based on the discounted ash flows. Discounted payback rule is better as it discounts the cash flows until the utlay is recovered. But it does not help much. It does not take into consideration the ntire series of cash flows. It can be seen in our example that if we use the NPV rule, roject Q (with higher discounted payback period) is better.

17.7 Discounted Rate and Certainty Equivalent Approach

The discount rate is the interest rate the Federal Reserve charges on loans it makes to banks and other financial institutions. The interest rate is used in determining the present value of future cash flows. In addition, the discount rate becomes the base interest rate for most of the consumers who take loan. In general, a bank uses the discount rate as a benchmark for the interest it charges on the loans it makes. Moreover, whenever the discount rate increases the interest rate that lenders charge on home mortgages and other loans also increases. Similarly, whenever the discount rate is lowered, the cost of consumer borrowing loan also ultimately decreases. Characteristically, the discount rate is:

(1) An interest rate used to bring future values into the present when considering the time value of money.

(2) The interest rate the Federal Reserve charges to loan money to banks.

If an item has been discounted by x per cent then its price has been reduced by x per cent. This price reduction is called discount.

Example: If an item costing ₹ 550 is discounted by 30 per cent then the discount is,

$$30\% \text{ of } ₹\ 550 = 30/100 \times 550$$
$$= ₹\ 165$$

Discounting is the process of finding the present value of an amount at some future date and along with compounding cash, forms the basis of time value of money calculations. The discounted value of a cash flow is determined by reducing its value by the appropriate *discount rate* for each unit of time between the times when the cash flow is to be valued to the time of the cash flow. Basically, the discount rate is expressed as an annual rate.

To calculate the present value of a single cash flow, it is divided by one plus the interest rate for each period of time that will pass. This is expressed mathematically as raising the divisor to the power of the number of units of time. The discount rate in financial calculations is considered to be equal to the cost of capital. Adjustments are made to the discounted rate to specify the risks associated with uncertain cash flows.

The characteristic reasons for high discount rates for startups are:

- Reduced marketability of ownerships because stocks are not traded publicly
- Limited number of investors willing to invest
- Startups face high risks
- Over-optimistic forecasts by enthusiastic promoters

The capital asset pricing model methodology is used to calculate the correct discount rate. The three variables which make up the discount rate are:

(a) **Risk Free Rate:** The percentage of return generated by investing in risk-free securities, such as government bonds.

(b) **Beta:** The measure of how a company's stock price reacts to a change in the market. Beta greater than 1 means excessive change in a share's price as

NOTES

compared to other shares in the same market, whereas Beta less than 1 means the share is stable and is not affected by changes in the market. Beta less than 0 signifies that the price of the share is moving in the opposite direction to that of market change.

(c) **Equity Market Risk Premium:** This is the return on investment that investors seek over and above the risk free rate. This excess return compensates investors for taking the relatively higher risk associated with the equity market. The premium amount will varies as the risk in a particular stock or in the stock market as a whole changes. Generally, high risk investments are compensated with a higher premium.

Discount Rate = Risk Free Rate + Beta × Equity Market Risk Premium

Discount Factor

The discount factor, $P(T)$, is the number which a future cash flow to be received at time T must be multiplied by in order to obtain the current present value. Thus, a fixed annually compounded discount rate is:

$$P(T) = 1 / (1+r)^T$$

For fixed continuously compounded discount rate,

$$P(T) = e - r^T$$

The market price also termed as effective price is the amount actually paid. The purpose of discounts is to increase short-term sales, move out-of-date stock, reward valuable customers and encourage distribution channel members to perform a function, or otherwise reward behaviours that benefit the discount issuer.

Certainty Equivalent Approach

Certainty equivalent is an amount that would be accepted today (risk free) in lieu of a chance to receive a possibly higher, but uncertain amount. In addition, it is a return on a risky event (such as, a future income) stated in terms of a risk-free event (such as, a cash payment). It is the lowest cash amount that will compensate an investor for participating in a venture with uncertain outcome. Certainty equivalent varies with an individual's attitude towards risk, also called certain equivalent. It is a small, zero-risk return an investor may trade for a larger potential return with an associated risk. Companies recommend certainty equivalent returns on certain investments and use their demand to determine the level of risk an investor will accept for a given return from the company. Thus it is a guaranteed return that someone would accept, rather than taking a chance on a higher, but uncertain return.

4.7.8 Expected Monetary Value and Hillier and Hertz's Models

The Expected Monetary (Money) Value (EMV) is the total of the weighted outcomes or payoffs associated with a decision, the weights reflecting the probabilities of the alternative events that produce the possible payoff. It is expressed mathematically as the product of an event's probability of occurrence and the gain or loss that will result. Future cash inflows can be considered as a random variable that assumes various values with time along with some probability attached to that specific event that it will assume a particular value. First the probabilities are allocated or assigned to the future events and then the expected monetary value is computed. This is obtained by multiplying the monetary values of the possible events by the probabilities.

Thus,

$$(\text{EMV})_t = \bar{A}_t = \sum_{i=1}^{n} A_{it}\, p_{it}$$

Here A_{it} is the cash inflow for the ith event in the tth time period.

Hillier and Hertz's Models

Hillier's Model: Hillier states that the uncertainty or the risk associated with a capital expenditure proposal can be illustrated by the standard deviation of the expected cash flows. Alternatively, the more certain a project is lesser would be the deviation of various cash flows from the mean cash flows. For this, Hillier has developed a specific model to evaluate the various alternative cash flows that may occur from a capital expenditure plan. He takes into account the mean of present value of the cash flows and the standard deviation of such cash flows which may be determined with the help of the formulae.

Hertz's Model: Hertz states that simulation technique of operations research may be used in capital budgeting problems.

Typically, Hillier and Hertz further developed the Discounted Cash Flow (DCF) method for probabilistic cash flows. Hillier postulates normal distribution for the expected NPV of cash flows of the previous years and also gives formulae to compute the standard deviation that can be used to compare the risk for various plans. Thus,

$$\text{Mean of NPV, } M = \sum_{t=0}^{n} (1+r)^{-t} M_t$$

$$\text{Variance of NPV, } s^2 = \sum_{t=0}^{n} (1+r)^{-2t} \cdot \sigma_t^2$$

Here M_t is the mean and σ_t is the standard deviation of the cash flow distribution (assumed normal) of the tth period and r is the discounting factor.

4.8 SUMMING UP

- Competition is a watchword of modern life. A competitive situation exists if two or more individuals are making decisions in situation that involves conflicting interests and in which the outcome is controlled by the decisions of all parties concerned.
- The theory of games is simply the logic of rational decisions. The term 'game' represents a conflict between two or more parties and so game theory is really the 'science of conflict.'
- In game theory, the number of competitors is finite and there is a conflict of interests between the participants.
- The outcome of the game is affected by choices made by all of the players; the choices are to be made simultaneously so that no competitor knows his opponent's choice until he is already committed to his own.
- In order to eliminate the necessity for written descriptions of the 'payoffs', a standard set of conventions has been established in game theory.

NOTES

Check Your Progress

11. What is investment analysis? Name its various methods.
12. Define the process of investment.
13. Why break-even analysis is used?
14. How the payback period is computed if the project generates constant annual cash inflows?
15. What is IRR?
16. What is discount factor?

NOTES

- The term 'strategy' is also used in game theory. It refers to the total pattern of choices employed by any player. It may be defined as a complete set of plans of action specifying precisely what the player will do under every possible future contingency that might occur during the play of the game. Strategy could be a pure or a mixed one.
- In two-person games the players may have many possible choices open to them for each play of the game but the number of players remains only two.
- The saddle point in a payoff matrix is one which is the smallest value in its row and the largest value in its column. It is also known as equilibrium point in the theory of games.
- The maximin-minimax principle is used for the selection of optimal strategies by two players.
- A game without saddle point can be solved by various solution methods.
- The concept of dominance is very useful for reducing the size of the game. Applying this concept a bigger game can be converted into a smaller game.
- Investment analysis is the study of the likely return from a proposed investment with the objective of evaluating the amount an investor may pay for it. There are various methods of investment analysis, including cash on cash return, payback period, internal rate of return and net present value.
- Break-even analysis or what is also known as profit contribution analysis is an important analytical technique used to study the relationship between the total costs, total revenue and total profits and losses over the whole range of stipulated output.
- The PayBack (PB) is one of the most popular and widely recognized traditional methods of evaluating investment proposals. Payback is the number of years required to recover the original cash outlay invested in a project.
- The Accounting Rate of Return (ARR), also known as the Return On Investment (ROI), uses accounting information, as revealed by financial statements, to measure the profitability of an investment.
- The Net Present Value (NPV) method is the classic economic method of evaluating the investment proposals. It is a DCF technique that explicitly recognizes the time value of money.
- The Internal Rate of Return (IRR) is the rate that equates the investment outlay with the present value of cash inflow received after one period. This also implies that the rate of return is the discount rate which makes NPV = 0.
- The discount rate is the interest rate the Federal Reserve charges on loans it makes to banks and other financial institutions. The interest rate is used in determining the present value of future cash flows.
- The discount factor, $P(T)$, is the number which a future cash flow to be received at time T must be multiplied by in order to obtain the current present value.
- Certainty equivalent is an amount that would be accepted today (risk free) in lieu of a chance to receive a possibly higher, but uncertain amount. In addition, it is a return on a risky event (such as, a future income) stated in terms of a risk-free event (such as, a cash payment).

NOTES

- The Expected Monetary (Money) Value (EMV) is the total of the weighted outcomes or payoffs associated with a decision, the weights reflecting the probabilities of the alternative events that produce the possible payoff. It is expressed mathematically as the product of an event's probability of occurrence and the gain or loss that will result.

4.9 KEY TERMS

- **Saddle point:** It is a position in the payoff matrix where the maximum of row minima coincides with the minimum of column maxima
- **Value of the game:** It is the payoff at the saddle point
- **Two-person game:** In this, the number of players is only two though they have many possible choices open to them for each play of the game
- **Zero-sum game:** It is a game in which the sum of the payments to all the competitors is zero for every possible outcome of the game
- **Games with finite number of moves:** These are the games where the number of moves is limited to a fixed magnitude before play begins
- **Accounting rate of return:** Ratio of the average after tax profit divided by the average investment. It is also called return on investment
- **Discounted payback period:** Number of periods taken to recover the investment outlay on the present value basis

4.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The term 'game' represents a conflict between two or more parties; game theory is really the 'science of conflict.' Game theory is not concerned with finding an optimum or winning strategy for a particular conflict situation but it provides general rules concerning the logic that underlies strategic behavior of all types.
2. Game theory applies to those competitive situations which are technically known as 'competitive games' or simply 'games'.
3. Any situation to be termed as a game must possess the following properties:
   (*i*) The number of competitors is finite.
   (*ii*) There is a conflict of interests between the participants.
4. The saddle point in a payoff matrix is one which is the smallest value in its row and the largest value in its column.
5. The maximin-minimax principle is used for the selection of optimal strategies by two players.
6. A totally mixed strategy is a mixed strategy in which the player assigns a strictly positive probability to every pure strategy.
7. Dominated rows or columns may be deleted which reduces the size of the game. Always look for dominance when solving a game.
8. (*i*) If all the elements in a column are greater than or equal to the corresponding elements in another column, then that column is dominated.

NOTES

(*ii*) Similarly, if all the elements in a row are less than or equal to the corresponding elements in another row, then that row is dominated.

9. Game theory has been used in business and industry to develop bidding tactics, pricing policies, advertising strategies and timing of the introduction of new models into the market.
10. The important limitations of game theory are:
    (*i*) Businessmen do not have all the knowledge required by the theory of games.
    (*ii*) There is a great deal of uncertainty that cannot conveniently be inserted into game-theory models.
    (*iii*) The minimax strategy implies that the businessman minimizes the chance of the maximum loss only.
11. Investment analysis is the study of the likely return from a proposed investment with the objective of evaluating the amount an investor may pay for it. There are various methods of investment analysis, including cash on cash return, payback period, internal rate of return and net present value.
12. The process of investment involves a series of activities leading to the purchase of securities or other investment alternatives. The investment process can be divided into five stages: (*i*) Framing of investment policy (*ii*) Investment analysis (*iii*) Valuation (*iv*) Portfolio construction (*v*) Portfolio evaluation.
13. Break-even analysis or what is also known as profit contribution analysis is an important analytical technique used to study the relationship between the total costs, total revenue and total profits and losses over the whole range of stipulated output.
14. Payback is the number of years required to recover the original cash outlay invested in a project. If the project generates constant annual cash inflows, the payback period can be computed by dividing cash outlay by the annual cash inflow.
15. The Internal Rate of Return (IRR) is the rate that equates the investment outlay with the present value of cash inflow received after one period. This also implies that the rate of return is the discount rate which makes NPV = 0.
16. The discount factor, $P(T)$, is the number which a future cash flow to be received at time T must be multiplied by in order to obtain the current present value.

4.11 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the meaning of the term 'game'.
2. Write the significance of the game theory.
3. What is a saddle point? Discuss its significance in the game theory.
4. What are the various methods through which the game without the saddle point can be solved?
5. What are mixed strategies?
6. State the rule of dominance.
7. What are the drawbacks of the game theory?

NOTES

8. List the features of investment decisions.
9. Why do investment decisions require special attention?
10. What are the merits of an ARR method?
11. What are the merits of an IRR method?
12. What are the characteristic reasons for high discount rates for startups?
13. What do you understand by the term Expected Monetary (Money) Value (EMV)?
14. What does Hillier and Hertz's models state?

Long-Answer Questions

1. For what type of business problems might game theory be helpful? Explain.
2. Given is the payoff matix in rupees for a hypothetical game:

Strategies Open to A	*Strategies Open to B*		
	y	*x*	*z*
a	3	4	15
b	11	8	13
c	13	6	4

Determine the optimum strategies for *A* and *B*, and also find the value of the game. Can it be described as a game of pure strategy? If so, why?

3. Firms *X* and *Y* are competing for business. Whatever *X* gains *Y* loses. The matrix shows the utilities to firm *X* for various market shares. Assume that the game is zero-sum.

X \ Y	No Advertising	Medium Advertising	Large Advertising
No Advertising	60	50	40
Medium Advertising	70	70	50
Large Advertising	80	60	75

Find an equilibrium solution to the above game.

4. The following is a 2 × 6 payoff matrix in rupees for *A*:

Strategies of *B*	Strategies of *A*					
	A_1	A_2	A_3	A_4	A_5	A_6
B_1	1	3	–1	4	2	–5
B_2	–3	5	6	1	2	0

Determine the optimal strategies for *A* and *B* and also work out the value of the game.

NOTES

5. The matrix of a certain, two-person zero-sum 3 × 3 game is given below:

$$\begin{array}{c} \\ A \end{array} \begin{array}{c} B \\ \begin{bmatrix} 1 & 0 & 2 \\ 3 & 0 & 0 \\ 0 & 2 & 1 \end{bmatrix} \end{array}$$

What should be the optimal strategies for A and B in the above game?

6. For each of the following game matrices, find the optimal strategy for each player. What is the saddle point and the value of each game? Is the game fair? Explain.

(a)

		Red		
		1	2	3
Blue	A	3	−3	−1
	B	0	4	−2
	C	1	3	0

(b)

		Black		
		1	2	3
White	A	−4	−3	4
	B	6	−2	3
	C	5	−4	−5

7. Given is a game matrix:

		Red		
		1	2	3
Black	A	−2	5	3
	B	−3	1	2
	C	4	0	1

Is it a game with a saddle point? If not, then determine the mixed strategies along with the value of the game.

8. Discuss the techniques and practices of capital investment.
9. Describe the concept of payback method and Net Present Value (NPV) Method.
10. Explain why NPV is important and how is it evaluated?

4.12 REFERENCES AND SUGGESTED READINGS

Swaroop, Kanti, P.K. Gupta and Man Mohan. 2001. *Operations Research*. New Delhi: Sultan Chand & Sons.

Vohra, N. D. 2006. *Quantitative Techniques in Management*, 3rd edition. New Delhi: Tata McGraw-Hill Education.

Taha, Hamdy A. 1992. *Operations Research: An Introduction*. New York: Macmillan.

Gupta, P.K. and D.S. Hira. 2002. *Introduction to Operations Research*. New Delhi: S. Chand & Co.

Sharma, S. D. 2006. *Operations Research*. Uttar Pradesh: Kedar Nath Ram Nath & Co.

UNIT 5 SYSTEM ANALYSIS AND DESIGN

NOTES

UNIT STRUCTURE

5.0 Introduction
5.1 Objectives
5.2 Concept of System
- 5.2.1 Characteristics of a System
- 5.2.2 Types of Systems
- 5.2.3 Information System

5.3 System Development Life Cycle and its Phases
- 5.3.1 Design Development Methodologies

5.4 System Analysis
- 5.4.1 System Planning and Initial Investigation
- 5.4.2 Information Gathering
- 5.4.3 Structured Analysis
- 5.4.4 Data Flow Diagram (DFD)
- 5.4.5 Data Dictionary
- 5.4.6 Decision Trees
- 5.4.7 Decision Tables
- 5.4.8 Structured English

5.5 Process and Stages of System Design
- 5.5.1 Input/Output and File Design
- 5.5.2 Logical and Physical Design
- 5.5.3 File Organization
- 5.5.4 Database Design
- 5.5.5 Form Design
- 5.5.6 Design Methodologies

5.6 System Implementation
- 5.6.1 System Testing
- 5.6.2 Quality Assurance
- 5.6.3 Hardware/Software Selection
- 5.6.4 System Security

5.7 Summing Up
5.8 Key Terms
5.9 Answers to 'Check Your Progress'
5.10 Questions and Exercises
5.11 References and Suggested Readings

5.0 INTRODUCTION

In this unit, you will learn about the basic concept of system analysis and design. Systems are created to solve problems. A collection of components that work together to realize some objectives forms a system. Basically there are three major components in every system, namely input, processing and output. Systems Development Life Cycle or SDLC is a series of steps agreed upon by software developers in creating special software. Developers adhere to certain steps to ensure that they have the right software for the right demand. In this unit, you will learn about the designing of a system. Data Flow Diagrams or DFDs are used to specify how the data flows between the functions of a system. The basic purpose of DFDs is to show how the system is currently implemented.

NOTES

A DFD is the starting point of the design phase that functionally decomposes the requirement specifications down to the lowest level of detail. Data dictionary is one of the most important reference works that contain information about data which is also known as metadata. The entire process of designing the software system is an exercise that specifies the 'what' and the 'how' of the system.

You will also learn about file organization and database design. File organization refers to the relationship of the key of the record to the physical location of that record in the computer file. The file organization may be either physical file or a logical file. A physical file is a physical unit, such as magnetic tape or a disk whereas a logical file on the other hand is a complete set of records for a specific application or purpose. Suitable database design is based on schema. One major advantage to the database schema is easy to make. Tables in the relational model are used to represent 'things in the real world'. Each table should represent only one thing. These things refer to entities and can be real world objects or events. In this unit, you will learn about the system testing and implementation. System testing provides stakeholders with information about the quality of the product or service under test. You will also learn about the factors used to determine the quality assurance. Conversion is a method of replacing the old software with a new one. A conversion plan includes a description of all those activities which must occur during the implementation of a new system. Finally, you will learn about system implementation, quality assurance and system security.

5.1 OBJECTIVES

After going through this unit, you will be able to:

- Describe the significance of system
- Explain the phases in system development life cycle
- Discuss system development methodologies
- Understand about the system analysis
- Learn about various information gathering tools
- Explain the process and stages of system design
- Learn the process of system implementation
- Discuss the basic concepts in system testing, quality assurance and system security

5.2 CONCEPT OF SYSTEM

A system is a set of organized principles, tasks, activities and procedures that defines the working of a particular establishment to achieve an objective. It is logically arranged and governed by rules, regulations and policies. Thus, a system is a process or a set of processes taking one or more inputs to produce outputs. These outputs are according to the objectives set by the policy makers. Thus, the fundamental components of a system are input, process and output which are depicted in Figure 5.1.

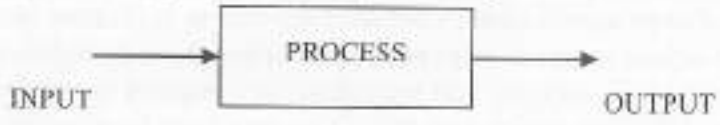


Fig. 5.1 Fundamental Components of a System

NOTES

A system is defined and determined by its boundaries and objectives. It is an arrangement of smaller systems in a logical order. When many smaller systems together make a larger system, the smaller systems are called the subsystems of the larger system and the larger system is called a super system.

The functioning of coordinated systems (process) includes receiving inputs from one system and giving output to other systems. A big system may have many subsystems and sub subsystems synchronized to produce the desired objectives which are outputs from the system. For example, an information system is a system that accepts data resources as input and processes them into informational products as output. It provides information when a set of organized procedures is executed.

Take the example of a garment manufacturing system where the inputs are fabric, thread, labour and utilities like electricity, etc. The process consists of taking measurements, cutting, sewing, etc. The output is in the form of shirts, trousers, jackets, etc., depending on products in which the organization is dealing. The whole business system related to the garment business has other subsystems like recruitment and training system, accounting system, procurement system, marketing system, etc. The bigger the business, the more the number of subsystems.

System Boundary

Things directly controlled by the system falls within the system boundary. Things falling outside the system boundary form the environment of the system. A system may be a closed system or an open system. In an open system, you can go on adding subsystems and it grows bigger and bigger. In a closed system, it is not so. Environment of a system means other systems that influence the system under consideration. This includes market, political systems and financial guidelines or statutes.

5.2.1 Characteristics of a System

The term 'system' is derived from the Greek word 'systema' (to combine) which means an organized relationship among functioning units or components. A system exists because it is designed to achieve one or more objectives. Following are the characteristics of a system:

- **Central Objective:** Well developed and user friendly systems are important to businesses in various ways. The use of computers is time saving for businesses and all financial information for the business which is well organized.
- **Components of a System:** A system is a collection of components organized to accomplish a specific function or a set of functions. A well developed system is an orderly arrangement of its components. The components of a system have specific predefined structure and order. The organization determines the flow of control, communications and chain of commands.
- **Time Saving:** A system is developed to save organization's time and money. A well developed system makes inputting information simple. Transactions are entered into the system and the system posts transactions accordingly.
- **Organization:** User friendly online system helps businesses in staying organized. When information is entered into the system, it makes finding the information easy, for example employees can look up any financial information whenever it is needed.

NOTES

- **Storage:** Storing information is vital to a business. After information is entered into the system, the information is stored indefinitely. Organizations arrange backups on the system regularly to avoid losing any information.
- **Distribution:** Computerized systems allow companies to distribute financial information easily. Financial statements are printed directly from the system and are distributed internally and externally to those needing the information.
- **Interdependence and Integration:** System integration is bringing together the component subsystems into one system which ensures that the subsystems function together as a system. In the field of system analysis and design, system integration is the process of linking together different computing systems and software applications physically or functionally to act as a coordinated whole. System integration also adds values and capabilities to the system to interact between the subsystems.
- **Maintainability and Dependability:** System supports critical characteristics known as 'maintainability' because software change is an evitable consequence of a changing business environment. Dependability characteristic includes readability, security and safety. A well developed system must support usability without undue effort by the type of user for whom it is designed. It means proposed system must have a suitable user interface and adequate documentation.

5.2.2 Types of Systems

A system is set of inter related tasks and procedures working in coordination towards achieving a common objective by receiving the inputs and providing the outputs in an organized transformation process. Such a system can be called as **dynamic system** and has three basic components called input, process and output. The input components capture and assemble the elements that enter the system for processing. The output components produce the processed elements in the form of information products to the other system for further processing or as a final objective. Figure 5.2 shows model of a dynamic system.

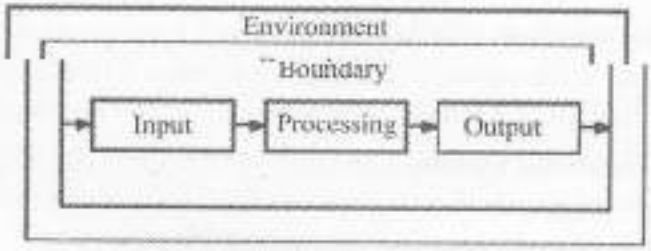


Fig. 5.2 Model of a System

All systems operate in an environment where a system is designed to achieve certain objectives, i.e., it sets the boundaries for itself. The environment influences the choice of inputs, the procedures for the processing and nature of the output. A system which includes two or more components, called feedback and control with input and output components is more effective. Such a self regulating and self monitoring system is called a **cybernetic system.**

NOTES

Business comprises of a set of activities pertaining to commerce, manufacturing or services resulting in producing economic output intended to make a profitable gain. A **business system** is a system with profit objectives or non-profit objectives. It can be private as in the case of a corporation or it may be public as in the case of a government unit. The manager is the control element in the system, keeping it on course as it moves towards its goals. The business firm is a physical system which is tangible. A **conceptual system** is a system that represents a physical system. The managers use a conceptual system to manage the physical one. The physical system is important for what it is, the conceptual system is important for the representation of the physical system.

A system in which the inputs, process and outputs are known with certainty is called a **deterministic system**. For example, when one condition is processed, a certain resultant condition will always occur. An accounting system is also an example of a deterministic system. A system is called probabilistic when the output can only be predicted in probabilistic terms. The nature and working of stock control and cash flow systems can be estimated however, few of the financial elements are variable and can be considered within the estimation of that system. These can be predicted in terms of probability of occurrence. A demand forecasting system can be called a **probabilistic system**.

A system, which is influenced by environmental changes and has an exchange with the environment, is called an **open system**. For example, any business organization can be considered an open system, as it is constantly affected by the changes in its environment. Its functions, development and results affect the environment in the same manner.

The data that is received about system performance is called feedback. Control is an important system function that monitors and evaluates the feedback to establish if the system is going towards achieving its goal. To ensure that the proper output is produced, the control component can make needed adjustments to the input and processing components of the system. A system performs towards its goal such a feedback is called positive feedback and if it is declining or deviating from its goal, then it can be called as a negative feedback. Figure 5.3 shows a cybernetic system.

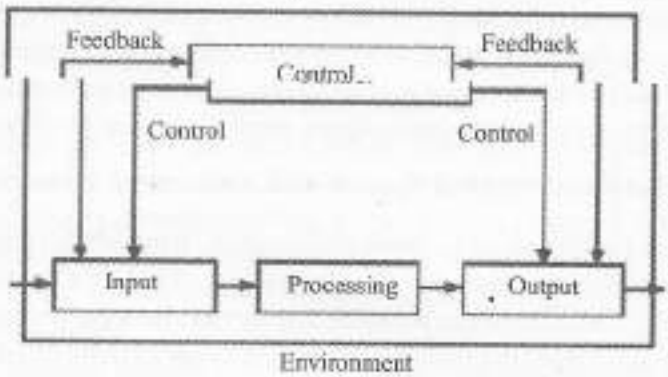


Fig. 5.3 A Cybernetic System

If a system has the ability to change by itself or its environment in order to survive, it is known as an **adaptive system**. In general life, human beings work towards fulfilling their daily requirements and achieving the goals of their life. This process makes them self adaptive. In similar way, the establishments and systems also work towards making themselves self adaptive by creating new systems and managing the organizations.

NOTES

An **Open Computer Based Information System (OCBIS)** is a type of system which applies an adequate open specifications for services, supporting formats and interfaces to allow engineered components that are proper to be used across a broad variety of systems with fewer alterations to interoperate with different components on remote and local systems and to interrelate with users in a fashion that make portability possible. Computer based information system include the Management Information System (MIS) and Decision Support System (DSS). **Man made information system** is the foundation or interface between the uses and the analyst. In determines the nature of relationships among decision-makes. **Open systems** promise a faster and more economical development of high quality systems that are technologically up-to-date.

If a system neither accepts nor provides any input or output to the external environment, such a system can be called a **closed system**. A closed system is not affected from the surrounding environment and can perform being independent of the happenings in the environment. The occurrences in an environment do not affect the nature and functioning of the closed system. A closed system is set up ensuring that all the raw and processed materials to be used can be recycled and there will be no loss of energy in the production process. A closed system can be compared to a closed or self contained society or community. Although, these closed systems or societies work with great perfection and yield results, these closed frameworks also encounter various problems. As the closed system is not adaptive to the happenings of the outside world and external influences, it becomes resistant to change. Also, in closed system, the organizations find it hard to adapt to the changes in the trends and preferences of the customers and clients and it affects the growth of the closed systems.

A **purposive system** is defined as a system which works towards achieving a set of related goals. Therefore, all types of systems can be termed as purposive system. An information system is an open purposive system that produces information using the input, process and output cycle. The simplest information consists of elements, such as people, procedures and data. People follow procedures to manipulate data to produce information. There are systems which consist of procedures that everyone is expected to follow.

When a policy or procedure is spelled out in writing, such a system can be called a **formal system**. Information provided by a computer system is the output of a formal system, such as information provided by weekly conferences and typed monthly reports. Systems that are not documented in writing are called informal systems. Examples of informal systems are telephone calls, letters, memos and special meetings.

Difference between Manual System and Automated System

SDLC refers to a methodology for developing systems. It provides a consistent framework of tasks and deliverables needed to develop systems. The SDLC methodology may be condensed to include only those activities appropriate for a particular project whether the system is automated or manual whether it is a new system, or an enhancement to existing systems. The SDLC methodology tracks a project from an idea developed by the user through a feasibility study, systems analysis and design, programming, pilot testing, implementation, and post-implementation analysis. Following are the difference between manual system and automated system:

- A manual system usually means done by hand. That could be typing all your information into Notepad whereas automated system refers to the use of control systems and information technologies to reduce the need for human work in the production of goods and services.

NOTES

- Manual system purely depends on the individual work capacity whereas automated system requires a skilled quality assurance professional with knowledge of the automation tool and the software being tested to set up the tests.
- In the manual system, everything from booking to other operations is to be done with instruction by the management manually. But, the automated systems provide a control system for the growing businesses, such as hotel and centralized real time supervision hence enabling more efficiency and control of alarms, air conditioning, electricity, heating, etc.
- Manual system refers to a system involving data processing which does not make use of stored program computing whereas automated system refers to the use of control systems and information technologies to reduce the need for human work in the production of goods and services. An automated system may have limited level of intelligence hence it is most likely susceptible to commit error.
- Manual system uses human skills whereas automated system uses human as well as computer system and its resources, such as printer, mouse, sharing the files and documents, networking facilities, etc.
- In computing, the purpose of manual system is to use human skill which is not based on cost and time whereas the purpose of automated system is to use broader productivity which is based on cost and time.

5.2.3 Information System

Information Systems (IS) are a special class of systems that are used for data storage, retrieval, processing, communication and security. Information systems that help management at different levels to take suitable decisions are called Management Information Systems (MIS). Information systems are housed in a computerized environment /platform to enable users to get accurate information immediately.

Information systems are data processing systems, which collect data from different sources, process that data and generate information from the data which can further be used for different applications within the organization. For example, in a business context, an information system collects data from various systems such as finance and sales systems from the supplier side. The information system processes the data and generates information for the customer. Customers provide feedback to the supplier depending on the information processed by the information system.

Information systems are basically systems that help to maintain and manage information. An information system helps to manage and store information to perform various functions such as decision-making, documentation of business activities and generation of reports for analysis of organizational operations. One needs to understand the concept of information and system to acquire a basic knowledge of information systems. Various terms used in information systems are as follows:

- *Data* is the raw material that can be a number, a fact, a sound, a picture or a statement gathered from different sources. In the real world, data can represent anything related to business processes and employee details.
- *Information* is meaningful data or processed data. It defines the relation between the different data.
- *System* is refers to a collection of components that helps in achieving a common objective. For example, in a human–machine system, the machine element consists of hardware and software to perform computation, and people make decisions based on this computation.

5.3 SYSTEM DEVELOPMENT LIFE CYCLE AND ITS PHASES

NOTES

The constant need to make accurate and efficient software has led to the formalization of certain stages and phases in the process of program creation. Systems Development Life Cycle (SDLC) is a series of steps agreed upon by software developers in creating special software. Developers adhere to certain steps to ensure that they have the right software for the right demand.

SDLC is an approach and a set of activities carried out by systems analysts and systems designers to develop an information system. It consists of the following activities, which are to:

- Identify problems, opportunities and objectives.
- Determine information requirements.
- Analyse the system needs.
- Design the recommended system.
- Develop and document software.
- Test and maintain the system.
- Implement and evaluate the system.

Remember that all subsystems in organizations are interrelated. A system cannot be implemented without considering its effect on the rest of the organization. Possible questions to be raised during project selection are as follows:

- Do we have support from management?
- Is this an appropriate time for project commitment?
- Does the project help the organization to achieve its goal?
- Do we have adequate resources and expertise?
- Is the development of the project worthwhile?

These requests are scrutinized by a committee to determine the orders that have to be approved. The approved system request is called planned system initiative. The requests that are not approved are called backlogged requests. A planned system initiative leads to the activities of SDLC.

Phases of the Systems Development Life Cycle

In many business organizations, all the phases and activities of software development are closely related. When a system owner or a user makes a request for a new system or some modifications in an existing system, the first system activity, i.e., identifying problems in the existing system, begins. Consider a scenario of a banking organization currently working in a manual working system. Now, the organization wants to computerize the entire working system. For this, the organization needs to follow all the phases of SDLC. The following are the steps that the organization needs to follow to implement the new system in the organization:

- Identify the problems in the existing manual system and the objectives of the proposed system.
- Determine the information required to implement the proposed system.

- Analyse the proposed system requirements.
- Develop the design to implement the proposed system.
- Implement and prepare the documentation of the proposed system.
- Test and maintain the working of the proposed system.
- Evaluate the performance of the proposed system.

Figure 5.4 shows the phases in the SDLC.

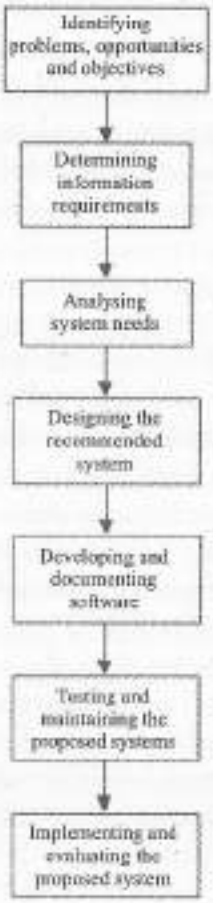


Fig. 5.4 Phases in SDLC

5.3.1 Design Development Methodologies

Every software process is characterized by basic concepts along with certain practices or methods. Methods represent the manner through which the concepts are applied. As new technology replaces older technology, many changes occur in the methods that are used to apply the concepts for the development of software. However, the fundamental concepts underlining the software design process remain the same, some of which are described here.

Abstraction

Abstraction refers to a powerful design tool, which allows software designers to consider components at an abstract level, while neglecting the implementation details of the components. **IEEE** defines abstraction as '*a view of a problem that extracts the essential information relevant to a particular purpose and ignores the remainder of the information.*' The concept of abstraction can be used in two ways: as a process

NOTES

NOTES

and as an entity. As a **process**, it refers to a mechanism of hiding irrelevant details and representing only the essential features of an item so that one can focus on important things at a time. As an **entity**, it refers to a model or view of an item.

Each step in the software process is accomplished through various levels of abstraction. At the highest level, an outline of the solution to the problem is presented whereas at the lower levels, the solution to the problem is presented in detail. For example, in the requirements analysis phase, a solution to the problem is presented using the language of problem environment and as we proceed through the software process, the abstraction level reduces and at the lowest level, source code of the software is produced.

There are three commonly used abstraction mechanisms in software design, namely, *functional abstraction*, *data abstraction* and *control abstraction*. All these mechanisms allow us to control the complexity of the design process by proceeding from the abstract design model to concrete design model in a systematic manner.

- **Functional Abstraction:** This involves the use of parameterized subprograms. Functional abstraction can be generalized as collections of subprograms referred to as 'groups'. Within these groups there exist routines which may be visible or hidden. Visible routines can be used within the containing groups as well as within other groups, whereas hidden routines are hidden from other groups and can be used within the containing group only.
- **Data Abstraction:** This involves specifying data that describes a data object. For example, the data object *window* encompasses a set of attributes (window type, window dimension) that describe the *window* object clearly. In this abstraction mechanism, representation and manipulation details are ignored.
- **Control Abstraction:** This states the desired effect, without stating the exact mechanism of control. For example, if and while statements in programming languages (like C and C++) are abstractions of machine code implementations, which involve conditional instructions. In the architectural design level, this abstraction mechanism permits specifications of sequential subprogram and exception handlers without the concern for exact details of implementation.

Architecture

Software architecture refers to the structure of the system, which is composed of various components of a program/system, the attributes (properties) of those components and the relationship amongst them. The software architecture enables the software engineers to analyze the software design efficiently. In addition, it also helps them in decision-making and handling risks. The software architecture does the following.

- Provides an insight to all the interested stakeholders that enable them to communicate with each other
- Highlights early design decisions, which have great impact on the software engineering activities (like coding and testing) that follow the design phase
- Creates intellectual models of how the system is organized into components and how these components interact with each other.

Currently, software architecture is represented in an informal and unplanned manner. Though the architectural concepts are often represented in the infrastructure (for supporting particular architectural styles) and the initial stages of a system configuration, the lack of an explicit independent characterization of architecture restricts

the advantages of this design concept in the present scenario. Note that software architecture comprises two elements of design model, namely, *data design* and *architectural design*. Both these elements have been discussed later in this chapter.

NOTES

Patterns

A pattern provides a description of the solution to a recurring design problem of some specific domain in such a way that the solution can be used again and again. The objective of each pattern is to provide an insight to a designer who can determine the following.

- Whether the pattern can be reused
- Whether the pattern is applicable to the current project
- Whether the pattern can be used to develop a similar but functionally or structurally different design pattern.

Types of Design Patterns

Software engineer can use the design pattern during the entire software design process. When the analysis model is developed, the designer can examine the problem description at different levels of abstraction to determine whether it complies with one or more of the following types of design patterns.

- **Architectural Patterns:** These patterns are high-level strategies that refer to the overall structure and organization of a software system. That is, they define the elements of a software system such as subsystems, components, classes, etc. In addition, they also indicate the relationship between the elements along with the rules and guidelines for specifying these relationships. Note that architectural patterns are often considered equivalent to *software architecture*.
- **Design Patterns**: These patterns are medium-level strategies that are used to solve design problems. They provide a means for the refinement of the elements (as defined by architectural pattern) of a software system or the relationship among them. Specific design elements such as relationship among components or mechanisms that affect component-to-component interaction are addressed by design patterns. Note that design patterns are often considered equivalent to *software components*.
- **Idioms**: These patterns are low-level patterns, which are programming-language specific. They describe the implementation of a software component, the method used for interaction among software components, etc., in a specific programming language. Note that idioms are often termed as coding patterns.

Modularity

Modularity is achieved by dividing the software into uniquely named and addressable *components*, which are also known as **modules**. A complex system (large program) is partitioned into a set of discrete modules in such a way that each module can be developed independent of other modules (see Figure 5.5). After developing the modules, they are integrated together to meet the software requirements. Note that larger the number of modules a system is divided into, greater will be the effort required to integrate the modules.

NOTES

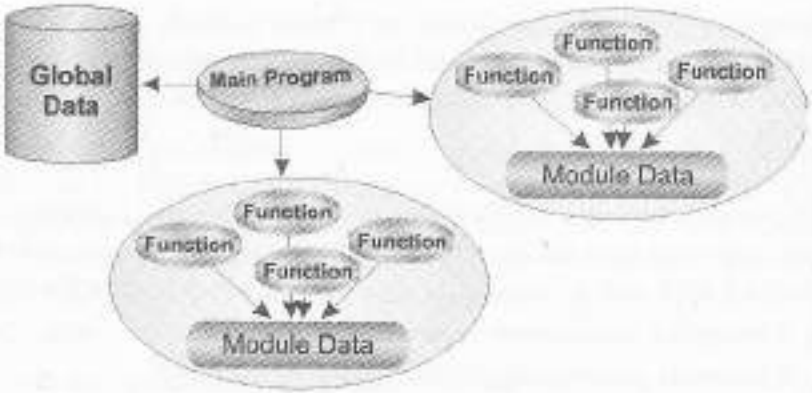


Fig. 5.5 Modules in Software Programs

Modularizing a design helps to plan the development in a more effective manner, accommodate changes easily, conduct testing and debugging effectively and efficiently, and conduct maintenance work without adversely affecting the functioning of the software.

Information Hiding

Modules should be specified and designed in such a way that the data structures and processing details of one module are not accessible to other modules (see Figure 5.6). They pass only that much information to each other, which is required to accomplish the software functions. The way of hiding unnecessary details is referred to as **information hiding. IEEE** defines information hiding as '*the technique of encapsulating software design decisions in modules in such a way that the module's interfaces reveal as little as possible about the module's inner workings; thus each module is a 'black box' to the other modules in the system.*'

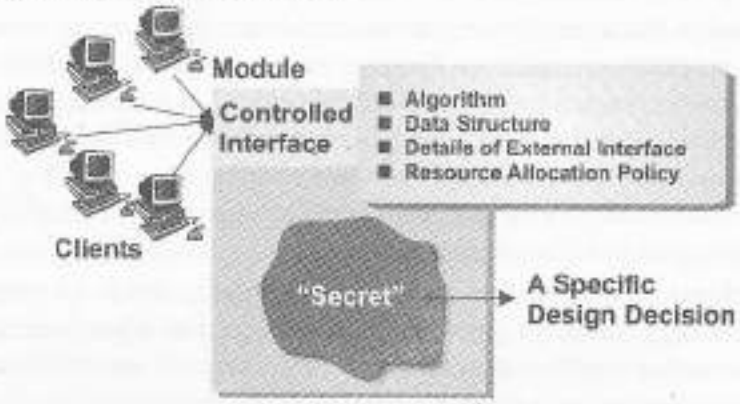


Fig. 5.6 Information Hiding

Information hiding is of immense use when modifications are required during the testing and maintenance phase. Some of the advantages associated with information hiding are listed below.

- Leads to low coupling.
- Emphasizes communication through controlled interfaces.
- Decreases the probability of adverse effects.
- Restricts the effects of changes in one component on others.
- Results in higher quality software.

NOTES

Stepwise Refinement

Stepwise refinement is a top-down design strategy used for decomposing a system from a high level of abstraction into a more detailed level (lower level) of abstraction. At the highest level of abstraction, function or information is defined conceptually without providing any information about the internal workings of the function or internal structure of the data. As we proceed towards the lower levels of abstraction, more and more details are available.

Software designers start the stepwise refinement process by creating a sequence of compositions for the system being designed. Each composition is more detailed than the previous one and contains more components and interactions. The earlier compositions represent the significant interactions within the system, while the later compositions show in detail how these interactions are achieved.

To have a clear understanding of the concept, let us consider an example of stepwise refinement. Every computer program comprises *input*, *process*, and *output*.

1. INPUT
   - Get user's name (string) through a prompt.
   - Get user's grade (integer from 0 to 100) through a prompt and validate.
2. PROCESS
3. OUTPUT

This is the first step in refinement. The input phase can be refined further as given here.

1. INPUT
   - Get user's name through a prompt.
   - Get user's grade through a prompt.
   - While (invalid grade)
     Ask again:
2. PROCESS
3. OUTPUT

Note: Stepwise refinement can also be performed for PROCESS and OUTPUT phase.

Refactoring

Refactoring is an important design activity that reduces the complexity of module design keeping its behaviour or function unchanged. Refactoring can be defined as a process of modifying a software system to improve the internal structure of design without changing its external behavior. During the refactoring process, the existing design is checked for any type of flaws like redundancy, poorly constructed algorithms and data structures, etc., in order to improve the design. For example, a design model might yield a component which exhibits low cohesion (like a component performs four functions that have a limited relationship with one another). Software designers may decide to refactor the component into four different components, each exhibiting high cohesion. This leads to easier integration, testing, and maintenance of the software components.

Structural Partitioning

When the architectural style of a design follows a hierarchical nature, the structure of the program can be partitioned either horizontally or vertically. In **horizontal partitioning**,

NOTES

the control modules (as depicted in the shaded boxes in Figure 5.7 (a)) are used to communicate between functions and execute the functions. Structural partitioning provides the following benefits.

- The testing and maintenance of software becomes easier.
- The negative impacts spread slowly.
- The software can be extended easily.

Besides these advantages, horizontal partitioning has some disadvantage also. It requires to pass more data across the module interface, which makes the control-flow of the problem more complex. This usually happens in cases where data moves rapidly from one function to another.

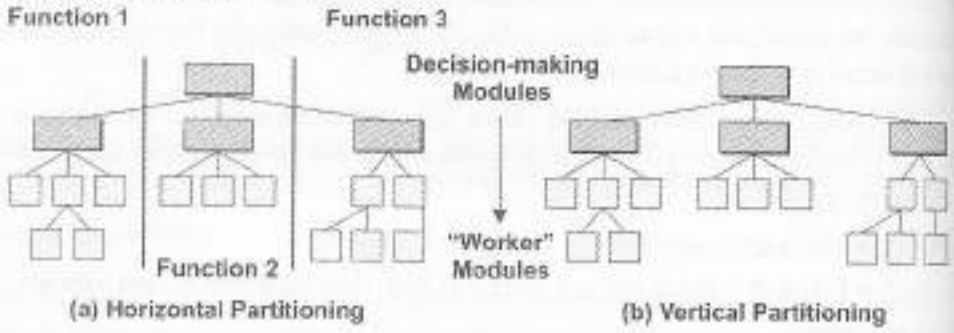


Fig. 5.7 Horizontal and Vertical Partitioning

In **vertical partitioning** (see Figure 5.7 (b)), the functionality is distributed among the modules in a top-down manner. The modules at the top level called **control modules** perform the decision-making and do little processing whereas the modules at the low level called **worker modules** perform all input, computation and output tasks.

Concurrency

Computer has limited resources and they must be utilized efficiently as much as possible. To utilize these resources efficiently, multiple tasks must be executed concurrently. This requirement makes concurrency one of the major concepts of software design. Every system must be designed to allow multiple processes to execute concurrently, whenever possible. For example, if the current process is waiting for some event to occur, the system must execute some other process in the mean time.

However, concurrent execution of multiple processes sometimes may result in undesirable situations such as an inconsistent state, deadlock, etc. For example, consider two processes A and B and a data item Q1 with the value '200'. Further, suppose A and B are being executed concurrently and firstly A reads the value of Q1 (which is '200') to add '100' to it. However, before A updates the value of Q1, B reads the value of Q1 (which is still '200') to add '50' to it. In this situation, whether A or B first updates the value of Q1, the value of Q1 would definitely be wrong resulting in an inconsistent state of the system. This is because the actions of A and B are not synchronized with each other. Thus, the system must control the concurrent execution and synchronize the actions of concurrent processes.

One way to achieve synchronization is mutual exclusion, which ensures that two concurrent processes do not interfere with the actions of each other. To ensure this, mutual exclusion may use locking technique. In this technique, the processes need to lock the data item to be read or updated. The data item locked by some process cannot be accessed by other processes until it is unlocked. It implies that the process, that needs to access the data item locked by some other process, has to wait.

5.4 SYSTEM ANALYSIS

NOTES

System analysis is a process in which data are collected and then interpreted to identify the problems within the system. Therefore, the collected information can then be used to recommend improvements in a system. In other words, system analysis includes identification, understanding and examining a system so that the objectives of the system can be achieved. The predetermined objectives of the system analysis include:

- Knowledge of the system operation.
- Identification of the user requirements in the proposed system.

The system analysis stage investigates the system operations and determines the solutions to solve the problem. Therefore, system analysis is considered as a logical process.

System analysis is really vital in the completion of the development process of a system. It may be possible that a user is aware of the problem but may not have the solution to it. In such a scenario, a system analyst works with the user so that a logical model of the system can be developed. A system analyst of technical background may move too quickly to the program design stage to make a system more physical. This approach is not desirable and should be avoided because it can affect the ultimate success of the system. A system analyst can work with a user to obtain the complete knowledge of a system. For having the complete information about a system, a logical model of the system is developed based on the detailed study of the system. The detailed study should be done using various modern tools and techniques, which include data flow diagrams, data dictionary and rough descriptions of the relevant algorithms. A set of system requirements for a proposed information system is the final product of the system analysis stage.

According to Orlikowski (1992), *'Nothing is more central to an organization's effectiveness than its ability to transmit accurate, relevant, understandable information amongst its employees. All of the advantages of an organization's economy of scale, financial and technical resources, diverse talents and contacts are of no practical value if the organization's employees are unaware of what other employees require of them.'*

The process of analysis involves a detailed study of the current system leading to specifications of a new system. The approach of analysis is a detailed study of various operations performed by a system and their relationships within and outside the system. During analysis, data are collected on the available files, decision points and transactions handled by the present system. Interviews, onsite observation and questionnaire are the tools used for system analysis. Using the following steps it becomes easy to draw the exact boundary of the new system under consideration:

- Keeping in view the problems and new requirements.
- Workout the pros and cons including new areas of the system.

All procedures, requirements must be analysed and documented in the form of detailed Data Flow Diagrams or DFDs, data dictionary, logical data structures and miniature specifications. System analysis also includes sub-dividing of complex process involving the entire system, identification of data store and manual processes. The features discussed in system analysis are as follows:

NOTES

- Specification of what the new system is to accomplish based on the user requirements.
- Functional hierarchy showing the functions to be performed by the new system and their relationship with each other.
- Function network which are similar to function hierarchy but they highlight those functions which are common to more than one procedure.
- List of attributes of the entities (data items) which need to be held about each entity, i.e., record.

The purpose of system analysis phase is to build a logical model of the new system. The first step is related to requirements modelling where you investigate business processes and documents what the new system must do to satisfy users. Requirements modelling continues the investigation which begins during the system planning phase. To understand the system, you perform information gathering, such as interviews, observations, surveys, document review, observation and sampling. You can use information gathering result produced by the tools to build business models, data and process models, and object models. The deliverable from the system analysis phase is the system requirement document. The system requirements document describes management and user requirements, costs and benefits, and outlines alternative development strategies. The system analysis phase placed in software development life cycle is shown in Figure 5.8.

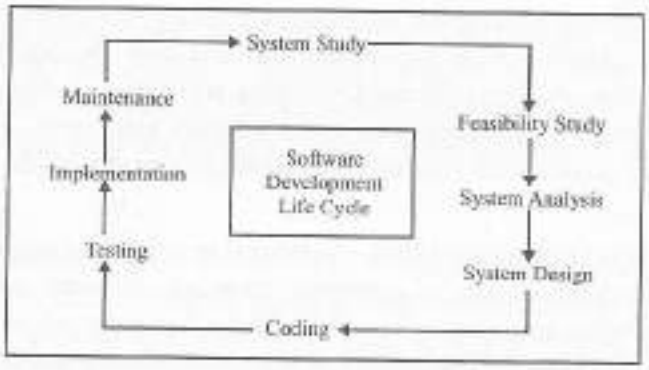


Fig. 5.8 System Analysis Phase in Software Development Life Cycle

System analysis is a process in which data are collected and then interpreted to identify the problems within the system. Therefore, the collected information can then be used to recommend improvements in a system. In other words, system analysis includes identification, understanding and examining a system so that the objectives of the system can be achieved. The predetermined objectives of the system analysis are as follows:

- It produces knowledge of the system operation.
- It identifies the user requirements for the proposed system.

The system analysis stage investigates the system operations and determines the solutions to solve the problem. Therefore, system analysis is considered as a logical process.

System analysis is really vital in the completion of the development process of a system. It may be possible that a user is aware of the problem but may not have the solution to it. In such a scenario, a system analyst works with the user so that a logical model of the system can be developed. A system analyst of technical background may move too quickly to the program design stage to make a system more physical. This

NOTES

approach is not desirable and should be avoided because it can affect the ultimate success of the system. A system analyst can work with a user to obtain the complete knowledge of a system. For having the complete information about a system, a logical model of the system is developed based on the detailed study of the given system. The detailed study should be done using various modern tools and techniques which includes data flow diagrams, data dictionary and rough descriptions of the relevant algorithms. A set of system requirements for a proposed information system is the final product of the system analysis stage.

5.4.1 System Planning and Initial Investigation

The system planning phase usually begins with a formal request known as system request which describes problems or desired changes in an information system or a business process. In many organizations, system planning is an integral part of overall system development life cycle. A system request comes from a top manager, a planning team or a department head. The purpose of this phase is to perform a preliminary investigation to evaluate business opportunities or problem. The preliminary or initial investigation is a critical step because the outcome will affect the entire development process. A key feature of preliminary investigation is a feasibility study that reviews anticipated costs and benefits and recommends the actions based on operational, technical, economic and time factors. Suppose you are a system analyst and you receive a request for system change or improvement. Then the first step of system analyst is to determine whether it launches a preliminary investigation at all. After an investigation, you might find that the information system functions properly but users need more training to be made aware of the systems functioning. But, the complete scale system review is necessary.

Preliminary investigation is the first step in SDLC. It begins once a project request is made. The purpose of this preliminary investigation is to evaluate the project request. However, this investigation does not completely leads to the description of the business system. Such investigation only helps in the collection of information for committee members to evaluate the merits of the project requests leading to an informed decision about the feasibility of the proposed project.

Systems analysts carry on the work of preliminary investigation with an aim to achieve the following objectives:

- Clarification and understanding of the project request.
- Determination of the size of the project.
- Cost assessment and decision on benefits of alternative approaches.
- Determination of the technical and operational feasibility of alternative approaches.
- Report of the investigation to the management.
- Recommendations on the acceptance or rejection of the proposal.

During the preliminary investigation, a systems analyst gathers data through the following three primary methods:

- Reviewing organization documents.
- Onsite observations.
- Conducting interviews.

Feasibility Study: It is an important outcome of the preliminary investigation which determines whether the system, for which the request has been made, is feasible or not. Feasibility study has three aspects as follows:

NOTES

- ***Technical Feasibility:*** This is related to the technicality of the project. This ponders over the question whether the work for the project can be done with the existing equipment, the current software package and the available human resources. In case, a new technology is required, it is essential to know whether it is possible to develop it further.
- ***Economic Feasibility:*** Every organization wants to develop a system that gives it some benefit in terms of financial return and that too at a reasonable cost. So economic feasibility addresses issues related to the creation of the system to make the costs affordable and acceptable. Another important question is whether the cost of not creating the system is so great that it is advisable to undertake the project.
- ***Operational Feasibility:*** This feature of feasibility study is related to its operational aspect wherein the working of the hardware, the software and the human resource is to be taken into account. So, the question of whether the system will work if it is developed and implemented comes naturally. Another aspect related to human resource is whether there will be resistance from users who may not agree to accept the changes.

5.4.2 Information Gathering

A systems analyst must obtain a clear understanding of an existing system in order to decide how much improvement can be made with the new system in the context of fact finding techniques. Many people understand the specific operations in a system but may not understand the entire system or vice versa. It is the responsibility of the analyst to understand the existing system before making any changes. This will help the analyst to evaluate the performance of the new system compared with the previous system. Figure 5.9 shows what the systems analyst has to decide.

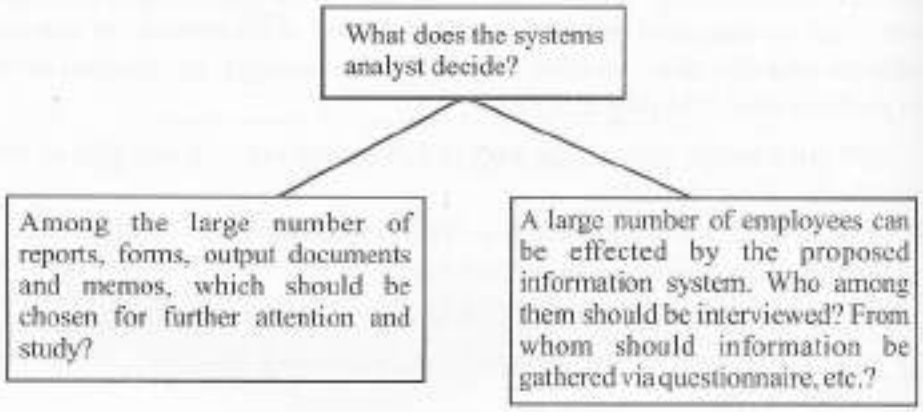


Fig. 5.9 Work Area of Systems Analyst

Various fact finding methods are available that help the systems analysts to collect information about the existing system. For example, the systems analyst can review documents, conduct interviews and questionnaires and perform onsite observations and Joint Application Development (JAD) operations. Systems analysts usually apply more than one of these techniques to ensure an accurate and comprehensive investigation of the existing system. Another important technique that helps in collecting information related to a software system is sampling. Sampling is a technique that helps collect information related to a large software system by investigating smaller and randomly selected items of the software. While examining these selected elements, it is assumed

NOTES

at the analysis will reveal useful information about the existing software system as a ole. Sampling reduces costs, speeds up the data gathering process and improves ectiveness. Figure 5.10 shows information gathering methods.

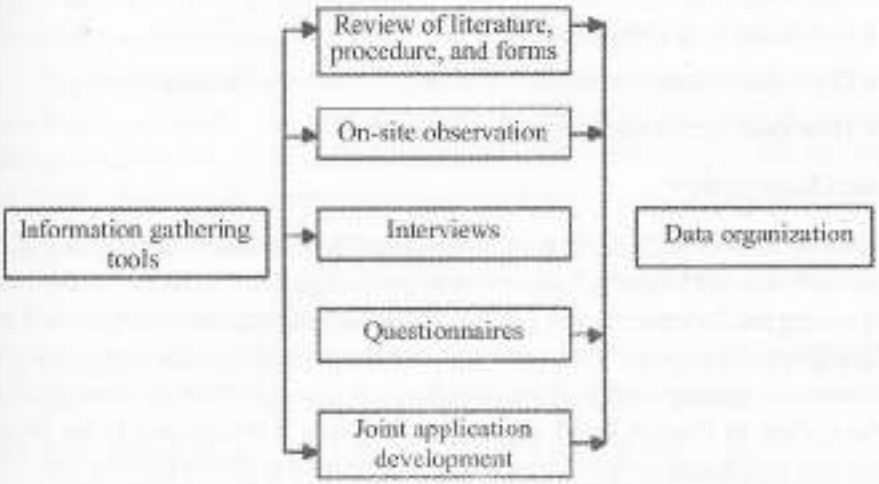


Fig. 5.10 Information Gathering Techniques

While collecting information about the existing system, the systems analyst should etermine how that system came into existence. After conducting the preliminary vestigation, the analyst must be in a position to answer the following questions:

- How was the system designed and installed?
- Who implemented it?
- Will anyone try to justify or defend the existing system and thus slow up the change to a new system?

nformation Gathering Tools

arious kinds of reports and records that store information related to the existing system re maintained in an organization. These reports and records can provide valuable formation regarding the organization, its operations and activities, to the systems analysts. review of available documentation is a logical starting point when seeking insight into system. It helps people gain some knowledge about the organization or operations by emselves. Reference to inputs, outputs and files implemented in any given procedure vill be helpful in documenting current operations.

Procedure manuals and forms are excellent sources of information because they escribe the format and functions of the present system. Update manuals can provide a t of information, which saves a lot of time. Recent developments in the area of Information Technology (IT) have made it possible to avail all the updated documentations about ost of the systems within a short span of time. The existing forms can provide a clear nderstanding about the transactions that are handled in the organization. An input form an be used to identify the various data items captured for processing. An output form or eport can be used to evaluate the performance of the process or to notice the different formation products that are being generated from the system. When the analyst analyses te existing forms, the following question must be pointed out:

- Who are the users of the forms?
- How important are these forms to the users?
- Do the forms contain all the required information?

NOTES

- Is it required to add or delete any item?
- What is the frequency of use of each form?
- How many users receive the forms?
- Is it readable or ambiguous?
- Does the information in the forms help the users to take decisions?
- How does it help the user in decision-making?

Onsite Observation

Observation allows analysts to gain information they cannot obtain by any other fact finding methods. With thorough observation, you can gain first hand information whereas by reviewing the documents, you get only secondary information. This method is really useful when analysts want to observe how transactions are handled, when and where the documents are captured, how the data flows or how activities are carried out, and so on. According to Harper Boyd and Ralph Westfall, 'Observation is the process of recognizing and noting people, objects and occurrences to obtain information.'

The objective of an onsite observation is to get as close as possible to the actual system being studied. As observers, analysts follow certain rules. While making observations, they are more likely to listen than talk with a sympathetic and genuine interest when information is conveyed. Care is taken not to argue with the persons being observed. Observation provides a more tangible perspective of what is described in the documentation. It also brings out the aspects of the documentation that are incomplete or outdated.

Interviews

Your personal ability to talk with other people and listen to them is probably the most important asset you have for getting along in the business world. As a systems analyst, it is one of the most important tools for gathering primary data for a personal interview. It is an excellent way to understand what is happening in a system by interviewing and talking directly with responsible individuals who perform various activities in an organization.

Interviewing is a common method used by systems analysts for collecting data on information requirements. An interview is a directed conversation with a specific purpose that uses a 'Question and Answer' format. Analysts use interviews to collect information from individuals or from groups. The respondents can be internal users as well as external users. That includes Chief Executive Officer (CEOs), Vice Presidents, Chief Technical Officers (CTOs), managers, supervisors, other technical staffs, remote users and mobile users, etc.

A real world situation is sometimes different from the case studies utilized to teach interview methods as the analyst develops an interviewing approach that best fits a situation.

Jerry Fitzgerald and Ardra Fitzgerald describe four ways of interviewing. This is referred to as 'FLIP' which represents the Formal, Legalistic, Informal and Political approaches.

The formal approach is purely businesslike which follows a systematically logical plan with no idle talk or undue friendship. The legalistic approach follows a 'Thumb Rule' approach. The interviewer always goes 'By The Book' or always follows rules

NOTES

The political approach is based on doing favours for people, performing only the jobs that make you look good and acting in a prudent, shrewd and diplomatic way. The informal or 'Play It By Ear' approach follows a plan but the interviewer is able to deviate as required without losing control of the situation. The informal approach may appear to the interviewee that the analyst is unstructured and disorganized.

The best approach is the one in which the interviewer is able to combine the above four approaches. The analyst can be formal when necessary, legalistic when challenged, play politics in necessary situations and be informal in unknown situations. An interview has the following advantages as compared with other techniques:

(i) It is flexible. The analyst can go for an informal interview where the subject/area is not much known to him; where he is does not know what questions to ask and how to organize questions.

(ii) It validates information. The analyst can observe not only what subjects people talk but also how they express it. Information can be verified.

(iii) It can handle complex subjects. An interview is the most effective technique for developing opinions about complex subjects and probing the sentiments underlying expressed opinions.

(iv) Many people enjoy it. Usually people enjoy being interviewed when all they have to do is talk, especially expressing their opinions, if the time permits.

(v) By seeking opinions, it is easier to discover a key problem.

(vi) Facts obtained from hard data explain past performances but goals project the organizational future.

(vii) Expressed feelings help the analyst to understand emotions and attitudes (may be you can get a hint whether the project will succeed or not).

Systems interviewing refers to just talking with people in order to find out how the system is functioning and how they expect their future systems to run. Interviewing must be adopted throughout the system study because it can produce up-to-date information whereas the organization charts and other documents are at least six months behind the times. During interviews, an analyst can observe how the system runs as well as the personality and attitude of the employee which is very important in many situations.

Interviewing is an art. The initial requirements for a successful interview are to develop a cordial atmosphere and put the respondent in a relaxed mood. Then, the interviewer proceeds with asking questions properly, obtaining valuable and reliable responses and recording them accurately and completely. Any approach can be adopted to conduct an interview.

Questionnaires

Interviews are one of the best tools to collect information from a limited number of people. However, sometimes the systems analyst needs to gather information from a large number of people regarding certain issues. In such situations, the most efficient way to obtain information is through questionnaires. A questionnaire asks people to respond to various written questions which are presented to them in a formal document.

The use of standardized questions in a questionnaire can help capture more data that is reliable. By using questionnaires, which is an information gathering technique, the systems analyst can collect valuable information from the people in the organization who may be affected by the current and proposed systems. Various tasks that are performed in the questionnaire technique are as follows:

NOTES

- Acquiring information before conducting the interview.
- Gaining information in order to prove facts found in the interview.
- Acquiring information on how users feel about the current system.
- Checking for problems that remain unsolved.
- Studying the expectations of people from a new or modifed system.

Advantages of Questionnaires

The following are some of the benefits provided by questionnaires:

Well designed and interpreted, questionnaires are very effective in surveying people's interests, attitudes, feelings and beliefs.

Open-ended questions can be used by the analyst to learn more about feelings, opinions and general experiences of the respondent. A closed-end questionnaire controls the frame of reference by presenting respondents with specific responses from which they can choose.

Questionnaires help in the collection of a huge amount of data about what problems they are experiencing with their work and about what they expect from a new or modified system. Developing a useful questionnaire takes quite some time, but in doing so, you can gather lots of information without spending time in face-to-face interviews. For example, a questionnaire would be the right technique if you want to know what percentage of users prefers 'online manuals' for learning about a new software package before using it.

However, in certain situations, such as doing an in-depth analysis of a manager's decision-making process, an interview is a better choice than a questionnaire. The following are the situations in which questionnaires should be used:

- The people who are to be questioned are widely dispersed, for example in different departments or branches of the same organization.
- The project involves a large number of people and you want to know what proportion of a given group approves or disapproves of a particular feature of the proposed system.
- The analyst wants to determine the overall opinion before the systems project is given any specific direction.

Types of Questions

The following are the types of questions:

(i) **Open-Ended:** These are questions that can be easily and correctly interpreted. They are more objective and less leading and mean to encourage a full answwer.

(ii) **Closed-Ended Questions:** These are the questions that are used when the systems analyst is able to effectively list all possible responses to the question. All possible responses of the closed questions should be mutually exclusive.

5.4.3 Structured Analysis

Structured Analysis and Design Technique (SADT) uses a graphical notation and is generally applied in information processing systems. The SADT language is known as the **language of Structured Analysis** (SA). SADT comprises two parts, namely, structured analysis and Design Technique (DT). SA describes the requirements with the help of diagrams, whereas DT specifies how to interpret the results.

NOTES

Structured analysis and the design technique provide a notation and a set of techniques, which facilitate to understand and record the complex requirements clearly and concisely. The top-down approach used in SADT helps to decompose high-level nodes into subordinate diagrams and to differentiate between the input, output, control, and mechanism for each node. In addition, this technique provides actigrams and datagrams and the management techniques to develop and review an SADT model. Note that SADT can be applied to all types of systems and is not confined only to software applications.

Essential model, environmental model, behavioral model and implementation model are considered as the 'elements of structured analysis and design'. Figure 5.11 illustrates the elements of structured analysis.

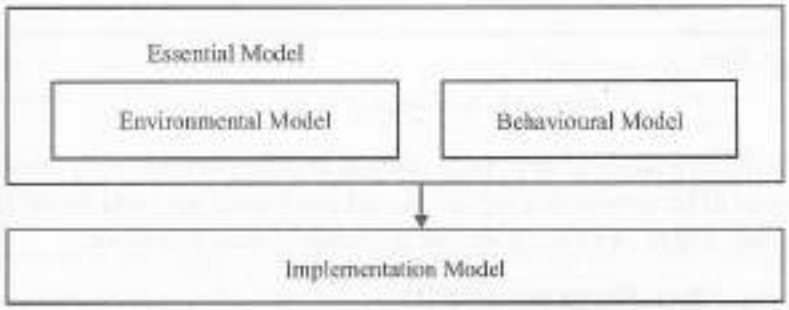


Fig. 5.11 Elements used in Structured Analysis

- **Essential Model:** This model refers to a model of what the system must do but it does not define how the system will accomplish its purpose. It is a combination of the environmental and behavioral model.
- **Environmental Model:** It defines the scope of the proposed system and the boundary and interaction between the system and the outside world. This model is composed of statement of purpose, context diagram and event list.
- **Behavioral Model:** This model of the internal behaviour and data entities of the system which models the functional requirements. It is composed of data dictionary, data flow diagram, entity relationship diagram, process specification, and state transition diagram.
- **Implementation Model:** This model maps the functional requirements to the hardware and software. It minimizes the cost of development and maintenance. It determines which functions should be manual vs. automated. This model can be used to discuss the cost-benefits of functionality with the users/stakeholders. It also defines the human-computer interface. Structure charts are the required tool used in implementation model.

Figure 5.12 illustrates the purpose of system design. A clear and concise textual description is used to define the purpose for the system. It is deliberately vague and hence intended for top-level management, user management and others who are not directly involved in the system, for example credit card system. The purpose of the credit card system is to provide a means for the company to extend credit to the customer. The system will handle details of credit application, credit management, billing, transaction capture, remittance, and management reporting. Information about transactions should be available to the corporate accounting system. The context diagram in system design is to highlight the boundary between the system and the outside world. It also highlights the people, organizations and outside systems to interact with the system under development.

NOTES

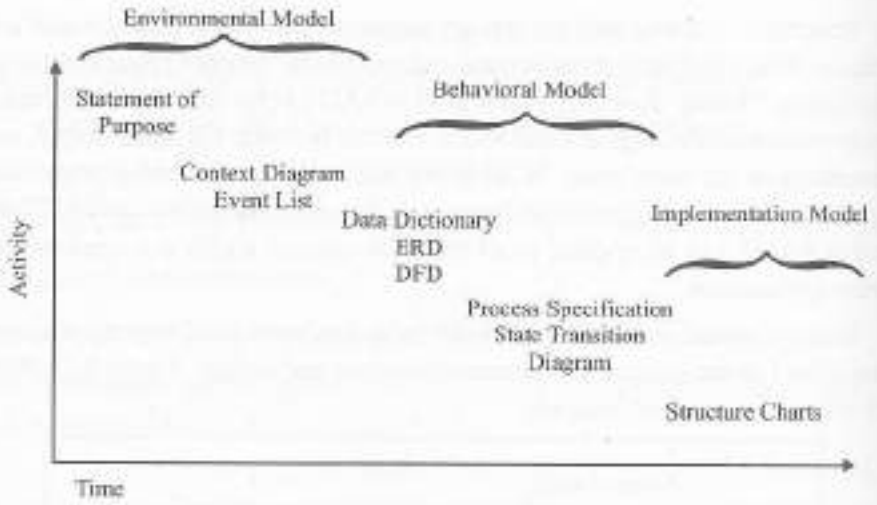


Fig. 5.12 Analysis and Design Process

The characteristics of structured analysis include graphical with supporting text, allow system to be viewed in a top-down and partitioned way and contain minimum redundancies. The reader should be able to predict system behaviour.

5.4.4 Data Flow Diagram (DFD)

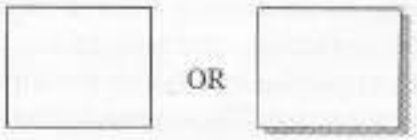
Data Flow Diagrams (DFDs) are used to specify how the data flows between the functions of a system. The basic purpose of DFDs is to show how the system is currently implemented. Larry Constantine first developed DFDs as a way of expressing system requirements in a graphical form. A DFD is the starting point of the design phase that functionally decomposes the requirement specifications down to the lowest level of detail.

Basic Elements of a DFD

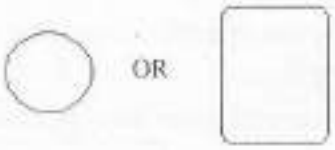
A data flow diagram illustrates the flow of data through a system and the work performed by that system. In data flow diagrams, the symbol set comprises of diagram entity, process, data store and data flow.

An entity is used to define the boundaries of the system. It is an external component of the system, for example a department, a person or a business that interacts with the existing system. It also provides data to the system and receives data from the system. An entity is represented by a rectangle. Table 5.1 shows the basic symbols used in data flow diagrams.

***Table 5.1** Basic Symbols used in Data Flow Diagram*

Symbol Name	Symbol	Symbol Meaning	Comments
Square	 OR	Source or destination of data	May be one customer or a number of customers with customer orders

NOTES

Arrow		Data flow	May be physically contained in a purchase order, invoice, etc.
Circle		Process that transforms that data flow	May be an accountant calculating discounts and preparing invoices
Open Rectangle		Data store	Can be a file, magnetic tape, a database on disk, etc.

A process is defined as a work or action performed by people, machines, etc., within a system. It is used to transform the input of a system into an output and is represented by bubbles.

A data store is used to store data. Data stores are represented by open boxes.

A data flow is any item that carries data to, within or from the system. That is, it is used to represent inputs and outputs of the system. Data flows are represented by arrows.

There are two types of data flows—physical DFD and logical DFD.

(i) **Physical DFD:** A physical DFD is an implementation dependent view of the current system, showing what functions are performed. A physical diagram provides details about hardware, software, files and people involved in the implementation of the system. Physical characteristics include names of the people, names of the departments, names of files, names of hardware, locations, names of procedures, form names, document names, etc.

(ii) **Logical DFD:** Logical DFD is an implementation independent view of a system that focuses only on the flow of data between different processes or activities. Logical diagrams show how the business operates and not how the system can be implemented. They explain the events of the system and the data required by each event of the system.

Physical DFD differs from the logical DFD in the following ways:

- Physical DFD is implementation dependent whereas logical DFD is implementation independent.
- Physical diagrams in the physical DFDs provide low level details, such as hardware and software requirements of a system whereas logical diagrams in the logical DFDs explain only the events involved in the system and the data required to implement each event of the system.

NOTES

Rules to Construct a DFD

The rules that must be followed to construct the DFD of a particular system are as follows:

- Each process involved in a system should be named and numbered for easy reference.
- The name of each process should be symbolic.
- The direction of flow of information in a DFD should be from the top to the bottom and from the left to the right.
- When a process is divided into several lower level processes, each low level process should be numbered.
- The names of the data stores, sources and destination should be in capital letters.

Level 0 DFD

Level 0 DFD provides more detail than the context diagrams by exploding them. By exploding context diagram processes into subprocesses, the systems analyst begins to fill in the details about the flow of data within the system. Each subprocess in the Level 0 DFD is numbered with an integer. Generally, this number starts from the upper left corner and proceeds towards the lower right corner of the Level 0 DFD. A Level 0 DFD, also called a fundamental system model or a context model, represents the entire software element as a single bubble with input and output data indicated by incoming and outgoing arrows respectively. Figure 5.13 shows the Level 0 DFD for customer order and credit verification system.

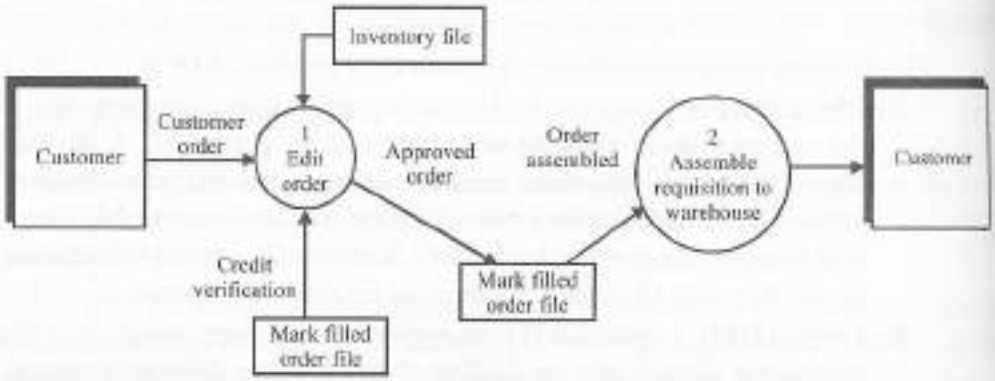


Fig. 5.13 Level 0 DFD for Customer Order and Credit Verification System Child Diagrams

Each process in the Level 0 DFD may, in turn, be exploded to create a more detailed child diagram. The process of Level 0 DFD that is exploded is called the 'parent process' and the resultant diagram is called the 'child diagram.'

The primary rule while creating child diagrams is that the child diagram cannot receive input or produce output that is not produced or received by its parent process. That means any data flowing in or out of the parent process should be shown flowing into or out of the child diagram.

The processes in the child diagram are numbered by using the parent process number, a decimal point and a unique number assigned for each child process of the

child diagram. Figure 5.14 shows a first Level DFD, elaborating customer order processing and shipment system.

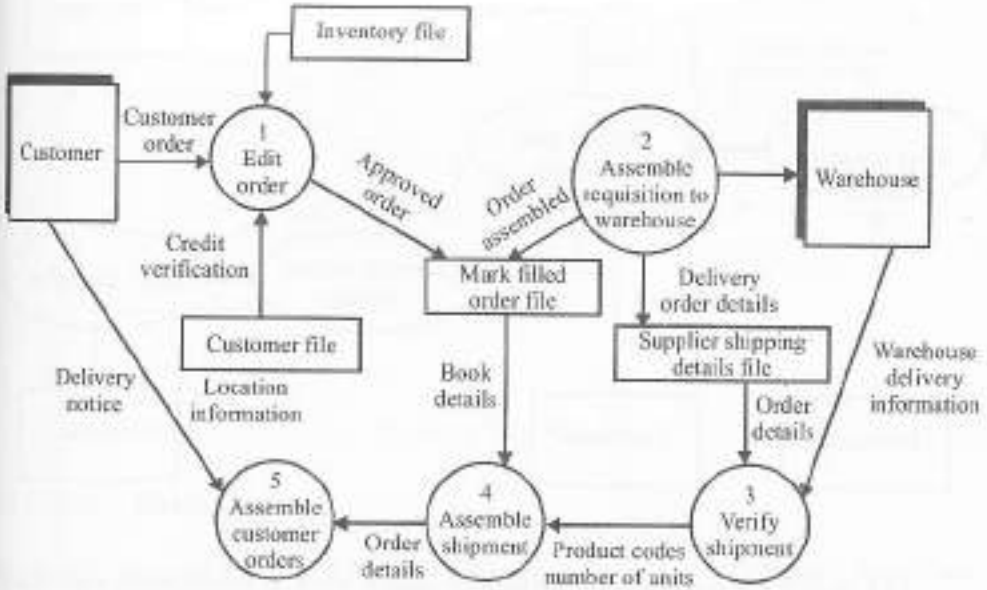


Fig. 5.14 First Level DFD, Elaborating Customer Order Processing and Shipment System

Developing a DFD Model of the System

Data flow modelling is a process of identifying the flow of information in a software system. Source information is provided as an input to the software system. A software system is a process that transforms the applied input into the desired output. In the software system, the input may pass through various states where each state performs a specific function on the applied input and produces a corresponding output. Each state in the software system is known as a processing step and the function performed by each state is known as transformation. For various processing steps in a software system, the transformation may execute in sequence or in parallel.

Data flow modelling uses a DFD for modelling the flow of information through a software system. In a DFD, circles represent the processing steps and an arrow line represents the flow of data between any two processing steps and rectangles represent the source and sink for information. The source for information provides the source input for the software system and the sink for information consumes the output from the software system.

Let us consider an example of a simple invoice system in an organization. The organization issues invoices to the customers for each purchase. At the end of the month, the organization rechecks the invoices and issues a receipt to the corresponding customers for the invoices that have been paid. For the invoices that have not been paid, the organization issues reminders to the respective customers. In this example of the data flow modelling, first you need to identify the various states, flow of information through these states and the function of each state. Now, a DFD is created for the invoice system.

NOTES

NOTES

Figure 5.15 shows the data flow diagram of an invoice system.

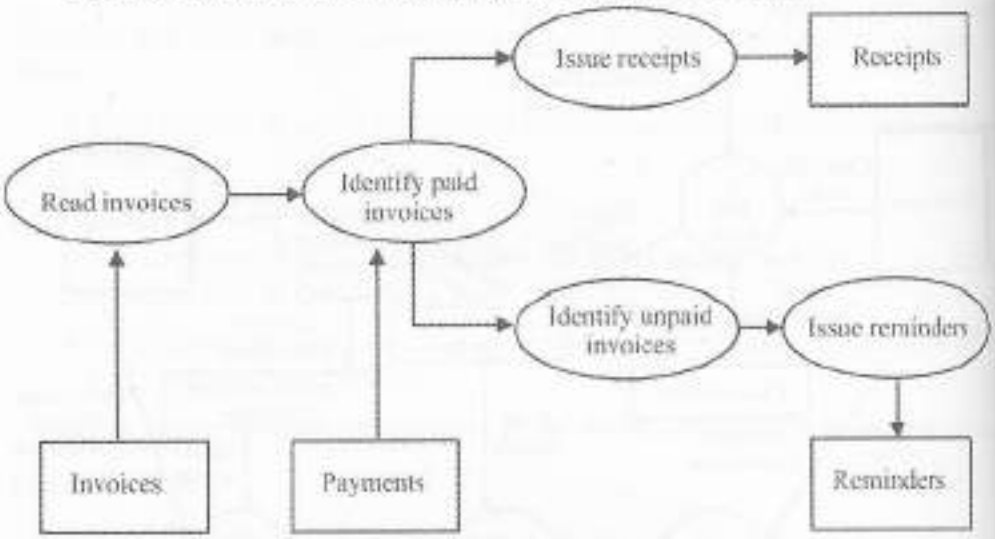


Fig. 5.15 Displaying a DFD of an Invoice System

In Figure 5.15, invoice and payment rectangle work as information source and receipt and reminder rectangle work as information sinks. The invoice information source provides information on all the invoices while the payments information source provides information only about the paid invoices.

The read invoices circle specifies the first state or processing step where invoices are read from the invoices information source. Then, the read invoices circle passes all the invoices to identify paid invoices circle. Identify paid invoices circle receives information from the payments information source and identifies the invoices that have been paid. After identifying the paid invoices, the 'identify paid invoices' circle passes the paid invoices to the issue receipts circle which issues the receipts for the paid invoices. In parallel to the issue receipts circle, the find due payment circle identifies the invoices that are not paid and passes these invoices to the issue reminders circle. The issue reminders circle issues the reminders for the unpaid invoices.

Using DFD in a Real Time System

The design process of the DFD for a real time system is complex as compared to the normal system. Real time systems are based on strict execution time constraints which make the designing of DFD more complex. You have to consider things, such as control flow and control processes for using the DFD for a real time system. Control flow is a form of pipeline, which is used to carry a binary signal, such as ON and OFF. Control processes are those processes which help coordinate and synchronize the execution of other processes in a real time system. Control processes use control flow as an input to obtain the information of completion of process execution and generate new control flows to initiate the execution of other processes. Control flows and control processes are represented by dashed lines in the DFD of a real time system as shown in Figure 5.16.

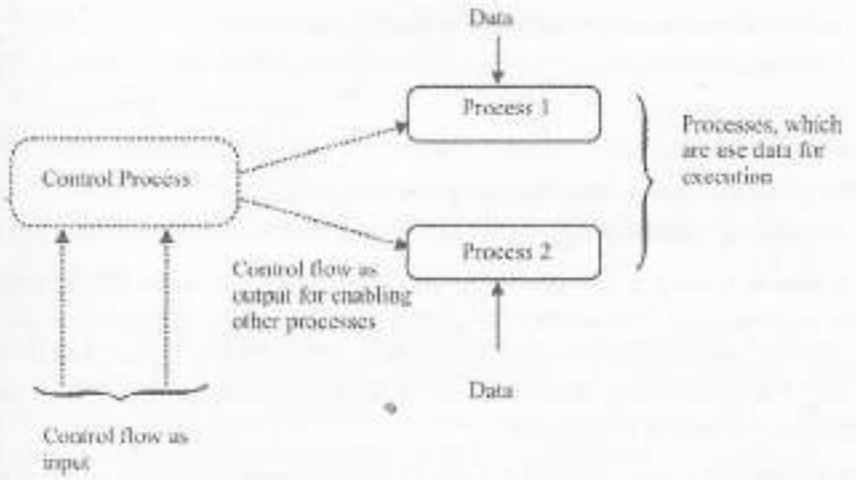


Fig. 5.16 DFD of a Real Time System

NOTES

5.4.5 Data Dictionary

It is one of the most important reference works that contain information about data which is also known as metadata. It is compiled by systems analysts and guides developers during analysis and design. It is created to assign specific meaning to every term used in the design, leaving no chance of ambiguity on the interpretation of the meaning of the term used. Contents of the data dictionary have descriptions and specifications on system data but it does not contain the data itself.

Data dictionary offers the following important functionalities:

- It provides basic documentation.
- It eliminates redundancy.
- It acts as starting point in the development of reports and screens.
- It validates DFD for accuracy and completeness.
- It aids the development of the logic for processes contained in DFD.

Using data dictionary, duplication of efforts is avoided and better communication is established between various departments of an organization that share a common database. This also makes maintenance easier and utilizes it as a standard for consistent data elements.

A data dictionary consists of the following elements:

- Name of the data item that the program handles.
- Alias, if any, for data item(s).
- A description for every data element in English.
- List of other data elements that are related to some specific data items.
- Acceptable values for every data item.

Naming is an important part in the creation of data dictionary. The name should be meaningful and distinct. There should be no duplication of the name. To derive maximum benefit from the data dictionary, it has to be tied to other programs within the system. This gives an advantage. When an item gets deleted/updated from data dictionary, it gets deleted/updated automatically from the database. It is strongly recommended to let the computer maintain dictionary by using cross referencing, whenever, there are changes made in the system.

NOTES

Data dictionaries can perform the following functions:

- Create reports.
- Create screens.
- Create forms.
- Generate source code for computer program.
- Analyse system design for its completion and detection of flaws in design.

A data dictionary is also known as database dictionary. It is a file defining basic database organization. This dictionary contains a list of every file with the number of records per file and information on each data field for names and types. Data dictionary, being one of the most vital data structure, is kept hidden from users for preventing accidental destruction of its contents.

An element in data dictionary is used in other important charts and documents. Data elements defined in a data dictionary are used in an Entity-Relationship (E-R) diagram.

Notations used in Data Dictionary

A data dictionary uses many symbolic notations carrying some meaning that are defined as:

'=' is composed of.

'+' and.

'()' element is optional.

'{ }' ' interation.

'[ ]' select one of a list of elements.

'|' separates choices of elements.

'**' comments.

'@' identifier for a store (unique id).

Creation of Data Dictionary

Creation of data dictionary is done with a view to define contents of each data store, data flow and every process that avoids any ambiguity on that data item or element.

Data elements are basic entries in the data dictionary that are also the simplest. For example, states (Delhi, Rajasthan, Haryana, West Bengal, Bihar, Jharkhand) or population for individual states are data elements and these are not further simplified for that data dictionary. A number of related items appearing in a data store or data flow are known as data structure. It may be simple like an address or it may be complex like an order form. Address may be composed of several data elements, such as address1, address2, location, urban/rural, state, postal code. Address1 may be for house number or flat number address2 may be street number, and so on. An order form may contain 10 structures at the minimum and it may have 40 to 60 elements in the dictionary. These structures that form collection of many elements are organized alphabetically in data dictionary. In this dictionary, every name has only one entry. For every element of data, data store or data flow documented information is similar.

Inclusion of the following is suggested:

- Item's name.
- Purpose (with description).

- Type of data (numerical, alphabetical, alphanumeric, date, time, etc.).
- Length of data (number of numeral digits, characters, decimal points, etc.).
- Mandatory characters.
- Display (whether rounded to integral number or up to two decimal places).
- Statements or examples on appearance of each data element within data modeling diagrams (data stores, processes, data flows).

NOTES

5.4.6 Decision Trees

Decision tree is another method for defining complex relationships. A decision tree has as many branches as logical alternatives within software. It depicts the logical structure based on a stated policy. Decision trees are used when combinations of conditions need to be present. When policies change, it is easy to change trees.

A decision tree approach represents a method for depicting alternative actions and conditions within a horizontal tree framework. These are used when complex branching occurs in a structured decision process. Trees are useful while keeping decisions in a particular sequence. It is important to distinguish between conditions and actions while drawing decision trees. A square node indicates an action and a circle indicates a condition. Figure 5.17 illustrates a decision tree structure.

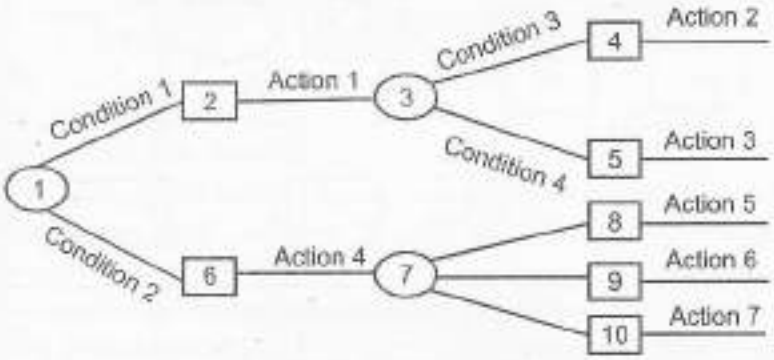


Fig. 5.17 Structure of a Decision Tree

Decision tree is a powerful tool which is popular as it is used for classification and prediction. It shows some rules for using a decision tree. It can also be directly used in query language, such as SQL for accessing and retrieving database and records that belong to a particular category.

Different tree diagrams are used to determine the optimum course of action, out of several possible alternatives with uncertain outcomes. Resulting chart or diagram looks like a cluster of tree branches and displays the structure of a particular decision. It also shows the interrelationships and interplay between various alternatives, decisions and possible outcomes.

In decision analysis, a decision tree is used to represent decisions and decision-making, visually and explicitly. When it is used for data mining, it describes only data and not decisions. Such a classification tree works as an input for decision-making.

For classifying an example, in a decision tree, start at the root of the tree and move through it until a leaf node is reached. This provides the classification (see Figure 5.18). By applying an inductive approach, decision tree is used to learn knowledge on classification. Decision tree is used for data mining to form a big database. Key requirements for mining by using decision trees are as follows:

NOTES

- **Attribute-Value Description:** The object is described as fixed collection of attributes. This requires conversion of continuous attributes in a discrete form or already provided in the algorithm.
- **Predefined Classes:** Categories must be established before assigning examples.
- **Discrete Classes:** Cases may or may not fall under a particular class. Number of cases have to be more than the number of classes.
- **Data Sufficiency:** Many cases are to be taken for this. Normally, hundreds of cases should be taken but the number may even be thousands.

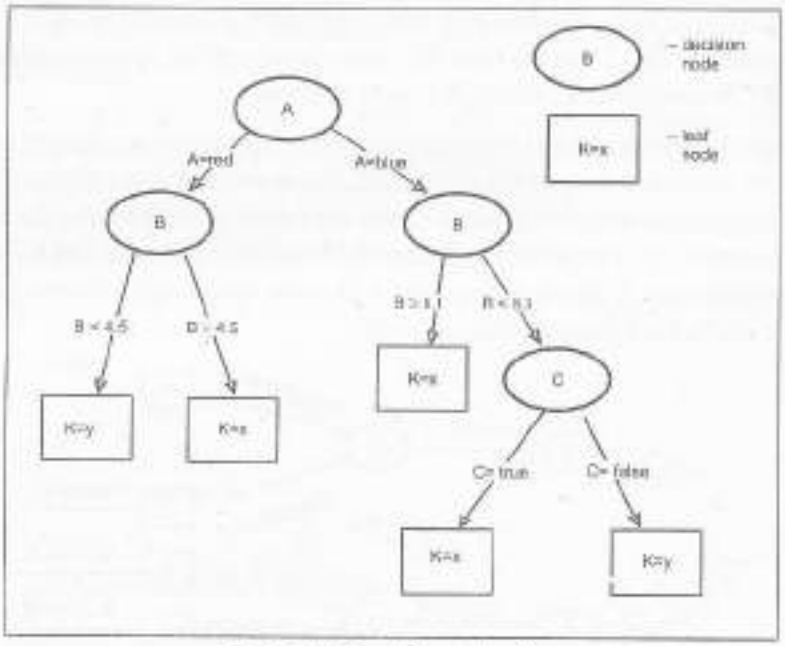


Fig. 5.18 Simple Decision Tree

Decision tree is significantly used in data mining. The main aim lies in the creation of a model that is capable of predicting the value of a target variable in response to the inputs. In a decision tree, an interior node represents inputs. Edges, extending to children nodes stand for values of input variable that may be possible. A leaf contains a value of the target variable. The path from the root to leaf represent values of the input variables.

A decision tree has three types of nodes:

1. **Decision Nodes:** Such nodes fall in-between terminal nodes (or leaf nodes) and root. These specify some kind of test that is to be carried out on a single attribute value, having one branch and a sub-tree for every possible outcome of the test. In a diagram, it is denoted by square.
2. **Chance Nodes:** These are also intermediate nodes and are represented by circles.
3. **End Nodes:** These nodes are also known as leaf nodes and indicate the value of the target attribute (class). These are represented by triangles in a diagram.

Decision trees are drawn from left to right. Such a tree shows splitting paths and no converging paths. Such trees can grow very big and become difficult to draw by hand.

NOTES

Practical Example of a Decision Tree

The manager of a golf club was facing problems with the attendance of his customers. On some days, the turnout of customers was too high that required overworking of the staff. However, on other days, the turnout was too low and the staff got too much free time. The manager wanted to optimize the staff time and for this he thought of predicting the days on which people would play. For this, he needed to understand the cause behind the decision of playing golf. He found that weather was an important factor and used the weather forecast for the upcoming week. He made a record of the following:

- The weather condition (sunny, overcast or raining)
- Ambient temperature
- Percentage humidity
- Windy or not
- Attendance of customer on that day

The manager then compiled the collected data in a tabular form as shown below.

Observations of the Weather

Independent Variable				Dependent Variable
Outlook	Temperature	Humidity	Windy	Play
Sunny day	82	88	F	No
Sunny day	81	90	T	No
Overcast	84	75	F	Yes
Raining	72	95	F	yes
Raining	70	82	F	Yes
Raining	66	78	T	No
Overcast	62	66	T	Yes
Sunny day	70	92	F	No
Sunny day	70	65	F	Yes
Raining	72	82	F	Yes
Sunny day	68	69	F	Yes
Overcast	70	88	T	Yes
Overcast	80	76	F	Yes
Raining	72	84	T	No

The manager applied a decision tree model and solved the problem as follows.

Using three categories of the variable outlook, three different groups were found:

- One that plays golf on sunny days.
- One that plays when it is overcast.
- One that plays during rain.

The manager's first conclusion was that people always play golf when the outlook is overcast. There are some who play golf even if it is raining. Then, he divided players playing on sunny days in two groups. He also found that people dislike playing if the percentage of humidity is more than seventy five.

Finally, he divided the category of people who play while it was raining into two and noted that people will also not play if the weather is windy.

Finally, a simple solution of the problem was worked out by the classification tree as shown in Figure 5.19.

NOTES

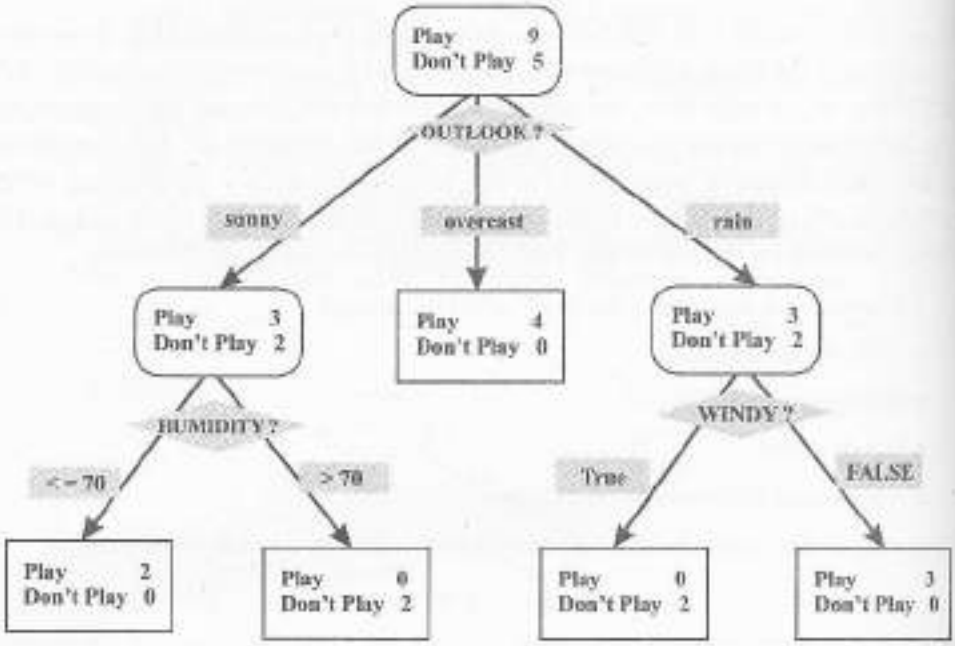


Fig. 5.19 Decision Tree for Playing Golf

The manager dismissed many and kept only few of the staff on days when the weather was sunny, humid, windy or rainy. He hired extra staff when more people would play golf. The decision tree had helped the manager to convert data of complex nature into an explanation for decision-making.

Thus, decision tree is a decision support tool, making use of a tree like structure or graph. It works as a decision model that shows possible consequences, chance event outcomes, resource costs and utility. Decision trees find use in operations research, as a support in decision analysis which helps in identifying a strategy most suitable to realize a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities

5.4.7 Decision Tables

Decision table is an excellent tool for expressing complex logical relationships in a highly understandable and precise manner. This table is a matrix containing rows and columns that are used to define relationships. A decision table consists of two parts, in which, one is STUB and the other is ENTRY. The STUB is further divided into the upper and lower quadrant. The upper quadrant is known as condition stub and the lower quadrant is known as action stub.

The ENTRY part is also subdivided into upper quadrant and lower quadrant. The upper quadrant is known as condition entry and the lower quadrant is known as action entry. In a decision tree:

- Y represents the existence of a condition.
- N represents the condition which is evaluated and is not satisfied.
- A blank space represents the condition involved that has been tested.

There are three major components in a decision table. They are conditions, actions and decision rules. Conditions are events or facts that determine the course of action to

ietaken. Actions are the processes that get activated under certain conditions. Decision ules express the relationships between combinations of conditions and courses of action.

Decision tables are useful in situations where the resulting actions depend on the ccurrence of one or several combinations of independent conditions.

Four quadrants of a decision table are summarized in Table 5.2.

Table 5.2 Quadrants of a Decision Table

Conditions or causes	Condition rules (alternatives) showing combinations of causes
Actions or effects	Action entries combination of effects

Conditions are put in quadrants at the upper left corner of the table and condition ules (alternatives) are placed in the upper right corner. The corner at quadrant on the ower left has actions needed and at the lower right there are rules for action. Conditions re also known as causes and actions are also known as effects.

- **Conditions *(Upper Left Quadrant)*:** Contains a list of all the possible conditions pertaining to a problem.
- **Condition Alternatives *(Upper Right Quadrant)*:** Contains the condition rules (set of possible values for each condition) of alternatives.
- **Actions *(Lower Left Quadrant)*:** Contains actions, which can be a procedure or operation to be performed.
- **Action Entries *(Lower Right Quadrant)*:** Contains the action rules, which specify the actions to be performed on the basis of the set of condition alternatives corresponding to that action entry.

Note that each column in the quadrant can be interpreted as a *processing rule*. The basic four-quadrant structure is common for all the decision tables; however, they may differ in the way of representation of the condition alternatives and action entries. For example, in some decision tables, condition alternatives are represented using true/ false values while in other tables, Yes/No, numbers (0 or 1), fuzzy logic or probabilistic representations may be used. Similarly, in decision tables, action entries are represented by placing check marks, cross marks, or numbering the actions to be performed.

Developing a Decision Table

In order to build decision tables, first the maximum size of the table is determined. Any impossible situations, inconsistencies or redundancies are then eliminated from the table. To develop a decision table, follow these steps:

1. Determine all the actions that can be associated with a specific module or procedure.
2. Determine all the conditions affecting the decision.
3. Associate the specified conditions with specific actions. In the simplest case, each condition in the decision table will be associated with two alternatives (Yes or No).
4. Define rules by indicating actions that occur for the specified conditions.

Let us consider an example to understand the concept of decision table by designing a program for a telephone company that sends monthly bills to its customers. If the

NOTES

customers pay their bills within two weeks, they are offered a discount as per the scheme given below.

- If amount > 350, then discount = 5%.
- If amount >= 200 and amount <= 350, then discount = 4%.
- If amount < 200, discount = none.

The decision table for this example is shown below:

Conditions	Rules			
	1	2	3	4
Amount paid within 2 weeks	Y	Y	Y	N
Amount over Rs 350	Y	N	N	-
200 < = Amount < = 350	N	Y	N	-
Amount < 200	N	N	Y	-
Actions				
5% discount	X			
4% discount		X		
No discount			X	X

5.4.8 Structured English

Structured English is used to analyse decision process. It has a restricted vocabulary and uses action-oriented verbs like add, multiply, and so on. It does not use a subject and makes use of the terms defined in data dictionary. It is a tool that enables one to state the rules precisely. It uses logical constructs and imperative sentences that are instructive. It is a special language that uses procedural logic derived from structured programming. Keywords of many programming languages are derived from such strong verbs in assembly language or other languages, sometimes using mnemonics. Words like do, add, sum, etc., are used. Shortened form, such as EQ for EQual is used in FORmula TRANslator (FORTRAN). Query languages make use of words of structured English, such as SELECT, JOIN, WHERE along with other combination of special characters, such as =, >, < etc.

Words, such as IF, THEN, ELSE, UNTIL, REPEAT and SO are used for decision-making. Structures show logical hierarchy and are most useful if it is required to take into account loops and sequences, in a program. It is also good for use in procedure manuals.

When the decision structure is not complex, structured English can be used. By using this tool, structured instructions are grouped and nested in an organized manner. To write structured English, the following should be kept in mind:

- Use sequential structures, iterations or decision structures while expressing logic.
- Make use of keywords that exist already. Such keywords are IF, THEN, ELSE, DO, DO WHILE, etc.
- Statement blocks should be properly indented to reflect hierachical structure or nesting.

- If phrases or words that you use are already defined in the data dictionary, underline them. This shows that they have special meaning.
- Before starting the coding stage, clarify logical statements.

Two basic building blocks are logic or instructions that are organized as grouped procedure and are nested and strong verbs that are used for actions. These verbs are move, add, multiply, etc. Structured English is used for labelling decision trees and creating decision tables. Four major considerations for using structured English in decision trees is follows:

- Identifying conditions.
- Identifying outcomes for every decision.
- Identifying actions.
- Identifying rules.

In decision tables, structured English words are used as they are needed when there are complex combinations, actions, rules and conditions on which actions are performed. Here, structured English provides a method that can help in avoiding contradictions, ambiguity due to duplication and impossible situations. Decision trees are useful when every condition is not relevant for every action.

Structured English is used in three basic constructs that are used in every programming language. These are as follows:

- Sequence
- Selection (If-else conditional)
- Iteration (Loops)

Program Design Language (PDL), also known as **structured English** or **pseudo-code** is a language that is used to specify the system design using natural language (English) along with the constructs of a programming language. **IEEE** defines program design language as *'a specification language with special constructs and sometimes verification protocols, used to develop, analyse, and document a program design.'* PDL can be used by the designers and programmers as a means to assure that the actual programming is expected to match with the design specifications.

This language is mainly used to communicate the system design and can be extended to describe the logic design. The benefit of using PDL as a design notation is that it requires the programmer to focus only on the algorithms without taking care of the syntax of the programming language. Other benefits of using PDL are listed below.

- It uses English-like statements that precisely describe specific operations.
- It supports the idea of iterative refinement.
- Program design is expressed in an easy-to-read format. Thus, the design can be converted easily into executable code.
- Reviews become easier since the source code is not examined.
- The language uses flowchart replacements, program documentation, and technical communication at all levels.
- Continued use and refinement of the program design language has established it as the medium of choice for either creating or refining a detailed program design.

NOTES

NOTES

Developing Program Design Language (PDL)

PDL uses the keywords provided for all structured constructs of a programming language and the syntax of natural language to describe processing details. Some keywords to denote various programming processes are listed below.

- Input: INPUT, GET, OBTAIN, PROMPT, and READ
- Output: DISPLAY, SHOW, and PRINT
- Compute: CALCULATE, DETERMINE, and COMPUTE
- Initialize: INIT and SET
- Add One: INCREMENT
- Subtract One: DECREMENT

Let us consider an example to understand the concept of program design language by designing a program to calculate interest of an account based on the specified rate. To develop a PDL for this program, consider the points listed below.

Keywords	*Description*
Known Values	Rate
Inputs	Principle, Time
Calculations	Interest = (Principle × Rate × Time) / 100
Outputs	Interest

Using the keywords and their description listed, the PDL for the said program is listed below.

```
PROMPT for Principle
READ Principle
PROMPT for Time
READ Time
CALCULATE Interest = (Principle x Rate x Time) / 100
DISPLAY Interest
```

Check Your Progress

1. How a system is defined and determined?
2. What is a closed system?
3. What do you understand by the term SDLC?
4. Name the commonly used abstraction mechanism in software design.
5. Define the term system analysis.
6. Name the commonly used method for collecting data on information requirements.
7. What is a decision table?

5.5 PROCESS AND STAGES OF SYSTEM DESIGN

The entire process of designing the software system is an exercise that specifies the 'what' and the 'how' of the system. This is not done in one step but is done in an iterative process that is based on 'what' the system will do as shown in the feasibility report. The systems design also tells how the system will do a particular job. This is given in the form of various programming constructs that designers specify. In any system there are three basic things: input, process and output. A customer who installs a system would want it to accept efficiently and reliably process the inputs provided by him so as to produce corresponding outputs. Programs or routines that are stored in various files (linked together) do the processing.

In general, there are five components of a system design process. These are as follows:

- **Output Design:** Systems requirement tells what outputs are required. Before asking for a system, a customer wants certain outputs. This is the starting point of the design process. The output design requires proper knowledge of the systems requirements.
- **Input Design:** This is the second most important part of a system design. After deciding on the output requirements and finalizing the same, data that is required as an input to produce the desired output has to be identified. The basic documents containing the details of input data which is a part of system requirement specification is used to serve as a guideline. It may become necessary to revise these documents and create new documents for this purpose. As far as possible, requirements should be finalized in Software Requirement Specification (SRS). Introducing fresh requirement at the design stage creates difficulties for systems designers.
- **File Design:** When a system accepts input data, it has to be stored for processing purposes, either for a short or a long period. Also, this data is stored in a structured form generally kept in the form of files in a logical manner. Once data is stored, it is retrieved by the computer program either to process it or to display it. Hence, systems designers devise techniques of storage and retrieval of data from these files.
- **Procedure Design:** This is the part that processes the data fed to the system and produces output. Procedures are written as programs which specify the manner in which processing will be carried out. Designers have to focus on various programming constructs to write efficient procedures or routines to process data. The two kinds of procedures adopted for this purpose are computer procedure and non-computer procedure. Computer procedure specifies functions that are to be carried out on computer. These procedures may have a different sequence in different programs that are run on the computer. These procedures are mostly application oriented. Non-computer procedures specify basic activities performed by the system to support computer procedure in feeding input data, receiving outputs, etc.
- **Control Design:** The control system is designed to properly sequence procedures or micro codes. It ties up procedures with data so that the design indicates the necessary procedures to ensure correct processing along with accuracy and timeliness of data. Then, seeing the results you feel assured that the system is functioning as per plan.

NOTES

5.5.1 Input/Output and File Design

Input design is an important phase in system design phase. Businesses use the new technology to speed up the input process, reduce costs and capture data in new forms such as digital signature. Digital signature follows authentication mechanism. A code is attached with messages in the process of digital signature. Primarily, the signature is generated by hashing the message and then later this message is encrypted with the sender's private key. Digital signature is based on public key encryption. A signature confirms that integrity and source of message is correct. National Institute of Standards and Technology (NIST) standard recognized the Decision Support System (DSS) standard that basically uses the Secure Hash Algorithm (SHA). Message authentication protects digital signature because in that mechanism messages are exchanged from the third party. Digital signature is analogous to manual signature. The main objective of input

NOTES

design is to ensure the quality, accuracy and timeliness of data. The main objectives of input design are to select a suitable input and data entry method, reduce input volume, design attractive data entry screens, use validations checks to reduce input errors, design required source documents as well as forms and develop effective input controls. Systems analysts apply business process engineering techniques when studying transactions and business operations to determine how and when data should enter the system. Data entry is the process of manually entering data into the information system usually in the form of keystrokes and mouse clicks. Data input is the operation of encoding the data and writing them to the database. From a user's point of view, the interface is the most critical part of the system design because it is considered as a checkpoint through which users's can interact with the system. For this, it is essential to construct models and prototypes for user approval. You can prepare initial input screen design to users in the form of storyboard which represents a rough layout showing the sample screen for proposed system.

Input data is determined while the time of input designs. It ensures the quality, accuracy and timeliness of input data. It is determined how data will be captured and entered into the system during input design. Data capture uses an automated or manually operated device to identify source data and convert it into computer readable form, for example credit card scanners, bar code readers, etc.

Key Tasks in Input Design

The following are the essential tasks involved in the input design:

- Designing data entry and input procedures.
- Designing source documents for data capture or devising other data capture methods.
- Designing input data records.
- Designing data entry screens.
- Designing screens for user interface.
- Designing security measures for the system and audit trails.

Output Design

A smart systems analyst has to choose among existing alternatives before implementing an output format in which the output may be desired. A suitable output format should be selected depending upon the requirement of the output. Several formats are available from which selection can be made to communicate information. Narrative format uses standard text, numbers and pictures. It is a format that has been exploited by word processing technology for reports, personalized form letters as well as business letters.

Tabular output is the format that uses columns of text and numbers to present information. This method is most common for presenting computer outputs. Zoned output places the text and numbers in the selected areas of a form or a screen. Graphic output uses graphs or charts to represent information. This format enables end users to better understand the relationships between data and trends. Pie chart, bar chart, column chart, line diagram and scatter diagrams are the most commonly used graphs.

Pie charts depict the relationship between parts to the whole at some specific period. Certain pie chart styles are used to emphasize a particular item. A pie chart should be used to compare seven or fewer portions.

NOTES

Bar chart is an output format that represents information using bars. Categories or classes for comparison are organized in a vertical manner whereas the values are arranged in a horizontal manner. A stacked bar chart portrays the relationship of individual items to its 'whole.'

Column chart is a simple variation of the bar chart. Column charts enable the user to depict comparisons among items and show the variation in an item over a period of time. This chart organizes categories horizontally and values vertically. Such appearance emphasizes variations over a period of time.

Line charts show trends of an item over a period of time. It organizes items on the horizontal axis and measurements on the vertical axis.

Those formats that plot data values of two different items that reflect uneven intervals of data are called scatter charts. Standard statistical methods are applied to establish the degree of existing correlation.

File Design

The file design includes the database design with reference to SDLC because in big organization the databases are required to maintain the various types of files. A file is a collection of data that can be read from or written to. A file can be a program. You can create a file to write the data and store in suitable device. Commands, printers, terminals and application programs are all stored in files which allow users to access diverse elements of the system in a uniform way and give the operating system great flexibility. No format is implied when a file is created. Files are used for all Input/Output (I/O) of information in the operating system. Input occurs when the content of a file is modified or read from or written to. Output occurs when the content of one file is read or transferred to another file. The file is to be considered for a computer as one dimensional sequence of bytes. The content of file can be of different types, such as integer values, texts, image pixels, audio/video content, multimedia effect, etc. The file represents the following interfaces:

- **Tree of Directories and Files:** This interface contains file and directory names which are named with path names.
- **Object:** It deals files and directories along with other kinds of objects, for example devices.
- **Path Name:** It contains a component name for each directory in the path.

File is a passive container of bytes on disk. The open file option is an active source or sink of bytes in a running program. These files are connected to a device. The file Open and file Read are the two frequently used options in computer applications. The file name extension represents the set of characters which help to make the operating system know what program is used to open it and what kind of information is kept into the file. The extension name appears at the end of the file name following a period and having three characters. Sometimes, you might occasionally encounter an error when copying a file with a very long file name to a location that has a longer path than its current location.

You must know the basic concept of file and other related terminologies. A number of main memory blocks are associated with message tags sent from client side. The tags are required for addressable memory. The file design technique considers basically the concept of caching, system calls and hierarchical file naming systems, objects and operations, and required connection in protocols mechanisms.

5.5.2 Logical and Physical Design

NOTES

Logical Design

The logical design of a system refers to an abstract representation of the data flows, inputs and outputs of the system. This is often conducted via modelling, using an over-abstract model of the actual system. In the context of systems design are included. Logical design includes E-R Diagrams or Entity Relationship Diagrams.

Physical Design

The physical design relates to the actual input and output processes of the system. It provides information about how data is input into a system, how it is verified/authenticated, how it is processed, and how it is displayed as in Physical design.

Table 5.3 summarizes the difference between logical and physical design.

Table 5.3 Difference between Logical Design and Physical Design

Logical Design	Physical Design
In logical design, decisions are made about how the interaction and integration will take place between these elements.	In physical design, decisions are made to implement the logical design successfully.
At the time of designing, all the logical errors can be checked while integrating the various system components.	The technology used to implement the design is decided in physical design phase. This approach supports the constraints found during the implementation.
During logical design, we identify the important objects that need to be represented in the database and the relationships between these objects.	During physical design, we decide how the logical design is to be physically implemented (as tables) in the target DBMS. Therefore, it is necessary to have selected the target DBMS before proceeding to physical design.
Logical design is the process of constructing a model of the data used in a company based on a specific data model but independent of a particular database and other physical considerations, such as files and documents.	Physical design is the process of producing a description of the implementation of the database on secondary storage.
It describes the base tables, file organizations, and indexes used to achieve efficient access to the data, and any associated integrity constraints and security restrictions.	It does not describe indexes used to achieve efficient access to the data.
The logical design phase of the methodology is divided into following main steps: • In Step 1, we create a data model and check that the data model has minimal redundancy and is capable of supporting user transactions. The output of this step is the creation of a logical data model which is a complete and accurate representation of the company or part of the company that is to be supported by the database. • In Step 2, the E-R model is mapped to a set of tables. The structure of each table is checked using normalization. Normalization is an effective means of ensuring that the tables are structurally consistent and logical with minimal redundancy. The tables are also checked to ensure that they are capable of supporting the required transactions. The required integrity constraints on the database are also defined.	Physical design is divided into six main steps: • Step 1 involves the design of the base tables and integrity constraints using the available functionality of the target data. • Step 2 involves choosing the file organizations and indexes. Typically, a number of alternative file organizations for data are provided which tend to have a fixed storage structure. • Step 3 involves the design of the user views originally identified in the requirements analysis and collection stage of the system development life cycle. • Step 4 involves designing the security measures to protect the data from unauthorized access. • Step 5 considers relaxing the normalization constraints imposed on the tables to improve the overall performance of the system. This is a step that you should undertake only if necessary because of the inherent problems involved in introducing redundancy while still maintaining consistency. • Step 6 is an ongoing process of monitoring and tuning the operational system to identify and resolve any performance problems resulting from the design and to implement new or changing requirements.

Logical design is concerned with the 'what'. Logical design is independent of implementation details, such as application programs, programming languages or any other physical considerations. The output of this process is a logical data model that includes a set of relational tables together with supporting documentation, such as a data dictionary. These represent the sources of information for the physical design process which provides you a mechanism for making trade-offs.	Physical design is concerned with the 'how'. However, physical design is not an isolated activity because there is often feedback between physical, logical and application design. For example, decisions taken during physical design to improve performance, such as merging tables together might affect the logical data model.
In the logical database design phase, the logical representation of the database system is built which includes identification of the important entities and relationships and then translate this representation to a set of tables.	In the physical database design phase we decide how the logical design is to be physically implemented in the target relational DBMS. This phase allows the designer to make decisions on how the database is to be implemented.

5.3 File Organization

The modular design of the software consists of a number of files. The file creation follows a predefined file structure. Figure 5.20 shows the various elements in hierarchy of a file structure.

Fig. 5.20 Hierarchical File Structure

Figure 5.20 shows the following elements of file structure:

- **Byte:** It is a set of eight bits that represent a character. It is the smallest addressable unit in a computer system.
- **Data Item:** Combination of more than one byte is called a data item or element which defines attributes of object. The data item is also called field.
- **Record:** The data items related to the object are combined into a record.
- **File:** It consists of a collection of related records. The size of a file depends upon the size of memory or the storage space of computer system. File has two types of characteristics which help determine the file structure. Following are characteristics of file structure:
  - **File Activity:** It specifies the percentage values of the records which are processed in single run.
  - **File Volatility:** This specifies that the properties of record change while processing.

NOTES

- **Database:** It is the upper layer in hierarchy of file structure. It is a set of interrelated files for real time processing. It contains the data, which is required for problem solving and can be used by several users.

File Organization

File organization is the methodology which is applied to structured computer files. Files contain computer records which can be documents or information. It refers primarily to the logical arrangement of data which can itself be organized in a system of records with correlation between the fields or columns in a file system. There are certain basic types of computer file which can include files stored as blocks of data and streams of data where the information streams out of the file while it is being read until the end of the file is encountered. Therefore, file organization refers to the way records are physically arranged on a storage device. The two main types of file organization are record type and record access. These types are described below:

- Record type refers to whether records in a file are all the same length are of varying length or use other conventions to define where one record ends and another begins.
- Record access refers to the method used to read records from or write records to a file regardless of its organization. The way a file is organized does not necessarily imply the way in which the records within that file will be accessed.

Following are the two prime types of file organizations:

Sequential File Organization

A sequentially organized file consists of records arranged in the sequence in which they are written to the file, i.e., the first record written is the first record in the file, the second record written is the second record in the file and so on. As a result, records can be added only at the end of the file. Attempting to add records at some place other than the end of the file will result in truncating the file at the end of the record. Sequential files are usually read sequentially, starting with the first record in the file. Sequential files with a fixed length record type that are stored on disk can also be accessed by relative record number (direct access). A sequential file contains records organized by the order in which they were entered. The order of the records is fixed. Records in sequential files can be read or written only sequentially.

Relative File Organization

Within a relative file are numbered positions, called cells. These cells are of fixed equal length and are consecutively numbered from 1 to n, where 1 is the first cell and n is the last available cell in the file. Each cell either contains a single record or is empty. Records in a relative file are accessed according to cell number. A cell number is a record's relative record number where its location relative to the beginning of the file. By specifying relative record numbers, you can directly retrieve, add or delete records regardless of their locations. Within the cells, you can store records of varying length as long as their size does not exceed the cell size.

It is necessary to organize the file properly in order to ensure that the records are available for processing. The two main characteristics that determine how the file is organized are activity and volatility. File activity determines the total number of records that can be processed in a single run. If the total number of records that can be accessed

at a specific time is small, then the file is organized on a disk or on a tape. File volatility focus on the properties of the records that changes.

5.5.4 Database Design

NOTES

Database is a collection of data stored for a specific purpose. Database Management System (DBMS) is a software system that allows you to define, construct and manipulate databases that are used for various purposes, such as for storing the customer and financial information of an organization. DBMS provides significant features that make a database system efficient and reliable. It also provides multiple user interfaces that help a user in interacting with the system. The interfaces provided by DBMS include the query language for users and the programming language interface for application programmers. DBMS prevents the data redundancy problems that occur in other management systems. DBMS provides security to the system by restricting the unauthorised users to access the data of an organization. The backup and recovery feature of DBMS allows you to recover data that is lost due to hardware or software failure. In addition, DBMS users can also perform a variety of operations, such as insert, delete, retrieve and update on the database.

Client-server architecture is used to retrieve information from the databases. Clients are the machines that request for the services and server is the machine that processes the requests send by the clients. Server processes the information based on the information that is stored either in flat file database or relational database. This unit also lists various versions of Structured Query Language or SQL Server available in the market. To utilize the information obtained from the SQL Server database, you need to install SQL Server clients. A SQL Server can be defined as database a collection of related data items that can be stored in database objects, such as tables. Databases provide an easy way to access, manage and update the data in a database. There are two types of databases:

- **Flat File Database:** Flat file databases are simple text files that store data in a single table.
- **Relational Database:** Relational databases are complex databases that store information in multiple tables. In relational database, every table that stores data is linked to another table using a common column.

Objectives of Database Design

Database design refers to a system that stores and retrieves data systematically from the database. However, first you must know the description of a few basic terms of DBMS so that you can design a suitable database as per the requirement of an organization. These are as follows:

- **Data:** Data means known facts that can be recorded and used to produce useful information, e.g., names, telephone numbers and addresses of the employees of an organization.
- **Database:** A database is a collection of interrelated data. For example, you may create a database by recording the names, telephone numbers and addresses of employees in an indexed addressed book or on a diskette, using a personal computer and software, such as Microsoft Access or Microsoft Excel. This recorded information is a collection of related data with an implicit meaning and hence is called a database.

NOTES

- **Defining a Database:** This involves specifying the data types, structures and constants for data to be stored in the database.
- **Constructing a Database:** This is a process of storing real data on some storage medium with the help of DBMS.
- **Manipulating a Database:** This involves performing functions, such as querying the database to retrieve specific data, updating the database to reflect changes made by the user and generating reports from data.
- **Database System:** The database and DBMS software together form a database system.

To understand the basics of DBMS design, you must know the terms and definitions that are used in DBMS technology. DBMS is a software program that may run on a user machine or a server computer. DBMS accepts queries from users and responds to these queries. The DBMS design has the following features:

- **Structured Data:** DBMS enables you to structure the data as tables, records or objects.
- **Query Language:** DBMS provides a query language, such as SQL to process the user requests.
- **Multi-user Access:** DBMS allows users to access data stored in a database. At the same time, it provides security features that restrict some users from viewing or manipulating data.
- **Data Dictionary:** DBMS provides a data dictionary that contains the structure of a database.
- **Data Abstraction:** It allows a DBA to logically separate data from the programs that use the data. There are three levels of abstraction in a database: external, conceptual and internal. The external level represents the user view of the database; the conceptual level allows you to map the internal and external levels; and the external level represents the operating system and DBMS level.

5.5.5 Form Design

Every rule related to the logical and physical data structures that has been obligated by the DBMS, must be followed by the entire database description managing the database. Therefore, steps related to, viewing the physical database design, transformation of the logical description with a physical model describing the organization of the database and the method of accessing the physical storage devices.

The user interface usually comprises of various documents, reports and different screens of a system and is the sole component of the system that is viewed by the user. Hence, for the user, user interface is a vital component the system. It should be designed in such a way that besides delivering information to the user it should also be visible.

For the purpose of designing an interface, the steps required are stated as follows:

1. Generation of the interface summary table for system use.
   a. Analysing the human – computer limit in the DFD.
   b. Filing the interface and analysing the design for the purpose of defining data elements that has to be added in every form/document, input screen and output report.
2. Designing the external input forms.

3. Designing the human – computer dialogues.
   a. Choosing the kind(s) of dialogue.
   b. Designing of every dialogue.
4. Designing input (data entry) screens.
5. Report designing
   a. Choosing the output technique for every report.
   b. Internal report designing.
   c. External report designing.

NOTES

Input Form Design

The excellence of system output is concluded by the quality of system input. After understanding such significant relation, it becomes necessary that the input forms and screens as well as output reports should be designed.

The basic objectives of the input form design must be achieved by a good designed input form and visual display screen which can only take place by using the essential design principles, information about which type of data is required to be used in the system and recognizing the way by which the users react to various different components of forms and screens.

The term 'effectiveness' basically refers that the input forms and screens play a particular role in the MIS (Management Information System), whereas the term 'accuracy', highlights the design that guarantees suitable conclusion.

Objectives

Extraction and capturing the needed information through the members of the organization which will be frequently added to the computer with the help of form designing. Forms act as the source documents regarding the data entry personnel with the help of this process.

Guidelines of Form Design

- Ease in filling out the forms.
- Provides certainty of the forms meeting the reason of their being.
- Makes the forms eye-catching.
- Design formats for confirming precise completion.

5.5.6 Design Methodologies

Based on the user requirements and the detailed analysis of a new system, the new system must be designed. This approach is known as is the 'phase of system designing'. It is a most crucial phase in the development of a system. Normally, the design proceeds in two stages are preliminary or general design and structure or detailed design. These stages are discussed below:

- **Preliminary or General Design:** In the preliminary or general design, the features of the new system are specified. The costs of implementing these features and the benefits to be derived are estimated. If the project is still considered to be feasible, the detailed design stage is determined.

NOTES

- **Structure or Detailed design:** In the detailed design stage, computer oriented work begins in earnest. At this stage, the design of the system becomes more structured. Structure design is a blue print of a computer system solution to a given problem having the same components and inter-relationship among the same components as the original problem. Input, output and processing specifications are drawn up in detail. In the design stage, the programming language and the platform in which the new system will run are also decided.

System hardware architecture describes the overall system hardware and organization. Include a list of hardware components (with a brief description of each item) and diagrams showing the connectivity between the components. If appropriate, use subsections to address each subsystem. System software architecture describes the overall system software and organization. Include a list of software modules (this could include functions, subroutines or classes), computer languages, and programming computer-aided software engineering tools (with a brief description of the function of each item). Use structured organization diagrams, object oriented diagrams that show the various segmentation levels down to the lowest level. All features on the diagrams should have reference numbers and names. Include a narrative that expands on and enhances the understanding of the functional breakdown. If appropriate, subsections are used to address each module.

Note: The diagrams should map to the data flow diagrams, providing the physical process and data flow related to the logical process and data flow.

Hardware Detailed Design

A hardware component is the lowest level of design granularity in the system. Depending on the design requirements, there may be one or more components per system. This approach provides detailed information about individual component requirements to correctly build and procure all the hardware for the system. If there are many components or if the component documentation is extensive, place it in an appendix or reference a separate document. Add additional diagrams and information, if necessary, to describe each component and its functions, adequately.

Software Detailed Design

A software module is the lowest level of design granularity in the system. Depending on the software development approach, there may be one or more modules per system. This approach provides detailed information about logic and data necessary to completely write source code for all modules in the system. If there are many modules or if the module documentation is extensive, place it in an appendix or reference a separate document. Add additional diagrams and information, if necessary, describe each module, its functionality and its hierarchy.

Internal Communications Detailed Design

If the system includes more than one component there may be a requirement for internal communications to exchange information, provide commands or support input/output functions. This approach provides enough detailed information about the communication requirements to correctly build and procure the communications components for the system.

Interface Detailed Design

For each system that provides information exchange with the system under development, there is a requirement for rules governing the interface. This section should provide enough detailed information about the interface requirements to correctly format, transmit, and/or receive data across the interface.

NOTES

5.6 SYSTEM IMPLEMENTATION

The implementation phase is concerned with making actual modifications in the software code, adding new features that support the specification of present software and finally installing the modified software. As shown in Figure 5.21, the *input* attribute comprises output of the design phase, the current source code and the updated project and system documentation.

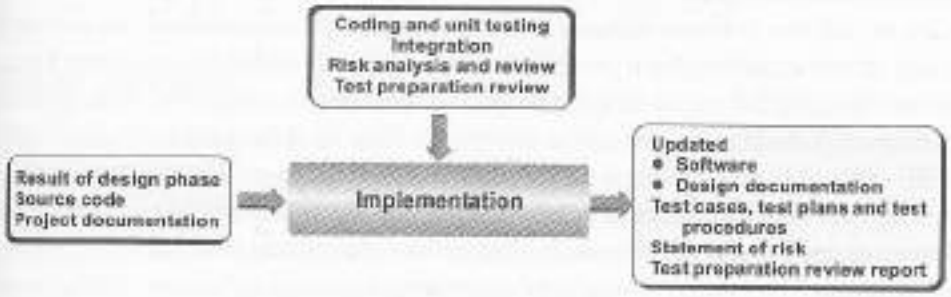


Fig. 5.21 Implementation Phase

The *process* attribute comprises the following subprocesses:

- ***Coding and Unit Testing:*** The user requirements as identified in design phase are implemented in the form of a software code. After the software code is written, unit testing on individual modules is performed.
- ***Integration:*** When the software code is written and all modules are tested, the modified software is integrated with the system. After integration of the software system, integration test and regression test are performed. These tests are conducted to identify the impact on the functionality, performance, usability, and security of the modified software system. Impacts that lead to serious errors, such as failure of software system, are recorded so that they can be removed after identifying and understanding the source of error.
- ***Risk Analysis and Review:*** In the implementation phase, the risk analysis and review are carried out periodically rather than on completion of this phase. Risks that commonly occur are risk for data failure, provision of back up and so on.
- ***Test Preparation Review:*** This review is conducted to evaluate the review for system test in accordance with IEEE standard.

The *control* attribute performs the following functions:

- Conducting software inspections of the code in compliance with IEEE standard.
- Ensuring that unit and integration testing are performed and properly documented.
- Ensuring that test documentation, such as test plan and test cases are either updated or created.

- Verifying that the new software is placed under Supply Chain Management (SCM) control.
- Verifying that the training and technical documentation have been updated.

NOTES

The *output* of this phase is in the form of updated (modified) software according to modification request, updated design documentation, updated test documentation, updated user documentation and updated training material.

5.6.1 System Testing

System testing is implied on a complete and integrated system to evaluate the system's compliance with its specified requirements. It uses the scope of black box testing and requires no knowledge of the inner design of the code or logic. System testing provides stakeholders with information about the quality of the product or service under test. It also provides an objective and independent view of the software to allow the business to appreciate and understand the risks of software implementation. A primary purpose of testing is to detect software failures so that defects may be discovered and corrected. Testing cannot establish that a product functions properly under all conditions but can only establish that it does not function properly under specific conditions. System testing is performed on the entire system in the context of a Functional Requirement Specification or FRS and a System Requirement Specification or SRS. System testing tests not only the design but also the behaviour as per expectations of the customer. It is also intended to test up to and beyond the bounds defined in the software and hardware requirements specifications. The system testing has three main purposes: verification, validation and defect finding which are discussed below:

- **Verification Process:** The verification process confirms that the software meets its technical specifications. The term 'specification' is a description of a function in terms of a measurable output value given a specific input value under specific preconditions.
- **Validation Process:** The validation process confirms that the software meets the business requirements. A simple example of a business requirement is after choosing a branch office name, information about the branch's customer account managers will appear in a new infrastructure. This approach will present manager identification and summary information about each manager's customer base as 'list of data elements'. Other requirements provide details on how the data will be summarized, formatted and displayed.
- **Defect Process:** A defect is a variance between the expected and actual result. The defect's ultimate source may be traced to a fault introduced in the specification, design or development or coding phases.

Test Plan

The release of a new application or an upgrade inherently carries a certain amount of risk that it will fail to do what to do. A good test plan goes a long way towards reducing this risk. By identifying areas that are riskier than others we can concentrate our testing efforts there. The risk areas require more certainty that they work properly, failing to correctly identify those risky areas leads to a misallocated testing effort. Table 5.4 summarizes the list of test plan.

Table 5.4 Summarizes the List of Test Plan

Component	Description	Purpose
Responsibilities	This component represents who are specific people and their assignments.	It assigns responsibilities and keeps everyone on track and focused.
Assumptions	This component represents code, systems status and availability.	It avoids misunderstandings about schedules.
Test	This component represents testing scope, schedule, duration and prioritization.	It outlines the entire process, maps and specific tests.
Communication	This component represents communications plan: who, what, when and how.	Everyone knows what and when they need to know it.
Risk Analysis	This component represents critical items that will be tested.	It provides focus by identifying areas that are critical for success.
Defect Reporting	This component states how defects will be logged and documented.	It tells how to document a defect so that it can be reproduced, fixed and retested.
Environment	This component represents the technical environment, data, work area and interfaces used in testing.	It reduces or eliminates misunderstandings and sources of potential delay.

NOTES

Testing Strategies

The purpose of a test strategy is to clarify the major tasks and challenges of the test project. Test approach and test architecture are other terms commonly used to describe the test strategy. It describes what kind of testing strategy needs to be done for a project, for example user acceptance testing, functional testing, load testing, performance testing, etc. Testing strategy is a general approach rather than a method of devising particular systems for component tests. Different strategies are applied depending on the type of system to be tested and the development process used. The testing strategies are as follows:

- **Thread Testing Strategy:** This testing strategy is used for systems with multiple processes where the processing of a transaction threads its way through these processes.
- **Stress Testing Strategy:** This testing strategy relies on stressing the system by going beyond its specified limits and hence testing how well the system can cope with overload situations.
- **Back-to-Back Testing Strategy:** This testing strategy is used when versions of a system are available. The systems are tested together and their outputs are compared.
- **Performance Testing Strategy:** This testing strategy is used to test the runtime performance of software.
- **Security Testing Strategy:** This testing strategy is used to verify that protection mechanisms built into system will protect it from improper penetration.
- **Recovery Testing Strategy:** This testing strategy forces software to fail in a variety of ways and verifies that recovery is properly performed.
- **Ad Hoc Testing Strategy:** Testing without a formal test plan or outside of a test plan. With some projects this type of testing is carried out as an adjunct to formal testing. If carried out by a skilled tester, it can often find problems that are not caught in regular testing. Sometimes, if testing occurs very late in the development cycle, this will be the only kind of testing that can be performed. Sometimes ad hoc testing is referred to as exploratory testing.

NOTES

- **Automated Testing Strategy:** Software testing that utilizes a variety of tools to automate the testing process and when the importance of having a person manually testing is diminished. Automated testing still requires a skilled quality assurance professional with knowledge of the automation tool and the software being tested to set up the tests.
- **Compatibility Testing Strategy:** This testing strategy is used to determine whether other system software components, such as browsers, utilities and competing software will conflict with the software being tested.
- **Configuration Testing Strategy:** This testing strategy is used to determine how well the product works with a broad range of hardware/peripheral equipment configurations as well as on different operating systems and software.
- **Functional Testing Strategy:** This testing strategy is used to test two or more modules together with the intent of finding defects, demonstrating that defects are not present, verifying that the module performs its intended functions as stated in the specification and establishing confidence that a program does what it is supposed to do.
- **Independent Verification and Validation or IV&V Strategy:** This testing strategy is used to check the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner. The individual or group doing this work is not part of the group or organization that developed the software. A term often applied to government work or where the government regulates the products, as in medical devices.
- **Installation Testing Strategy:** This testing strategy is used to test with the intent of determining if the product will install on a variety of platforms and how easily it installs.
- **Regression Testing Strategy:** This testing strategy is used to test with the intent of determining if bug fixes have been successful and have not created any new problems. Also, this type of testing is done to ensure that no degradation of baseline functionality has occurred.
- **Security Testing Strategy:** This testing strategy is used to test of database and network software in order to keep company data and resources secure from mistaken or accidental users, hackers and other attackers.
- **Software Testing Strategy:** This testing strategy is used to check the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner. The organization and management of individuals or groups doing this work is not relevant. This term is often applied to commercial products, such as Internet applications.

Complex and large systems are usually tested using a mixture of these strategies rather than any single approach. Different strategies may be needed for different parts of the system and at different stages in the testing process. Whatever testing strategy is adopted, it is always right to adopt an incremental approach to sub-system and system testing. Figure 5.22 illustrates the process of testing strategies.

NOTES

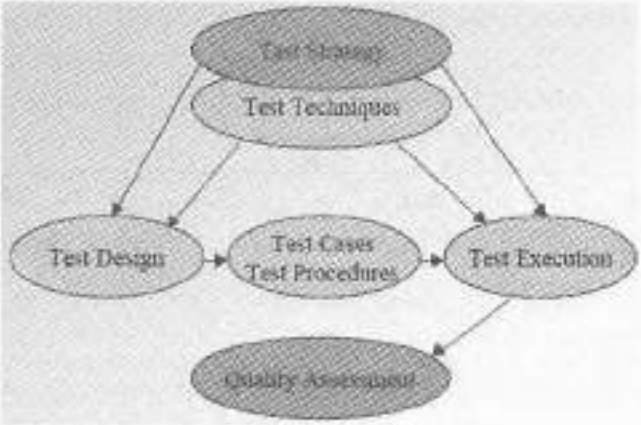


Fig. 5.22 Testing Strategy and Quality Assessment

Rather than integrate all components into a system and then start testing, the system should be tested incrementally. Each increment should be tested before the next increment is added to the system. This process should continue until all modules have been incorporated into the system. When a module is introduced at some stage in this process, tests which have previously been unsuccessful detect defects. These defects are probably due to interactions with the new module. Table 5.5 summarizes the types of testing, implementation techniques and their function.

***Table 5.5** Types of Testing, Implementation Technique and their Function*

Types of Testing	Implementation Techniques	Function
Unit Testing	Coding	This type of testing is used to focus on each module and whether it works properly. This testing makes heavy use of white box testing.
Integration Testing	Design	This type of testing is centered on making sure that each module works with another module. It is comprised of top-down and bottom-up integration. It mainly focusses on the design and construction of the software architecture.
Validation Testing	Analysis	This type of testing ensures the conformity with requirements.
Systems Testing	Systems Engineering	This type of testing is used to make sure that the software product works with the external environment, for example computer system, other software products.

5.6.2 Quality Assurance

The basic goal of system control and quality assurance is to check the sample of the developed system and implement methodologies to find errors, if any, in the processed data. The type and quality of the developed system needs to be assessed, planned and implemented. At this phase, the quality management team fixes the criteria to assess data quality and arrange training programs for effective implementation of system control and quality assurance. These steps ensure that the developed system is authentic and full of security safeguard; and it is properly documented to comply with the cost effectiveness of program requirements. The system control is basically oriented to prevent system deficiency, such as complexity, low quality of system execution and desired output screen.

NOTES

Fig. 5.23 Hierarchy of Quality Control Team

Figure 5.23 shows how the hierarchy of the quality control team works together to confirm quality assurance. The complete system is first scrutinized. The team then analyses the performance of the system and the team prepares a user manual to guide the users. This manual is prepared technically, i.e., it is scrutinized by gaining the Incoming Quality Control (IQC), Process Quality Control (PQC) and Outgoing Quality Control (IQC). Table 5.6 shows the 'FPA' phase known as Functionality → Processing → Adaptabilty mechanism.

***Table 5.6** Typical Software Quality Control Factors*

Functionality (Exterior Quality)	**Processing (Interior Quality)**	**Adaptability (Future Quality)**
Correctness	Efficiency	Flexibility
Reliability	Testability	Reusability
Usability	Documentation	Maintainability
Integrity	Structure of the system	

Process and product-level quality are part of software quality engineering. Process level quality is referred to as quality assurance whereas product level quality is referred to as testing. Software Quality Assurance (SQA) comprises the methods, procedures and tools applied by software professionals to ensure that the software fulfils the standards already specified and set. The purpose of SQA is to provide the administration with proper visibility of the software project and of the products being developed. It produces an efficient, optimized and acceptable software code as fault-free as possible by reviewing and auditing the software products and activities.

SQA monitors the methods and standards to confirm that the software specialists have properly implemented their expertise.

During the initial stages of the software development process, the SQA group along with the software team establishes plans, standards and procedures for the software that satisfy the objectives of the organization. The SQA group reviews and audits the software product and activities throughout the SDLC the following are the goals of SQA:

- To plan the SQA activities.
- To monitor the software development process and the final software developed.
- To ensure that the software project is implementing the standards and procedures set by the management.
- To notify groups and individuals about the SQA activities and results of these activities.

- To ensure that the issues that are not solved within the software are addressed by the upper management.
- To recognize any insufficiency in the product, process or the standards and fix them.

NOTES

SQA Activities

The following are the top-level activities performed by the SQA group:

- Prepare the SQA plan for the software according to a documented procedure.
- Ensure that the tasks performed by SQA are according to the plan.
- Help in preparation and review of the standards and procedures of the software development process.
- Review the activities of the software development team to confirm compliance.
- Inform the results of its activities to the software development team.
- Handle the difference found in the result of the software activities and the software product according to a documented procedure.

Levels of Quality Assurance or QA

In order to certify a software product, several levels of QA and testing have to be performed. All QA levels are not needed for all the different types of software described.

Level 1: Code Walk through

At this level, the software is checked for any violations of the official coding rules in the off-line software. In particular, emphasis is placed on the examination of the documentation and the level of in-code comments.

Level 2: Compilation and Linking

At this level, it is checked if the software can compile and link all official platforms and operating systems.

Level 3: Routine Running

Routine running assures that the software can run properly under a variety of conditions such as certain number of events and small and large event sizes, etc.

Level 4: Performance Test

At this final level, the software is evaluated for its performance. This test checks whether the performance of the software satisfies the previously specified performance level. The performance of the software can be evaluated by determining the time taken to reconstruct an event and the CPU time per event.

Quality assurance refers to the systematic measurement, comparison with a standard, monitoring of processes and an associated feedback loop that confers error prevention. Quality assurance activities include a planned system of review procedures conducted by personnel not directly involved in the inventory compilation and development process. Reviews, preferably by independent third parties, should be performed upon a finalized inventory following the implementation of QC procedures. Reviews verify that data quality objectives have been used to ensure that the inventory represents the best possible estimates of emissions given the current state of scientific knowledge and data available, and support the effectiveness of the QC program. Therefore, quality assurance

NOTES

is a set of activities designed to ensure that the development and maintenance process is adequate to ensure a system will meet its objectives. The quality control is a set of activities designed to evaluate a developed work product. These activities can always include the audits conducted by the QA team to assess the cost of correcting defects, documentation, etc. The quality control in software system impacts significantly to control the version and manage the changes in a fast-paced iterative environment. The status report sometimes comes critical due to quality check. The lack of configuration status wreaks any project with many occurrences of change. Therefore, the configuration management is an important phase in the system maintenance cycle. In a nutshell, it incorporates how the iterative life cycle increases opportunity to interject test activities. If errors are found at this stage, the best real time is calculated. The role of quality assurance engineers to check the quality of the application more effectively because the quality checking team nurtures the application in various design iterations, integration tests, platform deployments configuration management and software reviews at regular interval. The significant advantage of this phase is to increase interaction between system quality and software engineers. The two approaches in testing maintenance are involved as follows:

- The testing maintenance is used to avoid defects in final system.
- It is also used in fault tolerance methods which that enable a system to continue operating properly in the event of the failure.

The above two approaches are the major criteria to operate the system in the presence of latent faults. They assess with reference to projected product reliability, budget and schedule constraints. Both mechanisms involve analyzing the software development process, executing the system and simulating the real process. They are used to forecast the impact of software reliability (effectiveness), cost and development time (efficiency). The fault tolerance practices mask the remaining defects. Testing with operational phase uncover the faults which are encountered in the operational phase. If failure is encountered, the corresponding faults and defects are identified known as fault identification and corrected known as rework testing. Basically, the regression testing is used to detect periodically for finding new defects (bad fixes). Factors, such as conversion and operation plans used to determine the quality assurance are discussed below.

5.6.3 Hardware/Software Selection

The hardware and software selection is a very crucial and important step in SDLC phase. The success rate of system design and implementation phase depends on this factor. The SDLC phase works out with vendor selection and benchmarking schemes. Selecting software is a crucial task to determine the issues of objectives, goals, needs, plans, etc. Basically, the development team mainly focuses on the software functions. The information based system finalizes the information requirements to check out the vendor. The developed system is not complete until software is installed properly, tested-up and running. The hardware and software selection plays an important role in SDLC because of the data security and system maintenance. The categories of hardware and software are considered as per functional requirements. The selection evaluates basically software vendors, implementers, Application Service Providers (ASP), technology trends and impact for the organization for which complete system is being developed. The complete scenario is explained with the help of comprehensive case studies. If a company plans to select and implement the Local Area Network (LAN) setting for the organization,

NOTES

he various types of software are delivered to the computer as per selection. The erformance and acceptance criteria plays a business scale for successful implementation f system. For this, system analysts and vendors are quite aware for taking step wise decision. It is also necessary for both to be sure that the selected software is compatible with specific server that fulfils the requirements. If company wants that selected software s to be available on each station then the proper arrangements are implemented by nalysts, such as costing factor is checked to install all the systems for the specified Web site. The basic task is to determine what software is required for hardware equisition.

'At the clubhouse, young people become designers and creators not just consumers of computer based products. Participants use leading edge software to create their own artwork, animations, simulations, multimedia presentations, virtual worlds, musical creations, Web sites and robotic constructions.'

[Mitch Resnick]

The software selection is not easy task because selections should always be reflective of the company's mission. It is believed that no guarantee is preferred for any software and it in the commercial circulation performs the best of its specific types. It is also crucial to select hardware equipments because it is hard to know whether increased memory in hardware is able to drive the software configurations. Most software packages will list the minimum memory and a hardware requirement to run the software properly therefore it is needed to be sure to check these requirements before purchasing the software. The servers are installed to enhance and renovate the system design. For this, system analyst keeps in the mind about memory capacity and other hardware system limitations when selecting vendors and software packages.

Hardware Selection

The hardware selection is an essential task in the phase of system designing because system analyst checks thoroughly the comprehensive platform for developed system. For example, if the system is developed for online air ticket booking system then analyst must pay the attention to various Web tools, enhanced browsers, etc. The system application server software requires the hardware and software requirements. There must be 35MB free space for temporary files and directories and 250MB free space is required for installation. If you remove the application server from system unit in which software is installed, maximum space is required to install a new version. The various ports are used for default domain. Some of the default ports are required for developed systems. The port 8080 is required for HTTP, port 3920 is required for Secure Socket Layer (SSL) with mutual authentication and port 4848 is required for administration server. Firewall software must be installed with system application server software because most of the ports are disabled by default. Software needs an installation program for available ports that can enable the ports for programs. The hardware and software requirements during the system installation phase enable networking capabilities and fast Internet connection that are required to access a HyperText Transfer Protocol (HTTP) server. It simulates the Web page accessing. Table 5.7 summarizes operating system with their memory requirements.

NOTES

Table 5.7 Platform for Selected Language

Operating system	Minimum memory	Recommended memory	Minimum disk space	Recommended disk space
Sun Solaris 9, 10 (SPARC) Solaris 9, 10(x86)	512MB	512MB	250MB free	500MB free
64-bit Sun Solaris 10 SPARC, x86)	512MB	512MB	250MB free	500MB free
Red Hat Enterprise Linux 3.0 U1, 4.0	512MB	1GB	250MB free	500MB free
SUSE Linux Enterprise Server 10SP1	512MB	1GB	250MB free	500MB free
64-bit SUSE Linux Enterprise server 10SP1	512MB	1GB	250MB free	500MB free
Ubuntu Linux, Hardy Release	512MB	1GB	250MB free	500MB free
AIX 5.2, 5.3	512MB	1GB	250MB free	500MB free
Windows Server 2003 Windows XP Pro Windows Vista	1GB	2GB	500MB free	1GB free
Macintosh (Intel, Power)	512MB	512MB	250MB free	500MB free

In Table 5.7, Solaris 9, 10(x86, SPARC) uses the Solaris patch cluster that supports security features and runtime environment for programs and applets. The connectivity requirements contain 64-bit hardware communication of Transmission Control Protocol/ Internet Protocol (TCP/IP). The TCP/IP for Internet Protocol version 4 (IPv4) and Internet Protocol version 6 (IPv6) protocols are released with software program so that programs and applets make connection easily with Internet. The hardware contains Random Access Memory (RAM), Field Programmable Gate Array (FPGA) and Read Only Memory (ROM) which are tagged with host controller in which host system contains finalizer, scheduler, clock, idler and memory. The drivers and databases need to get the hardware and software compatibility requirements.

Note: The acronym SUSE came from the German name Software und System Entwicklung (Software and System Development)

Software Selection

For online system, the software selection is crucial task because you cannot use FoxPro database for back end support instead you can use MySQL or Oracle for back end database. In the same way, ASP.NET or HyperText Preprocessor (PHP) scripting language is required for that. The Word processors, computer language programs, Notepad, Text Pad, language editors, various software, such as Java, C++, Ada, scripting languages, Web development tools, Oracle, etc., must be installed in the system to write, edit and debug the coding. The application server is designed to support the connectivity of additional DBMS that corresponds to system driver. Table 5.8 summarizes the requirement of supported database servers for driver.

NOTES

Table 5.8 Supported Database Server

Vendor of Java Database Connectivity (JDBC) Driver	Supported Database Server
Derby Network Client	Derby 10.2
DataDirect 3.6.x, 3.7.x	Oracle 10g Oracle 9i MySQL 2005
MySQL Connector	MySQL 5.0
Oracle 10G	MySQL 5.0
PostGres	8.2x
Oracle Call Interface (OCI)	Oracle 9i
International Business Machines (IBM) DB2	DB2 9.1

The output files can be displayed in proper display unit. The system unit must have 1024 pixels × 768 pixels. The header and navigation area needs 760 pixels × 570 pixels for navigation and header areas that provide complete visibility for graphics and pictures. The development environment for various programs provides runtime environment that is necessary to execute the Java programs. It facilitates plug-ins for browsers that linked with the programs are to be fetched by browsers. This software toolkit comes in two types, such as a precompiled binary and a source package. The Web development tools must be installed in Windows machine therefore, Windows operating system is required. Table 5.9 summarizes the software and hardware requirement for this release.

Table 5.9 Component and Requirement for System Development

Component	Requirement
Operating System	Solaris 10 on Scalable Processor ARChitecture (SPARC), x86 and x64 based systems are required to include the whole root local and sparse root zones. Red Hat Enterprise Linux 5.0 Server, 32-bit and 64-bit update versions and Red Hat Enterprise Linux 3 and 4 update versions, Advanced Server (32-bit and 64-bit versions) and Enterprise Server (32-bit and 64-bit versions) are also required. Windows operating system, Windows 2000 Advanced Server, Data Center Server version SP4 on x86, Windows 2003 Standard (32-bit and 32-bit versions), Enterprise (32-bit and 32-bit versions), Data Center Server (32-bit version) on x86 and x64 based systems and Windows XP Professional SP2 on x86 based systems are considered as required operating system.
Java 2 Standard Edition (J2SE) platform	J2SE platform supports Update 9 (HP-UX 1.5.0.03), 1.4.2 Update 11 and Update 5.0 and Update 12 for Java Enterprise System.
Directory Server	Directory server represents Access Manager information tree which includes Directory Server Enterprise Edition 6.1 (compatible with Access Manager P7.1 Patch 1), Sun Java system directory Server 6.0 and Sun Java system directory server 5.2 2005Q4. Access Manager Identity Repository contains directory server enterprise edition 6.1 and 6.2 which is compatible with Access manager 7.1 Patch 1, Sun Java system directory server 5.2 and 6.0 and Microsoft active directory.
Web Containers	Sun Java system Web server 7.0 and 7.0 update 1 supported platform/OS combinations are run on the Web server instance in a 64-bit Java Virtual Machine (JVM). The supporting platforms are required for Web containers in Java language, for example Solaris 9/SPARC, Solaris 10/SPARC, Solaris 10/AMD64, Red Hat AS or ES 3.0/AMD64, Red Hat AS or ES 4.0/AMD64.
RAM	The RAM capacity contains basic testing of system performance as 512 Mbytes. The actual deployment should be 1 Gbyte for threads and Access Manager can be facilitated by Software Development Kit (SDK), HyperText Transfer Protocol (HTTP) server and other internals.
Disk Space	The disk space would be of 512 Mbytes that is suitable for Access Manager and Java associated applications.

NOTES

5.6.4 System Security

System security can be divided into different categories according to security issues, which are related to the file structure, data structure and access procedures of system. Following are the various system security issues:

- **System Reliability:** Refers to the technical innovation and procedures applied to the hardware and operating system to protect system from damage.
- **System Integrity:** Refers to the proper working process of hardware, software, appropriate physical security and safety from external security threat.
- **Data Integrity:** Refers that data should be same as its original form and should not be disclosed, altered or destroyed due to unsecure access.
- **System Privacy:** Refers to the rights of the users or organization to determine which information should be shared with or accepted from other and how can an organization protect their information system from unfair use.
- **Protection Against Viruses:** Refers to the methodes that can be used to protect system from virus infection. For this you can use diffrernt anti virus softwares to protect system.

Threats to System Security

A procedure of system security makes sure that the facility is physically secure and provides a recovery capability. But system security can be damaged due to external threats. Various external threats are as follows:

- Errors generated due to people making mistakes
- Unauthorized users
- Natural disasters
- Physical attack

Risk Analysis

The risk analysis helps to analyse the probability of problem occurrence, cost of each possibility of system damage and the preventive measures to include as part of security plan. Two key factors of risk analysis are the value of potential loss and the probability of loss.

Control Measures

After determining system security issues and risk associated with system security you need to identify the control measures that can be internal and external. Following are the various control measures:

- **Identifying the Users:** Helps to detremine the user of system with their defined access right.
- **Access Control:** Helps in controlling the system access. Various activities can be performed to control the access of system, such as encoded card system and encryption method.

Protection Against Virus

A virus refers to a computer program which enters in a computer without the knowledge of the user. In a virus infected file, virus is hidden inside a program contained in the file. You can protect your system against viruses using anti virus software and recovery method.

Antivirus Software

Antivirus software contains computer programs that identify and eliminate computer viruses from computer. Following methods are used by anti virus software to detect viruses in a computer as follows:

- Scanning all the files available in a computer to search for viruses.
- Detecting suspicious behaviour of any program in a computer which might be a virus.

Recovery Methods

Recovery methods are used to recover the data infected by virus in a computer. A computer infected with virus is unsafe to use without reinstalling the operating system. The various recovery methods are as follows:

- Operating system reinstallation helps recover the data infected by virus.
- Virus removal tools are provided by antivirus software which helps recover the data.

System Audit

System audit refers to a technical assessment of a system or application. Personal computers, servers and mainframes are the example of system whereas database and Web services are the examples of applications. System audit is performed using two methods, automated and manual. Automated system audit uses software for automated assessment. This system audit performs various activities, such as generating audit reports or monitor and report changes made to various files stored in the system. Manual system audit includes various activities, such as interviewing staff, reviewing system and security vulnerability scans.

5.7 SUMMING UP

- A system is a set of organized principles, tasks, activities and procedures that defines the working of a particular establishment to achieve an objective. It is logically arranged and governed by rules, regulations and policies.
- A system is defined and determined by its boundaries and objectives. It is an arrangement of smaller systems in a logical order.
- A system may be a closed system or an open system. In an open system, you can go on adding subsystems and it grows bigger and bigger. In a closed system, it is not so. Environment of a system means other systems that influence the system under consideration.
- The term 'system' is derived from the Greek word 'systema' (to combine) which means an organized relationship among functioning units or components. A system exists because it is designed to achieve one or more objectives.
- A system is set of inter related tasks and procedures working in coordination towards achieving a common objective by receiving the inputs and providing the outputs in an organized transformation process. Such a system can be called as dynamic system.
- A manual system usually means done by hand. That could be typing all your information into Notepad whereas automated system refers to the use of control

NOTES

Check Your Progress

8. Name the five components of system design process.
9. What are the concerns in implementation phase?
10. What is specification?
11. What is the purpose of test strategy?
12. Write the basic goal of system control and quality assurance.
13. What is the significance of hardware/software selection?
14. What is the use of risk analysis?

NOTES

systems and information technologies to reduce the need for human work in the production of goods and services.

- Information systems are data processing systems, which collect data from different sources, process that data and generate information from the data which can further be used for different applications within the organization.
- Systems Development Life Cycle (SDLC) is a series of steps agreed upon by software developers in creating special software. Developers adhere to certain steps to ensure that they have the right software for the right demand.
- When a system owner or a user makes a request for a new system or some modifications in an existing system, the first system activity, i.e., identifying problems in the existing system, begins.
- System analysis is a process in which data are collected and then interpreted to identify the problems within the system. Therefore, the collected information can then be used to recommend improvements in a system.
- The process of analysis involves a detailed study of the current system leading to specifications of a new system. The approach of analysis is a detailed study of various operations performed by a system and their relationships within and outside the system.
- The purpose of system analysis phase is to build a logical model of the new system. The first step is related to requirements modelling where you investigate business processes and documents what the new system must do to satisfy users.
- The system planning phase usually begins with a formal request known as system request which describes problems or desired changes in an information system or a business process.
- Preliminary investigation is the first step in SDLC. It begins once a project request is made. The purpose of this preliminary investigation is to evaluate the project request.
- Feasibility study is an important outcome of the preliminary investigation which determines whether the system, for which the request has been made, is feasible or not.
- Various kinds of reports and records that store information related to the existing system are maintained in an organization. These reports and records can provide valuable information regarding the organization, its operations and activities, to the systems analysts.
- Procedure manuals and forms are excellent sources of information because they describe the format and functions of the present system. Update manuals can provide a lot of information, which saves a lot of time.
- Interviewing is a common method used by systems analysts for collecting data on information requirements. An interview is a directed conversation with a specific purpose that uses a 'Question and Answer' format.
- A questionnaire asks people to respond to various written questions which are presented to them in a formal document.
- The use of standardized questions in a questionnaire can help capture more data that is reliable.
- Open-ended questions can be used by the analyst to learn more about feelings, opinions and general experiences of the respondent. A closed-end questionnaire

NOTES

controls the frame of reference by presenting respondents with specific responses from which they can choose.

- Structured Analysis and Design Technique (SADT) uses a graphical notation and is generally applied in information processing systems. The SADT language is known as the language of Structured Analysis (SA). SADT comprises two parts, namely, structured analysis and Design Technique (DT).
- Essential model, environmental model, behavioral model and implementation model are considered as the 'elements of structured analysis and design'.
- Data Flow Diagrams (DFDs) are used to specify how the data flows between the functions of a system. The basic purpose of DFDs is to show how the system is currently implemented.
- Decision tree is another method for defining complex relationships. A decision tree has as many branches as logical alternatives within software. It depicts the logical structure based on a stated policy.
- Decision table is an excellent tool for expressing complex logical relationships in a highly understandable and precise manner. This table is a matrix containing rows and columns that are used to define relationships.
- Structured English is used to analyse decision process. It has a restricted vocabulary and uses action-oriented verbs like add, multiply, and so on. It does not use a subject and makes use of the terms defined in data dictionary. It is a tool that enables one to state the rules precisely.
- The entire process of designing the software system is an exercise that specifies the 'what' and the 'how' of the system. This is not done in one step but is done in an iterative process that is based on 'what' the system will do as shown in the feasibility report.
- Input design is an important phase in system design phase. Businesses use the new technology to speed up the input process, reduce costs and capture data in new forms, such as digital signature.
- The file design includes the database design with reference to SDLC because in big organization the databases are required to maintain the various types of files. A file is a collection of data that can be read from or written to.
- File organization is the methodology which is applied to structured computer files. Files contain computer records which can be documents or information. It refers primarily to the logical arrangement of data which can itself be organized in a system of records with correlation between the fields or columns in a file system.
- Database is a collection of data stored for a specific purpose. Database Management System (DBMS) is a software system that allows you to define, construct and manipulate databases that are used for various purposes, such as for storing the customer and financial information of an organization.
- Database design refers to a system that stores and retrieves data systematically from the database.
- Based on the user requirements and the detailed analysis of a new system, the new system must be designed. This approach is known as is the 'phase of system designing'. It is a most crucial phase in the development of a system.

NOTES

- The implementation phase is concerned with making actual modifications in the software code, adding new features that support the specification of present software and finally installing the modified software.
- System testing is implied on a complete and integrated system to evaluate the system's compliance with its specified requirements. It uses the scope of black box testing and requires no knowledge of the inner design of the code or logic.
- The purpose of a test strategy is to clarify the major tasks and challenges of the test project. Test approach and test architecture are other terms commonly used to describe the test strategy.
- The basic goal of system control and quality assurance is to check the sample of the developed system and implement methodologies to find errors, if any, in the processed data. The type and quality of the developed system needs to be assessed, planned and implemented.
- The hardware and software selection is a very crucial and important step in SDLC phase. The success rate of system design and implementation phase depends on this factor.
- System security can be divided into different categories according to security issues, which are related to the file structure, data structure and access procedures of system.
- Recovery methods are used to recover the data infected by virus in a computer. A computer infected with virus is unsafe to use without reinstalling the operating system.
- System audit refers to a technical assessment of a system or application. Personal computers, servers and mainframes are the example of system whereas database and Web services are the examples of applications.

5.8 KEY TERMS

- **System:** A set of organized principles, tasks, activities and procedures that defines the working of a particular establishment to achieve an objective
- **System analysis:** A process in which data are collected and then interpreted to identify the problems within the system
- **Manuals:** Qualitative documents that pertains to the working of any system
- **SDLC:** A series of steps agreed upon by software developers in creating special software
- **Economic feasibility:** The measure of the effectiveness of a project or solution in terms of cost
- **Interviewing:** A common method used by system analysts for collecting data on information requirements
- **System design:** A blueprint to define the tasks that need to be performed to build a system
- **Data flow diagram:** It is a graphical representation of the flow of data through an information system which models its process aspects
- **Data dictionary:** An organized collection of all relevant system data elements

- **Decision table:** An excellent tool used to express complex logical relationships in precise manner
- **Record:** The data items related to the object
- **File:** A collection of related records
- **Database:** The upper layer in hierarchy of file structure

NOTES

5.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A system is a set of organized principles, tasks, activities and procedures that defines the working of a particular establishment to achieve an objective. A system is defined and determined by its boundaries and objectives. It is an arrangement of smaller systems in a logical order.
2. If a system neither accepts nor provides any input or output to the external environment, such a system can be called a closed system. A closed system is not affected from the surrounding environment and can perform being independent of the happenings in the environment.
3. Systems Development Life Cycle (SDLC) is a series of steps agreed upon by software developers in creating special software.
4. The three commonly used abstraction mechanisms in software design are functional abstraction, data abstraction and control abstraction.
5. System analysis is a process in which data are collected and then interpreted to identify the problems within the system. Therefore, the collected information can then be used to recommend improvements in a system. In other words, system analysis includes identification, understanding and examining a system so that the objectives of the system can be achieved.
6. Interviewing is a common method used by systems analysts for collecting data on information requirements. An interview is a directed conversation with a specific purpose that uses a 'Question and Answer' format.
7. Decision table is an excellent tool for expressing complex logical relationships in a highly understandable and precise manner. This table is a matrix containing rows and columns that are used to define relationships. A decision table consists of two parts, in which, one is STUB and the other is ENTRY.
8. The five components of system design process are output design, input design, file design, procedure design and control design.
9. The implementation phase is concerned with making actual modifications in the software code, adding new features that support the specification of present software and finally installing the modified software.
10. The term 'specification' is a description of a function in terms of a measurable output value given a specific input value under specific preconditions.
11. The purpose of a test strategy is to clarify the major tasks and challenges of the test project. Test approach and test architecture are other terms commonly used to describe the test strategy.
12. The basic goal of system control and quality assurance is to check the sample of the developed system and implement methodologies to find errors, if any, in the

NOTES

processed data. The type and quality of the developed system needs to be assessed, planned and implemented.

13. The hardware and software selection is a very crucial and important step in SDLC phase. The success rate of system design and implementation phase depends on this factor.
14. The risk analysis helps to analyse the probability of problem occurrence, cost of each possibility of system damage and the preventive measures to include as part of security plan. Two key factors of risk analysis are the value of potential loss and the probability of loss.

5.10 QUESTIONS AND EXERCISES

Short- Answer Questions

1. Write one of the characteristics of a system.
2. What are the pre-determined objectives of the system analysis?
3. Write the functions of memos.
4. What are the advantages of questionnaire?
5. Write the function of data dictionary.
6. Write the key requirements for mining by using decision tree.
7. Name the major components in a decision table.
8. What do you mean by input design of a system?
9. Write the difference between logical and physical design.
10. What are the four methods to access a file?
11. Define the term database design.
12. Name the subprocesses that comprise the process attribute.
13. How can you say that hardware and software selection plays a crucial role in SDLC?
14. What are the various system security issues?

Long-Answer Questions

1. Explain the term 'system' with the help of examples and illustrations.
2. What are the types of a system? Explain.
3. Explain the information gathering tools which are essentially required to develop a system.
4. Discuss the system development life cycle and its phases with the help of examples and illustrations.
5. Describe the steps involved in feasibility analysis.
6. Discuss the elements of structured analysis and characteristics with the help of examples and illustrations.
7. Explain the data flow diagrams with the help of examples and illustrations.
8. Discuss the rules to construct a DFD.

9. How to create a data dictionary and decision table? Explain with the help of example.
10. Describe the process of system design with the help of examples.
11. Describe the concept of sequential file organization and relative file organization with the help of illustrations.
12. Discuss the difference between flat file databases and relational databases with the help of examples.
13. How to design a good database? Explain with the help of example.
14. Discuss the testing strategies with the help of examples.
15. What are the trends in system testing? Explain.
16. Explain the concept of quality assurance with the help of illustrations.

NOTES

5.11 REFERENCES AND SUGGESTED READINGS

Kendall, Kenneth E. and Julie E. Kendall. 2010. *Systems Analysis and Design*, 8th edition. New Jersey: Prentice Hall.

Awad, E.M. 1991. *System Analysis and Design*. New Delhi: Galgotia Publications.

Date, C.J. 2006. *An Introduction to Database Management Systems*, 8th edition. New Jersey: Pearson Education (Addison–Wesley).

UNIT 6 INTRODUCTION TO E-COMMERCE

NOTES

Structure

6.0 Introduction
6.1 Objectives
6.2 Electronic Commerce (E-Commerce): Fundamentals
- 6.2.1 Processes Involved in E-Commerce
- 6.2.2 Types/Categories of E-Commerce
- 6.2.3 Requirements of an E-Commerce System
- 6.2.4 Working of E-Commerce
- 6.2.5 Lure of E-Commerce
- 6.2.6 E-Business versus E-Commerce

6.3 Encryption
- 6.3.1 Components of Cryptography
- 6.3.2 Symmetric Key Encryption
- 6.3.3 Public Key Encryption
- 6.3.4 Comparison Between Symmetric and Public Key Encryption

6.4 Authentication
- 6.4.1 Authentication in E-Commerce

6.5 Digital Signatures
- 6.5.1 Attacks on Digital Signatures
- 6.5.2 Types of Digital Signature

6.6 Key Certification
- 6.6.1 Digital Certificates

6.7 Securing Channels of Communication
- 6.7.1 Secure Socket Layer (SSL)
- 6.7.2 Secure Electronic Transaction (SET)
- 6.7.3 SSL versus SET

6.8 Cryptography
- 6.8.1 Cryptographic Algorithms

6.9 Biometrics
- 6.9.1 Factors that Affect Biometric Systems
- 6.9.2 Design Issues in Biometric Systems

6.10 Future of E-Commerce
6.11 Summing Up
6.12 Key Terms
6.13 Answers to 'Check Your Progress'
6.14 Questions and Exercises
6.15 References and Suggested Readings

6.0 INTRODUCTION

In this unit, you will learn about the significance of e-commerce. Traditionally, commerce is seen as the exchange or buying and selling of goods and services, which involves exchange of money, and sometimes transportation of goods from one place to another. We see commercial activities all around us in many forms. For instance, at a grocery store, we can see two main activities that form the essence of commerce, the buying (people purchasing goods or service) and selling (people offering goods and services to buyers). With the passage of time and advancement in technology, commerce has added

NOTES

a new dimension in the form of electronic trade, a more effective and convenient manner than the traditional commerce which is revolutionizing the way business activities are carried out across the world. In today's business world, where time and money are considered as the most important factor for success, e-commerce plays an important role in bridging the gap between buyers and sellers.

A typical e-commerce transaction involves information regarding products or services, order, delivery and payment (credit card details). All this sensitive information is transferred through the Internet, which is a public and un-trusted network. Though Internet is an efficient means for communication, it brings with itself many security risks, such as virus intrusion, hacking, password cracking, spoofing and sniffing, phishing and identity theft and Denial of Service (DoS) attacks. Thus, security is an important issue for companies doing online businesses.

There exists a variety of tools that can be used to protect online transactions. These tools include encryption, digital signatures, digital certificates, security protocols including Secure Socket Layer (SSL) and Secure Electronic Transaction (SET), firewalls. Finally, you will learn about various cryptographic algorithms and the significance of biometric systems.

6.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the fundamentals of e-commerce
- Discuss categories of e-commerce
- Describe requirements of an e-commerce system
- Explain working of e-commerce, and its advantages and disadvantages
- Understand the concept of cryptography
- Discuss the tools used in cryptography
- Explain the role and need of digital signatures and key certification
- Discuss authentication and various protocols used for implementing authentication
- Discuss protocols, such as SSL and SET, needed to implement security in e-commerce transactions
- Explain various cryptographic algorithms
- Discuss the significance of biometric systems

6.2 ELECTRONIC COMMERCE (E-COMMERCE): FUNDAMENTALS

Electronic commerce or **e-commerce** refers to the electronic means of carrying business communications and transactions over the Internet or any other form of network. It does not refer only to buying and selling of goods and services on the Internet; rather it combines the business and electronic infrastructures, allowing traditional business transactions to be conducted electronically (see Figure 6.1). Here,

NOTES

transaction refers to an exchange taking place between two entities where one sells products or services to the other.

Both traditional commerce and e-commerce are used as a means of carrying out business. The traditional commerce uses outdated methods whereas e-commerce adopts easy and fast business methods. The comparison between traditional commerce and e-commerce is listed in Table 6.1.

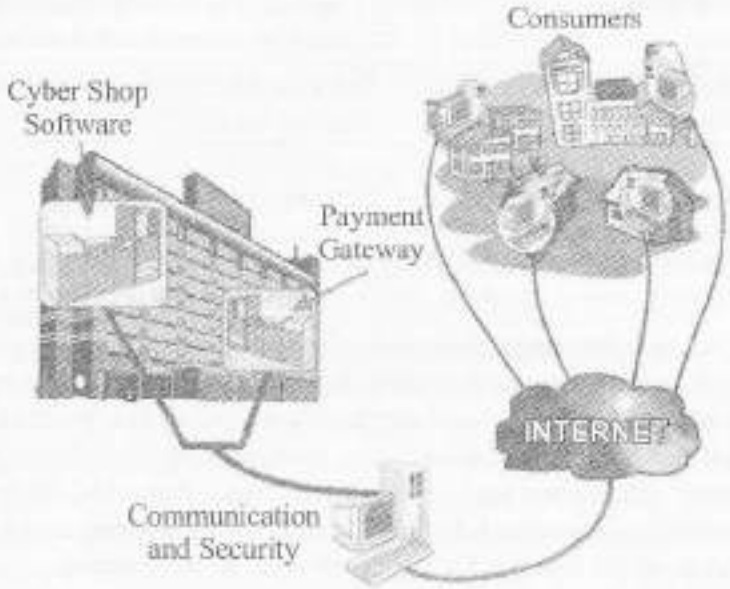


Fig. 6.1 Electronic Commerce and Business Processes

Table 6.1 Comparison between Traditional Commerce and E-Commerce

Traditional Commerce	E-Commerce
It refers to buying and selling of goods and services, involving exchange of money and transportation of goods from one place to another by physical means.	It refers to carrying business communications and transactions (selling and purchasing of products or services between two entities) using computers over Internet or any other form of network.
The customer should be physically present during the exchange of goods.	The customer needs not to be present physically.
It is very time consuming process.	It is fast and efficient process as compared to its counterpart.
The information about the products or services can be acquired through catalogs, magazines or flyers.	Product information can be acquired through Websites, such as *www.amazon.com* and *www.onlineshopping.com*.
Orders are sent or received via fax or e-mail.	Orders can be sent or received through e-mail or EDI.
It involves heavy overhead cost due to requirement of maintaining a storefront.	Overhead costs are nominal.

NOTES

Middleman, such as sales representatives need to be hired for interaction between the buyer and purchaser.	The buyer and purchaser can communicate directly with each other to carry out transactions. This eliminates the need of middleman what is known as **disintermediation**.
Businesses cannot be moved easily. It requires a huge cost.	It is easier, faster and usually less expensive to move the business and reaching the customers worldwide.
The business is confined to a small geographic area.	An e-commerce business caters to the global market.

6.2.1 Processes Involved in E-Commerce

Generally, there exist five important processes in an e-commerce lifecycle: *information sharing*, *ordering*, *online payment*, *fulfilment* and *support and services*.

- **Information Sharing:** Sharing of information on the Internet is the foremost process in e-commerce activities. It makes the customer aware of all the products and services offered by the online vendor. This particular activity is carried out by means of advertising, marketing on the Internet and through customer information gathering process. Apart from this, there are various networked communities (chat room, multi-party conference, and newsgroup) available on the Internet for the distribution of information.

  World Wide Web (WWW) provides one of the effective mediums for communication with the customers. Websites are designed to include product catalogues. One can obtain the data for products that are requested. At the same time, it also allows Web visitors to provide information about them. Apart from this, timely notification, product upgradation and new features are sent through e-mail. Large numbers of Websites aim at providing information and data to the general public, however, there exist some significant sites on the Internet that cater to the business communities. These sites are called **intermediaries** and their sole purpose lies in making buyers to interact, trade, and make purchases from the sellers.

- **Ordering:** This process involves customers to electronically place order for the company's goods and services. Placing of order requires use of client/server application where order of goods or services is placed on the electronic form placed on the company's Website.

- **Online Payment:** This process is crucial in e-commerce. The payment activity requires the use of high security to protect online money transactions from any unauthorized use. In most cases, data is transferred in encrypted form on a secured medium. Different mechanisms for payment are available on the Internet and the customers can use any of them.

- **Fulfillment:** Modern businesses are highly dependent on the massive amount of information and precious data, which is continuously being transferred over the computer network across the world. Many online trading companies make use of Internet for transferring funds, generating money, and providing information product (newsletter, articles, stock prices, reports). However, when product takes the shape of physical entity, support of EDI is required to

NOTES

clear the sale of product to the buyer with up-to-date-delivery notification issued to the buyer's e-mail.

- **Support and Services:** Support and services are regarded as one of the most important processes in e-commerce activity. The continuing relationship of an organization with customer depends on the kind of support and services provided to the customers. Relationship of an organization with the customer starts with the sale of the product. It might happen that customer will require assistance with product and services in future. Therefore, items, such as technical notes, FAQs, and support documentation should be available on the company's Website.

6.2.2 Types/Categories of E-Commerce

There exist a variety of types of e-commerce and these types can be characterized in different ways. Four major types of e-commerce are discussed as follows:

- **Business-to-Consumer (B2C):** In this type of e-commerce, the online businesses (businesses over the Internet) sell to individual consumers. It refers to indirect trading between company and consumer where the consumer can directly buy the products or services from the supplier's Website. An example of B2C type e-commerce is virtual Internet shopping malls; the customer can shop online the required items anytime instead of spending time in moving around in crowded supermarkets and have the goods delivered at their doorstep.
- **Business-to-Business (B2B):** This type of e-commerce takes place between two companies, where one sells products to another. For example, manufacturers selling products to the distributors and wholesalers selling products to retailers. It is the largest form of e-commerce that provides a way for businesses to trade with each other. B2B is one of the cost effective means of selling products worldwide.
- **Consumer-to-Consumer (C2C):** C2C system refers to e-commerce activities, which uses an auction style model. This model consists of person-to-person transaction that completely excludes businesses from the equation. Though C2C system enables customers to directly deal with each other, the transactions cannot be fulfilled without the help of use of online market maker, such as an auction site *eBay* or *Amazon*. The consumer prepares the products and puts it on the auction site for sale. The market maker is responsible for providing product catalog, search engine and transaction clearing capabilities so that the product can be displayed, searched and paid for.
- **Peer-to-Peer (P2P):** P2P system of e-commerce is similar to C2C e-commerce in the sense that the communication is directly between the two users on the Internet— no intermediate party is required. P2P is not only an e-commerce but also a technology that allows users to share computer resources and computer files over the Internet without the involvement of a central Web server. *BitTorrent* and *eDonkey* are the most common examples of P2P e-commerce which are used for downloading large video files and audio files, respectively. These two P2P programs account for approximately half of the Internet network worldwide.

NOTES

6.2.3 Requirements of an E-Commerce System

There are several basic requirements that should be met for an e-commerce system. These requirements are discussed as follows:

- **Merchant Bank Account:** During online business transactions, the payments are made generally through credit cards. In order to accept credit cards over the Internet, you need an account with a merchant account provider (referred to as **acquiring bank**). The acquiring bank is responsible for online authorization of credit cards and payment processing. Today, an endless amount of acquiring banks are available each of which charges a varying processing fee per each transaction processed. You can easily search for the desired one through the Internet and open an account with it.
- **Customer Issuing Bank:** It is a financial institution that issues a payment instrument to the customer and is responsible for debt payment of credit card's holder.
- **Credit Card Association:** It is a financial institution that offers credit card services, such as MasterCard and Visa card, discussed in Unit 3. These services are branded and distributed by the customer issuing bank.
- **Web Hosting Package:** To host your Website on the Internet, a Web hosting package is required. There are two possibilities for hosting your Website. First, you can set up or maintain your own Web server and second, you can choose some Internet Service Provider (ISP) for Web hosting administration. ISP is a company that provides Internet access and some hard drive space on their Web servers to host the Website. As it proves quite expensive to set up and maintain your own Web servers, this option is chosen by usually large organizations.
- **Shopping Cart Application:** It is the software that is set up by the merchant for displaying the products and services offered as well as for pricing, shipping charges and taxes calculations. This application executes on an **online transaction server** (provided by an **online transaction provider**) that is located elsewhere on the Internet. An online transaction provider may either sell or rent the use of an online shopping cart (**Web store**) to you for your business. An example of shopping cart application is *SoftCart* by *Mercantec*.
- **Registered Domain Name:** A domain name is a unique name that identifies a Website on the Internet and represents the name of the Web server where the Web pages reside. For example, *www.yahoo.co.in* and *www.nic.in* are some valid domain names. Before you can be online, you need to register for a domain name on the Internet through a domain name registrar. The domain name registrar offers you the available domain names from which you can choose the one you want others to use for finding your Website. Note that you also need to pay a registration fee to the registrar for keeping your Website active on the Internet.
- **Payment Gateway:** It is a service that connects merchants, customers and financial networks with one another for processing credit card authorization and payment. Usually, a third party, such as VeriSign is responsible for the operation of payment gateway.
- **Processor:** It is the large data center that is responsible for processing the credit card transactions and settling funds to merchants' accounts. It is

NOTES

connected to the merchant's site via payment gateway on the behalf of acquiring bank.

- **Digital Certificate:** A digital certificate, also referred to as **Secure Sockets Layer (SSL) server certificate**, is an electronic document that enables SSL on the Web server. It provides the basis for secure credit card transactions over the Internet by enabling each person involved in a transaction to verify the identity of others easily and quickly. Every e-commerce company that offers online Web store needs to have SSL to enable secure communication.

6.2.4 Working of E-Commerce

To understand the working of e-commerce, consider A is a customer and B is a merchant. A visits the B's Website through the Internet where he/she finds various products and services offered by B and decides to make a purchase. In a typical scenario, the following steps will be involved in making, such transaction to happen.

- A visits the Web store by clicking on a link or button, (such as Buy Now, Shop Online or image of a shopping cart button) on the Web page. As A does this, he/she gets connected to the online transaction server running the shopping cart application.
- To make a purchase, A inputs his/her credit card details.
- A's Web browser encrypts the information and sends it to the B's Web server. This is done via SSL encryption.
- From the B's Web server, the transaction information is forwarded to the payment gateway. This is also an SSL encrypted connection.
- The payment gateway then sends the transaction information to the processor used by B's acquiring bank.
- The processor forwards the transaction information to the credit card association which further sends the information to A's issuing bank.
- The issuing bank checks the authorization of credit card and sends the response (either approved or declined) back to the processor.
- The processor sends back the response to the payment gateway which further passes the response to B's Website.
- Depending on the result received, B accepts or rejects the transaction accordingly.
- After the payment has been authorized, the settlement process begins where the money is transferred from A's account to B's account. To settle the transaction, B sends a request to the payment gateway.
- The payment gateway sends details of the transaction to be settled to the processor.
- The processor further passes these details to A's issuing bank as well as to B's acquiring bank. The issuing bank admits B's charge on A's credit card statement and debts A's account, while the acquiring bank credits B's account.

6.2.5 Lure of E-Commerce

E-commerce has added important values in today's business. Some of the business values are explained below:

NOTES

- **Streamlined Processes:** E-commerce automates and streamlines business transaction processing and produces more accurate information. It also allows the companies to coordinate their operations and sales, production and distribution decisions in a better way. The result is lower overall costs.
- **Lower Prices:** With automated tools, e-commerce allows the companies to reduce warehousing costs, labour cost and administrative time as well as increase inventory turns. By reducing expenses, they maintain lower prices of products and services.
- **Lower Costs:** With electronic exchange of business documents, pre-printed forms, envelopes and postage can be eliminated. With a reduction of data entry and other manual processing, staff can be freed to do other value-added work. Expenses associated with the processing of business transactions, therefore, can be reduced. In a recent report in *BusinessWeek*, it was noted that it costs banks only a penny to process an Internet-based transaction. Comparable costs are $1.07 for a teller and 27 cents for an ATM.
- **Inventory:** E-commerce inherently requires just-in-time processing environments. Therefore, companies that need supplies no longer have to carry large safety stocks and thus inventory costs are reduced. Inventory levels are aligned with order lead-time, demand and seasonal or environmental changes. E-commerce reduces lead-time because the supplier receives purchase orders within minutes. Order data is received accurately, assuring correct stock requirements. Also, inventory adjustments for items affected by seasonal or environmental changes can be made quickly. The supplier can make the goods to order because you can place an order quickly. With less stock needed, warehousing costs are significantly reduced.
- **Accuracy:** E-commerce data is computer-generated, thus, it is received exactly and accurately. As a result, you should have fewer product returns, lost orders, backorders, and credit/debit memos. Companies can trust that when a customer's order arrives, it will be accurate, complete and timely.
- **Better Information:** New types of information stored in inventory and Web site are available to customers and suppliers. Access to real-time inventory levels allows the buyers to make wiser purchasing decisions. Customers can get the most competitive prices by searching for the same item on multiple Web sites. The information related to the capability to access sales and usage makes it possible to adjust buying decisions based on factors such as climate changes and fashion trends.
- **Improved Cash Flow:** Online payment options allow the vendors to receive payments in real time rather than days or months later. In most cases, companies use e-commerce technologies such as EDI and EFT to automate the payment process and to improve cash flow. Electronically transmitting the customer's payment remittance details to a supplier allows for quick reconciliation with their accounts receivable system. Advance knowledge of incoming funds also improves cash management and cash flow.
- **Increased Revenue Stream:** E-commerce can potentially increase revenues by offering a new way to market the existing products. In addition to traditional sales channels, companies can sell products and services

NOTES

using the Internet. E-commerce also gives companies the opportunity to sell new products. For example, in addition to buying CDs from the local music retailer, you might be able to download your favourite songs from a Web site for a fee and burn your own CD at home. E-commerce offers suppliers the opportunity to better serve their customers through newer service delivery channels. For instance, customers can check account balances on the Web.

- **New Customers:** The Web and e-commerce technologies increase the exposure to customers and provide better prospects in the company.

  E-commerce allows start-ups and small companies to compete with large organizations. By attracting new customers, they extend their reach into markets and geographic areas that were previously beyond their access. In addition, some companies gain additional exposure and sales by participating in online e-markets and auction sites.
- **Enhanced Marketing:** Although traditional marketing approaches using print, television and radio are still effective tools, the Internet offers e-commerce sellers a new opportunity to do one-on-one marketing. Efforts include opt-in e-mails where periodic promotions are sent to the buyer. For example, major airlines e-mail their weekly e-fare special or other personalized advertisement campaigns. According to the Yankee Group, Internet marketing is estimated to be 60 to 65 per cent cheaper than traditional direct-mail marketing.
- **Increased Customer Satisfaction:** E-commerce technologies provide quicker access to the information that the customers want and provide satisfaction in creating an opportunity to serve customers. Furthermore, a benefit of using Web-based, customer-service functions is that it can save anywhere from 10 to 50 per cent of the costs associated with customer service.
- **Analysis of Data for Quality Assurance:** E-commerce technologies aid the production and delivery processes. With e-commerce, a retailer can capture daily sales and return information and send it directly to corporate headquarters and the supplier to analyze trends and identify possible quality problems. Using bar coding, retailers can capture and analyze data about a product throughout its manufacturing life cycle, isolating any problem areas.
- **Fulfilment of Customer Demands**: Many companies use e-commerce to meet rising customer demands. The e-commerce system can anticipate frequently asked questions or offer product using fix-it-yourself instructions. With E-commerce, companies can go to great lengths to meet customer needs such as custom ordering online and online billing. For example, Dell encourages customers to place a custom order for a PC. GM and Ford are looking to offer customers the same sort of capability by having custom ordered cars ready in two to three days.
- **Enhanced Partnership**: E-commerce requires buyers and sellers to redefine business partnerships because the business functions of one company depend heavily on other companies. E-commerce requires trust, coordination and cooperation in order to bring companies together into true collaboration.
- **Competitive Positioning**: Some companies get on board with e-commerce

NOTES

because their competitors do. Without e-commerce accessibility, long-term customers potentially turn to the competition. Others implement e-commerce to gain a competitive advantage over a competitor. In an e-commerce world, customers and suppliers can exchange key business information quickly to engineer, manufacture, purchase and sell products more efficiently. These capabilities improve internal procedures and external relations of a company.

- **Marketing Strategies in the E-Commerce Business**: With the help of a marketing strategy, the company manages the platform on which the community exchanges take place. This way, the company is not a member of the community but simply facilitates its existence by offering a Web site for communication purposes. For the development of this concept, the main step should be an analysis of the real possibilities of creating a community via the Internet. These possibilities derive either from the particular characteristics of a product or the pre-existence of the community off-line.

6.2.6 E-Business versus E-Commerce

E-Business and e-commerce are closely related, have similar meanings and can be used in place of each other. Both these terms are widely used to refer to commercial activities conducted over the Web. E-Commerce is a more general term of which e-business is a part. Table 6.2 lists the differences between e-business and e-commerce.

Table 6.2 Differences Between e-Business and e-Comerce

S. No.	E-Business	E-Commerce
1.	Explicitly involves business transactions.	Involves all types of commerce.
2.	It is a closed system.	It is an open system.
3.	It is secured.	It is not much secured.
4.	E-Business uses concept of EDI.	E-Commerce does not involve the concept of EDI.
5.	E-Business operates on Intranet or Extranet.	E-Commerce always operates on Internet.
6.	E-Business is used for bulky transactions.	E-Commerce is used for small and large transactions.

6.3 ENCRYPTION

The communication over the Internet (online communication) can be secured by adopting certain security mechanism. The most basic of these mechanisms is the cryptography. The term cryptography is derived from a Greek word *kryptos* which means "secret writing". In simple terms, **cryptography** is the process of altering messages in a way that their meaning is hidden from the adversaries who might intercept them. It allows a sender to disguise a message to prevent it from being read or altered by intruder as well as it enables receiver to recover the original message from disguised one.

Cryptography relies upon two basic components: an algorithm (or cryptographic methodology) and a key. **Algorithms** are the complex mathematical formulae and **keys**

are the strings of bits. For two parties to communicate over a network (Internet), they must use the same algorithm (or algorithms that are designed to work together). In some cases, they must also use the same key.

NOTES

6.3.1 Components of Cryptography

Cryptography allows a sender to disguise a message to prevent it from being read or altered by intruder as well as it enables receiver to recover the original message from disguised one.

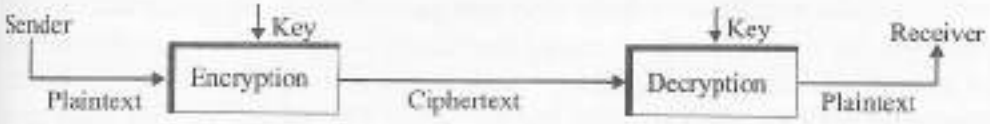


Fig. 6.2 Cryptography Components

Various components are involved in cryptography (see Figure 6.2) which are described as follows:

- **Plaintext:** It refers to the original unencrypted message that the sender wishes to send.
- **Ciphertext:** It refers to the encrypted message that is received by the receiver.
- **Encryption:** It is the process of encrypting the plaintext so that ciphertext can be produced. The plaintext is transformed to ciphertext using the encryption algorithm.
- **Decryption:** It is opposite of the encryption process. In this process, the ciphertext is converted back to plaintext using a decryption algorithm.
- **Ciphers:** The encryption and decryption algorithms are together known as ciphers. Ciphers need not necessarily be unique for each communicating pair; rather a single cipher can be used for communication between multiple pairs of sender and receiver.
- **Key:** A key is usually a number or a set of numbers on which the cipher operates. Encryption and decryption algorithms make use of a key to encrypt or decrypt messages respectively. At the sender's end, the encryption algorithm and encryption key are required to convert the plaintext to ciphertext. At the receiver's end, decryption algorithm uses the decryption key to convert ciphetrext back to the plaintext. The longer the key, the harder it is for an intruder to decrypt the message.

6.3.2 Symmetric Key Encryption

Traditionally, cryptography involved the use of same key for encrypting or decrypting the messages. This is referred to as **symmetric key (private key or secret key) encryption**. A single shared key (secret key) is used for both encryption and decryption of data. Thus, it is obvious that the key must be known to both the sender and the receiver. As shown in Figure 6.3, the sender uses the shared key and the encryption algorithm to transform the plaintext into ciphertext. The ciphertext is then sent to the receiver via a communication network. The receiver applies the same key and the decryption algorithm to decrypt the ciphertext and recover the plaintext.

NOTES

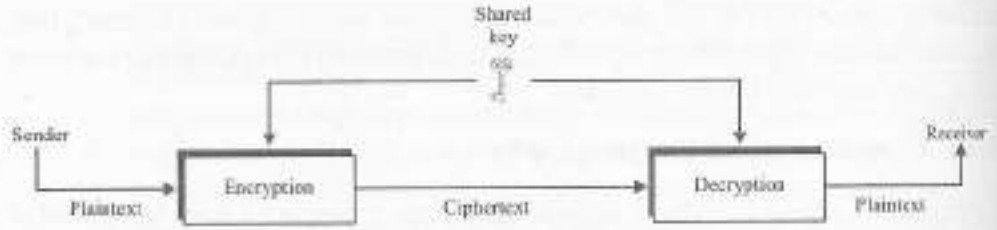


Fig. 6.3 Message Exchange using Secret Key

The main problem in symmetric key encryption is getting the sender and receiver to agree on the secret key without anyone else finding it out. If the key is compromised, the security offered by symmetric key encryption is severely reduced or eliminated. Symmetric key encryption assumes that the parties who share a key rely upon each other not to disclose the key and protect it against modification. If they are in separate physical locations, they must trust on a medium, such as courier or a phone system, to prevent the disclosure of the secret key. Anyone who overhears or intercepts the key in transit can later read, modify, and forge all messages encrypted or authenticated using that key.

The most common example of symmetric key algorithm is Data Encryption Standard (DES) that was first published in 1977 by National Institute of Standards and Technology (NIST). DES uses a 56-bit encryption to encrypt a 64-bit block of plaintext, producing a 64-bit ciphertext block. An improved and more secure version of DES is **triple-DES** that performs encryption in three stages and a different key (of 56 bits) is used for encryption/decryption in each stage. The most commonly used symmetric key algorithm today is Advanced Encryption Standard (AES) which is potentially, the most secure encryption method published by NIST. AES offers three different key sizes: 128, 192 and 256 bits.

6.3.3 Public Key Encryption

Modern cryptography involves the use of different keys for encryption and decryption. This is referred to as **public key (asymmetric key) encryption**. It was introduced by Diffie and Hellman in 1976 to overcome the problem found in symmetric key encryption. It involves the use of two different keys for encryption and decryption. These two keys are referred to as the **public key** (used for encryption) and the **private key** (used for decryption). Each authorized user has a pair of public and private keys. The public key of each user is known to everyone, whereas, the private key is known to its owner only.

Now suppose that a user A wants to transfer some information to user B securely. The user A encrypts the data by using public key of B and sends the encrypted message to B. On receiving encrypted message, B decrypts it by using his private key. Since decryption process requires private key of user B, which is only known to B, the information is transferred securely. Figure 6.4 illustrates the whole process.

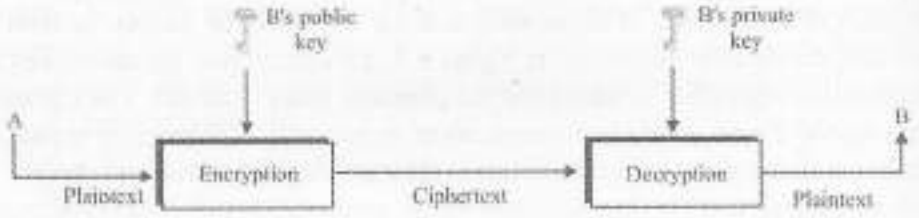


Fig. 6.4 Message Exchange using Public Key

NOTES

The main advantage of public key encryption is that the need for the sender and receiver to share the secret key is eliminated and all communications involve only public keys. Thus, no private key is ever transmitted or shared. Anyone can send a confidential message using the public key, but the message can only be decrypted with a private key, which is in the sole possession of the intended recipient.

A well known example of public key algorithm is RSA, the name derived from the initials of its three discoverers Ron Rivest, Adi Shamir, and Len Adleman. The algorithm requires keys of at least 1024 bits for good security.

6.3.3 Comparison Between Symmetric and Public Key Encryption

Some differences between symmetric and public encryption are listed in the Table 6.3.

Table 6.3 Differences between Symmetric and Public Key Encryption

Symmetric Key Encryption	Public Key Encryption
1. It uses a single key for both encryption and decryption of data.	1. It uses two different keys—public key for encryption and private key for decryption.
2. Both the communicating parties share the same algorithm and the key.	2. Both the communicating parties should have at least one of the matched pair of keys.
3. The process of encryption and decryption is very fast.	3. Encryption and decryption process is slower as compared to symmetric key encryption.
4. Key distribution is a big problem.	4. Key distribution is not a problem.
5. The size of encrypted text is usually same or less than the original text.	5. The size of encrypted text is usually more than the size of the original text.
6. It can only be used for confidentiality, that is, only for encryption and decryption of data.	6. It can be used for confidentiality of data as well as for integrity and non-repudiation checks (that is, for digital signatures).

6.4 AUTHENTICATION

As already discussed, the process of verifying the identity of a user is termed as **authentication**. In other words, it refers to establishing validity of the claim. The vendor or the dealer that sells the products or services needs to use certain authentication mechanism for recognizing the genuine buyers. Authentication presents several challenges, such as collecting authentication data, transmitting the data securely, and identifying the same person who was earlier authenticated and is still using the computer system.

Kerberos Protocol

Kerberos is an authentication protocol which has been designed to be used in an open distributed environment, where the users at workstations request for the services on the servers distributed throughout the network. It basically provides a centralized authentication server which is responsible for authenticating users to servers and

NOTES

servers to users. Its name has been derived from Greek mythology after the name of a three-headed dog that was used as a guard at the gates of Hades. Kerberos has become very popular as it can also acts as a KDC in addition to an authentication protocol. It was originally developed as a part of Project Athena at MIT. Several versions of Kerberos have been evolved. However, the most commonly used version of Kerberos is version 4. The Kerberos protocol has four components, namely, *client workstation, Authentication Server (AS), Ticket-Granting Server (TGS)* and *real server*.

- Client workstation is an entity that wants to access services from a server.
- Authentication Server (AS) acts as a KDC in the Kerberos protocol. It is responsible for the verification of identity of users during login. To get verified, all the users first need to register themselves with the AS. Each registered user is assigned a unique ID and a password. All the IDs and their corresponding passwords are stored in a centralized database of AS. In addition to verifying the users, it also issues a session key to the requesting user which is to be used between the user and the TGS, and sends a ticket for the TGS. AS also shares a unique secret key with each server, and hence, all the servers also need to register themselves with AS.
- Ticket-Granting Server (TGS) is responsible for issuing tickets for establishing connection with the real server. It also issues the session key, which is to be used between the user and the real server. Once a user has got verified by AS, he/she can contact TGS any number of times to obtain the tickets for different real servers.
- Real server provides the required services to the user. Here, the user uses its own process (client process) to access the server process using a client-server program (for example FTP).

The working of Kerberos is shown in Figure 6.5. To understand how Kerberos protocol operates, consider a user A that wants to access the services of a real server B. The user A can access the processes running on B by performing the following steps.

Authentication Procedures of X.509

The X.509 standard defines three authentication procedures, namely, *one-way authentication, two-way authentication* and *three-way authentication*. These authentication procedures make use of public key signatures and can be used across a variety of applications. While working with these procedures, it is assumed that two communicating nodes know each other's public key. These procedures are discussed as follows.

- **One-Way Authentication:** It involves single transfer of information from user A to user B [see Figure 6.5 (a)]. This information is transferred in order to confirm that the message is generated by user A, and that the message is intended for user B and has not been altered or replayed by any intruder. Thus, this procedure only verifies the identity of A. At a minimum, the message includes the following and is signed with private key of A.
  - **Timestamp (t_A):** It indicates the time when the message was generated along with the validity or expiration time of the message.
  - **Nonce (r_A):** It is a random number used to protect the message from replay attacks. The value of nonce must be unique during the validation

NOTES

period. This is done so that B can discard any new message having the same nonce value.

- **Identity of B (ID_B):** It is included so that the user B can access the X.509 directory and be sure that the message is intended for him/her only.
- **Signed Data (sgnData):** It is the information which is actually conveyed between the two users. This information is included within the scope of the signature thus, ensuring its authenticity and integrity.
- **Encrypted Data ():** It includes the session key K_{ss}, encrypted with B's public key (B_{pu}). This key is used after the authentication process gets over.

- **Two-Way Authentication:** It involves two-way transfer of information; first from user A to B, and then from user B to A. This two-way transfer allows both users to verify each other's identity. First, A sends a message containing the same fields as described in one-way authentication to B. Then, B replies to A with a message including t_B, r_B, ID_A, r_A, sgnData and optionally and the message is signed by B [see Figure 6.5 (b)]. The inclusion of nonce r_A in the reply message ensures that the reply is valid and has come from B, whereas r_B protects the information being transferred by B from replay attacks.
- **Three-Way Authentication:** It involves three-way transfer of information from A to B. In this procedure, a third message from A to B is included in addition to two messages which are same as described in two-way authentication. The third message contains a signed copy of the nonce r_B signed by A [see Figure 6.5 (c)]. Since both the nonces are echoed back by the other side, each user can check the returned nonce to detect replay attacks. This approach is useful when synchronized clocks are not available.

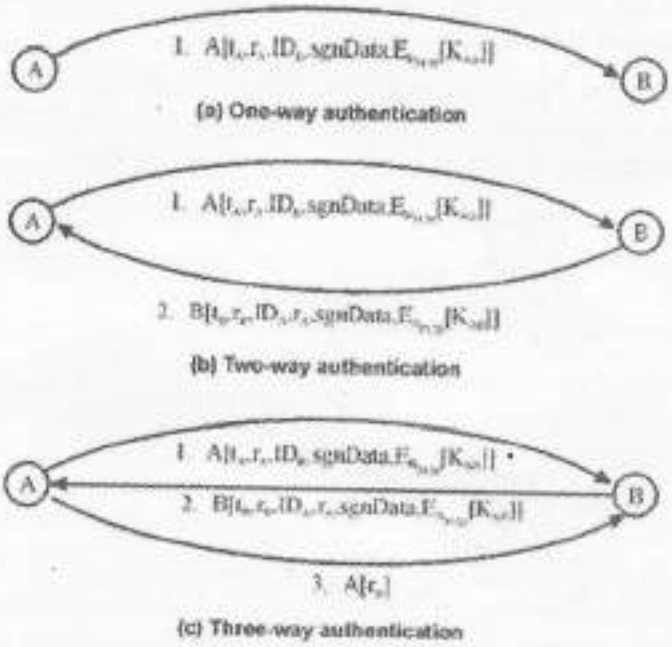


Fig. 6.5 Authentication Procedures

NOTES

6.4.1 Authentication in E-Commerce

An eCommerce site is usually made up of several integral components, including the normal network components such as routers, hubs, and switches. There are other lesser acquainted components: firewalls, IDSs, Web servers, load balancers, database servers, and financial processing servers.

1. **Firewall:** A firewall is a device used to provide access controls for a network or segment. Think of this system as a network traffic cop, allowing or disallowing traffic into a network based on who the requestor is and the type of connection they are asking for. Figure 6.6 shows the basic firewall.

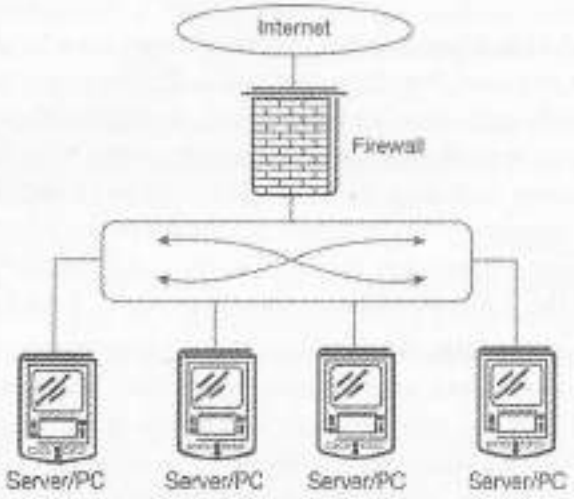


Fig. 6.6 Firewall Model

2. **Intrusion Detection System:** An IDS can be network based or host based, or both. These tools are very flexible; they can monitor and manage data and make content filtering decisions.
3. **Web Servers:** This is the most common server in an eCommerce site. This system's job is to serve up the Web pages or content that the consumers using your site request. Figure 6.7 is basic web server example.

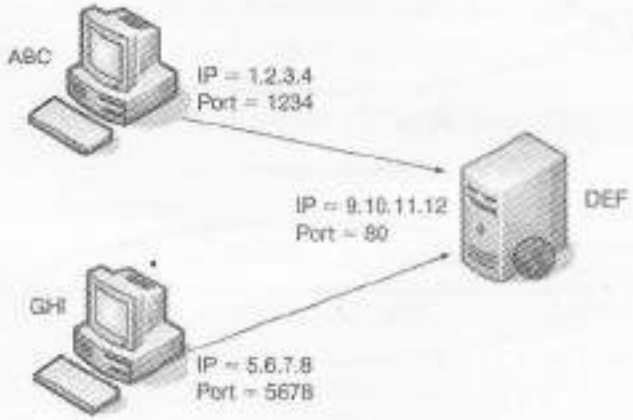


Fig. 6.7 A Web Server

4. **Load Balancers:** These specialized devices are used to regulate the traffic flow to the web servers, ensuring that the workload is balanced between the multiple systems that perform the work of the desired site.

NOTES

5. **Database Servers:** These systems are used to store the information the site depends on for business, including catalogs, product descriptions, consumer data, and all the other bits of information that are needed to do business. If these servers have consumer information on them, they must be protected even more carefully than systems just serving your site's data to the Web. Figure 6.8 shows Database server connectivity.

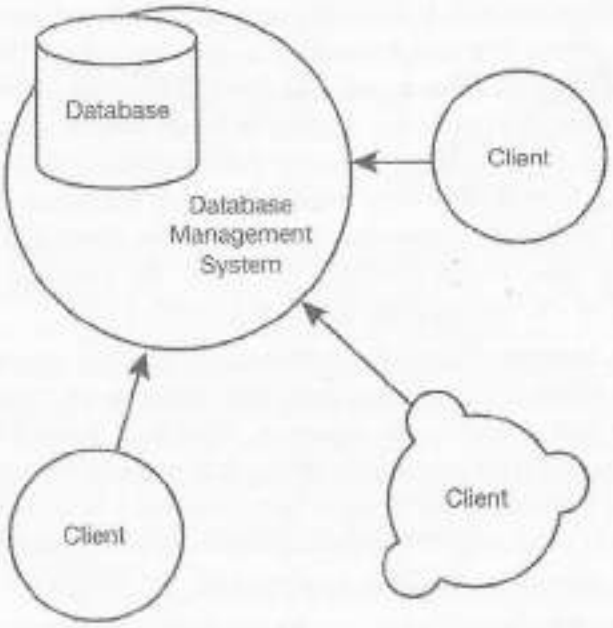


Fig. 6.8 Database Connectivity

6. **Financial Processing Servers:** These servers are used to store and process customer and vendor financial information. They are often the end-line goal of most attackers, so they must be given utmost care of any of the systems on your network.

Losing the information in these servers could spell the doom of business, so they must be treated with caution. A site may have additional components, or redundant sets of these types of devices, but these are the basic commonalities across the board.

6.5 DIGITAL SIGNATURES

The historical legal concept of 'signature' is defined as any mark made with the intention of authenticating the marked document. The authentication mechanism used to verify the authenticity of electronic documents is referred to as **digital signature**. In other words, digital signatures play the role of physical signatures in verifying electronic documents.

Digital signature allows the sender to attach an electronic code with the message in order to ensure its authenticity and integrity. This electronic code acts as a signature of the sender and hence, named digital signature. Digital signatures use public key cryptography technique. The sender uses his private key and a signing algorithm to create a digital signature, and the signed document can be made public. The receiver, on the other hand, uses the public key of the sender and a verifying algorithm to verify the digital signature.

Check Your Progress

1. Define the term e-commerce.
2. Give an example of B2C e-commerce.
3. What is a payment gateway?
4. Define the term encryption.
5. Name the most common example of symmetric key algorithm.
6. What do you understand by the term authentication?

NOTES

Digital signatures are analogous to physical handwritten signatures and provid the following security services.

- **Message Authentication:** A normal message authentication scheme protects the two communicating parties from the attack of the third party (intruder). However, a secure digital signature scheme protects the two parties against each other also. Suppose A wants to send his signed message (message with A's digital signature) to B through a network. For this, A encrypts the message using his private key, which results in a signed message. The signed message is then sent through the network to B. Now, B attempts to decrypt the received message using A's public key in order to verify that the received message has really come from A. If the message gets decrypted, B can believe that the message is from A. However, if the message or the digital signature has been modified during the transmission, it cannot be decrypted using A's public key. From here, B can conclude that either the message transmission has tampered or the message has not been generated by A.
- **Message Integrity:** Digital signatures also provide message integrity. If a message bears a digital signature, then any change in the message after signature will invalidate the signature. That is, it is not possible to get the same signature if the message is changed. Moreover, there is no efficient way to modify a message and its signature, such that a new message with a valid signature is produced. These days a hash function is used in the signing and verifying algorithms that help in preserving the integrity of the message.
- **Non-Repudiation:** Digital signatures also ensure non-repudiation. For example, if A has sent a signed message to B, then in future A cannot deny about the sending of the message. B can keep a copy of the message along with A's signature. In case A denies, B can use A's public key to generate the original message. If the newly created message is same as that of message initially sent by A, it is proved that the message has been sent by A only. In the same way, B can never create a forged message bearing A's digital signature because only A can create his digital signatures with the help of his private key.
- **Message Confidentiality:** Digital signatures do not provide message confidentiality directly because anyone knowing the sender's public key can decrypt the message. Thus, to achieve message confidentiality, we need to encrypt the message along with the signature using either secret key encryption or public key encryption scheme. For example, if we use public-key encryption scheme, then at A's end, first the message is encrypted using A's private key and then a second encryption is performed using the B's public key. Similarly, at B's end, first the message is decrypted using B's private key and then a second decryption is performed using A's public key. With this mechanism, only B can decrypt the encrypted message received from A because only he knows his own private key.

Digital Signature Process

The digital signature process is shown in Figure 6.9. Suppose user A wants to send his signed message to B through network. To achieve this communication, these steps are followed.

NOTES

(i) A uses his private key (E_A), applied to a signing algorithm, to sign the message (M).

(ii) The message (M) along with A's digital signature (S) is sent to B.

(iii) On receiving the message (M) and the signature (S), B uses A's public key (D_A), applied to the verifying algorithm, to verify the authenticity of the message. If the message is authentic, B accepts the message; otherwise he rejects it.

Fig. 6.9 Digital Signature Process

The private and public keys used in digital signatures are different from the private and public keys used in public-key encryption. In the former case, the private and public keys of the sender are used to create and verify the digital signature, respectively. However, in public-key encryption, the public and private keys of the receiver are used for encryption and decryption of the message, respectively. Digital signatures ensure message integrity, authentication and non-repudiation. They do not ensure message confidentiality directly. For this, they must be used along with encryption.

Properties and Requirements of Digital Signatures

A digital signature is used in those situations where there is a lack of trust between the sender and receiver. For example, suppose a user A transfers funds to B electronically. Now, B in future increases the amount of funds transferred and claims that the larger amount had arrived from A. Thus, to achieve secure communication between the two users and to resolve their disputes if any, the digital signature must have the following properties.

- It must be able to verify the author and the date and time of the signature.
- It must be able to authenticate the contents of the message at the time of the signature.
- There must be some third (trusted) party that can verify the digital signature to resolve disputes between the sender and receiver.

Thus, we can say that the authentication function is included within the digital signature function. Based on the above-mentioned properties, we can devise the following requirements for a digital signature.

- The signature must be in a form of a bit pattern and relative to the message being signed.
- The signature must contain some information that is unique to the sender so that forgery and denial can be avoided.
- The process of creating the digital signature must be comparatively easy.
- The process of recognizing and verifying the digital signature must also be comparatively easy.
- A high computational effort must be required to forge a digital signature. That is, it must be infeasible for an intruder to create a new message for an existing signature or to create a fake digital signature for an existing message.
- The copy of a digital signature must be retained in some storage mechanism.

NOTES

6.5.1 Attacks on Digital Signatures

There are three kinds of attacks possible on digital signatures, namely, *key-only*, *known message* and *chosen-message attack*. To get idea of these attacks, consider A and B as the communicating parties and C as the opponent.

- **Key-Only Attack:** In this type of attack, the intruder knows only the public information of A. With the help of this information, the intruder tries to create A's digital signature. He then forges the message by putting newly created A's signature and sends the signed message to B pretending that it has come from A.
- **Known-Message Attack:** In this type of attack, the intruder has access to some messages signed by A. That is, he already has one or more message-signature pairs. With the help of these known message-signature pairs, the intruder tries to create another message and forge A's digital signature on it.
- **Chosen-Message Attack:** In this type of attack, the intruder manages to let A sign one or more messages for him. In other words, the intruder himself chooses a message-signature pair. Later, he creates a new message with the contents he wants and forges A's digital signature on it, and sends it to B pretending that it has come from A.

6.5.2 Types of Digital Signature

The digital signature scheme has got many variations and additions to its main concept. Some of the variants of digital signature are discussed as follows:

- **Timestamped Signature:** Some digital signatures include a timestamp value in order to prevent the replay attacks. This is what we call as timestamped signatures. In a replay attack, the documents can be replayed by a third party. For example, if A signs a request to the bank C to transfer a certain amount of money to B, B can intercept the document and replay it if there is no timestamp on the document. One way to handle, such problems is to include the actual date and time. However, this may create a problem if the time zones are different or the clocks are not synchronized. Another solution is to use a nonce, (which is a randomly generated number that can be used only once). The receiver makes a note that the particular nonce is used by the sender and cannot be used again. That is, a new nonce defines the present time and a used nonce defines the past time.
- **Blind Signature:** Blind signature scheme was developed by David Chaum. The concept of blind signature is used when the sender does not want to reveal the contents of the message to the signer and just get it signed by him. Blind signatures are typically used in those situations where the signer and the message author are completely different parties. Examples include electronic voting systems and electronic payment systems. The basic idea behind blind signatures is that first the sender A creates a message and blinds it. He then sends the blinded message to the signer B. B signs the blinded message and returns the signature on the blinded message to A. A unblinds the signature to obtain a signature on the original message.

  In the blind signature scheme, there are more chances of fraud. That is, it is possible that A gets a blind message signed by B that may later hurt him/her. For example, A may get a sign from B on a document that contains B's will. To deal

NOTES

with, such type of situations, some laws have been passed that protect B in case he has signed a blind document that is against his interest.

- **Undeniable Digital Signature:** The digital signature schemes discussed so far consist of signing algorithm and verifying algorithm. That is, a signer can only create his signature using his private key, and anyone having the signer's public key can verify the signatures. Thus, the verification can be performed without the signer's consent or involvement. In some cases, a signer may not like that anyone could verify his signatures. Thus, in order to increase the privacy of the signer, another scheme called undeniable digital signature scheme was proposed by David Chaum and Hans van Antwerpen in 1989. An **undeniable digital signature scheme** is a non-self-authenticating signature scheme in which no signatures can be verified without the signer's cooperation and notification.

The main advantage of this scheme is that if a signature is invalid, a fraud signer cannot prove it valid because he will not be able to successfully complete the verification protocol. In the same way, a fraud signer cannot deny a signature which is valid as he will not be able to successfully complete the disavowal protocol.

6.6 KEY CERTIFICATION

In e-commerce world, one of the major problems that often occur is how one can be sure that the person sending an encrypted message is really who he/she claims to be. Such assurance can be provided in the digital world by certifying the public key of the users with the help of public key (digital) certificates.

6.6.1 Digital Certificates

A digital certificate is an electronic document that signifies the association between the user and his/her public key. It is digitally signed and issued by a trusted third party, referred to as **Certificate Authority (CA)**. Every user who wants to send an encrypted message applies for a digital certificate from a CA by presenting his/her public key. In response, the CA issues an encrypted digital certificate containing the name of the user, his/her public key, a digital certificate number, issuance date, expiration date, the digital signature of CA and other identification information. The CA also makes its public key readily available on the Internet or via other means.

After obtaining a certificate from CA, the user can publish his/her certificate. The recipient of the encrypted message decodes the attached digital certificate using the public key of CA to verify that the certificate has been generated by the CA only and is not a fake certificate. Then, the recipient can obtain the public key of the sender and other identification information held within the certificate to verify the identity of the sender. This way the digital certificates enable the e-commerce participants to validate each other's identities at the point of purchase, thus, facilitating secure online transactions.

X.509 Standard

X.509 was designed by ITU-T to describe the public key certificates in a structured way. It is a part of X.500 series of recommendations that define a directory service.

NOTES

A **directory** is basically a server or a set of servers distributed over a network that maintains a database of information about the users, such as mappings from user names to network addresses. X.509 provides authentication services to its users through X.500 directory that may act as the repository of public-key certificates associated with each user. Each certificate contains the public key of the user and is signed by the private key of the trusted Certification Authority (CA). Note that only CA can create the user certificates and place them in the directory. The directory server itself is not responsible for creating public keys or certificates; it simply acts as a warehouse of certificates for the users.

X.509 also defines a set of alternative authentication protocols based on the use of public key certificates. The format of the certificate and the authentication protocols defined in X.509 are used in several contexts, such as IP Security, S/MIME, SSL/TLS and SET, and hence it is considered to be an important standard.

The first version of X.509 was issued in 1988. Then, in 1993 its second version came out to address some of the security concerns. A third version was released in 1995, which was revised in 2000. X.509 makes uses of ASN.1 (Abstract Syntax Notation One) protocol, which defines the general format of the X.509 certificate with several fields as shown in Figure 6.10. The description of these fields is as follows:

- **Version Number:** This field differentiates among the various versions of certificate format. That is, the version of X.509 is specified in this field. The first version number was 0 and the current version (third version) is 2.
- **Serial Number:** This field contains an integer value assigned to each certificate. This value is unique within the issuing CA.
- **Signature Algorithm ID:** This field specifies the algorithm used by CA to sign the certificate, together with any associated parameters.
- **Issuer name:** This field specifies the name of the CA that has created and signed the certificate. The name is usually specified as hierarchy of strings which include the name of the country, state, organization, department, and so on.
- **Validity Period:** This field specifies the lifetime of the certificate. It includes the earliest time and latest time; the certificate is invalid before the earliest time and after the latest time.
- **Subject Name:** This field specifies the entity to which this certificate refers. That is, this certificate certifies the public key of the user who is holding the corresponding private key. This field is also specified as a hierarchy of strings; one of the strings is called the *common name*, which is the actual name of the user.
- **Subject Public Key:** This field specifies the main content of the certificate, which is the public key. It also specifies the corresponding public-key algorithm and its associated parameters.
- **Issuer Unique Identifier:** This is an optional field which specifies the unique identifier of the authority issuing the certificate. This field allows having the same value in the *issuer name* field of two different certificates provided they have different values in the *issuer unique identifier* field.
- **Subject Unique Identifier:** It is also an optional field which specifies the unique identifier of the subject to which the certificate issued. This field allows having

NOTES

the same value in the *subject name* field of two different certificates provided they have different values in the *subject unique identifier* field.

- **Extensions:** This field is used by the issuers to add more private information to the certificate. The extensions were added in the third version of X.509. This is also an optional field.
- **Signature:** This field is divided into three sections as discussed below:
  - First section includes all other fields in the certificate.
  - Second section contains the digest of the first section which is encrypted with the private key of CA.
- Third section includes the algorithm identifier which has been used to create the second section.

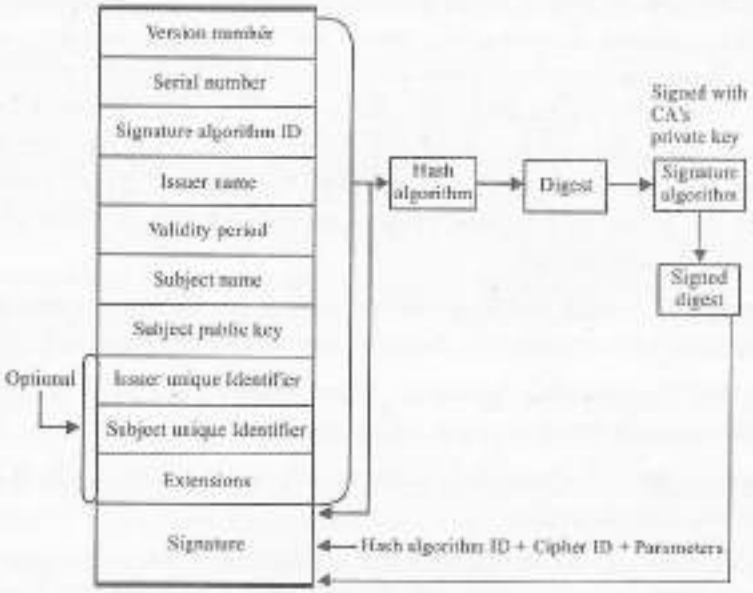


Fig. 6.10 X.509 Certificate Format

The X.509 certificates (see Figure 6.10) have a validity associated with them. If everything goes on smoothly, then the certificates can be renewed by the CA. A new certificate is issued before the validity of the old one gets over. This process is known as **certificate renewal**. However, sometimes due to some reasons the certificate has to be revoked before the validity of the certificate expires. This is what is known as **certificate revocation**. The certificates can be revoked due to several reasons. Some of them are as follows:

- The private key of the user might have been compromised.
- The private key of the CA which can verify the certificates might have been compromised. In this case, the CA must revoke all unexpired certificates.
- The certificate authority is no longer willing to issue certificate to the user

6.7 SECURING CHANNELS OF COMMUNICATION

For ensuring secure online transactions, an important concern is the security of channel via which the communication is being taken place. For this purpose, two protocols are commonly used, which are discussed in this section.

NOTES

6.7.1 Secure Socket Layer (SSL)

SSL protocol was developed by Netscape Corporation in 1994 to provide exchange of information between a Web browser and a Web server in a secured manner. Like other protocols, its main aim is to provide entity authentication, message integrity and confidentiality. SSL is an additional layer located between the application layer and the transport layer of the TCP/IP protocol suite. All the major Web browsers support SSL. It comes in three versions: 2, 3, and 3.1. Among these, version 3 is the most popular version which was released in 1995.

To start with, the Web browser (client) sends a request for secure session to the merchant's server. If the server can grant the request, the session ID and method of encryption to be used are negotiated between the two. The client and server exchange their digital certificates to establish one another's identity. The client generates a **session key**—a unique symmetric encryption key chosen for a single secure session—and creates a **digital envelope** (the container of session key) encrypted using the public key of the server. Then, it sends the digital envelope to the server. The server decrypts the envelope using its private key to retrieve the session key. After the session key has been exchanged, the encrypted transmission using the session key begins; the URL of the documents requested by the client, its contents, cookies exchanged and contents of forms are encrypted using the session key.

The term SSL session should not be confused with a SSL connection, as there are certain differences between the two. Some of these differences are as follows:

- A session is established between a client and a server, while a connection is established between two peers having equal roles.
- A session can consist of many connections, while a connection is associated with only one session.
- A connection can be terminated and reestablished within the same session. When a connection is terminated, the session between the two parties may or may not be terminated. However, when the session is terminated all the connections of that session also terminate. A session may be suspended or resumed again.
- Data can be exchanged only when the connection between the two parties is established; mere establishment of session is not enough for exchange of data.
- To create a new session, the communicating parties have to go through a negotiation process. But, to create a new connection within the same session, the negotiation process can be skipped.

6.7.2 Secure Electronic Transaction (SET)

SET protocol is used for secured credit card transactions over the Internet. SET itself is not a payment system; rather it contains the security protocols and formats which are used to provide security to the credit card payments made by a user on a public network, such as Internet. The SET concept was started as early as the year 1996, but the first product came to be available in the year 1998. SET mainly provides three services which are as follows:

- It provides a secured medium of communication for all the parties involved in a communication.

NOTES

- It provides trust by the use of X.509v3 digital certificates.
- It has got complete privacy as the parties involved in the communication can access the information only when and where necessary.

To understand the working of SET, it is important to know the participants of the SET system. The main participants of SET system are:

- **Cardholder:** The person who owns the card (such as Mastercard, Visa) and does the purchasing on the Internet.
- **Merchant:** The organization or the individual who sells goods or offers services to the cardholder. Usually, these goods and services are offered via Website.
- **Issuer:** The financial institutions, (such as banks) that issue the credit cards and are responsible for the payment of purchases made by the cardholder.
- **Acquirer:** It is also a financial institution that is responsible for establishing an account with the merchant and processing payment card authorizations and payments. Generally, merchant accepts many credit cards of different banks. Since, the merchant cannot deal with all the banks, or it cannot have account in all the banks, therefore, it needs an acquirer who provides authorization to the merchant that the card account is valid and active. The acquirer also transfers the payments electronically to the merchant's account. The issuer subsequently reimburses the acquirer over some kind of payment network for electronic funds transfer.
- **Acquirer Payment Gateway:** It is an interface between SET and the computer networks of banks for authorization and payment functions.
- **Certification Authority (CA):** It is an organization which offers different classes of certificates for cardholders, merchants, and payment gateways.

Working of SET

Before using SET, both the cardholder and the merchant must register with the CA. After the registration process, the working of SET involves many steps which are as follows:

1. The customer browses the merchant's Website to determine the products offered by the merchant. He/she then selects the products to be purchased and adds them to the shopping cart.
2. The customer then uses a single message to communicate with the merchant and payment gateway. The message has two parts, namely, *purchase order* which is used by the merchant and *card information* which is used by the merchant's bank (acquirer).
3. The card information is then forwarded to the acquirer.
4. The acquirer contacts the issuer and checks about payment authorization.
5. If the purchase is authorized, the issuer sends the authorization to acquirer.
6. The copy of the authorization is also forwarded to the merchant.
7. The merchant completes the order and informs the customer about it.
8. Merchant captures the transaction from its bank.
9. Finally, the credit card invoice is printed by the issuer and provided to the customer.

NOTES

6.7.3 SSL versus SET

SSL and SET are both Internet security protocols. However, they differ in some aspects as SET is more secured than SSL. The authentication mechanism used in SET is very complex, which makes it almost impossible for both the sender and receiver to commit any kind of fraud. In nutshell, SET has been specifically designed for secured e-commerce transactions involving online payments, while SSL has been designed only for exchanging messages over the Internet. The main differences between SSL and SET are listed in the Table 6.4.

Table 6.4 Differences between SSL and SET

Issue	SSL	SET
Main Objective	To allow exchange of data in an encrypted form.	To support e-commerce related payment mechanism.
Certification	The certificates are exchanged between the two parties	A trusted third party certifies all the parties involved in the communication process.
Authentication	The authentication mechanism is not very strong.	The authentication mechanism is very strong.
Risk of Merchant Fraud	It is prone to merchant fraud as financial data is provided to the merchant.	It is free from this fraud as financial data is given to the payment gateway only.
Risk of Customer Fraud	It is prone to this kind of fraud as customer can refuse to pay later; there is no mechanism that can prevent, such kind of fraud.	The payment instructions are digitally signed by the customer. Thus, there is less chance of such fraud.
Action in Case of Customer Fraud	Merchant is responsible if a customer later refuses to pay.	Payment gateway is responsible in case of customer fraud
Practical Usage	High	Less

6.8 CRYPTOGRAPHY

Cryptography can be defined as the conversion of data into a scrambled code that can be deciphered and sent across a public or private network, without unauthorized access to it.

In simple words Cryptography is the conversion of information from a readable state to cypertext (nonsense), where, the sender retains the ability to decrypt the information and therefore avoids unwanted persons being able to read it. It is the art of using and building cryptosystems. A cryptosystem disguises messages so that only intended people can see through the mask. Messages that are directly readable by humans are called plaintext. Cryptosystems can be simple or robust, but each one relies on a protocol or a set of well defined rules that enable it to operate. 'Cryptanalysis' is also used to refer to any attempt to circumvent the security of other types of cryptographic algorithms and protocols in general, and not just encryption. Encryption is the act of passing a

plaintext message through a mathematical formula to create ciphertext. Decryption turns ciphertext messages back into plaintext messages.

How Cryptography is Done?

NOTES

For most information technology occupations, knowledge of cryptography is a very small part of a broader skill set, and is generally limited to relevant application. One can simply say that this is why the Internet is so extraordinarily plagued with so many security breaches. The majority of IT administrators, software programmers, and hardware developers are barely cognizant of the power of true cryptography. Overburdened with battling the plague that they inherited, they can't afford to devote the time or resources needed to implement a truly secure strategy. Just as it is said that everyone has an evil twin somewhere in the world, for every cryptographer there is a de-cryptographer working just as diligently to decipher a new encryption algorithm.

Traditionally, cryptography has consisted of any means possible whereby communications may be encrypted and transmitted. This could be as simple as using a language with which the opposition is not familiar.

Cryptanalysis is the Converse Process of Cryptography

Cryptography is science and art of making the message a secret, where as cryptanalysis is the art of cracking the codes and breaking the security. The person who performs the cryptanalysis is known as the cryptanalyst. The cryptanalyst's job is to decode the secrets either by hook or crook, he keeps on trying different combinations. The performance of a cryptanalyst depends upon the time he/she takes to decode the data. It means that the lesser the time a cryptanalyst takes; the better is his/her performance.

In cryptanalysis, the cryptanalysts firstly analyse the data and try those combinations which they feel are the most probable ones, if that does not comes off then they start on trying all possible combinations. The task of a cryptanalyst is based on the hit and trial method.

Methods of Encrypting

Cryptography contains the following elements:

- An original message
- A cipher text
- Resultant encryption

All true cryptography is composed of an original message being encoded into a cipher text and then finally being decoded by the user. Cryptanalysts analyse and decipher secret coding systems and decode messages for political, military, or law enforcement agencies or organizations.

Substitution Method

The substitution method replaces each letter of each word with a different letter using a scheme that's only known between the sender and the receiver. The most common example of the substitution method is known as Rotate or ROT 'n' system.

With today's modern computer systems, a message encrypted using the Rotate cryptosystem could be cracked rather quickly simply by trying all possible rotation

NOTES

values and inspecting the results. This technique iscalled a brute-force attack, which succeeds more often than you'd like to think. Almost no one uses the original Rotate system on its own today because, frankly, it's not very secure.

Transposition Method

Transposition methods for cryptography are used to 'scramble' the contents of a message using a shared secret between the sender and receiver(s).

6.8.1 Cryptographic Algorithms

Data Encryption Standard

The algorithm for the Data Encryption Standard (DES) was developed by IBM in the early 1970s. The United States National Institute of Standards and Technology (NIST) adopted the algorithm (as FIPS publication 46) for DES in 1977 after it was examined, modified, and approved by NSA. The standard was reaffirmed in 1983, 1988, 1993, and 1999.

The Data Encryption Standard uses a 56-bit key. The key uses seven bits of eight 8-bit bytes (the eighth bit of each byte is used for parity). DES is a block cipher that operates on one 64-bit block of plaintext at a time (see Figure 6.11 for a block diagram of the algorithm). There are 16 rounds of encryption in this method with a different sub-key used for each round.

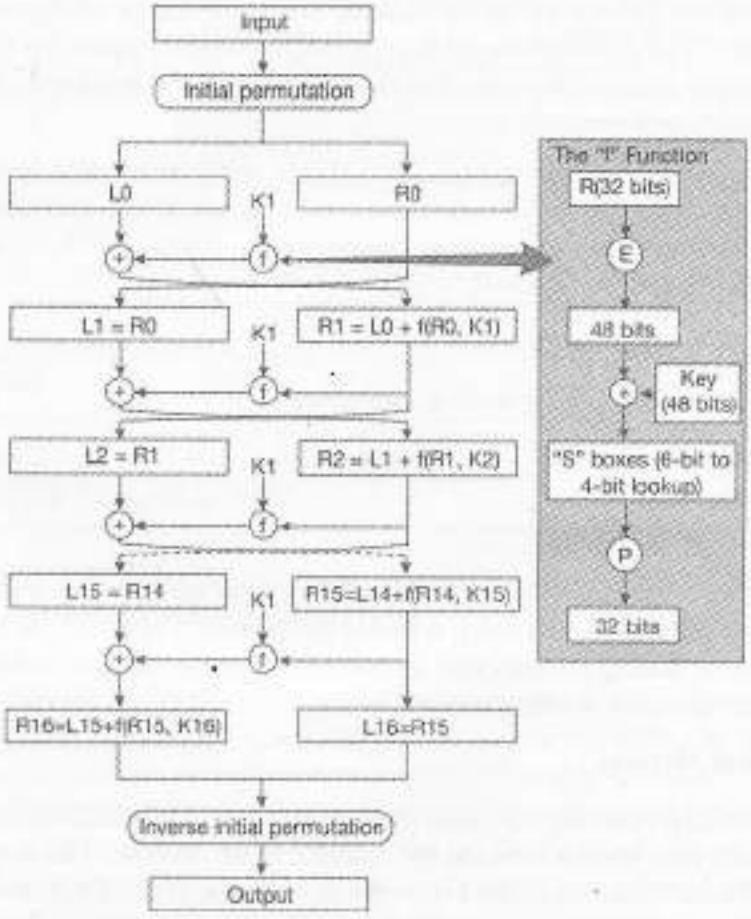


Fig. 6.11 DES Block Diagram

NOTES

The key goes through its own algorithm to derive the 16 sub-keys (see Figure 6.12).

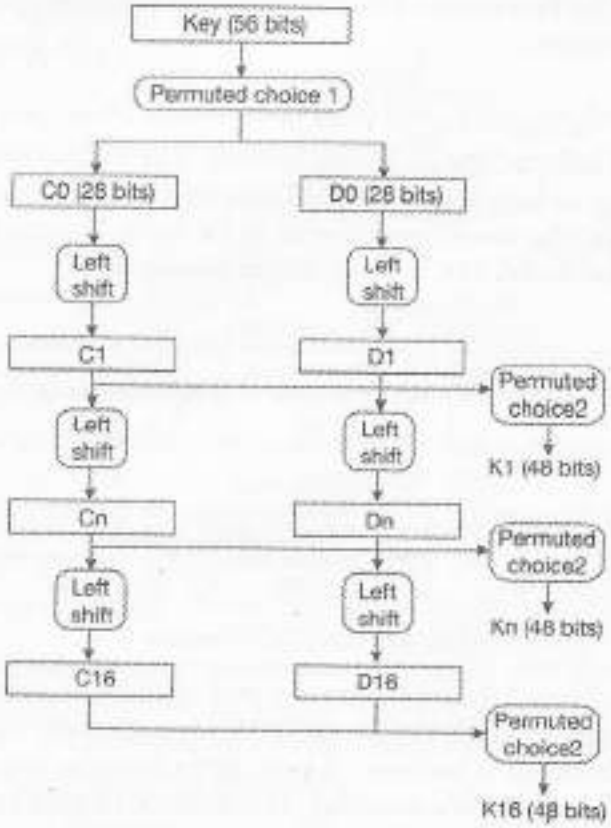


Fig. 6.12 DES Sub-Key Generation Algorithm

The block diagram in Figure 6.12 has several blocks where permutations occur. The standard defines a specific rearrangement of bits for each permutation. The same is true for the sub-key generation algorithm. There are specific bit rearrangements for permuted choice 1 and 2. You can also find a call out of the function 'f.' Within the function, there is a block that says 'S' boxes. The 'S' boxes are table lookups (also defined in the standard) that change a 6-bit input into a 4-bit output.

There are four modes of operation for DES:

- **Electronic Code Book:** This is the basic block encryption where the text and the key are combined to form the cipher text. Identical input produces identical output in this mode.
- **Cipher Block Chaining:** In this mode, each block is encrypted as in electronic code book but a third factor, derived from the previous input, is added. In this case, identical input (plaintext) does not produce identical output.
- **Cipher Feedback:** This mode uses previously generated cipher text as input to DES. The output is then combined with plaintext to produce new cipher text.
- **Output Feedback:** This mode is similar to cipher feedback but uses DES output and does not chain cipher text.

There are no known attacks against the DES algorithm. However, the 56-bit key has become a weakness. This key provides a total of 255 potential keys (including few keys that are found weak and cannot be used). With today's computer systems, this entire key space can be examined within a small amount of time.

NOTES

In 1997, the Electronic Frontier Foundation (EFF) announced a computer system that can find a DES key in four days. This system cost $2,50,000 to build. With today's hardware systems, the time to brute-force a DES key is far too short to protect information that must be kept secret.

Triple DES

In 1992, research indicated that DES could be used multiple times to create a stronger encryption. Thus was born the concept of Triple DES (TDES). Figure 6.13 shows how TDES works. The second operation as in the figure, is actually a decryption. This is the key that makes TDES stronger than normal DES.

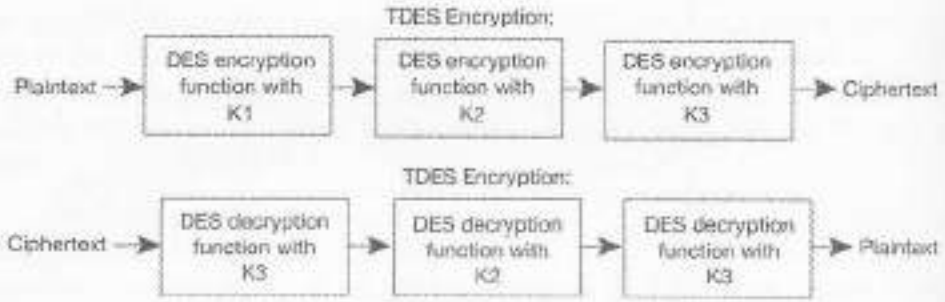


Fig. 6.13 The TDES Function

TDES can be used with either three keys or two keys. In the case of two keys, K1 and K3 are equal and K2 is different. TDES is a relatively fast algorithm as it can still be implemented in hardware. It takes three times the overall time as DES since there are three operations occurring. TDES should be used instead of DES for most applications.

RSA

In 1978, Ron Rivest, Len Adleman, and Adi Shamir released the Rivest-Shamir-Adleman (RSA) public key algorithm. Unlike the Diffie-Hellman algorithm, RSA can be used both for encryption and decryption. Also unlike Diffie-Hellman, the security of RSA is based on the difficultly of factoring large numbers. This is considered a major problem when the numbers are very large (512 bits or larger).

The basic algorithm for confidentiality is very simple:

$Ciphertext = (Plaintext)^e \bmod n$

$Plaintext = (Ciphertext)^d \bmod n$

$Private\ Key = \{d, n\}$

$Public\ Key = \{e, n\}$

The difficulty in calculating d given e and n provides the security. It is assumed that the owner of the key pair keeps the private key a secret and that only the public key is published. Therefore, if information is encrypted with the public key, only the owner can decrypt it. It should also be noted that the algorithm can be reversed to provide authentication of the sender. In this case, the algorithm would be

$Ciphertext = (Plaintext)^d \bmod n$

$Plaintext = (Ciphertext)^e \bmod n$

$Private\ Key = \{d, n\}$

$Public\ Key = \{e, n\}$

NOTES

For authentication, the owner encrypts the information with the private key. Only the owner could do this since the private key is kept secret. Anyone can now decrypt the information and verify that it could have only come from the owner of the key pair.

Generating RSA Keys

Care must be taken in the generation of RSA keys. These are the following text to generate an RSA key pair:

1. Choose two prime numbers a and b and keep them a secret
2. Calculate $n = a \times b$
3. Calculate $(n) = (a - 1)(b - 1)$.
4. Select e such that e is relatively prime to (n).
5. Determine d such that $(d)(e) = 1 \bmod (n)$ and that $d < (n)$.

The number n should be on the order of a 200-digit number or larger. Therefore, both a and b should be at least 100-digit numbers. Keys for real-world use should be at least 1,024 bits. For sensitive information, 2,048 bits and larger keys should be considered.

Elliptic Curves

The way that the elliptic curve operations are defined is what gives ECC its higher security at smaller key sizes.An elliptic curve is defined in a standard, two dimensional x,y Cartesian coordinate system by an equation of the form: $y^2 = x^3 + ax + b$. Figure 6.14 shows an elliptic curve.

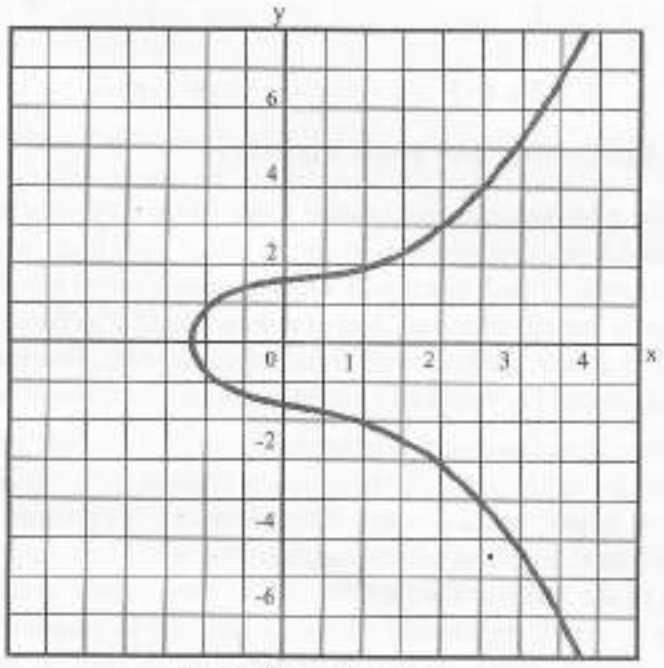


Fig. 6.14 An Elliptic Curve

In elliptic curve cryptosystems, the elliptic curve is used to define the members of the set over which the group is calculated, as well as the operations between them which define how math works in the group. If you feel that this statement was not enough to make you understand the ECC system, you need not worry, it has been further illustrated by an example.

NOTES

Elliptic Curve Cryptography

The dominant operation in ECC cryptographic schemes is point multiplication. This operation is the key to the use of elliptic curves for asymmetric cryptography. This critical operation is fairly simple. ECC arranges itself in such a way so that when you wish to perform an operation the cryptosystem makes it easy. Whether it is encrypting a message with the public key, or decrypting it with the private key, the operation you are actually performing at the back end is nothing but point multiplication. Thus, this section is one of the most important ones to be discussed in ECC (see Figure 6.15).

Point multiplication is simply calculating kP, where k is an integer and P is a point on the elliptic curve defined in the prime field.

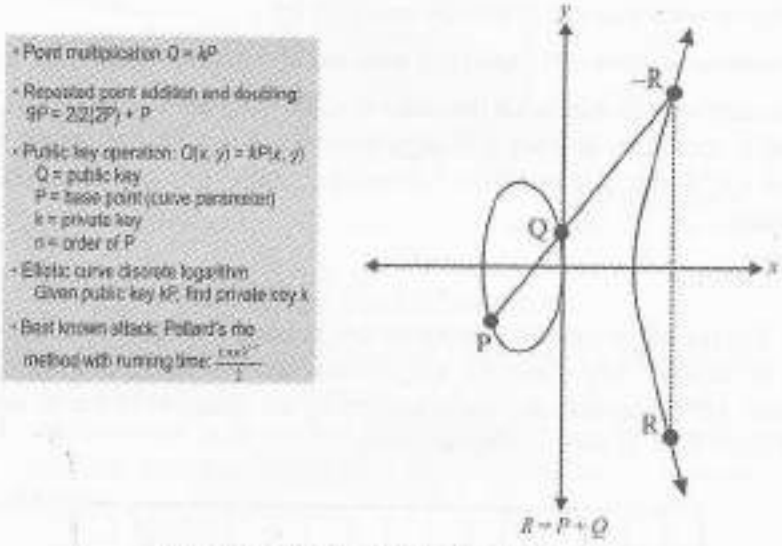


Fig. 6.15 Elliptical Curve Cryptography

Message Authentication and Hash Functions

In cryptography, a Message Authentication Code (often MAC) is a short piece of information used to authenticate a message. A MAC algorithm, sometimes called a keyed (cryptographic) hash function, accepts as input a secret key and an arbitrary-length message to be authenticated, and outputs a MAC. The MAC value protects both a message's data integrity as well as its authenticity, by allowing verifiers (who also possess the secret key) to detect any changes to the message content.

While MAC functions are similar to cryptographic hash functions, they possess different security requirements. To be considered secure, a MAC function must resist existential forgery under chosen plaintext attacks. This means that even if an attacker has access to an oracle which possesses the secret key and generates MACs for messages of the attacker's choosing, the attacker cannot guess the MAC for other messages without performing infeasible amounts of computation.

MACs differ from digital signatures as MAC values are both generated and verified using the same secret key. This implies that the sender and receiver of a message must agree on the same key before initiating communications, as is the case with symmetric encryption. For the same reason, MACs do not provide the property of non-repudiation offered by signatures specifically in the case of a network-wide shared secret key: any user who can verify a MAC is also capable of generating MACs for other messages. In contrast, a digital signature is generated using the private

NOTES

key of a key pair, which is asymmetric encryption. Since this private key is only accessible to its holder, a digital signature proves that a document was signed by none other than that holder. Thus, digital signatures do offer non-repudiation.

An alternative authentication technique uses a secret key for generating a small fixed-size block of data, known as a cryptographic checksum or MAC that is appended to the message. In this authentication technique, it is assumed that two communicating parties, for example, X and Y, share a common secret key. When A sends a message to B, it calculates the MAC as a function of the message and the key:

$MAC = MC(K, M)$

Where,

M: Input message

C: MAC function

K: Shared secret key

MAC: Message Authentication Code

Then the message and the MAC are transmitted to the intended recipient. The recipient performs the same calculation on the received message, using the same secret key, to generate a new MAC. The received MAC is then compared to the calculated MAC.

A MAC function is almost similar to encryption. The only difference is that the MAC algorithm need not be reversible, as it must be for decryption. In general, the MAC function is a many-to-one function. The domain of the function consists of messages of some arbitrary length, whereas the range consists of all possible MACs and all possible keys. If an n-bit MAC is used, then there are 2^n possible MACs, whereas there are possible messages with $N >> 2^n$. Furthermore, with a k-bit key, there 2^k are possible keys.

Hash Functions are also popularly known as message digests and one-way encryption. Hash functions, unlike Secret Key Cryptography and Public Key Cryptography, do not use a key.

MD5 Algorithm

MD in MD5 stands for 'Message Digest'. MD5 is one of a series of message digest algorithms designed by Professor Ronald Rivest of MIT (Rivest, 1994). MD5 is a one-way hash algorithm. It is most commonly used to ensure the authenticity of the data received over the network.

No two distinct inputs (data) can have the same fingerprint. This fact forms the basis of the MD5 algorithm. If you have downloaded a file from a remote system or from another system within your own network, then you can use the MD5 hash algorithm to ensure that the downloaded required destination file is authentic and not a corrupt, malicious one (affected by an intruder, worm or virus).

What does MD5 do?

- The MD5 one-way algorithm accepts a message or chunk of data of an arbitrary length as an input. In this case I have used an example which considers a file to be the input which has been downloaded from a host within one's own local network.
- This input data is then processed using MD5 hash algorithm to produce a fixed length 128-bit fingerprint. Hence, in the example, the MD5 algorithm uses the file to generate a 128-bit fingerprint of the file.

NOTES

- The host from which the file is downloaded possesses the original fingerprint of the file. Now this newly-generated fingerprint of the file is compared with the original fingerprint of the file. If this newly-generated fingerprint is the same as the original fingerprint, then the authenticity of the file is confirmed.

SHA-1

The Secure Hash Algorithm (SHA) was developed by the National Institute of Standards and Technology (NIST) and published as a federal information processing standard (FIPS 180) in 1993. A revised version was issued as FIPS 180-1 in 1995 and is generally referred to as SHA-1. SHA-1 is called secure because it is computationally infeasible to find a message which corresponds to a given message digest, or to find two different messages which produce the same message digest. Any change to a message in transit will, with very high probability, result in a different message digest, and the signature will fail to verify. Thus, SHA is used for the authentication of the message. SHA-1 is employed in a large variety of popular security applications and protocols, including TLS, SSL, PGP, SSH, S/MIME, and IPSec. It was considered to be the successor to MD5, an earlier, widely-used hash function. Both are reportedly compromised.

SHA-1 Algorithm

The overall processing of the message is very similar to that of MD5 developed by Rivest. A message with a maximum length of less than 2^64 bits is taken and a 160-bit message digest is produced as the output.

- The message is padded so that its length in bits is congruent to 448 modulo 512 (length = 448 modulo 512). It means that the length of the padded message must be 64-bits less than an integer multiple of 512-bits. Padding is always added such that if the length of the plaintext is 448-bits, it is padded by 512-bits. Hence, the number of padding bits is in the range of 1 to 512. Padding consists of a single 1-bit followed by the necessary number of 0-bits.
- In the second step, the target is to make the message obtained from step 1 an integral multiple of 512. Thus, the low order 64-bits of the original message (before the padding) is appended to the result of step 1 (least significant byte first).
- In the third step, a 160-bit buffer is initialized which is used to hold the intermediate and final results of the hash function. The buffer can be represented as five 32-bit registers (A, B, C, D, E). These registers are initialized to the following 32-bit integers (hexadecimal values).

  A = 67452301

  B = EFCDAB89

  C = 98BADCFE

  D = 10325476

  E = C3D2E1F0

  These values are actually stored in big-endian format, which states that the most significant byte of a word must occupy the low address byte position. Thus, the initialization values (in hexadecimal) are:

  A = 67 45 23 01

  B = EF CD AB 89

C = 98 BA DC FE

D = 10 32 54 76

E = C3 D2 E1 F0

- A sequence of logical functions f(0), f(1),..., f(79) is used in SHA-1. Each f(t), 0 <= t <= 79, operates on three 32-bit words B, C, D and produces a 32-bit word as output. f(t;B,C,D) is defined as follows: for words B, C, D,

  f(t;B,C,D) = (B AND C) OR ((NOT B) AND D) (0 <= t <= 19)

  f(t;B,C,D) = B XOR C XOR D (20 <= t <= 39)

  f(t;B,C,D) = (B AND C) OR (B AND D) OR (C AND D) (40 <= t <= 59)

  f(t;B,C,D) = B XOR C XOR D (60 <= t <= 79).

  A sequence of constant words K(0), K(1), ... , K(79) is used in the SHA-1. In hex these are given by

  K(t) = 5A827999 (0 <= t <= 19)

  K(t) = 6ED9EBA1 (20 <= t <= 39)

  K(t) = 8F1BBCDC (40 <= t <= 59)

  K(t) = CA62C1D6 (60 <= t <= 79).

- The message digest is computed using the message padded as described in section 4. The computation is described using two buffers, each consisting of five 32-bit words, and a sequence of eighty 32-bit words. The words of the first 5-word buffer are labelled A,B,C,D,E. The words of the second 5-word buffer are labelled H0, H1, H2, H3, H4. The words of the 80-word sequence are labelled W(0), W(1),..., W(79). A single word buffer TEMP is also employed. To generate the message digest, the 16-word blocks M(1), M(2),...,M(n) defined in section 4 are processed in order. The processing of each M(i) involves 80 steps. Before processing any blocks, the H's are initialised as follows: in hex,

  H0 = 67452301

  H1 = EFCDAB89

  H2 = 98BADCFE

  H3 = 10325476

  H4 = C3D2E1F0.

Now M(1), M(2), ... , M(n) are processed. To process M(i), we proceed as follows:

(a) Divide M(i) into 16 words W(0), W(1), ... , W(15), where W(0) is the left-most word.

(b) For t = 16 to 79 let
W(t) = S^1(W(t–3) XOR W(t–8) XOR W(t–14) XOR W(t–16)).

(c) Let A = H0, B – H1, C – H2, D = H3, E = H4.

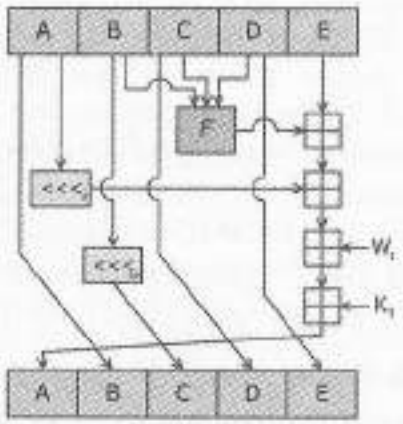
(d) For t = 0 to 79 do
TEMP = S^5(A) + f(t;B,C,D) + E + W(t) + K(t);
E = D; D = C; C = S^30(B); B – A; A = TEMP;

(e) Let H0 = H0 + A, H1 = H1 + B, H2 = H2 + C, H3 = H3 + D, H4 = H4 + E.

NOTES

NOTES

After processing M(n), the message digest is the 160-bit string represented by the 5 words H0 H1 H2 H3 H4.



6.9 BIOMETRICS

Biometrics are another recent innovation in identifying users. User authentication can in general be based upon:

- Knowledge [i.e., something a person knows (e.g., a password or PIN)].
- A token [i.e., something a person owns (e.g., a smart card or passport)].
- Or, a personal characteristic [i.e., something a person naturally has or generates (e.g., a fingerprint or signature)].

The third type of authentication mechanism is the subject of biometrics. Applications using biometric methods have forensic uses (e.g., criminal investigation), civilian uses (e.g., passport), security uses (e.g., access control), and commercial uses (e- commerce and m-commerce applications). Many companies such as MasterCard, IBM, and American Express are studying the use of biometric technologies in eCommerce and security. Figure 6.16 shows a biometric machine.

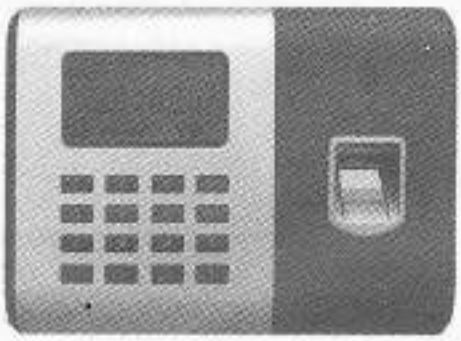


Fig. 6.16 A Biometric Machine

Biometric identification can be defined as the process of identifying an individual on the basis of his/her distinguishing physiological and/or behavioral characteristics. It is essentially a matter of pattern recognition. In the enrollment phase, the biometric characteristic of an individual is scanned, processed, and stored in digital form as a template. The template can be stored in a central database or on a smart card. In the recognition phase, the biometric characteristic is scanned and processed again, and then compared to the template. In the recognition mode, the person to be recognized claims

NOTES

no particular identity. The system searches the entire template database to find a match, which obviously may take a long time. The verification mode is generally much faster because the person claims a specific identity (e.g., by using a smart card) so that the system can immediately find the right tem- plate and compare it to the newly scanned data. Passwords or PINs can easily be forgotten. They can be told to other people, or even acquired in a fraudulent way. In the latter case it is not possible to differentiate between an authorized person and an impostor. Smartcards or passports can be lost or stolen. Biometric methods offer a simpler means of authentication, especially in combination with smart cards, although they are not necessarily faster or more secure. Figure 5.2 shows a Biometric machine.

Criteria for Selecting an Ideal Biometric System

There are several important criteria that should be fulfilled by any biometric method based on a specific characteristic:

- **Universality,** which means that every person must possess the characteristic.
- **Uniqueness,** which means that no two (or more) persons may have the same characteristic.
- **Permanence,** which means that the characteristic does not change significantly over time.
- **Unfakeability,** which means that the characteristic cannot be presented in a fraudulent way.
- **Acceptability,** which means that most people would have no objections to using the method.
- **Collectability,** which means that the characteristic must be easily measurable by affordable technical equipment.
- **Performance,** which means that the system should be accurate, fast, robust, and require no more than a reasonable amount of resources.

Care must be taken when biometrics data is transferred over insecure links. If it is stolen, it cannot be replaced like a password. The emerging Bio API standards will provide interfaces for secure networking and encryption. Calabrese proposes to always use challenge-response protocols for authentication in such a way that the biometrics data is never sent over the network and thus exposed to attacks. Instead, the authenticator sends a random challenge, and the biometric device (e.g., a smart card) responds with a secure hash of the biometric data concatenated with the challenge. The approach of using body characteristics to encrypt or scramble data is also called biometric encryption. Figure 6.17 shows a fingerprint reader.

Fig. 6.17 A Fingerprint Reader

NOTES

One of the main problems with biometrics is that scanning results may vary to a greater or lesser extent (i.e., be dispersed), and thus differ from the reference template. The probability of the system's accepting an impostor is referred to as the False Match Rate (FMR), and the probability of its rejecting an authorized individual is known as the False Non-match Rate (FNR). High-security applications require a small FMR because less damage is done if an authorized individual is rejected than if an unauthorized one is accepted. FNR and FMR can be influenced by adjusting the limit values of the allowed scanning result dispersions.

6.9.1 Factors that Affect Biometric Systems

Biometrics, as we know, is a field dependant on the biological aspects of people. There are two basic factors that affect the usage of biometric systems:

(a) Physiological Characteristics

(b) Behavioral Characteristics

These characteristics may require different kinds of biometric systems for producing accurate results. Table 6.5 shows the various factors that affect biometrics.

***Table 6.5** Factors Affecting Biometrics*

Physiological Characteristics	*Behavioural Characteristics*
Face	Keystroke
Fingerprint	Signature
Hand	Speech/voice
Iris	
DNA	

Physiological Characteristics

Face recognition is one of the most active areas of biometric research. It is typically based on location, shape, and spatial relationships of eyes, eyebrows, nose, lips, and other facial attributes. The method is completely contact-less, but often requires a simple background or special illumination and is heavily view-dependent. Furthermore, a face can change considerably over time, for example, through a new haircut, makeup, or glasses. The template is at least 500 bytes. FNR is quite high, about 10%. Facial thermogram is a pattern produced by the underlying vascular system in the human face and emitted from the skin when heat passes through the facial tissue. It has two advantages over face recognition: it does not change even after plastic surgery, and it does not need special illumination. It has not been proven, however, that facial thermo grams are sufficiently discriminative.

Hand geometry includes measurements of the shape of the human hand, lengths and widths of the fingers, and sometimes the vein pattern. It has been in use for some fifteen years now. The template size is very low, about 10 to 30 bytes, but FNR and FMR may be up to 1%. Also, verification may take up to 9 seconds. Retinal pattern is the specific arrangement of the veins under the retinal surface of an eye. The template is small (40 to 80 bytes), and FNR and especially FMR are rather low. The method is not well accepted by some people, however, out of fear of infectious diseases or eye damage through infrared light. Also, contact lenses may cause problems because they are not completely transparent to infrared light.

NOTES

Behavioral Characteristics

Behavioral characteristics are more likely to change over time than physiological characteristics, so they require adaptive methods that can modify the reference template accordingly. Currently there are three biometric methods in this category: keystroke dynamics (i.e., typing rhythm), speech (i.e., voice) recognition, and signature.

Keystroke dynamics methods are based on measuring the intervals between key strokes. A person to be authenticated is required to type between 100 and 150 alphanumerical characters, using all ten fingers, which is the biggest disadvantage of this method.

Speech recognition can be based on either text-dependent or text- independent speech input. Text-dependent methods are not secure enough because they are based on the utterance of a fixed predetermined phrase, which can also be played from an audio tape. Text-independent methods are much more complex. FMR and FNR are about 1%, which makes the method suitable only for low-security applications. Template size may be up to 1K. Finally, signature methods may be static or dynamic.

Static methods use only the geometry of a signature, but it is very difficult to differentiate between a genuine and a copied signature. Dynamic methods use geometry, but also acceleration, velocity, pressure, and trajectory profiles of a signature. The method is well known and well accepted, but FMR and FNR are relatively high (up to 1%). Template size is up to 1K. Table 6.6 gives a brief overview of biometric characteristics.

Table 6.6 Overview of Biometric Characteristics

	Unique-ness	Perma-nence	Accept-ability	Template size (byte)	FNR (%)	FMR (%)
Physiological characteristics						
Face recognition	–	+	++	500	10	1
Facial thermogram	++	++	+	NA	NA	NA
Fingerprint	++	++	–	300–800	0.01	10–5
Hand geometry	+	+	+	10–30	0.8	0.8
Retinal pattern	++	+	–	40–80	0.005	10–9
Iris	++	++	+	256–1000	NA	10–10
Behavioural characteristics						
Keystroke dynamics	+	+	+	NA	NA	NA
Speech recognition	+	–	++	100–1000	1	1
Dynamic signature	+	–	++	40–1000	1	0.5

Access Control

Traditionally users, also known as employees, have been the weakest link in a company's defensive armor. Though necessary to the organization, they can be a nightmare waiting to happen to your network.

Authentication, Authorization and Accounting

Authentication is simply proving that a user's identity claim is valid and authentic. Authentication requires some form of 'proof of identity.' In network technologies, physical proof (such as a driver's license or photo ID) cannot be employed, so you have to get something else from a user. That typically means having the user respond to a challenge to provide genuine credentials at the time he requests access. For our

NOTES

purposes, credentials can be something the user knows, something the user has, or something they are. Once they provide authentication, there also has to be authorization, or permission to enter. Finally, you want to have some record of users' entry into your network—username, time of entry, and resources. That is the accounting side of the process.

What the User Knows?

Users know a great many details about their own lives—birthdays, anniversaries, first cars, their spouse's name and many will try to use these nuggets of information as a simple form of authentication. What they do not realize is just how insecure those pieces of information are. These pieces of information are often used as fixed passwords and PINs as they are easy to remember. Unless some strict guidelines are established on what form a password or PIN can take, a password will offer little to no real security.

Is the User Authorized?

Authorization is independent of authentication. A user can be permitted entry into the network but not be authorized to access a resource. You do not want an employee having access to HR information or a corporate partner getting access to confidential or proprietary information. Authorization requires a set of rules that dictate the resources to which a user will have access. These permissions are established in your security policy.

Keeping Current

One of the best ways to stay ahead is to not fall behind in the first place. New systems with increasing sophistication are being developed all the time. They can incorporate a more intelligent and autonomous process in the way the system handles a detected threat, a faster and more easily accomplished method for updating threat files, and configuration flexibility that allows for very precise customization of access rules, authentication requirements, user role assignment, and how tightly it can protect specific applications. Register for newsletters, attend seminars and network security shows, read white papers, and, if needed, contract the services of network security specialists. The point is that you shouldn't go cheap on network security. The price you pay to keep ahead will be far less than the price you pay to recover from a security breach or attack.

6.9.2 Design Issues in Biometric Systems

Biometric systems, by their way nature are complex systems with responsive decision making involved in terms of physical access controls. The two most critical issues that designers of biometric systems face are: storing and protecting the template and ensuring accuracy of the biometric systems. They are briefly discussed as follows:

1. **Storage and Protection of the Template:** Biometric systems have to scan, store/ retrieve a template and match. It is important to note that depending on the design of the system, the match can be performed in different locations: on the processor that is used to acquire the biometrics sample data, on a local PC or a remote server or on a portable medium, such as a smart card (equipped with a strong enough processor). In addition, the reference template may be stored on the same three media leaving us with five different combinations and resulting in five different level of trust. Also, there can be three different modes of protection that may be used for the template: no protection, data encryption or digital signature. This means that there can be 15 possible

NOTES

configurations. Each use of combination has its own advantages and disadvantages; the choice of the combination is clearly application-dependent (based on risk and requirement analysis).

2. **Accuracy of Biometric Systems Steps:** The evaluation of a biometric system has to be based on the evaluation of all components: the recognition system performance, the communication interface, the matching and decision step and other key factors, such as ease of use, acquisition speed and processing speed. The performance a biometric system ultimately depends on the accuracy of the end decision only.

In the case of verification system there are two possible types of error: false non-match (also known as false negative or false rejection i.e rejection of a legitmate user) and false match (also known as false positive or false acceptance i.e acceptance of an impostor).

Interoperability Issues

This is another major concern related to the biometric systems that designers need to worry—the interoperability in cross-application of biometrics. Many consider the use of word interoperability to refer to purely technical matters. The greater relevance of interoperability, however lies in the interoperability of a process and where applicable, supporting legislation. This is especially relevant to international situations such as information system for border control and the use of nationally issued documents in other countries. Let us consider, for an example a biometric ID verification check that returns negative result. Is this result understood and interpreted in the same way throughout Europe and the America or in the Asia Pacific region? If not, what are the consequences of such regional interpretation? No doubt, therefore, that so much attention is being paid, to the interoperability issues in the biometric systems.

There are several approved standard that should be used for interoperability between systems of/for both identification and verification functions. The American National Standard Institute (ANSI) standard, Data Format for the Interchange of Fingerprints, Facial and Scar Mark and Tattoo (SMT) Information (ANSI/NIST-ITL 1-2000), formats fingerprint data to perform background searches against any Automated Fingerprint Identification (AFIS) criminal file. The Common Biometric Exchange File Format (CBEFF) accommodates any biometric technology and recognizes the ANSI/National Institute of Standard and Technology (NIST) fingerprint format. Finally, the smart card interoperability specification insures interoperability at the higher interface layers and defines a standard CBEFF-compliant container for biometric data on the card.

6.10 FUTURE OF E-COMMERCE

Over the past ten years, the e-commerce revolution has transformed the way of shopping to the point where it is not necessary to make contact with another human to find a great deal. Although the amount of time and money spend online grows every year, the e-commerce experience has changed very little since its inception. The huge growth of virtual communities—people getting together in ad hoc interest groups online - promises to shift the balance of economic power from the manufacturer to the consumer.

E-commerce proposes substantial prospects for developing countries like India. In India, e-commerce is still in growing stage but forecast indicates a major growth in coming years. The low cost of computers, increase in the Internet usage and an escalating

NOTES

competition in the Internet Service Provider (ISP) market are considered as the major factors that promotes the growth of e-commerce in India. Rediff.com was the first e-commerce site in India and was one of the most trafficked portals. There is a rise in the number of companies that enables e-commerce and Web solution services in past two years. Rather depending on advertising revenues major Indian portal Websites have shifted towards e-commerce.

The content of these portals managed in such a way that they target the consumer/buyer effectively and they can purchase everything from various gadgets to clothing. In spite of RBI regulation low internet usage e-commerce Websites have popped up everywhere hawking things like groceries, bakery items, gifts, books, audio and video cassettes, computer, etc. None of the major players have been dissuade by the low PC penetration and credit card.

India is developing rapidly and if development is to be measured, how can we ignore the role of e-commerce in it. As compared the Internet user base with US or UK, it is still merely 100 million which is much less in India but increasing rapidly. It is believed that in coming years customary retailers will switch to online business as the number of entrant in this area is increasing reaching its peak. This fact is more prominently revealed by the insights into increase in demands for broadband, busy lifestyle, availability of wider product range and reduced prices, etc. Due to the explosion of mobile devices and the services available over the Internet at special discounts, it is expected that the digital downloads segment will grow in the Indian e-commerce market. Certain unique attributes of the e-commerce industry in India, such as cash on delivery mode of payment and direct imports that lower costs considerably are probably going to bring about a speedy growth in this industry in coming years.

Many modern researches state that the e-commerce market in India is set to grow the fastest within the Asia-Pacific region at a CAGR (Compound Annual Growth Rate) of over 57% between 2012-16. Indian e-shoppers can get great deals and services online. The report found that nearly 60 per cent of citizens in India visited a retail site in November 2011, with the number of online shoppers increasing by 18 per cent in the past year. As e-commerce is the cheapest medium to reach out new market and can became an integral part of sales strategy, it proposes a great way of development and doing e-commerce attribute to the successful implementation to carefully understanding the products and services, customers and the business process, easy-to-use system to extend the business on the Web.

The e-commerce platforms maximize its reach to the potential customers and provide them with a convenient, satisfying and secure shopping experience.

Check Your Progress

7. Why digital signatures do not provide message confidentiality?
8. Who designed X.509 standard?
9. What is SET?
10. What is a MAC algorithm?
11. Define the term biometric identification.

6.11 SUMMING UP

- Traditionally, commerce is seen as the exchange or buying and selling of goods and services, which involves exchange of money, and sometimes transportation of goods from one place to another.
- Electronic commerce or e-commerce refers to the electronic means of carrying business communications and transactions over the Internet or any other form of network.
- Transaction refers to an exchange taking place between two entities where one sells products or services to the other.

NOTES

- Generally, there exist five important processes in an e-commerce lifecycle: information sharing, ordering, online payment, fulfillment, and support and services.
- Information sharing is carried out by means of advertising, marketing on the Internet and through customer information gathering process.
- World Wide Web (WWW) provides one of the effective mediums for communication with the customers.
- Ordering involves customers to electronically place order for the company's goods and services.
- Online payment is the most crucial activity in e-commerce. It requires the use of high security to protect online money transactions from any unauthorized use.
- There exist a variety of types of e-commerce and these types can be characterized in different ways. Four major types of e-commerce are Business-to-Consumer (B2C), Business-to-Business (B2B), Consumer-to-Consumer (C2C), and Peer-to-Peer (P2P).
- B2C refers to indirect trading between company and consumer where the consumer can directly buy the products or services from the supplier's Website.
- B2B type of e-commerce takes place between two companies, where one sells products to another.
- C2C refers to e-commerce activities, which uses an auction style model. This model consists of person-to-person transaction that completely excludes businesses from the equation.
- P2P is not only an e-commerce but also a technology that allows users to share computer resources and computer files over the Internet without the involvement of a central Web server.
- The communication over Internet (online communication) can be secured by adopting certain security mechanism. The most basic of these mechanisms is the cryptography.
- In simple terms, cryptography is the process of altering messages in a way that their meaning is hidden from the adversaries who might intercept them.
- Cryptography relies upon two basic components: an algorithm (or cryptographic methodology) and a key. Algorithms are the complex mathematical formulae and keys are the strings of bits.
- Various components involved in cryptography are plaintext, ciphertext, encryption, decryption, ciphers and keys.
- Traditionally, cryptography involved the use of same key for encrypting or decrypting the messages. This is referred to as symmetric key (private key or secret key) encryption.
- Modern cryptography involves the use of different keys for encryption and decryption. This is referred to as public key (asymmetric key) encryption.
- A digital certificate is an electronic document that signifies the association between the user and his/her public key. It is digitally signed and issued by a trusted third party, referred to as Certificate Authority (CA).
- The process of verifying the identity of a user is termed as authentication. In other words, it refers to establishing validity of the claim.

NOTES

- The historical legal concept of 'signature' is defined as any mark made with the intention of authenticating the marked document. The authentication mechanism used to verify the authenticity of electronic documents is referred to as digital signature.
- Digital signature allows the sender to attach an electronic code with the message in order to ensure its authenticity and integrity. This electronic code acts as a signature of the sender and hence, named digital signature.
- Digital signatures are analogous to physical handwritten signatures and provide the security services namely, message authentication, message integrity, non-repudiation and, message confidentiality.
- There are three kinds of attacks possible on digital signatures, namely, key-only, known message and chosen-message attack.
- The digital signature scheme has got many variations and additions to its main concept. The two variants of digital signature are timestamped signature, blind signature and undeniable digital signature.
- There are several approaches which have been proposed for the digital signature function. These approaches are categorized into two types: direct and arbitrated.
- For ensuring secure online transactions, an important concern is the security of channel via which the communication is being taken place. For this purpose, two protocols are commonly used, which include Secure Socket Layer (SSL) and Secure Electronic Transaction (SET).
- SSL is a protocol developed by Netscape Corporation in 1994 to provide exchange of information between a Web browser and a Web server in a secured manner. Its main aim is to provide entity authentication, message integrity and confidentiality.
- SET is a protocol used for secured credit card transactions over the Internet. SET itself is not a payment system; rather it contains the security protocols and formats which are used to provide security to the credit card payments made by a user on a public network, such as the Internet.
- Cryptography can be defined as the conversion of data into a scrambled code that can be deciphered and sent across a public or private network, without unauthorized access to it.
- Cryptanalysis' is also used to refer to any attempt to circumvent the security of other types of cryptographic algorithms and protocols in general, and not just encryption.
- Encryption is the act of passing a plaintext message through a mathematical formula to create ciphertext. Decryption turns ciphertext messages back into plaintext messages.
- All true cryptography is composed of an original message being encoded into a cipher text and then finally being decoded by the user.
- Cryptanalysts analyse and decipher secret coding systems and decode messages for political, military, or law enforcement agencies or organizations.
- In elliptic curve cryptosystems, the elliptic curve is used to define the members of the set over which the group is calculated, as well as the operations between them which define how math works in the group.

NOTES

- In cryptography, a Message Authentication Code (often MAC) is a short piece of information used to authenticate a message. A MAC algorithm, sometimes called a keyed (cryptographic) hash function, accepts as input a secret key and an arbitrary-length message to be authenticated, and outputs a MAC.
- Biometric identification can be defined as the process of identifying an individual on the basis of his/her distinguishing physiological and/or behavioral characteristics.

6.12 KEY TERMS

- **Electronic commerce:** The electronic means of carrying business communications and transactions over the Internet or any other form of network
- **Acquiring bank:** A bank which is responsible for online authorization of credit cards and payment processing
- **Secure Sockets Layer (SSL) server certificate:** A digital certificate, also referred to as Secure Sockets Layer (SSL) server certificate, is an electronic document that enables SSL on the Web server
- **Cryptography:** The process of altering messages in a way that their meaning is hidden from the adversaries who might intercept them
- **Digital signature:** An authentication mechanism which allows the sender to attach an electronic code with the message in order to ensure its authenticity and integrity
- **Digital certificate:** An electronic document that signifies the association between the user and his/ her public key
- **Authentication:** The process of verifying the identity of a user

6.13 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Electronic commerce or e-commerce refers to the electronic means of carrying business communications and transactions over the Internet or any other form of network.
2. An example of B2C type e-commerce is virtual Internet shopping malls; the customer can shop online the required items anytime instead of spending time in moving around in crowded supermarkets and have the goods delivered at their doorstep.
3. Payment gateway is a service that connects merchants, customers and financial networks with one another for processing credit card authorization and payment. Usually, a third party, such as VeriSign is responsible for the operation of payment gateway.
4. Encryption the process of encrypting the plaintext so that ciphertext can be produced. The plaintext is transformed to ciphertext using the encryption algorithm.
5. The most common example of symmetric key algorithm is Data Encryption Standard (DES) that was first published in 1977 by National Institute of Standards and Technology (NIST). DES uses a 56-bit encryption to encrypt a 64-bit block of plaintext, producing a 64-bit ciphertext block.

NOTES

6. The process of verifying the identity of a user is termed as authentication. In other words, it refers to establishing validity of the claim. The vendor or the dealer that sells the products or services needs to use certain authentication mechanism for recognizing the genuine buyers.
7. Digital signatures do not provide message confidentiality directly because anyone knowing the sender's public key can decrypt the message. Thus, to achieve message confidentiality, we need to encrypt the message along with the signature using either secret key encryption or public key encryption scheme.
8. X.509 was designed by ITU-T to describe the public key certificates in a structured way. It is a part of X.500 series of recommendations that define a directory service.
9. SET or Secure Electronic Transaction protocol is used for secured credit card transactions over the Internet. SET itself is not a payment system; rather it contains the security protocols and formats which are used to provide security to the credit card payments made by a user on a public network, such as the Internet.
10. In cryptography, a Message Authentication Code (often MAC) is a short piece of information used to authenticate a message. A MAC algorithm, sometimes called a keyed (cryptographic) hash function, accepts as input a secret key and an arbitrary-length message to be authenticated, and outputs a MAC.
11. Biometric identification can be defined as the process of identifying an individual on the basis of his/her distinguishing physiological and/or behavioral characteristics.

6.14 QUESTIONS AND EXERCISES

Short-Answer Questions

1. How e-commerce is different from traditional commerce?
2. Discuss the major categories of e-commerce?
3. What are the advantages and disadvantages of e-commerce?
4. What are the various components involved in cryptography?
5. Write the difference between symmetric and asymmetric key encryption.
6. What is the significance of Kerberos protocol?
7. How digital signatures ensure security?
8. What is a digital certificate?
9. How SSL is different from SET?
10. How cryptography is done?
11. What are the criteria for selecting an ideal biometric system?

Long-Answer Questions

1. Discuss the steps involved in e-commerce lifecycle.
2. Explain the basic requirements that should be met for an e-commerce system.

3. Describe the working of e-commerce with the help of a suitable example.
4. Describe digital certificates used for key certification.
5. Explain different approaches of digital signatures.
6. Explain the working of SET protocol.
7. Explain Kerberos protocol in detail.
8. Explain X.509 protocol used for authentication.
9. Discuss the key dimensions of security in e-commerce environment.
10. Differentiate the following:
    (a) Symmetric key and public key encryption
    (b) SSL and SET protocol
11. Explain different cryptographic algorithms.
12. Describe the various design issues in biometric systems.

NOTES

6.15 REFERENCES AND SUGGESTED READINGS

Diwan, Parag and Sunil Sharma. 2002. *E-Commerce: A Manager's Guide to E-Business*. New Delhi: Excel Books.

Kosiur, David R. 1997. *Understanding Electronic Commerce*. United States: Microsoft Press.