NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF SCIENCE AND TECHNOLOGY

COURSE CODE: CIT 758

COURSE TITLE: WIRELESS COMMUNICATION II

Course Code: **CIT 758**

Course Title: **Wireless Communication II**

Course Developer/Writer: A. J. Ikuomola
Department of Computer Science
College of Natural Science
University of Agriculture Abeokuta,
Ogun State, Nigeria

Course Editor:

Programme Leader:

Course Coordinator: **Afolorunso A.A.**
National Open University of Nigeria,
Victoria – Island, Lagos

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria

Headquarters
National Open University of Nigeria
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Branch office
245 Samuel Adesujo Ademulegun Street,
Central Business District
Opposite Arewa Suites
Abuja
E-mail: centrainfo@nou.edu.ng
URL:www.nou.edu.ng

First published
ISBN

Course Guide

Contents

Introduction

Introduction to Wireless Communication II is a three unit post graduate course in Information Technology.

Wireless is a term used to describe telecommunications in which electromagnetic waves (rather than some form of wire) carry the signal over part or the entire communication path. Some monitoring devices, such as intrusion alarms, employ acoustic waves at frequencies above the range of human hearing; these are also sometimes classified as wireless. Wireless communication is the transfer of information over a distance without the use of electrical conductors or "wires". Wireless communication is generally considered to be a branch of telecommunications.

The past decade has seen many advances in physical-layer wireless communication and their implementation. This course takes a view of wireless communication system. Topic covers include: brief review of wireless communication I, satellite communication, coding and error control, IEEE 802.11 wireless, Bluetooth and radio frequency identification.

What You will Learn in this Course

This Course Guide is the starting point for this course. It tells you briefly what the course is about, what course materials you will be using and how you can work your way through these materials. It also gives you guidance on your tutor-marked assignments as well as describes what you need to do in order to pass this course. There will be regular tutorial classes that are related to the course. It is advisable for you to attend these tutorial sessions. The course will prepare you for the challenges you will meet in the field of wireless communication.

Course Aims

The aim of this course is to provide you with an understanding of wireless communication, its use, application and the technology behind it; it also aims to provide you with solutions to problems in wireless/cellular communication system. This will be achieved by:

- introducing you to what the wireless communication is and what it consists of;

- explaining to you the cellular network generation: evolution of GSM technology
- enabling you to understand the basic concepts of satellite communication
- enabling you to understand the various transmission error detection and error control codes
- enabling you to understand the IEEE 802.11 architecture and standards
- explaining to you the concept of Bluetooth
- enabling you to understand the operating principle and application of radio frequency identification

Course Objectives

To achieve the aims set out above, the course has a set of objectives. Each unit has specific objectives which are included at the beginning of each unit. You should read these objectives before you study the unit. You may wish to refer to them during the study to check on your progress. On successful completion of this course, you should be able to:

- explain the concept of wireless communication
- discuss on the cellular network generation
- explain the concept of satellite communication
- state how satellite are used
- describe various satellite orbit and frequency band
- discuss on different types of error detection and error correction
- outline the areas of application of error detection and error correction
- discuss the architecture behind IEEE 802.11
- state areas of application of Bluetooth
- describe Bluetooth specification
- describe the operating principle of radio frequency identification
- list the application areas of radio frequency application

Working through this course

To complete this course you are required to read each study unit, the textbooks, related materials you find on the internet and other materials which may be provided by the National Open University of Nigeria.

Each unit contains self-assessment exercises, and at a point in the course, you are required to submit assignments for assessment purposes. At the end of this course there is a final examination. The course will take about 18 weeks to complete. Below, you will find listed all the components of the course, what you have to do and how you should allocate your time to each unit in order to complete the course successfully on time.

The Course Materials

The main components of the course are:

1. The Course Guide
2. Study Units
3. Presentation Schedule
4. Assignment
5. References/Further Readings

Study Units

There are twenty one study units in this course. Each unit should take you 2-3 hours to work through. The twenty one units are divided into four modules. Three modules contain 5 units each while the last module contains six units. This is arranged as follows: The study units in this course are as follows:

Module 1 Brief review of major concept of Wireless Communication I

Unit 1 Overview of Wireless Communication I
Unit 2 Mobile Radio Propagation
Unit 3 Modulation, Diversity and Multiple Access Techniques
Unit 4 Cellular Wireless – Evolution of GSM Technology

Module 2 Satellite Communication
Unit 1 Basic concept of Satellite Communication
Unit 2 Types of Satellite Communication: Orbits
Unit 3 Types of Satellite Communication: Frequency Band
Unit 4 Capacity Allocation
Unit 5 Application areas of Satellite Communication

Module 3 Coding and Error Control
Unit 1 Error Detection
Unit 2 Error Control
Unit 3 Applications of Error Control Codes

Module 4 IEEE 802.11 Wireless
Unit 1 Overview of IEEE 802.11
Unit 2 IEEE 802.11 Architecture
Unit 3 IEEE 802.11 Architecture

Module 5 Bluetooth
Unit 1 Basic Concept of Bluetooth
Unit 2 Bluetooth Specification
Unit 3 Technical information on Bluetooth
Module 6 Radio Frequency Identification
Unit 1 Radio Frequency Identification: Operating Principle
Unit 2 Radio Frequency Identification: Sensing Applications
Unit 3 Radio Frequency Identification: Challenges

In Module 1: The first unit focuses on the concept of wireless communication and cellular system Second unit deals with the mobile radio propagation models. The third unit concentrates on various techniques used in modulation, diversity and multiple access. Unit 4 deals with the evolution of GSM technology

In Module 2: The first five units concentrate on meaning of satellite, satellite communication, advantage and disadvantage of satellite, types of satellite communication orbits and frequency bands, capacity allocation and application areas of satellite communication.

In Module 3: The first three units focus on the meaning error detection, types of error detection , meaning of error control, types of error controls as well as the application of error control codes.

In Module 4: The first three units discuss on the overview of IEEE 802.11 and standards, IEEE 802.11 component and layer description.

In Module 5: The first three units concentrate on Bluetooth profile, application, Bluetooth devices, Bluetooth specification, protocol stack as well as setting up connection.

In Module 6: The first three unit focus on overview of radio frequency identification, operating principle of RFID using inductive coupling and backscatter coupling. It also focus on sensing application of RFID and the likely challenges.

Presentation schedule

Your course materials have important dates for the early and timely completion and submission of your TMAs and attending tutorials are necessary. You should remember that you are required to submit all your assignments by the stipulated time and date. You should guard against falling behind in your work. The assignment file will be made available to you in due course. In this file, you will find all the details of the work you must submit to your tutor for marking. The marks you obtain for these assignments will count in computing the final mark you obtain for this course.

Assessment

There are three aspects to the assessment of the course: the self-assessment exercises, the tutor-marked assignments and the written examination/end of course examination.

You are advised to do the exercises. In tackling the assignments, you are expected to apply information, knowledge and techniques you gathered during the course. The assignments must be submitted to your facilitator for formal assessment in accordance with the deadlines stated in the presentation schedule and the assignment file. The work you submit to your tutor for assessment will count for 30% of your total course work. At the end of the course you will need to sit for a final or end of course examination of about three hour duration. This examination will count for 70% of your total course mark.

Tutor-Marked Assignment

The TMA is a continuous assessment component of your course. It accounts for 30% of the total score. You will be given four (4) TMAs to answer. Three of these must be answered before you are allowed to sit for the end of course examination. The TMAs would be given by your facilitator and returned after you have done the assignment. Assignment questions for the units in this course are contained in the Assignment File. You will be able to complete your assignments from the information and materials contained in your reading and study units, references and the internet. When you complete each assignment, send it, together with the TMAs, to your tutor.

Make sure that each assignment reaches your tutor/facilitator on or before the deadline given in the presentation schedule and assignment file.

Final Examination and Grading

The end of course examination for Wireless Communication will be for about 3 hours and it has a value of 70% of the total course work. The final examination covers information from all parts of the course i.e. all areas of the course will be assessed. You might find it useful to review your self-tests, tutor-marked assignments and comments on them before the examination.

Course Marking Scheme

The following table lays out how the actual course marking is broken down.

Assignment	**Marks**
Assignment 1 – 4	Four assignments, best three marks of the four count at 10% each – 30% of course marks
End of course examination	70% of overall course marks
Total	100%

Facilitators/Tutors and Tutorials

There are 16 hours of tutorials provided in support of this course. You will be notified of the dates, times and location of these tutorials as well as the name and phone number of your facilitator, as soon as you are allocated a tutorial group.

Your facilitator will mark and comment on your assignments, keep a close watch on your progress and any difficulties you might face and provide assistance to you during the course. You are expected to mail your Tutor Marked Assignment to your facilitator before the schedule date (at least two working days are required). They will be marked by your tutor and returned to you as soon as possible.

Do not delay to contact your facilitator by telephone or e-mail if you need assistance. The following might be circumstances in which you would find assistance necessary, hence you would have to contact your facilitator if:

- You do not understand any part of the study or the assigned readings.
- You have difficulty with the self-tests or exercises.
- You have a question or problem with an assignment or with the grading of an assignment.

You should try to attend the tutorials. This is the only chance to have face to face contact with your course facilitator and to ask questions which are answered instantly. You can raise any problem encountered in the course of your study.

To gain maximum benefit from course tutorials, prepare a question list before attending them. You will learn a lot from participating actively in discussions.

References/Further Readings

Albrecht K. and McIntyre L. (2005). Spychips : How Major Corporations and Government Plan to Track Your Every Move with RFID. Nelson Current Publishing.

Alhoniemi E. (1998). Error Detection and Control in Data Transfer

Andrews K (2007). The Development of Turbo and LDPC Codes for Deep-Space Applications, Proceedings of the IEEE, 95(11).

Avoine G. (2006). Security and Privacy in RFID Systems Bibliography. Available at: http://lasecwww.epfl.ch/~gavoine/rfid/.

Bluetooth SIG (2008). "How Bluetooth Technology Works". Available Online at http://bluetooth.com/Bluetooth/Technology/Works/

Bluetooth SIG (2008). "Specification Documents". Available Online at: http://www.bluetooth.com/Bluetooth/Technology/Building/Specifications/.

Bluetooth.com. (2010). "Profiles Overview". Available Online at http://www.bluetooth.com/English/Technology/Works/Pages/Profiles_Overview.aspx

Bluetooth.com. (2010). "Specification Documents". Available Online at http://www.bluetooth.com/Specification%20Documents/AssignedNumbersServiceDiscovery.pdf

Brenner P. (1996). A Technical tutorial on the IEEE802.11protocol. Breezecom Wireless Communication

Bridgelall R. (2004). RADAR Technology for Commodity Goods.

Chomienne D. and Eftimakis M. (2010). "Bluetooth Tutorial" (PDF). Available Online at: http://www.newlogic.com/products/Bluetooth-Tutorial-2001.pdf.

Cisco Systems, Inc. (2007) "Channel Deployment Issues for 2.4 GHz 802.11 WLANs". http://www.cisco.com/en/US/docs/wireless/technology/channel/deployment/guide/Channel.html.

Clark G. C. and Cain J. B (1981). *Error-Correction Coding for Digital Communications*. New York: Plenum Press. ISBN 0-306-40615-2.

Dana M. (2005). "Communication speed nears terminal velocity". *New Scientist* 187 (2507): 38–41. ISSN 0262-4079.

Dietrich C. (2000): Adaptive Arrays and Diversity Antenna Configurations for Handheld Wireless Communication Terminals‖.

Elbert, B. R. (1987). Introduction to Satellite Communication, Norwood, MA: Artech House

Elbert, B. R. (1992). Networking Strategies for Information Technology, Norwood, MA: Artech House

Elbert, B. R. (2001). The Satellite Communication Ground Segment and Earth Station Handbook, Norwood, MA: Artech House.

Elbert, B. R. (2004). Satellite Application Handbook, Norwood, MA: Artech House.

Feldman J., Abou-Faycal I. and Frigo M. (2002). "A Fast Maximum-Likelihood Decoder for Convolutional Codes". *Vehicular Technology Conference* 1: 371–375.

Finkenzeller K (2003). RFID-Handbook; John Wiley & Sons.

Flock, W. L. (1987). Propagation Effects on Satellite Systems at Frequencies Below 10 GHz, Second Edition, NASA Reference Publication 1108(2), National Aeronautics and Space Administration.

Fred H. (1998). Data Communication, Computer Networks and Open Systems, Fourth Edition, Addison-Wesley Publishhing Company Inc., United States of America

Gilbert W. J. and Nicholson W. K. (2004), *Modern Algebra with Applications* (2nd ed.), John Wiley

Glaise, René J. (1997). "A two-step computation of cyclic redundancy code CRC-32 for ATM networks". *IBM Journal of Research and Development* (Armonk, NY: IBM) 41 (6): 705. Available online at http://www.research.ibm.com/journal/rd/416/glaise.html.

Goldsmith A. (2005). Wireless Communications, Cambridge University Press. ISBN-13:9780521837163

Handy M.. RFID Technology. Institute of Applied Microelectronics & CS University of Rostock, RFID-Workshop, 30.9./1.10.04, Berlin

Harmon C.K. (2003). Basics of RFID Technology, MIT RFID Privacy Workshop, Cambridge, MA..

Heidi M. (1999). "Bluetooth Technology and Implications". SysOpt.com. http://www.sysopt.com/features/network/article.php/3532506.

Heute Technologie Von Morgen Beherrschen (2011). Physical Principles of the RFID-Handbook. Available online at http://RFID-handbook.com

Hodges S. and McFarlane D. (2005). Radio Frequency Identification: Technology, Applications and Impact. Uk: Auto-ID Lab, Cambridge University

Hong J. and Vetterli M. (1995), "Simple Algorithms for BCH Decoding", *IEEE Transactions on Communications* 43 (8): 2324–2333

Huffman W. and Pless V. (2003). *Fundamentals of error-correcting codes*. Cambridge University Press. ISBN 13: 9780521782807.

Ibid. (2003). Wal-Mart Expands RFID Mandate. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/539/1/1/.

Ibid. (2004). RFID for Item Management. ISO/IEC 18000.

Ibid. (2005). AmEx Adds RFID to Blue Credit Cards. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/1646/1/1/.

Information Age (2007). "The Bluetooth Blues". Available Online at: http://web.archive.org/web/20071222231740/http://www.information-age.com/article/2001/may/the_bluetooth_blues.

Intelligraphic (2010). Introduction to IEEE 802.11.

International Organization for Standardization (ISO). (2003). Identification cards -Contactless integrated circuit(s) cards -- Vicinity cards. ISO/IEC 14443.

Jim K. (2008). "How Bluetooth got its name". Available online at http://www.eetimes.eu/scandinavia/206902019?cid=RSSfeed_eetimesEU_scandinavia

Juels A. and Pappu R. (2003). Squealing Euros: Privacy-Protection in RFID-Enabled Banknotes. Financial Cryptography. Lecture Notes in Computer Science. Volume 2742,103-121.

Juels A., Rivest R. L. and Szydlo M. (2003). The Blocker Tag: Selective Blocking of RFID Tags for Consumer Privacy, ACM Press.

Juels, A. (2006). RFID Security and Privacy: A Research Survey. Selected Areas of Cryptography.

Kadish, J. E., and T. W. R. East, Satellite Communications Fundamentals, Norwood, MA:

Kolawole M.O. (2004). Satellite Communication Engineering New York: Marcel Dekker Inc.

Lemmon, J.J. (2002). Wireless link statistical bit error model. US National Telecommunications and Information Administration (NTIA) Report 02-394

Lin S. and Costello D. J.(1983). *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-283796-X.

Linnartz, J.P.M.G. (1995). Cellular Telephone Network. (Online) Available at http://wireless.per.nl/reference/about.htm

Linnartz‘s J.M.G. (1996). Wireless Communication, Baitzer Science Publishers

Lint J. H. (1992). *Introduction to Coding Theory*. GTM. (2nd ed.). Springer-Verlag. ISBN 3-540-54894-7.

Lior Baruch (2008). Mobile Computing Definition –Wireless.

M.S. Ryan and G.R. Nudd. (1993). The Viterbi Algorithm. Technical Report University of Warwick RR-238.

Macario, R. C. V. (1991). Personal and Mobile Radio Systems, London, England: Peter Peregrinus

McAuley A. J. (1990) Reliable Broadband Communication Using a Burst Erasure Correcting Code, ACM SIGCOMM

Mirabito, M. And Morgenstern, B. (2004). Satellites: Operations and Applications. The New Communication Technologies (fifth edition). Burlington: Focal Press.

Molisch A.F. (2005): Wireless Communications, Wiley and Associated

Nokia. (2004). Nokia Mobile RFID Kit. Available at: http://www.nokia.com/nokia/0,,55738,00.html.

O'Brien, J. & Marakas, G.M.(2008). Management Information Systems. New York, NY: McGraw-Hill Irwin.

Olla P. (2011). Evolution of GSM Network Technology.

Peebles and Peyton Z. Jr. (1998). Radar Principles, John Wiley and Sons, Inc.

Peterson and Davie (2003). *Computer Networks: A Systems Approach*, Third Edition

Pister K. (2004). Smart Dust: Autonomous Sensing and Communications in a Cubic Millimeter. Available at: http://robotics.eecs.berkeley.edu/~pister/SmartDust/.

Porter, M. E.(1980). Competitive Strategy, New York: The Free Press.

Rappaport T.S. (2005). Wireless Communication :Principle and Practice. 2nd Edition, India : Prentice Hall

Reed I. S. and Chen X. (1999), *Error-Control Coding for Data Networks*, Boston, MA: Kluwer Academic Publishers

Reynolds M.(2003). Physics of RFID, MIT RFID Privacy Workshop, Cambridge, MA

RFID Journal. (2003). Gillette Confirms RFID Purchase. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/258/1/1/.

Rudra A. (2010).*CSE 545, Error Correcting Codes: Combinatorics, Algorithms and Applications*, University at Buffalo, Available online at: http://www.cse.buffalo.edu/~atri/courses/coding-theory/

Ryan W.E. and Lin S. (2009). *Channel Codes: Classical and Modern*. Cambridge University Press. ISBN 978-0-521-84868-8.

Saeed F. A. (2006). "Capacity Limit Problem in 3G Networks". Purdue School of Engineering.

Sharma S. (2006). Wireless & Cellular Communications, New Delhi: S.K. Kataria & Sons.

Shea J. (2000). Brief History of Wireless Communication.

Shu Lin, Daniel J. Costello, Jr. (1983). Error Control Coding: Fundamentals and Applications. Prentice Hall. ISBN 0-13-283796-X.

Simon Haykin and Michael Moher (2004), Modern Wireless Communications, Prentice Hall, USA.

Sklar, B. (2001). Digital Communications—Fundamentals and Applications, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2001.

Stockman, H. (1948). Communication by Means of Reflected Power. Proceedings of the Institute of Radio Engineers. October. Pages 1196-1204.

Tanenbaum, A. S. (1988). *Computer Networks*, Second Edition. Prentice Hall.

Tom F. (2007). "The Cell-Phone Revolution". *American heritage of invention & technology.* New York: American Heritage, 22 (3): 8–19.

ISSN 8756-7296. OCLC 108126426. BL Shelfmark 0817.734000. http://www.americanheritage.com/events/articles/web/20070110-cell-phone-att-mobile-phone-motorola-federal-communications-commission-cdma-tdma-gsm.shtml.
Tozer T.C. and Grace D. (2001). High Altitude Platforms for Wireless Communications, Electronics & Communication Engineering Journal.
Tutorial_Reports.com (2007). IEEE802.11 Architecture.
Uniform Code Council. (2006). Webpage. Available at: http://www.uc-council.org.
United States Department of Defense. (2006). Radio Frequency Identification. Available at: http://www.acq.osd.mil/log/rfid/index.htm.
Weis S. A. RFID (Radio Frequency Identification): Principles and Applications
Wicker S. B. (1995). *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-200809-2.
Wikipedia (2011). Bluetooth. Available online at: http://en.wikipedia.org/wiki/Bluetooth
Wikipedia (2011). Communication Satellite. Available online at http://en.wikipedia.org/wiki/Communications_satellite
Wikipedia (2011). IEEE 802.11. Available online at http://en.wikipedia.org/wiki/IEEE_802.11
Wikipedia (2011). Radio Frequency Identification. Available online at http://en.wikipedia.org/wiki/Radio-frequency_identification
William S. (2005). Wireless communications and networks. Upper Saddle River, NJ: Pearson Prentice Hall.
Williams, Ross N. (1996). "A Painless Guide to CRC Error Detection Algorithms V3.00" Available online at: http://www.repairfaq.org/filipg/LINK/F_crc_v3.html.
Wilson S. G. (1996). *Digital Modulation and Coding*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-210071-1.
Wireless Networks Spring (2005). Coding and Error Control
Zhili Sun(2005). Satellite Networking, Principles and Protocols, West Sussex, England, Wiley.

Summary

Wireless Communication II is a course that describes communications in which electromagnetic waves or RF (rather than some form of wire) carry a signal over part or the entire communication path. Upon completing this course, you will be equipped with the basic knowledge of Wireless Communication. In addition, you will be able to answer these questions.

- What does wireless communication means?
- What are the application areas of wireless technology
- Identify four generations of wireless network
- Explain the concept of satellite communication
- Explain the types of satellite orbit
- Describe how transmission error can be detected with parity checking
- Describe block code
- Discuss on the three types of control frame used to facilitate the exchange of data frame between stations
- List the type of Bluetooth Profile
- Describe Bluetooth specification
- Explain on the backscatter coupling operating principle of radio frequency identification

Of course, the list of questions that you can answer is not limited to the above list. We wish you success in the course and hope that you will find it both interesting and useful and wishing you every success in your future

MODULE 1: BRIEF REVIEW OF MAJOR CONCEPTS IN WIRELESS COMMUNICATION I

UNIT 1: OVERVIEW OF WIRELESS COMMUNICATION I

1.0 Introduction

Wireless communication is the transfer of information over a distance without the use of electrical conductors or "wires". The distances involved may be short (a few meters as in television remote control) or long (thousands or millions of kilometers for radio communications). Wireless communication is generally considered to be a branch of telecommunications.

It encompasses various types of fixed, mobile, and portable two-way radios, cellular telephones, personal digital assistants (PDAs), and wireless networking. Other examples of wireless technology include GPS units, garage door openers, wireless computer mice, keyboards and headsets, satellite television and cordless telephones.

Wireless communication can be via:

- radio frequency communication,
- microwave communication, for example long-range line-of-sight via highly directional antennas, or short-range communication, or
- infrared (IR) short-range communication, for example from remote controls or via Infrared Data Association (IrDA).

2.0 Objectives

At the end of this unit, you should be able to

(i) define wireless communication
(ii) mention examples of wireless communication equipment
(iii) state the application area of wireless communication
(iv) identify the use of wireless technology
(v) state the classes of wireless
(vi) explain a basic cellular system

3.0 Main Content

3.1 Concept of Wireless Communication

Wireless is a term used to describe telecommunications in which electromagnetic waves (rather than some form of wire) carry the signal over part or the entire communication path.

3.1.1 Wireless Equipment

Common examples of wireless equipment in use today include:

- Cellular phones and pagers: these provide connectivity for portable and mobile applications, both personal and business
- Global Positioning System (GPS): allows drivers of cars and trucks, captains of boats and ships, and pilots of aircraft to ascertain their location anywhere on earth.
- Cordless computer peripherals: the cordless mouse is a common example; keyboards and printers can also be linked to a computer via wireless.
- Cordless telephone sets: these are limited-range devices, not to be confused with cell phones.
- Home-entertainment-system control boxes: the VCR control and the TV channel control are the most common examples; some hi-fi sound systems and FM broadcast receivers also use this technology.

- Remote garage-door openers: one of the oldest wireless devices commonly use by consumers which are usually operated at radio frequencies.
- Two-way radios: this includes Amateur and Citizens Radio Service, as well as business, marine, and military communications.
- Baby monitors: these devices are simplified radio transmitter/receiver units with limited range.
- Satellite television: allows viewers in almost any location to select from hundreds of channels.
- Wireless LANs or local area networks: provide flexibility and reliability for business computer users

3.1.2 Examples of Wireless Communication and control

More specialized and exotic examples of wireless communications and control include:

- Global System for Mobile Communication (GSM): a digital mobile telephone system
- General Packet Radio Service (GPRS): a packet-based wireless communication service that provides continuous connection to the Internet for mobile phone and computer users.
- Enhanced Data GSM Environment (EDGE): a faster version of the Global System for Mobile (GSM) wireless service.
- Universal Mobile Telecommunications System (UMTS): a broadband, packet-based system offering a consistent set of services to mobile computer and phone users no matter where they are located in the world.
- Wireless Application Protocol (WAP): a set of communication protocols to standardize the way that wireless devices, such as cellular telephones and radio transceivers, can be used for Internet access.
- i-Mode: the world's first "smart phone" for Web browsing, first introduced in Japan; provides colour and video over telephone sets.

3.1.3 Classification of Wireless

Wireless can be divided into the following classes:

- Fixed wireless: the operation of wireless devices or systems in homes and offices, and in particular, equipment connected to the Internet via specialized modems
- Mobile wireless: the use of wireless devices or systems aboard motorized, moving vehicles; examples include the automotive cell phone and PCS (personal communications services)
- Portable wireless: the operation of autonomous, battery-powered wireless devices or systems outside the office, home, or vehicle; examples include handheld cell phones and PCS units
- IR wireless: the use of devices that convey data via IR (infrared) radiation; employed in certain limited-range communications and control systems

3.1.4 Applications of Wireless Technology

- **Security systems:** Wireless technology may supplement or replace hard wired implementations in security systems for homes or office buildings.
- **Television remote control:** Modern televisions use wireless (generally infrared) remote control units but now radio waves are also used.
- **Cellular telephony (phones and modems):** These instruments use radio waves to enable the operator to make phone calls from many locations world-wide. They can be used anywhere

there is a cellular telephone site to house the equipment that is required to transmit and receive the signal that is used to transfer both voice and data to and from these instruments.

- **WiFi :** Wi-Fi (for wireless fidelity) is a wireless LAN technology that enables laptop PC‘s, PDA‘s, and other devices to connect easily to the internet. Wi-Fi is less expensive and nearing the speeds of standard Ethernet and other common wire-based LAN technologies.
- **Wireless energy transfer:** Wireless energy transfer is a process whereby electrical energy is transmitted from a power source to an electrical load that does not have a built-in power source, without the use of interconnecting wires.
- **Computer Interface Devices:** Answering the call of customers frustrated with cord clutter, many manufactures of computer peripherals turned to wireless technology to satisfy their consumer base. Originally these units used bulky, highly limited transceivers to mediate between a computer and a keyboard and mouse, however more recent generations have used small, high quality devices, some even incorporating Bluetooth

3.2 Cellular System

A cellular system is a radio network made up of a number of radio cell (or just cells) each served by at least one fixed-location transceiver known as a cell site or base station. These cells cover different land areas to provide radio coverage over a wider area than the area of one cell, so that a variable number of portable transceivers can be used in any one cell and moved through more than one cell during transmission.

Cellular systems offer a number of advantages including but not limited to the following over alternative solutions:

- Increased capacity
- Reduced power usage
- Larger coverage area
- Reduced interference from other signals

3.2.1 Basic Cellular System

A basic cellular system is made up of three parts:

(i) a mobile unit

(ii) a cell site

(iii) Mobile Telephone Switching Office (MTSO) and with connections to link the three subsystems:

- **Mobile Unit** : A mobile telephone unit contains a control unit, a transceiver and an antenna system
- **Cell Site:** The cell site provides interface between the MTSO and the mobile unit. It has control unit, radio cabinets, antennas, a power plant and data terminals.
- **MTSO:** The switching office, the central coordinating element for all cell sites, contains the cellular processor and cellular switch. It interfaces with telephone company zone offices; controls call processing and handle billing activities.
- **Connections:** As a matter of fact, the radio and high speed data links connect the three subsystems. Each mobile unit can only use one channel at a time for its communication link.

However, the channel is not fixed i.e. it can be any one in the entire band assigned by the serving area, with each site having multi-channel capabilities that can connect simultaneously to many mobile units.

In addition to the above:

- The MTSO is the heart of the cellular mobile system. Its processor provides central coordination and cellular administration.
- The cellular switch, which can be either analog or digital, switches calls to connect mobile subscribers to other mobile subscribers and to the nationwide telephone network. It uses voice trucks similar to telephone company interoffice voice trucks. It also contains data links providing supervision links between the processor and the switch and between the cell sites and the processor.
- The radio link carries the voice and signal between the mobile unit and the cell site.
- The high-speed data links cannot be transmitted over the standard telephone trucks and therefore must use either microwave links or T-carriers (wire lines).
- Microwave radio links or T-carriers carry both voice and data between the cell site and the MTSO.

3.2.2 Advantages of Cellular Systems

The advantages of operating in a cellular arrangement are listed below:

(i) the use of low power transmitter, and
(ii) an allowance for frequency reuse

Frequency reuse needs to be structured so that co-channel interference is kept at an acceptable level. As the distance between co-channel cells increases co-channel interference will decrease. If the cell size is fixed, the average signal-to-co-channel interference ratio will be independent of the transmitted power of each cell. The distance between any two co-channel cells can be examined by making use of the geometry of hexagonal cells.

3.2.3 Interference and System Capacity

Interference is the major limiting factor in the performance of the cellular radio systems. Sources of the interference include the following cases:

- Another mobile in the same cell
- A call in progress in the neighbouring cell
- Other base stations operating in the same frequency band
- Any non-cellular system which inadvertently leaks energy into cellular frequency band.

Interference on the voice channels causes crosstalk, where the subscriber hears interference in the background due to an undesired transmission. On control channels, interference leads to missed and blocked call due to errors in the digital signaling. Interference is more severe in urban areas due to the greater radio frequency (RF) noise floor and the large number of base stations and the mobiles. Interference has been recognized as a major bottleneck in increasing capacity.

The two major types of system generated cellular interference are:

(i) Co-channel interference
(ii) Adjacent channel interference

4.0 Conclusion

This unit has introduced you to the concept and applications of wireless communication. You have also been informed on cellular system.

5.0 Summary

The main points in this unit include the following:

- Wireless communication is the transfer of information over a distance without the use of electrical conductors or "wires".
- It encompasses various types of fixed, mobile, and portable two-way radios, cellular telephones, personal digital assistants (PDAs), wireless networking, etc.
- A cellular system is a radio network made up of a number of radio cell (or just cells) each served by at least one fixed-location transceiver known as a cell site or base station.
- A basic cellular system is made up of three parts: a mobile unit, cell site, and mobile Telephone Switching Office (MTSO) and with connections to link the three subsystems

6.0 Tutor-Marked Assignment

(i) What is wireless communication
(ii) Mention five examples of wireless communication equipment.
(iii) List five uses of wireless technology
(iv) Briefly explain the application area of wireless communication
(v) Outline the four classes of wireless
(vi) What do you understand about a basic cellular system?

7.0 References/Further Readings

Goldsmith A. (2005). Wireless Communications, Cambridge University Press. ISBN-13:9780521837163

Lior Baruch (2008). Mobile Computing Definition –Wireless.

O'Brien, J. & Marakas, G.M.(2008). Management Information Systems. New York, NY: McGraw-Hill Irwin.

Sharma S. (2006): Wireless & Cellular Communications, New Delhi: S.K. Kataria & Sons.
Shea J. (2000). Brief History of Wireless Communication.

UNIT 2: MOBILE RADIO PROPAGATION

1.0 Introduction

Mobile communication is burdened with particular propagation complications, making reliable wireless communication more difficult than fixed communication between carefully positioned antennas. The antenna height at a mobile terminal is usually very small, typically less than a few meters. Hence, the antenna is expected to have very little 'clearance', so obstacles and reflecting surfaces in the vicinity of the antenna have a substantial influence on the characteristics of the propagation path. Moreover, the propagation characteristics change from place to place and, if the terminal moves, from time to time.

2.0 Objectives

At the end of this unit, you would be able to:

(i) discuss the statistical propagation models
(ii) describe the propagation mechanism

3.0 Main Content

3.1 Statistical propagation models

In generic system studies, the mobile radio channel is usually evaluated from 'statistical' propagation models: no specific terrain data is considered and channel parameters are modeled as stochastic variables. Three mutually independent, multiplicative propagation phenomena can usually be distinguished: 'large-scale' path loss, shadowing and multipath fading.

3.1.1 Path-loss

The 'large-scale' effect causes the received power to vary gradually due to signal attenuation determined by the geometry of the path profile in its entirety. This is in contrast to the local propagation mechanisms, which are determined by terrain features in the immediate vicinity of the antennas.

Free Space Propagation Model for Mobile Communication

The free space propagation model is a model which is used to predict received signal strength at a particular location when the transmitter and receiver have a clear, unobstructed line-of-sight path between them. For example, satellite communication systems and microwave line-of-sight radio links, typically, undergo free space propagation. Like other large-scale radio-wave propagation models, the free space-radio propagation model predicts that the received power decays as a function of the transmitter-receiver separation distance raised to some power (i.e. a power law function). It states that power falls off proportional to distance (d).

Figure 2.1: Line-of-sight

3.1.2 Shadowing

This is a 'medium-scale' effect: field strength variations occur if the antenna is displaced over distances larger than a few tens or hundreds of metres.

Shadowing is the effect that the received signal power fluctuates due to objects obstructing the propagation path between transmitter and receiver. If there are any objects (such buildings or trees) along the path of the signal, some part of the transmitted signal is lost through absorption, reflection, scattering, and diffraction. This effect is called shadowing. As shown below, if the base antenna were a light source, the middle building would cast a shadow on the subscriber antenna. Hence, the name shadowing.

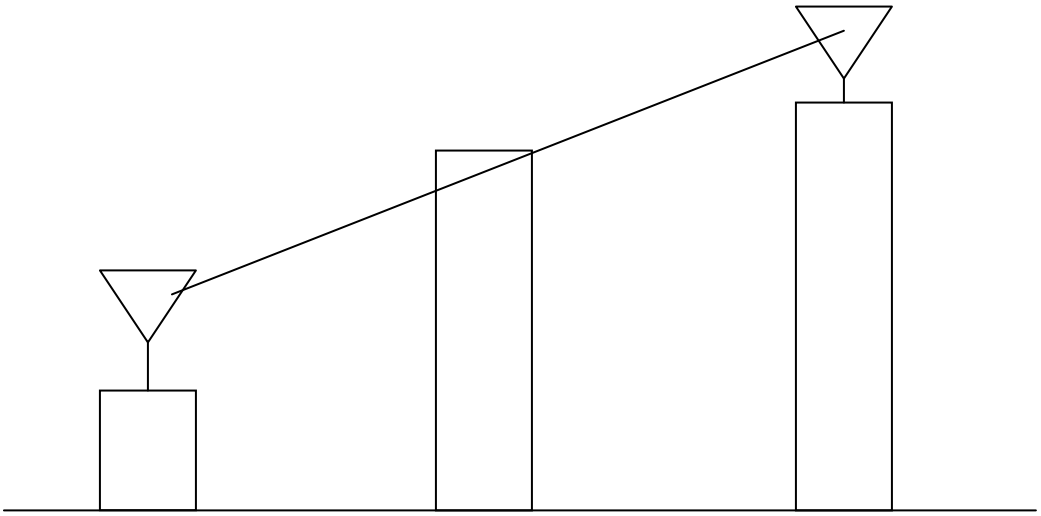
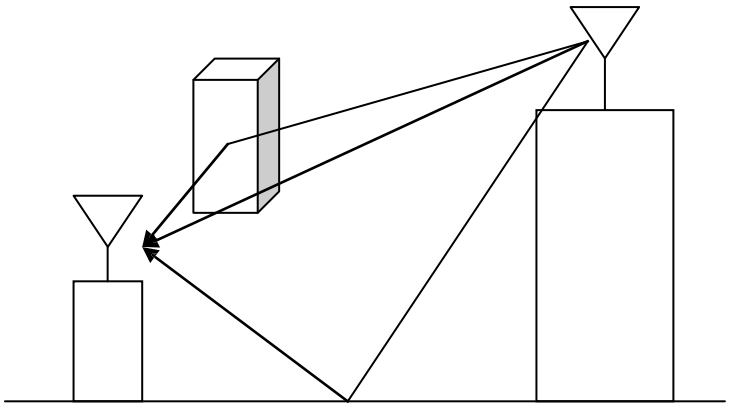


Figure 2.2: Shadowing

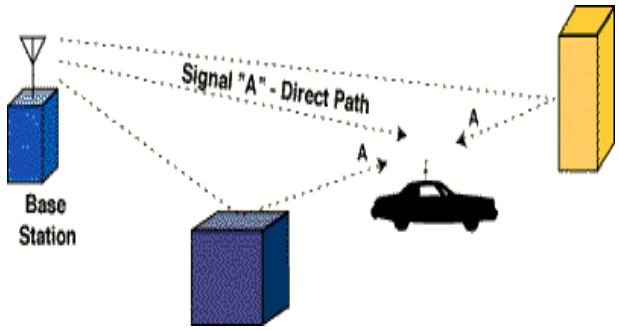
3.1.3 Multipath-propagation

Small scale fading or in simple word fading is used to describe the rapid fluctuations in amplitudes, phases, or mutipath delays of a radio signal over a small period of time or travel distance, so that large-scale path loss effects may be ignored. Multipath fading thus has a 'small-scale' effect.

The objects located around the path of the wireless signal reflect the signal. Some of these reflected waves are also received at the receiver. Since each of these reflected signals takes a different path, it has a different amplitude and phase.



(a)



(b)

Figure 2.3 (a &b): Example of Multipath

3.2 The Three Basic Propagation Mechanisms

In order to describe radio propagation, three basic mechanisms are generally considered. These mechanisms are:

- **Reflection:** occurs when a propagating electromagnetic wave impinges upon an object which is very large in dimensions when compared to the wavelength of the propagation wave e.g. the surface of the earth, buildings, walls, etc. These mechanisms often dominate radio propagation in indoor applications. In outdoor urban areas, this mechanism often loses its importance because it involves multiple transmissions that reduce the strength of the signal to negligible values.
- **Diffraction:** occurs when the radio path between the Transmitter (Tx) and Receiver (Rx) are obstructed by a surface that has irregularities (edges). The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacle, giving rise to a bending of waves around the obstacle, even when a line-of-sight path does not exist between transmitter and receiver. At high frequencies, diffraction, like reflection, depends on the geometry of the object, as well as the amplitude, phase, and polarization of the incident wave at the point of diffraction.

- **Scattering:** occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength and where the number of obstacles per unit volume is large. Scattered waves are produced by rough surfaces, small objects, or by other irregularities in the channel. In practice, foliage, street signs, and lamp posts induce scattering in a mobile communications system."

4.0 Conclusion

In this unit, you have learnt about the propagation model; pathloss, shadowing and multipath propagation. You have also learnt on the basic propagation mechanism

5.0 Summary

The main points in this unit are:

- The 'large-scale' effect causes the received power to vary gradually due to signal attenuation determined by the geometry of the path profile in its entirety.
- Shadowing is the effect that the received signal power fluctuates due to objects obstructing the propagation path between transmitter and receiver.
- Shadowing has a ‗medium-scale' effect.
- Multipath fading has a 'small-scale' effect.
- Reflection occurs when a propagating electromagnetic wave impinges upon an object which is very large in dimensions when compared to the wavelength of the propagation wave
- Diffraction: occurs when the radio path between the Transmitter (Tx) and Receiver (Rx) are obstructed by a surface that has irregularities (edges).
- Scattering: occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength, and where the number of obstacles per unit volume is large.

6.0 Tutor-Marked Assignment

(ii) discuss the statistical propagation models
(ii) describe three basic propagation mechanism

7.0 References/Further Readings

Goldsmith A. (2005): Wireless Communications, Cambridge University Press

Linnartz's J.M.G. (1996). Wireless Communication, Baitzer Science Publishers

Molisch A.F. (2005): Wireless Communications, Wiley and Associated

Rappaport T.S. (2005). Wireless Communication :Principle and Practice. 2nd Edition, India : Prentice Hall

Sharma S. (2007): Wireless and Cellular Communication. New Delhi: S.K. Kataria & sons.

UNIT 3: MODULATION, DIVERSITY AND MULTIPLE ACCESS TECHNIQUES

1.0 Introduction

This unit provides you with a brief review on modulation, diversity and multiple access techniques

2.0 Objectives

At the end of the unit, you would be able to

(i) define a modulation
(ii) outline the types of modulation
(iii) state the classes of diversity
(iv) explain Multiple Access concept
(v) discuss on the types of Multiple Access Techniques

3.0 Main Content

3.1 Modulation

3.1.1 Definition of Modulation

Modulation is the process of encoding information from a message source into manner suitable for transmission. It generally involves translating a baseband message signal (called the source) to a bandpass signal at frequencies that are very high when compared to the baseband frequency. The **bandpass** signal is called the modulated signal and the **baseband** signal is called the modulating signal. Modulation may be done by varying the amplitude, phase or frequency of a high frequency carrier in accordance with the amplitude of the message signal.

3.1.2 Types of Modulation Techniques

- Analog modulation: The aim of analog modulation is to transfer an analog baseband (or lowpass) signal, for example an audio signal or TV signal, over an analog passband channel, for example a limited radio frequency band or a cable TV network channel.
- BandPass Digital modulation: The aim of digital modulation is to transfer a digital bit stream over an analog passband channel, for example over the public switched telephone network (where a bandpass filter limits the frequency range to between 300 and 3400 Hz), or over a limited radio frequency band.
- Digital baseband modulation or line coding modulation: The aim of digital baseband modulation methods, also known as line coding, is to transfer a digital bit stream over a baseband channel, typically a non-filtered copper wire such as a serial bus or a wired local area network.
- Pulse shaping modulation: The aim of pulse modulation methods is to transfer a narrowband analog signal, for example a phone call over a wideband baseband channel or, in some of the schemes, as a bit stream over another digital transmission system.

3.2 Diversity

3.2.1 Diversity Concept

Diversity reception reduces the probability of occurrence of communication failures (outages) caused by fades by combining several copies of the same message received over different channels. In telecommunications, diversity scheme is referred to as a method for improving the reliability of a message signal by using two or more communication channels with different characteristics. Diversity plays an important role in combating fading and co-channel interference and avoiding error bursts. Diversity techniques may exploit the multipath propagation, resulting in a diversity gain, often measured in decibels.

3.2.2 Classes of Diversity Scheme

The following classes of diversity schemes can be identified:

(a) **Time diversity**: Multiple versions of the same signal are transmitted at different time instants. Alternatively, a redundant forward error correction code is added and the message is spread in time by means of bit-interleaving before it is transmitted. Thus, error bursts are avoided, which simplifies the error correction. In time diversity, the time difference between two transmissions should be large compared to the time it takes the mobile antenna to move half a wavelength. In systems with stationary antennas, such as indoor wireless communication, time diversity will be less effective as the channel characteristics do not change very much with time. However, time diversity may be helpful if uncorrected interference signals are experienced during successive attempts.

(b) **Frequency diversity**: In frequency diversity, the same message is transmitted more than once, respectively at different carrier frequencies. The difference in carrier frequency should be more than the coherence bandwidth to achieve effective diversity. Digital cellular system can use slow frequency hopping (SFH) for diversity reason: each block of bits is transmitted at a different carrier.

(c) **Angle diversity:** The desired message is received simultaneously by several directed antennas pointing in widely different directions. The received signal consists of scattering wave coming from all directions. It has been observed that the scattered signals associated with the different (non-overlapping) directions are uncorrelated. Angle diversity can be viewed as a special case of space diversity since it also requires multiple antennas.

(a) **Multiuser diversity**: Multiuser diversity is obtained by opportunistic user scheduling at either the transmitter or the receiver. Opportunistic user scheduling is as follows: the transmitter selects the best user among candidate receivers according to the qualities of each channel between the transmitter and each receiver. In Frequency Division Duplex (FDD) systems, a receiver must feed back the channel quality information to the transmitter with the limited level of resolution.

(e) **Co-operative diversity**: is a co-operative multiple antenna techniques which exploit user diversity by decoding the combined signal of the relayed signal and the direct signal in wireless multihop networks. A conventional single hop system uses direct transmission where a receiver decodes the information only based on the direct signal while regarding the relayed signal as interference, whereas the cooperative diversity considers the other signal as contribution. That is, cooperative diversity decodes the information from the combination of two signals. Hence, it can be seen that cooperative diversity is an antenna diversity that uses distributed antennas belonging to each node in a wireless network.

(f) **Space diversity**: also known as **Antenna diversity** is any one of several wireless diversity schemes that use two or more antennas to improve the quality and reliability of a wireless link. Often, especially in urban and indoor environments, there is no clear line-of-sight (LOS) between transmitter and receiver. Instead the signal is reflected along multiple paths before finally being received. Each of these bounces can introduce phase shifts, time delays, attenuations and even distortions that can destructively interfere with one another at the aperture of the receiving antenna. Antenna diversity is especially effective at mitigating these multipath situations.

3.3 Multiple Access

3.3.1 Definition of Multiple Access

Multiple Access is a signal transmission situation in which two or more users wish to simultaneously communicate with each other using the same propagation channel. This is precisely the uplink transmission situation in a wireless communication. In the uplink or reverse channel, multiple users will want to transmit information simultaneously. Without proper coordination among the transmitting users, collisions will occur when two or more users transmit simultaneously. Access methods that incur collision are referred to as random access and variants of random access. Multiple access method allows several terminals connected to the same multi-point transmission medium to transmit over it and to share its capacity.

3.3.2 Multiple Access Techniques

(a) **Frequency Division Multiple Access (FDMA)**

FDMA is based on frequency-division multiplex (FDM). It gives users an individual allocation of one or several frequency bands, or Channels. FDMA provides different frequency bands to different users or nodes. In FDMA, the total bandwidth is divided into non-overlapping frequency subband. Each user is allocated a unique frequency subband for the duration of the connection, whether the connection is an active or idle state. Othogonality among transmitted signals from different mobile users is achieved by bandpass filtering in the frequency domain. This type of multiple access support is narrowband, and is not suitable for multimedia communications with various transmission rates. In addition, it incurs a waste of bandwidth when the user is in a dormant state. An example of FDMA systems were the first-generation (1G) cell-phone systems. A related technique is wave-length division multiple access (WDMA), based on Wavelength division multiplex (WDM), where different users get different colors in fiber-optical communication.

(b) **Time-Division Multiple Access (TDMA)**

TDMA is based on time-division multiplex (TDM). It allows several users to share the same frequency channel by dividing the signal into different time slots. The users transmit in rapid succession, one after the other, each using his own time slot. This allows multiple stations to share the same transmission medium (e.g. radio frequency channel) while using only a part of its channel capacity.

TDMA provides different time-slots to different transmitters in a cyclically repetitive frame structure. In a TDMA, the channel time is partitioned into frames. The length of a frame is long enough so that every user in service has an opportunity to transmit once per frame. To achieve this, a TDMA frame is further partitioned into time slots. Users have to transmit in their assigned slots from frame to frame. For example, user 1 may use time slot 1, user 2 time slot 2, etc until the last user. Then it starts all over again. TDMA is used in the digital 2G cellular systems such as Global System for Mobile Communications (GSM), IS-136, Personal Digital Cellular (PDC) and in the Digital Enhanced Cordless Telecommunications (DECT) standard for portable phones.

It is also used extensively in satellite systems, and combat-net radio systems. Multi-Frequency Time Division Multiple Access (MF-TDMA) is one of the types of TDMA.

(c) Code Division Multiple Access (CDMA), or Spread spectrum multiple access (SSMA)

CDMA is a spread spectrum multiple access method. The principle of spread spectrum communications is that the bandwidth of the baseband information-carrying signals from the different users is spread by different signals with a bandwidth much larger than that of the baseband signals. Ideally, the spreading signals used for different users are orthogonal to each other. Thus, at the receiver, the same spreading signal is used as the dispreading signals to coherently extract the baseband signal from the target user, while suppressing the transmissions from any other users. An example of CDMA is the 3G cell phone system.

(d) Space Division Multiple Access (SDMA)

SDMA enables creating parallel spatial pipes next to higher capacity pipes through spatial multiplexing and/or diversity, by which it is able to offer superior performance in radio multiple access communication systems. In traditional mobile cellular network systems, the base station has no information on the position of the mobile units within the cell and radiates the signal in all directions within the cell in order to provide radio coverage. These results in wasting power on transmissions when there are no mobile units to reach, in addition to causing interference for adjacent cells using the same frequency, so called co-channel cells. Likewise, in reception, the antenna receives signals coming from all directions including noise and interference signals. By using smart antenna technology and by leveraging the spatial location of mobile units within the cell, space-division multiple access techniques offer attractive performance enhancements. The radiation pattern of the base station, both in transmission and reception is adapted to each user to obtain highest gain in the direction of that user. This is often done using phased array techniques.

4.0 Conclusion

In this unit, you have learnt about modulation, diversity and multiple access as well as their various techniques

5.0 Summary

The main points in this unit are:

- modulation is the process of encoding information from a message source into manner suitable for transmission.
- modulation techniques are: analog modulation, bandpass digital modulation, digital baseband modulation and pulse shaping modulation
- diversity scheme is referred to as a method for improving the reliability of a message signal by using two or more communication channels with different characteristics.
- diversity plays an important role in combating fading and co-channel interference and avoiding error bursts.
- multiple access is a signal transmission situation in which two or more users wish to simultaneously communicate with each other using the same propagation channel.
- multiple access techniques are FDMA, TDMA, CDMA and SDMA

6.0 Tutor-Marked Assignment

(i) Write on the following: modulation, diversity and multiple access

(ii) Mention four types of modulation techniques
(iii) State the classes of diversity
(iv) Discuss on any three types of multiple access techniques

7.0 References/Further Readings

Bloomfield L.A. (2000). How Things Work: The Physics of Everyday Life. Second Edition, New York: John Wiley & Sons.

Davidovits P. (1972). Communication. New York: Holt Rinehart and Winston, inc.

Dietrich C. (2000): Adaptive Arrays and Diversity Antenna Configurations for Handheld Wireless Communication Terminals‖.

Laneman J. N., Tse D. N. C., and Wornell G. W.(2004). ―Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior,‖ IEEE Trans. Inform. Theory, 50(12), 3062-3080.

Linnartz, J.P.M.G. (1995). Cellular Telephone Network. (Online) Available at http://wireless.per.nl/reference/about.htm

Linnartz J.P.M.G (2001). ―Performance Analysis of Sychronous MC-CDMA in mobile Rayleigh Channels with both Delay and Doppler Spreads‖, IEEE, 50(6), 1375-1387.

Moon J. and Kim. Y. (2003) ―Antenna Diversity Strengthens Wireless LANs.‖ Communication Systems Design.

Sendonaris A, Erkip E, and Aazhang B. (2003). ―User Cooperation Diversity Part I and Part II,‖ IEEE Trans. Communication, 51(11),1927–48.

Sharma S. (2006): Wireless & Cellular Communications, New Delhi: S.K. Kataria & Sons.

UNIT 4: CELLULAR WIRELESS: EVOLUTION OF GSM TECHNOLOGY

1.0 Introduction

The explosive growth of Global System for Mobile (GSM) Communication services over the last two decades has changed mobile communications from a niche market to a fundamental constituent of the global telecommunication markets. GSM is a digital wireless technology standard based on the notion that users want to communicate wirelessly without limitations created by network or national borders. In a short period of time, GSM has become a global phenomenon. The explanation for its success is the cooperation and coordination of technical and operational evolution that has created a virtuous circle of growth built on three principles: interoperability based on open platforms, roaming and economies of scale.

2.0 Objectives

At the end of this unit, you should be able to:

(i) discuss on the cellular network generation

(ii) state the technology behind each generation

3.0 Main Content

3.1 Generation of Cellular Wireless

The generations of cellular wireless are:

First Generation (1G)

Second Generation (2G)

Third Generation (3G)

Fourth Generation (4G)

3.2 First-Generation (1G) Networks

1G (or 1-G) refers to the first-generation of wireless telephone technology, mobile telecommunications. The first-generation (1G) cellular systems were the simplest communication networks deployed in the 1980s. The first-generation networks were based on analogue-frequency-modulation transmission technology. The technological development that distinguished the First Generation mobile phones from the previous generation was the use of multiple cell sites and the ability to transfer calls from one site to the other as the user travelled between cells during a conversation.

Figure 4.1: Analog Motorola DynaTAC 8000X mobile phone as of 1983

Disadvantages
Challenges faced by the operators included:

- inconsistency, frequent loss of signals and low bandwidth
- 1G network was also expensive to run due to a limited customer base.

3.3 Second-Generation (2G) Networks

The second-generation (2G) cellular systems were the first to apply digital transmission technologies for voice and data communication. The data transfer rate was in the region of 10s of Kbps. Other examples of technologies in 2G systems include frequency-division multiple access (FDMA), time-division multiple access (TDMA) and code-division multiple access.

Figure 4.2: Two 1991 GSM mobile phones with several AC adapters

3.3.1 2G Technologies

2G technologies can be divided into TDMA-based and CDMA-based standards depending on the type of multiplexing used. The main 2G standards are:

- GSM (TDMA-based)
- IS-95 aka cdmaOne
- PDC (TDMA-based)
- iDEN (TDMA-based)
- IS-136 aka D-AMPS (TDMA-based)

3.3.2 Capacity

Using digital signals between the handsets and the towers increases system capacity in two key ways:

- digital voice data can be compressed and multiplexed much more effectively than analog voice encodings through the use of various codecs, allowing more calls to be packed into the same amount of radio bandwidth.
- the digital systems were designed to emit less radio power from the handsets. This meant that cells could be smaller, so more cells could be placed in the same amount of space. This was also made possible by cell towers and related equipment getting less expensive.

3.3.3 Advantages

- The lower power emissions helped address health issues.
- Going all-digital allowed for the introduction of digital data services, such as SMS and email.
- Greatly reduced fraud. With analog systems it was possible to have two or more "cloned" handsets that had the same phone number.

- Enhanced privacy. A key digital advantage not often mentioned is that digital cellular calls are much harder to eavesdrop on by use of radio scanners. While the security algorithms used have proved not to be as secure as initially advertised, 2G phones are immensely more private than 1G phone, which have no protection against eavesdropping.
- Three primary benefits of 2G networks over their predecessors were that:
  - phone conversations were digitally encrypted;
  - 2G systems were significantly more efficient on the spectrum allowing for far greater mobile phone penetration levels; and
  - 2G introduced data services for mobile, starting with SMS text messages.

3.3.4 Disadvantages

- In less populous areas, the weaker digital signal may not be sufficient to reach a cell tower. This tends to be a particular problem on 2G systems deployed on higher frequencies, but is mostly not a problem on 2G systems deployed on lower frequencies. National regulations differ greatly among countries which dictate where 2G can be deployed.
- Analog has a smooth decay curve, digital a jagged steppy one. This can be both an advantage and a disadvantage. Under good conditions, digital will sound better. Under slightly worse conditions, analog will experience static, while digital has occasional dropouts. As conditions worsen, though, digital will start to completely fail, by dropping calls or being unintelligible, while analog slowly gets worse, generally holding a call longer and allowing at least a few words to get through.
- While digital calls tend to be free of static and background noise, the lossy compression used by the codecs takes a toll; the range of sound that they convey is reduced. You'll hear less of the tonality of someone's voice talking on a digital cell phone, but you will hear it more clearly.

3.3.5 Evolution

The second-generation networks deliver high-quality and secure mobile voice and basic data services such as fax and text messaging along with full roaming capabilities across the world.

2.5G

To address the poor data transmission rates of the 2G network, developments were made to upgrade 2G networks without replacing the networks. These technological enhancements were called 2.5G technologies and include networks such as General Packet Radio Service (GPRS).
GPRS-enabled networks deliver features such as always-on, higher capacity, Internet-based content and packet-based data services enabling services such as colour Internet browsing, e-mail on the move, visual communications, multimedia messages, and location-based services.
Another 2.5G network enhancement of data services is high-speed circuit-switched data (HSCSD). This allows access to non-voice services 3 times faster than conventional networks, which means subscribers are able to send and receive data from their portable computers at speeds of up to 28.8 Kbps; this is currently being upgraded in many networks to 43.2 Kbps. The HSCSD solution enables higher rates by using multiple channels, allowing subscribers to enjoy faster rates for their Internet, e-mail, calendar and file-transfer services.

2.75G (EDGE)

GPRS networks evolved to Enhanced Data Rates for GSM Evolution (EDGE) networks with the introduction of 8PSK encoding. This network upgrade offers similar capabilities as those of the GPRS network. The EDGE provides a potential three-fold increase in capacity of GSM/GPRS networks. The specification achieves higher data-rates (up to 236.8 kbit/s) by switching to more sophisticated methods of coding (8PSK), within existing GSM timeslots.

3.4 Third-Generation (3G) Networks

The most promising period is the advent of third-generation (3G) networks. These networks are also referred to as the universal mobile telecommunications systems (UMTSs). The global standardization effort undertaken by the International Mobile Telecommunications (ITU) is called IMT-2000. The aim of the group was to evolve today‘s circuit-switched core network to support new spectrum allocations and higher bandwidth capability. Over 85% of the world‘s network operators have chosen 3G as the underlying technology platform to deliver their third-generation services.

The main technological difference that distinguishes 3G technology from 2G technology is the use of packet switching rather than circuit switching for data transmission. In addition, the standardization process focused on requirements more than technology (2 Mbit/s maximum data rate indoors, 384 kbit/s outdoors, for example).

Application services include wide-area wireless voice telephone, mobile Internet access, video calls and Multimedia Broadcast Multicast Service |mobile TV. Compared to the older 2G and 2.5G standards, a 3G system must provide peak data rates of at least 200 kbit/s according to the IMT-2000 specification. Recent 3G releases, often denoted 3.5G and 3.75G, also provide mobile broadband access of several Mbit/s to laptop computers and smart phones.

The following standards comply with the IMT2000/3G standard:

- EDGE, a revision by the (3GPP) to 2G GSM's existing GPRS transmission methods, utilizing the exact same hardware and frequencies as GPRS. This makes it unique compared to the other common UMTS and CDMA2000 standards, as GSM is a standard utilizing TDD for encoding data.
- Universal Mobile Telecommunications System, created and revised by the 3GPP (originally released in 2001). The family is a full revision of GSM encoding methods and hardware, used primarily in Europe and Asia-Pacific. The cell phones used utilize UMTS in combination with 2G GSM standards and bandwidths, but not EDGE.
- the CDMA2000 system or IS-2000, updating the IS-95 CDMA system, first offered in 2002, standardized by 3GPP2 (differing from the 3GPP, which is responsible for GSM and UMTS), used especially in North America and South Korea, sharing infrastructure with the IS-95 2G standard. The cell phones are typically CDMA2000 and IS-95 hybrids. The latest release EVDO Rev B offers downstream peak rates of 14.7 Mbit/s.

3.5 Future Trends: Fourth-Generation Mobile Networks

The fourth-generation (4G) systems are expected around 2010 to 2015. They will be capable of combining mobility with multimedia-rich content, high bit rates and Internet-protocol (IP) transport.

The benefits of the fourth-generation approach are voice-data integration, support for mobile and fixed networking and enhanced services through the use of simple networks with intelligent terminal devices. The fourth-generation networks are expected to offer a flexible method of

payment for network connectivity that will support a large number of network operators in a highly competitive environment.
Currently, the Internet is used solely to interconnect computer networks; IP compatibility is being added to many types of devices such as set-top boxes, automotive systems, and home electronics. The large-scale deployment of IP-based networks will reduce the acquisition costs of the associated devices. The future vision is to integrate mobile voice communications and Internet technologies, bringing the control and multiplicity of Internet-applications services to mobile users.
The creation and deployment of IP-based multimedia services (IMSs) allows person-to-person real-time services, such as voice over the 3G packet-switched domain. IMS enables IP interoperability for real-time services between fixed and mobile networks, solving current problems of seamless, converged voice-data services. Service transparency and integration are key features for accelerating end-user adoption. Two important features of IMS are IP-based transport for both real-time and non-real-time services and a multimedia call-model based on the session-initiation protocol (SIP). The deployment of an IP-based infrastructure will encourage the development of voice-over-IP (VoIP) services.
One of the main ways in which 4G differed technologically from 3G was in its elimination of circuit switching, instead employing an all-IP network. Thus, 4G ushered in a treatment of voice calls just like any other type of streaming audio media, utilizing packet switching over internet, LAN or WAN networks via VoIP.

4.0 Conclusion

In just over two decades, mobile network technologies have evolved from simple 1G network to today‘s 3G networks, which are capable of high-speed data transmission allowing innovative applications and services. The evolution of the communication networks is fueling the development of the mobile Internet and creating new types of devices. In the future, 4G networks will supersede 3G.
The fourth-generation technology supports broadly similar goals to the third-generation effort, but starts with the assumption that future networks will be entirely packet-switched using protocols evolved from those in use in today‘s Internet. Today‘s Internet telephony systems are the foundation for the applications that will be used in the future to deliver innovative telephony services.

5.0 Summary

In this unit you have learnt that:

- the first-generation networks were based on analogue-frequency-modulation transmission technology.
- the second-generation (2G) cellular systems were the first to apply digital transmission technologies for voice and data communication.
- the main technological difference that distinguishes 3G technology from 2G technology is the use of packet switching rather than circuit switching for data transmission.
- one of the main ways in which 4G differed technologically from 3G was in its elimination of circuit switching, instead employing an all-IP network.

6.0 Tutor-Marked Assignment

Write short note on the following cellular network generation:

(i) first generation
(ii) second generation
(iii) third generation and
(iv) fourth generation

7.0 References/Further Readings

Olla P. (2011). Evolution of GSM Network Technology. http://encyclopedia.jrank.org/articles/pages/6603/Evolution-of-GSM-Network-Technology.html"

Saeed F. A. (2006). "Capacity Limit Problem in 3G Networks". Purdue School of Engineering. http://www.ece.iupui.edu/~dskim/Classes/ECE695MWN/2006-saeed-Capacity_Limit_Problem_in_3G_Networks.ppt.

Sharma S. (2006). Wireless & Cellular Communications, New Delhi: S.K. Kataria & Sons.

Tom F. (2007). "The Cell-Phone Revolution". *American heritage of invention & technology.* New York: American Heritage, 22 (3):8–19. ISSN 8756-7296. OCLC 108126426. BL Shelfmark 0817.734000. http://www.americanheritage.com/events/articles/web/20070110-cell-phone-att-mobile-phone-motorola-federal-communications-commission-cdma-tdma-gsm.shtml.

Online Citation

http://www.cdg.org/worldwide/index.asp?h_area=0&h_technology=999&h_frequency=1.
http://www.elisa.com/english/index.cfm?t=6&o=6532.50.

MODULE 2: SATELLITE COMMUNICATION

UNIT 1: BASIC CONCEPTS OF SATELLITE COMMUNICATION

1.0 Introduction

Satellite communication systems are now a major part of most telecommunications networks as well as every-day lives through mobile personal communication systems and broadcast television. A fundamental understanding of such systems is therefore important for a wide range of system designers, engineers and users.

What is a Satellite?

A Satellite is an object that orbits another object like planet.

What is Satellite Communication?

A Satellite Communication is an artificial satellite stationed in space that is used for the purpose of telecommunications. It is a specialized wireless receiver/transmitter — receiving radio waves from one location and transmitting them to another (also known as a ‒bent pipe‖) — that is launched by a rocket and placed in orbit around the earth. Today, there are hundreds of commercial satellites in operation around the world. Those satellites are used for such diverse purposes as wide-area network communications, weather forecasting, television and radio broadcasting, amateur radio communications, Internet access and the Global Positioning System. Satellites have many important uses, not just communications. Most modern weather reports rely on satellite information. Global Positioning systems work because of a linked set of satellites. Scientific studies of our planet, the atmosphere and the universe all rely on satellites.

2.0 Objectives

At the end of this unit, you should be able to:

(i) explain a satellite communication
(ii) mention the advantages and disadvantages of satellite communication
(iii) state how satellite are used.

3.0 Main Content

3.1 Satellite Communication

Most satellites communications are in geostationary. A single geostationary satellite can cover as much as 40 percent of the earth's surface; so, in theory, three such satellites can provide global

coverage. To ensure accurate and strong coverage of a specific region, continent or country, the transponders are often ―shaped‖ to focus transmission and increase signal strength for a service area.

A satellite‘s job in the communications network is to serve as a repeater. That is, it receives a signal from one location and broadcasts same to another station. Reception and retransmission are accomplished by a transponder. A single transponder on a geostationary satellite is capable of handling approximately 5,000 simultaneous voice or data channels. A typical satellite has 32 transponders.

Each of the Transponders work on a specific radio frequency wavelength, or ―band.‖ Satellite communications work on three primary bands: C, Ku and Ka. C was the first band used and, as a longer wavelength, requires a larger antenna. Ku is the band used by most current VSAT systems. Ka is a new band allocation that isn‘t yet in wide use. Of the three, it has the smallest wavelength and can use the smallest antenna.

Because of attenuation and business competition, there are far more than three GEO satellites. Satellites of similar frequency can be as close as 3 degrees apart without causing interference. Since there are 360 degrees in a circle, that means 120 satellites of a specific frequency can be placed in GEO orbits.

The combination of individual transponder volumes and the number of transponders in orbit means today's communication satellites are an ideal medium for transmitting and receiving almost any kind of content, from simple data to the most complex and bandwidth-intensive video, audio and data content.

Figure 1.1: U.S. military WGSS communications satellite

Satellite communications are:

- highly survivable (physical survivability and robustness)
- independent of terrestrial infrastructure
- able to provide the load sharing and surge capacity solution for larger sites
- best for redundancy: they add a layer of path diversity and link availability

3.2 Satellite systems perform effectively when:

- terrestrial infrastructure is damaged, destroyed, or overloaded
- interconnecting widely distributed networks
- providing interoperability between disparate systems and networks
- providing broadcasting services over very wide areas such as a country, region, or entire hemisphere
- providing connectivity for the ―last mile‖ in cases where fiber networks are simply not available
- providing mobile/transportable wideband and narrow-band communications
- natural disasters or terrorist attacks occur. satellites are the best and most reliable platform for communications in such situations — fiber networks or even terrestrial wireless can be disrupted by tsunamis, earthquakes, or hurricanes. Satellites are instant infrastructure.

3.2 How do Satellites Work

Two Stations on Earth want to communicate through radio broadcast but are too far away to use conventional means.

- The two stations can use a satellite as a relay station for their communication
- One **Earth Station** sends a transmission to the satellite. This is called an **Uplink**.
- The satellite **Transponder** converts the signal and sends it down to the second earth station. This is called a **Downlink**.

3.3 Advantages of Satellites

The advantages of satellite communication over terrestrial communication are:

- Ubiquitous Coverage: A group of satellites can cover virtually all the Earth‘s surface. The coverage area of a satellite greatly exceeds that of a terrestrial system.
- Instant Infrastructure: Satellite service can be offered in areas where there is no terrestrial infrastructure and the costs of deploying a fiber or microwave network are prohibitive. It can also support services in areas where existing infrastructure is outdated, insufficient, or damaged.
- Independent Of Terrestrial Infrastructure: Satellite service can provide additional bandwidth to divert traffic from congested areas, provide overflow during peak usage periods and provide redundancy in the case of terrestrial network outages.
- Temporary Network Solutions: For applications such as news gathering, homeland security or military activities, satellite often provide the only practical, short-term solution for getting necessary information in and out.
- Rapid Provisioning of Services: Since satellite solutions can be set up quickly, communications networks and new services can be quickly recovered and reconfigured. In addition, you can expand services electronically without traditional terrestrial networks. As a result, you can achieve a high level of communications rapidly without high budget expenditures.
- Transmission cost of a satellite is independent of the distance from the center of the coverage area.
- Satellite to Satellite communication is very precise.
- Higher Bandwidths are available for use.

3.4 Disadvantages of Satellites

The disadvantages of satellite communication:

- Launching satellites into orbit is costly.
- Satellite bandwidth is gradually becoming used up.
- There is a larger propagation delay in satellite communication than in terrestrial communication.
- Satellite systems have a number of physical and electronic security issues.

3.5 Factors in satellite communication

- Elevation Angle: The angle of the horizontal of the earth surface to the center line of the satellite transmission beam.
  - This effects the satellites coverage area. Ideally, you want a elevation angle of 0 degrees, so the transmission beam reaches the horizon visible to the satellite in all directions.
  - However, because of environmental factors like objects blocking the transmission, atmospheric attenuation, and the earth electrical background noise, there is a minimum elevation angle of earth stations.
- Coverage Angle: A measure of the portion of the earth surface visible to a satellite taking the minimum elevation angle into account.
- $R/(R+h) = \sin(\pi/2 - \beta - \theta)/\sin(\theta + \pi/2)$
  $= \cos(\beta + \theta)/\cos(\theta)$
  $R = 6370$ km (earth's radius)
  h = satellite orbit height
  β = coverage angle
  θ = minimum elevation angle
- Other impairments to satellite communication:
  - The distance between an earth station and a satellite (free space loss).
  - Satellite Footprint: The satellite transmission's strength is strongest in the center of the transmission, and decreases farther from the center as free space loss increases.
  - Atmospheric Attenuation caused by air and water can impair the transmission. It is particularly bad during rain and fog.

3.6 How Satellites are used

Service Types

- Fixed Service Satellites (FSS)
  - Example: Point to Point Communication
- Broadcast Service Satellites (BSS)
  - Example: Satellite Television/Radio
  - Also called Direct Broadcast Service (DBS).
- Mobile Service Satellites (MSS)
  - Example: Satellite Phones

4.0 Conclusion

This unit focused on the concepts underlying satellite communication systems; the advantages and disadvantage of satellite communication, factors in satellite communication and how satellite are used.

5.0 Summary

In this unit you have learnt that:

- satellite communication is an artificial satellite stationed in space that is used for the purpose of telecommunications.
- satellite communication is highly survivable
- satellite system can perform effectively when providing broadcasting services over very wide areas such as a country, region, or entire hemisphere
- two stations can use a satellite as a relay station for their communication

6.0 Tutor-Marked Assignment

List the main advantages and disadvantages of satellite communications.

7.0 References/Further Readings

Brown, C. (2002). Elements of Spacecraft Design, American Institute of Aeronautics and Astronautics, Reston, VA.

Elbert, B. R. (1987). Introduction to Satellite Communication, Norwood, MA: Artech House

Elbert, B. R. (1992). Networking Strategies for Information Technology, Norwood, MA: Artech House

Elbert, B. R. (2004). Satellite Communication Application Handbook. Norwood, MA: Artech House

Elbert, B. R. and B. Martyna (1994) Client/Server Computing—Architecture, Applications, and Distributed Systems Management, Norwood, MA: Artech House

Foley T. (1998). ―Meteors and Solar Wind: Serious Threat or Hot Air‖, Communications

Kolawole M.O. (2004). Satellite Communication Engineering New York: Marcel Dekker Inc.

Macario, R. C. V. (1991). Personal and Mobile Radio Systems, London, England: Peter Peregrinus

Porter, M. E.(1980). Competitive Strategy, New York: The Free Press.

Wikipedia (2011). Communication Satellite. Available online at http://en.wikipedia.org/wiki/Communications_satellite

UNIT 2: TYPES OF SATELLITES: ORBITS

1.0 Introduction

In unit 1, basic concept of satellite communication was introduced but in this unit you will learn on the types of satellites.

2.0 Objectives

At the end of this unit, you should be able to

(i) mention the types of satellite
(ii) describe the types of satellite orbit
(iii) state the advantage and disadvantage of any of the satellite orbit
(iv) differentiate between two orbit

3.0 Main Content

3.1 Types of Satellite

Satellite is a microwave repeater in the space. There are about 750 satellites in the space; most of them are used for communication.

The types of satellite are:

(a) Satellite Orbits
- GEO
- LEO
- MEO
- Molniya Orbit
- HEO

(b) Frequency Bands

3.2 Satellite Orbits

Modern communications satellites use a variety of orbits including geostationary orbits, low (polar and non-polar) earth orbits, medium earth orbit, Molniya orbits and High elliptical orbiting.

3.2.1 Geostationary orbits (GEO)

GEO satellites orbit the earth directly over the equator, approximately 35 400 km (22 000 miles) up. At that altitude, one complete trip (orbit) around the earth takes 24 hours. Thus, the satellite remains over the same spot on the surface of the earth (geo) at all times and stays fixed in the sky (stationary) from any point on the surface from which it can be seen. A satellite in a geostationary orbit appears to be in a fixed position to an earth-based observer.

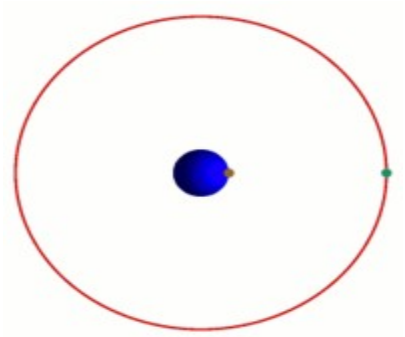


Figure 2.1: Geostationary Orbit

The geostationary orbit is useful for communications applications because ground based antennas which must be directed toward the satellite can operate effectively without the need for expensive equipment to track the satellite‘s motion. Especially for applications that require a large number of ground antennas (such as direct TV distribution), the savings in ground equipment justify the extra cost and onboard complexity of lifting a satellite into the relatively high geostationary orbit.
Geostationary orbits are useful because they cause a satellite to appear stationary with respect to a fixed point on the rotating earth, allowing a fixed antenna to maintain a link with the satellite.

Advantages

- A GEO satellite‘s distance from earth gives it a large coverage area, almost a fourth of the earth‘s surface.
- GEO satellites have a 24 hour view of a particular area.
- These factors make it ideal for satellite broadcast and other multipoint applications.

Disadvantages

- A GEO satellite‘s distance also cause it to have both a comparatively weak signal and a time delay in the signal, which is bad for point to point communication.
- GEO satellites, centered above the equator have difficulty broadcasting signals to near polar regions

Derivation of geostationary altitude

In any circular orbit, the centripetal force required to maintain the orbit is provided by the gravitational force on the satellite. To calculate the geostationary orbit altitude, begins with this equivalence and use the fact that the orbital period is one sidereal day.

$$F_c = F_g$$

By Newton's second law of motion, the forces **F** is replace with the mass m of the object multiplied by the acceleration felt by the object due to that force:

$$ma_c = mg$$

Note that the mass of the satellite m appears on both sides — geostationary orbit is independent of the mass of the satellite. Calculating the altitude means calculating the point where the magnitudes of the centripetal acceleration required for orbital motion and the gravitational acceleration provided by Earth's gravity are equal.

The centripetal acceleration's magnitude is:

$$|a_c| = \omega^2 r$$

where ω is the angular speed, and r is the orbital radius as measured from the Earth's center of mass.

The magnitude of the gravitational acceleration is:

$$|g| = \frac{GM}{r^2}$$

where M is the mass of Earth, 5.9736×10^{24} kg and G is the gravitational constant, $6.67428 \pm 0.00067 \times 10^{-11}$ $m^3\ kg^{-1}\ s^{-2}$.

Equating the two accelerations gives:

$$r^3 = \frac{GM}{\omega^2} \rightarrow r = \sqrt[3]{\frac{GM}{\omega^2}}$$

The product GM is known with much greater precision than either factor alone; it is known as the geocentric gravitational constant $\mu = 398{,}600.4418 \pm 0.0008$ $km^3\ s^{-2}$:

$$r = \sqrt[3]{\frac{\mu}{\omega^2}}$$

The angular speed ω is found by dividing the angle travelled in one revolution (360° = 2π rad) by the orbital period (the time it takes to make one full revolution: one sidereal day, or 6,164.09054 seconds). This gives:

$$\omega \approx \frac{2\pi\ rad}{86164\ s} \approx 7.2921 \times 10^{-5}\ rad/s$$

The resulting orbital radius is 42,164 kilometres (26,199 mi). Subtracting the Earth's equatorial radius, 6,378 kilometres (3,963 mi), gives the altitude of 35,786 kilometres (22,236 mi).

Orbital speed (how fast the satellite is moving through space) is calculated by multiplying the angular speed by the orbital radius:

$$\upsilon = \omega r \approx 3.0746\ km/s \approx 11068\ km/h \approx 6877.8\ mph$$

Now, by the same formula, let us find the geostationary orbit of an object in relation to Mars (an areostationary orbit). The geocentric gravitational constant GM (which is μ) for Mars has the value of 42,828 km^3s^{-2} and the known rotational period (T) of Mars is 88,642.66 seconds. Since $\omega = 2\pi/T$, using the formula above, the value of ω is found to be approximately 7.088218×10^{-5} s^{-1}. Thus, $r^3 = 8.5243 \times 10^{12}$ km^3, whose cube root is 20,427 km; subtracting the equatorial radius of Mars (3396.2 km) we have 17,031 km.

3.2.2 Low-Earth-orbiting satellites

A Low Earth Orbit (LEO) is between 500 and 1500 km above the earth surface and revolving around the earth at a period of about 90 minutes. These satellites are only visible from within a radius of roughly 1000 kilometres from the sub-satellite point because of their low altitude. In addition, satellites in low earth orbit change their position relative to the ground position. So

even for local applications, a large number of satellites are needed if the mission requires uninterrupted connectivity.
Low earth orbiting satellites are less expensive to launch into orbit than geostationary satellites and due to proximity to the ground; it does not require high signal strength (recall that signal strength falls off as the square of the distance from the source, so the effect is dramatic). Thus there is a trade off between the number of satellites and their cost. In addition, there are important differences in the onboard and ground equipment needed to support the two types of missions.

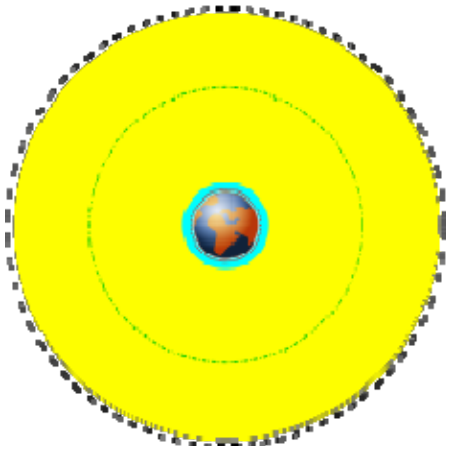


Figure 2.2: Various earth orbits to scale; light blue represents low earth orbit.

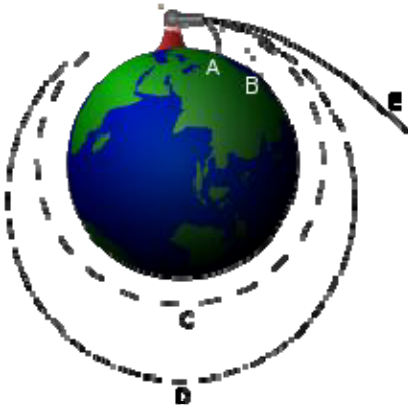


Figure 2.3: An orbiting cannon ball showing various sub-orbital and orbital possibilities.

Advantages

- A LEO satellite‘s proximity to earth compared to a GEO satellite gives it a better signal strength and less of a time delay, which makes it better for point to point communication.
- A LEO satellite‘s smaller area of coverage is less of a waste of bandwidth.

Disadvantages

- A network of LEO satellites is needed, which can be costly
- LEO satellites have to compensate for Doppler shifts cause by their relative movement.

- Atmospheric drag affects LEO satellites, causing gradual orbital deterioration.

3.2.3 Medium Earth Orbit (MEO)

MEO is defined simply as the area between LEO and GEO. The primary satellite systems there are the GPS (Global Positioning System) satellite constellations. A MEO satellite is in orbit somewhere between 8,000 km and 18,000 km above the earth‘s surface.

- MEO satellites are similar to LEO satellites in functionality.
- MEO satellites are visible for much longer periods of time than LEO satellites, usually between 2 to 8 hours.
- MEO satellites have a larger coverage area than LEO satellites.

Advantage

A MEO satellite‘s has longer duration of visibility and wider footprint means fewer satellites are needed in a MEO network than a LEO network.

Disadvantage

A MEO satellite‘s distance gives it a longer time delay and weaker signal than a LEO satellite, though not as bad as a GEO satellite.

3.2.4 Molniya satellites

A **Molniya orbit** is a type of highly elliptical orbit with an inclination of 63.4 degrees and an orbital period of precisely one half of a sidereal day.

As mentioned, geostationary satellites are constrained to operate above the equator. As a consequence, they are not always suitable for providing services at high latitudes: at high latitudes, a geostationary satellite will appear low on the horizon, affecting connectivity and causing multipath (interference caused by signals reflecting off the ground and into the ground antenna).

Molniya orbits can be an appealing alternative in such cases. The Molniya orbit is highly inclined, guaranteeing good elevation over selected positions during the northern portion of the orbit. (Elevation is the extent of the satellite‘s position above the horizon. Thus, a satellite at the horizon has zero elevation and a satellite directly overhead has elevation of 90 degrees).

Furthermore, the Molniya orbit is designed so that the satellite spends the great majority of its time over the far northern latitudes, during which its ground footprint moves only slightly. Its period is one half day, so that the satellite is available for operation over the targeted region for eight hours every second revolution. In this way a constellation of three Molniya satellites (plus in-orbit spares) can provide uninterrupted coverage.

Molniya satellites are typically used for telephony and TV services over Russia. Another application is to use them for mobile radio systems (even at lower latitudes) since cars travelling through urban areas need access to satellites at high elevation in order to secure good connectivity, e.g. in the presence of tall buildings.

Molniya orbits can be computed for any celestial body for which the dominant effects on bodies orbiting it are due to:

- Secular variations in the longitude of the ascending node;
- the argument of perigee due to the celestial body's oblateness.

Derivation

In order to ensure that the position of the apogee is not severely affected by orbit perturbations, an inclination close to 63.4 degrees is chosen. This results in the argument of perigee remaining nearly constant for a long period of time.
The change per day of argument of perigee for earth orbits is as follows (just considering the effect of the Earth's oblations (J_2) on the orbit - which is the dominating perturbation):

$$\Delta\omega_{day} = 4.98^\circ \left(\frac{R_E}{a}\right)^{\frac{7}{2}} \frac{5\cos^2 i - 1}{(1-\varepsilon^2)^2}$$

where:

- R_E - earth's radius,
- a - length of semi-major axis,
- i - the inclination, and
- ε - orbital eccentricity.

The equation becomes zero for an inclination of 63.4 degrees

3.2.5 High elliptical orbiting satellite (HEO)

HEO satellite is a specialized orbit in which a satellite continuously swings very close to the earth, loops out into space, and then repeats its swing by the earth. It is an elliptical orbit approximately 18,000 to 35,000 km above the earth‘s surface, not necessarily above the equator. HEOs are designed to give better coverage to countries with higher northern or southern latitudes. Systems can be designed so that the apogee is arranged to provide continuous coverage in a particular area. By definition, an apogee is the highest altitude point of the orbit, that is, the point in the orbit where the satellite is farthest from the earth.
By geometry,

$$S_m = a\sqrt{1-e^2}$$

$$S_p = \frac{S_m}{a} = a(1-e^2)$$

where the eccentricity, or the amount by which the ellipse departs from a circle, is

$$e = \frac{S_f}{a}$$

The general equation of an ellipse can thus be written as

$$r = \frac{a\left(1-e^2\right)}{1+e\cos\theta}$$

It is apparent from the equation that if e = 0, the resulting locus is a circle.

4.0 Conclusion

In this unit, you have learnt about the types of satellite and satellite orbit.

5.0 Summary

The main points in this unit include the following:

Geostationary Earth Orbit (GEO)

- These satellites are in orbit 35,863 km above the earth's surface along the equator.
- Objects in Geostationary orbit revolve around the earth at the same speed as the earth rotates. This means GEO satellites remain in the same position relative to the surface of earth.

Low Earth Orbit (LEO)

- LEO satellites are much closer to the earth than GEO satellites, ranging from 500 to 1,500 km above the surface.
- LEO satellites don't stay in fixed position relative to the surface, and are only visible for 15 to 20 minutes each pass.
- A network of LEO satellites is necessary for LEO satellites to be useful

Medium Earth Orbit (MEO)

A MEO satellite is in orbit somewhere between 8,000 km and 18,000 km above the earth's surface.

- MEO satellites are similar to LEO satellites in functionality.
- MEO satellites are visible for much longer periods of time than LEO satellites, usually between 2 to 8 hours.

Molniya Orbit Satellites

- Used by Russia for decades.
- Molniya Orbit is an elliptical orbit. The satellite remains in a nearly fixed position relative to earth for eight hours.
- A series of three Molniya satellites can act like a GEO satellite.
- Useful in near polar regions.

High elliptical orbiting satellite (HEO)

HEO satellite is a specialized orbit in which a satellite continuously swings very close to the earth, loops out into space, and then repeats its swing by the earth.

6.0 Tutor-Marked Assignment

(i) Mention two major types of satellite
(ii) Explain briefly the following types of satellite orbit: GEO, LEO and MEO
(iii) State the advantage and disadvantage of GEO and LEO
(iv) Differentiate between LEO and MEO

7.0 References/Further Readings

Elbert, B. R. (2004). Satellite Communication Application Handbook. Norwood, MA: Artech House

Kolawole M.O. (2004). Satellite Communication Engineering New York: Marcel Dekker Inc.

Simon Haykin and Michael Moher (2004), Modern Wireless Communications, Prentice Hall, USA.

Terrestrial Digital Microwave Communications, Artec House, Ferdo Ivanek.

Tozer T.C. and Grace D. (2001). High Altitude Platforms for Wireless Communications, Electronics & Communication Engineering Journal.

Wikipedia (2011). Communication Satellite. Available online at http://en.wikipedia.org/wiki/Communications_satellite
Zhili Sun(2005). Satellite Networking, Principles and Protocols, West Sussex, England, Wiley.

UNIT 3: TYPES OF SATELLITES: FREQUENCY BANDS

1.0 Introduction

In the previous unit, you learnt about the types of satellite orbits but in this unit you will also learn about the types of frequency bands.

2.0 Objectives

At the end of this unit, you should be able to
(i) List the types of frequency bands
(ii) Explain the different types of frequency bands

3.0 Main Content

3.1 Frequency Bands Tradeoff

Satellite communication is a form of radio or wireless communication and therefore must compete with other existing and potential uses of the radio spectrum. During the initial 10 years of development of these applications, there appeared to be more or less ample bandwidth, limited only by what was physically or economically justified by the rather small and low powered satellites of the time. In later years, as satellites grew in capability, the allocation of spectrum has become a domestic and international battlefield as service providers fight among themselves, joined by their respective governments when the battle extends across borders. So, all factors must be considered when selecting a band for a particular application.

3.2 Types of Frequency Band

Different kinds of satellites use different frequency bands which are:

(i) L–Band: 1 to 2 GHz, used by MSS

L band refers to four different bands of the electromagnetic spectrum: 40 to 60 Ghttp://en.wikipedia.org/wiki/Giga (NATO), 1 to 2 GHz (IEEE), 1565 nm to 1625 nm (optical), and around 3.5 micrometres (infrared astronomy).

The NATO L band is defined as the frequency band between 40 and 60 GHz (5–7.5 mm).

The IEEE L band (20-cm radar long-band) is a portion of the microwave band of the electromagnetic spectrum ranging roughly from 1 to 2 GHz. It is used by some communications satellites and terrestrial Eureka 147 digital audio broadcasting (DAB). The amateur radio service also has an allocation between 1240 and 1300 MHz (23-centimeter band). The L band refers to the frequency range of 950 MHz to 1450 MHz. Satellite modems and television receivers work in this range and the signal is translated to and from the band the satellite uses by either dedicated up-converters/down-converters or a solid-state Low-noise block converter and Block up-converter.
L band is also used in optical communications to refer to the wavelength range 1565 nm to 1625 nm.
In infrared astronomy, the L band refers to an atmospheric transmission window centred on 3.5 micrometres (in the mid-infrared).

(ii) S-Band: 2 to 4 GHz, used by MSS, NASA, deep space research

S-band is that part of the frequency spectrum between 2 GHz and 4 GHz. Many satellites transmit at S-band frequencies. This S-band catalogue includes a significant number of satellites in geosynchronous orbit and a quantity in inclined; highly eccentric orbits (HEOs). However, it is much used for low orbiting satellites also.

(iii) C-Band: 4 to 8 GHz, used by FSS

The C-band is a name given to certain portions of the electromagnetic spectrum including wavelengths of microwaves that are used for long-distance radio telecommunications. The IEEE C-band and its slight variations - contains frequency ranges that are used for many satellite communications transmissions; Wi-Fi devices; cordless telephones; and weather radar systems. For satellite communications, the microwave frequencies of the C-band perform better in comparison with K_u band (11.2 GHz to 14.5 GHz) microwave frequencies, under adverse weather conditions, which are used by another large set of communication satellites. The adverse weather conditions collectively referred to as rain fade, all have to do with moisture in the air, including rain and snow.

C-Band Variations Around The World		
Band	**Transmit Frequency (GHz)**	**Receive Frequency (GHz)**
Standard C-Band	5.850–6.425	3.625–4.200
Extended C-Band	5.850–6.725	3.400–4.200
INSAT / Super-Extended C-Band	6.725–7.025	4.500–4.800
Russian C-Band	5.975–6.475	3.650–4.150
LMI C-Band	5.7250–6.025	3.700–4.000

C-Band
Downlink: 3.7 – 4.2 GHz
Uplink: 5.9 – 6.4 GHz

Advantages

- Less disturbance from heavy rain fade
- Cheaper Bandwidth

Disadvantages

- Needs a larger satellite dish (diameters of minimum 2-3m)
- Powerful (expensive) RF unit
- More expensive hardware
- Possible Interference from microwave links

(iv) X-Band: 8 to 12.5 GHz

The frequency range from 8.0 – 12.0 GHz. The X-band frequency enables high power operations with very small terminals. Applications include COTM, manpacks, emergency communications and airborne and shipboard platforms. X-band is also less vulnerable to rain fade and adjacent satellite side lobe interference than other frequencies.

(v) Ku-Band: 12.5 to 18 GHz: used by FSS and BSS (DBS)

The K_u band is a portion of the electromagnetic spectrum in the microwave range of frequencies. In radar applications, it ranges from 10.95 - 14.5 GHz according to the formal definition of radar frequency band nomenclature in IEEE Standard 521-2002.

The Ku-band frequency range is allocated to be exclusively used by satellite communication systems, thereby eliminating the problem of interference with microwave systems. Due to higher power levels at new satellites Ku-band allows for significantly smaller earth station antennas and RF units to be installed at the VSAT location.

K_u band is primarily used for satellite communications most notably for fixed and broadcast services and for specific applications such as NASA's Tracking Data Relay Satellite used for both space shuttle and ISS communications. K_u band satellites are also used for backhauls and particularly for satellite from remote locations back to a television network's studio for editing and broadcasting. The band is split into multiple segments that vary by geographical region by the International Telecommunication Union (ITU). NBC was the first television network to uplink a majority of its affiliate feeds via K_u band in 1983.

Some frequencies in this radio band are used for vehicle speed detection by law enforcement, especially in Europe.

Ku-Band
Downlink: 11.7 – 12.2 GHz
Uplink: 14.0 – 14.5 GHz

Advantages

- No interference from microwave links and other technologies

- Needs less power - cheaper RF unit. Compared with C-band, K_u band is not similarly restricted in power to avoid interference with terrestrial microwave systems and the power of its uplinks and downlinks can be increased. This higher power also translates into smaller receiving dishes and points out a generalization between a satellite‘s transmission and a dish‘s size. As the power increases, the dish‘s size can decrease. This is because the purpose of the dish element of the antenna is to collect the incident waves over an area and focus them all onto the antenna's actual receiving element, mounted in front of the dish (and pointed back towards its face); if the waves are more intense, less of it need to be collected to achieve the same intensity at the receiving element.
- The K_u band also offers a user more flexibility. A smaller dish size and a K_u band system‘s are free from terrestrial operations simplify finding a suitable dish site.
- For the End users K_u band is generally cheaper and enables smaller antennas (both because of the higher frequency and a more focused beam).
- K_u band is also less vulnerable to rain fade than the K_a band frequency spectrum.
- The satellite operator's Earth Station antenna do require more accurate position control when operating at K_u band than compared to C band. Position feedback accuracies are higher and the antenna may require a closed loop control system to maintain position under wind loading of the dish surface.

Disadvantages

- The disadvantage of K_u band system is that at frequencies higher than 10 GHz in heavy rain fall areas, a noticeable degradation occurs, due to the problems caused by and proportional to the amount of rainfall (commonly known as "rain fade").
  This problem can be mitigated, however, by deploying an appropriate link budget strategy when designing the satellite network and allocating a higher power consumption to compensate rain fade loss. The K_u band is not only used for television transmission, which some sources imply, but also very much for digital data transmission via satellites and for voice/audio transmissions.
- The higher frequency spectrum of the K_u band is particularly susceptible to signal degradation more than C-band satellite frequency spectrum. A similar phenomenon, called "snow fade" (where snow or ice accumulation significantly alters the focal point of a dish) can also occur during winter precipitation.
- Also, the $K_u$band satellites typically require more power to transmit than the C-band satellites. Under both "rain fade" and "snow fade" conditions, Ka and Ku band losses can be marginally reduced using super-hydrophobic Lotus effect coatings.
- More expensive capacity

(vi) Ka-Band: 26.5 to 40 GHz: used by Fixed Satellite Service (FSS)

The Ka band is an electromagnetic frequency range which covers 26.5 – 40 GHz. The Ka band is a portion of the K microwave band, which ranges from around 18 to 40 GHz. _Ka‘ is short for _K-above,‘ denoting that this range approximately covers the upper third of the entire K band. The term _Ka‘ frequently refers to the band with the recommended operating frequency of the WR-28 waveguide (within 26.5 – 40 GHz) .

The IEEE K band is segmented into three secondary bands:

- Ka (K-above) band – ranging from 26.5–40 GHZ, which is primarily used in experimental communications and radar
- K band – ranging from 18–27 GHz
- Ku (K-under) band – ranging from 12–18 GHz, mainly for radar, satellite communications, and terrestrial microwave transmissions

Downlink within the 18.3–18.8 or 19.7–20.2 GHz bands, communications satellites, and high-resolution close-range targeting radar (aboard military aircraft) use the 30/20 GHz band. The uplink for the 30/20 systems are around 30 GHz. This radar range is used for vehicle speed identification required by law enforcement.

The Ka band uplink frequencies are within 27.5–31 GHz, while the downlink frequencies are within 18.3–18.8 and 19.7–20.2 GHz.

Ka band satellites usually transmit using more power than C band satellites, although C band dishes are bigger than Ka band satellites. Ka band dishes range from 2 feet to 5 feet in diameter, while C band dishes range from 7 feet to 12 feet.

C band dishes are also known as BUDs (Big Ugly Dishes) due to their relative size. Conversely, due to the higher frequency range and smaller dish size, the signals are more prone to signal disruptions and quality problems caused by adverse weather conditions such as rainfall (recognized as rainfade).

4.0 Conclusion

Commercially it is fact that hardware for C Band is significantly more expensive while the capacity is cheaper. So users with large bandwidth requirements preferably choose this technology. Ku Band on the other hand operates with small antennas and less expensive equipment, while the capacity price is higher than C Band.

5.0 Summary

The L and S band frequencies from 1-4 GHz become increasingly useful for satellite communications because the high frequencies allow high capacities (although still much less than C and Ku band) and propagation is line-of-sight with little man-made noise and relatively low absorption by the atmosphere (although ionospheric scintillation and polarisation rotation must be taken into account). L and S band are shared with terrestrial services such as industrial and educational television and studio-to-television transmitter links and other space links such as radio astronomy applications and NASA‘s space probes. These bands are therefore not very suitable for GEO operation as they do not have enough capacity and are very difficult to coordinate with terrestrial services over large areas.

Ku-Band is typically used for broadcasting and 2-way Internet connections. It operates with a smaller satellite dish (diameters from 0.9m) - cheaper and easy to install. The problem of Ku-band is that it is sensitive to heavy rain fade (significant attenuation of the signal) and possibly can be managed by appropriate dish size or transmitter power.

6.0 Tutor-Marked Assignment

(i) Mention at least five types of frequency bands

(ii) Explain any three types of frequency bands

7.0 References/Further Readings

Donald, M. (2006). "Electromagnetic Frequency Spectrum".
Elbert, B. R. (2001). The Satellite Communication Ground Segment and Earth Station Handbook, Norwood, MA: Artech House.
Elbert, B. R. (2001). Introduction to Satellite Communication, 2nd ed., Norwood, MA: Artech
Elbert, B. R. (2004). Satellite Communication Application Handbook. Norwood, MA: Artech House
Flock, W. L. (1987). Propagation Effects on Satellite Systems at Frequencies Below 10 GHz, Second Edition, NASA Reference Publication 1108(2), National Aeronautics and Space Administration.
Kadish, J. E., and T. W. R. East, Satellite Communications Fundamentals, Norwood, MA: Artech House, 2000.
Kolawole M.O. (2004). Satellite Communication Engineering New York: Marcel Dekker Inc.
Mirabito, M. And Morgenstern, B. (2004). Satellites: Operations and Applications. The New Communication Technologies (fifth edition). Burlington: Focal Press.
Peebles and Peyton Z. Jr. (1998). Radar Principles, John Wiley and Sons, Inc.
Sklar, B. (2001). Digital Communications—Fundamentals and Applications, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2001.
Wikipedia (2011). Communication Satellite. Available online at http://en.wikipedia.org/wiki/Communications_satellite

UNIT 4: CAPACITY ALLOCATION

1.0 Introduction

In this unit you will learn on the capacity allocation in satellite communication.

2.0 Objectives

At the end of this unit, you should be able to:

(i) explain the capacity allocation in satellite communication
(ii) outline the ways in which frequency division multiple access can be performed
(iii) state the advantages of time division multiple access

3.0 Main Content

3.1 Capacity Allocation

3.1.1 FDMA

Frequency-division multiple access (FDMA) in satellite communications divides a satellite communications channel amongst users who are each given their portion of the available channel bandwidth for their permanent use. Frequency division multiple access is commonly used in satellite communications because it minimizes the coordination required between different earth stations.

- Satellite frequency is already broken into bands and is broken into smaller channels in Frequency Division Multiple Access (FDMA).
- Used extensively in the early telephone and wireless multiuser communication systems.
- Overall bandwidth within a frequency band is increased due to frequency reuse (a frequency is used by two carriers with orthogonal polarization).
- If a channel, such as a cable has a transmission bandwidth W Hz, and individual users require B Hz to achieve their required information rate, then the channel in theory should be able to support W/B users

- Near-Far problem
- The number of sub-channels is limited by three factors:
  - Thermal noise (too weak a signal will be affected by background noise).
  - Intermodulation noise (too strong a signal will cause noise).
  - Crosstalk (cause by excessive frequency reusing).
- FDMA can be performed in two ways:
  - **FAMA-FDMA**

    Fixed-assigned multiple access (FAMA) in satellite communications is one of the two main techniques for allocating satellite channels to users. In fixed-assigned multiple access (FAMA) in satellite communications, each user is allocated a channel permanently, whether they use it or not. Fixed-assigned multiple access is inefficient and many satellite multiple-access systems use demand-assigned multiple access (DAMA) in which the available channels are allocated on an as-required basis to users.
  - **DAMA-FDMA**

    Demand-assigned multiple access (DAMA) in satellite communications is one of the two main techniques for allocating satellite channels to satellite communications users. In fixed-assigned multiple access (FAMA), each satellite communications user is allocated a channel permanently, whether they use it or not. This is inefficient and most multiple-access systems use demand-assigned multiple access (DAMA) in which the available channels on the satellite are allocated on an as-required basis to users. When the user is finished with the demand-assigned multiple access (DAMA) channel it is relinquished and made available for another user.

3.1.2 TDMA

Time Division Multiple Access (TDMA) is a way of sharing a channel by assigning different time slots to different users.

- TDMA (Time Division Multiple Access) breaks a transmission into multiple time slots, each one dedicated to a different transmitter.
- TDMA is increasingly becoming more widespread in satellite communication.
- TDMA uses the same techniques (FAMA and DAMA) as FDMA does.

Advantages of TDMA over FDMA.

- Digital equipment used in time division multiplexing is increasingly becoming cheaper.
- There are advantages in digital transmission techniques. Ex: error correction.
- Lack of inter-modulation noise means increased efficiency.

3.1.3 CDMA

In recent years, the interference immunity of CDMA for multi-user communications, together with its very good spectral efficiency characteristics, has been seen to offer distinct advantages for public cellular-type communications.

There are two very distinct types of CDMA system, classified as direct sequence CDMA and frequency hopping CDMA. Both of these systems involve transmission bandwidths that are many times that required by an individual user, with the energy of each user's signal spread with time throughout this wide channel. Consequently these techniques are often referred to as spread spectrum systems.

4.0 Conclusion

In this unit you have learnt about the capacity allocation in satellite communication**.**

5.0 Summary

The main points in this unit include the following:

Frequency-division multiple access (FDMA) in satellite communications divides a satellite communications channel amongst users who are each given their portion of the available channel bandwidth for their permanent use. FDMA can be performed in two ways:

- Fixed-assignment multiple access (FAMA): The sub-channel assignments are of a fixed allotment. Ideal for broadcast satellite communication.
- Demand-assignment multiple access (DAMA): The sub-channel allotment changes based on demand. Ideal for point-to-point communication.

Time Division Multiple Access (TDMA) is a way of sharing a channel by assigning different time slots to different users.

6.0 Tutor-Marked Assignment

(i) Explain briefly the capacity allocation in satellite communication
(ii) What are the ways in which FDMA can be performed
(iii) State the advantages of TDMA

7.0 References

Elbert, B. R. (2001). The Satellite Communication Ground Segment and Earth Station Handbook, Norwood, MA: Artech House.

Elbert, B. R. (2001). Introduction to Satellite Communication, 2nd ed., Norwood, MA: Artech

Elbert, B. R. (2004). Satellite Communication Application Handbook. Norwood, MA: Artech House

Flock, W. L. (1987). Propagation Effects on Satellite Systems at Frequencies Below 10 GHz, Second Edition, NASA Reference Publication 1108(2), National Aeronautics and Space Administration.

Kadish, J. E., and T. W. R. East, Satellite Communications Fundamentals, Norwood, MA: Artech House, 2000.

Kolawole M.O. (2004). Satellite Communication Engineering New York: Marcel Dekker Inc.

Sklar, B. (2001). Digital Communications—Fundamentals and Applications, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2001.

Wikipedia (2011). Communication Satellite. Available online at http://en.wikipedia.org/wiki/Communications_satellite

UNIT 5: APPLICATION AREAS OF SATELLITE COMMUNICATION

1.0 Introduction

Applications in satellite communications have evolved over the years to adapt to competitive markets. Evolutionary development is a natural facet of the technology because satellite communication is extremely versatile.

2.0 Objectives

At the end of this unit, you should be able to

(i) List the application areas of satellite communication
(ii) Explain the application areas of satellite communication

3.0 Main Content

Application Areas

(a) Telephone

The first and historically most important application for communication satellites was in intercontinental long distance telephony. The fixed Public Switched Telephone Network relays telephone calls from land line telephones to an earth station, where they are then transmitted to a geostationary satellite. The downlink follows an analogous path. Improvements in submarine communications cables, through the use of fiber-optics, caused some decline in the use of

satellites for fixed telephony in the late 20th century, but they still serve remote islands such as Ascension Island, Saint Helena, Diego Garcia, and Easter Island, where no submarine cables are in service. There are also regions of some continents and countries where landline telecommunications are rare to nonexistent, for example large regions of South America, Africa, Canada, China, Russia, and Australia. Satellite communications also provide connection to the edges of Antarctica and Greenland.
Satellite phones connect directly to a constellation of either geostationary or low-earth-orbit satellites. Calls are then forwarded to a satellite teleport connected to the Public Switched Telephone Network

(b) Satellite television

Satellite television is television delivered by the means of communications satellite and received by an outdoor antenna, usually a parabolic mirror generally referred to as a satellite dish and as far as household usage is concerned, a satellite receive either in the form of an external set-top box or a satellite tuner module built into a TV set. Satellite TV tuners are also available as a card or a USB stick to be attached to a personal computer. In many areas of the world satellite television provides a wide range of channels and services, often to areas that are not serviced by terrestrial or cable providers.
As television became the main market, its demand for simultaneous delivery of relatively few signals of large bandwidth to many receivers being a more precise match for the capabilities of geosynchronous comsats. Two satellite types are used for North American television and radio: Direct Broadcast Satellite (DBS) and Fixed Service Satellite (FSS)

(c) Fixed Service Satellite

Fixed Service Satellite (FSS), is the official classification (used in North America) for geostationary communications satellites used for broadcast feeds for television, radio stations, networks as well as telephony and data communications. Fixed Service Satellites use the C band, and the lower portions of the K_u bands. They are normally used for broadcast feeds to and from television networks and local affiliate stations (such as program feeds for network and syndicated programming, live shots, and backhauls), as well as being used for distance learning by schools and universities, business television (BTV), Videoconferencing, and general commercial telecommunications. FSS satellites are also used to distribute national cable channels to cable television headends. Free-to-air satellite TV channels are also usually distributed on FSS satellites in the K_u band. The Intelsat Americas 5, Galaxy 10R and AMC 3 satellites over North America provide a quite large amount of FTA channels on their K_u band transponders.

(d) Direct broadcast satellite (DBS)

Direct broadcast satellite (DBS) is a term used to refer to satellite television broadcasts intended for home reception. It is a communications satellite that transmits to small DBS satellite dishes (usually 18 to 24 inches or 45 to 60 cm in diameter). It generally operate in the upper portion of the microwave K_u band. DBS technology is used for DTH-oriented (Direct-To-Home) satellite TV services, such as DirecTV and DISH Network in the United States, Bell TV and Shaw Direct in Canada, Freesat and Sky Digital in the UK, the Republic of Ireland, and New Zealand.

(e) Mobile satellite technologies

Initially available for broadcast to stationary TV receivers, by 2004 popular mobile direct broadcast applications made their appearance with that arrival of two satellite radio systems in the United States: Sirius and XM Satellite Radio Holdings. Some manufacturers have also introduced special antennas for mobile reception of DBS television. Using Global Positioning System (GPS) technology as a reference, these antennas automatically re-aim to the satellite no matter where or how the vehicle (on which the antenna is mounted) is situated. These mobile satellite antennas are popular with some recreational vehicle owners. Such mobile DBS antennas are also used by JetBlue Airways for DirecTV (supplied by LiveTV, a subsidiary of JetBlue), which passengers can view on-board on LCD screens mounted in the seats.

(f) Satellite radio

Satellite radio is an analogue or digital radio signal that is relayed through one or more satellites and thus can be received in a much wider geographical area than terrestrial FM radio stations. Satellite radio offers audio services in some countries, notably the United States. Mobile services allow listeners to roam a continent, listening to the same audio programming anywhere. A satellite radio or subscription radio (SR) is a digital radio signal that is broadcast by a communications satellite, which covers a much wider geographical range than terrestrial radio signals.

(g) Amateur radio

Amateur radio operators have access to the OSCAR satellites that have been designed specifically to carry amateur radio traffic. Most such satellites operate as spaceborne repeaters and are generally accessed by amateurs equipped with UHF or VHF radio equipment and highly directional antennas such as Yagis or dish antennas. Due to launch costs, most current amateur satellites are launched into fairly low Earth orbits and are designed to deal with only a limited number of brief contacts at any given time. Some satellites also provide data-forwarding services using the AX.25 or similar protocols.

(h) Satellite Internet

After the 1990s, satellite communication technology has been used as a means to connect to the Internet via broadband data connections. This can be very useful for users who are located in very remote areas and cannot access a broadband connection.

(i) Military uses

Communications satellites are used for military communications applications such as Global Command and Control Systems. Examples of military systems that use communication satellites are the MILSTAR, the DSCS and the FLTSATCOM of the United States, NATO satellites, United Kingdom satellites, and satellites of the former Soviet Union. Many military satellites operate in the X-band and some also use UHF radio links while MILSTAR also utilizes Ka band

4.0 Conclusion

In this unit, you have learnt on different application areas of satellite communication.

5.0 Summary

The different application areas of satellite communication are: telephone, satellite television,

fixed service satellite, direct broadcast satellite, mobile satellite technologies, satellite radio, amateur radio, satellite Internet and military uses

6.0 Tutor-Marked Assignment

Mention and explain briefly the application areas of satellitecommunication.

7.0 References/Further Readings

Elbert, B. R. (1987). Introduction to Satellite Communication, Norwood, MA: Artech House.

Elbert, B. R. (1992). Networking Strategies for Information Technology, Norwood, MA: Artech House.

Elbert, B. R. (2004). Satellite Application Handbook, Norwood, MA: Artech House.

Elbert, B. R. and Martyna B. (1994). Client/Server Computing—Architecture, Applications, and Distributed Systems Management, Norwood, MA: Artech House.

Kolawole M.O. (2004). Satellite Communication Engineering New York: Marcel Dekker Inc.

Macario, R. C. V.(1991). Personal and Mobile Radio Systems, London, England: Peter Peregrinus.

Porter, M. E. (1980). Competitive Strategy, New York: The Free Press.

Wikipedia (2011). Communication Satellite. Available online at http://en.wikipedia.org/wiki/Communications_satellite

MODULE 3: CODING AND ERROR CONTROL

UNIT 1: ERROR DETECTION

1.0 Introduction

The general idea for achieving error detection and correction is to add some redundancy (i.e., some extra data) to a message, which receivers can use to check consistency of the delivered message and to recover data determined to be erroneous. Error-detection and correction schemes can be either systematic or non-systematic: In a systematic scheme, the transmitter sends the original data and attaches a fixed number of check bits (or parity data), which are derived from the data bits by some deterministic algorithm. If only error detection is required, a receiver can simply apply the same algorithm to the received data bits and compare its output with the received check bits; if the values do not match, an error has occurred at some point during the transmission. In a system that uses a non-systematic code, the original message is transformed into an encoded message that has at least as many bits as the original message.

Good error control performance requires the scheme to be selected based on the characteristics of the communication channel. Common channel models include memory-less models where errors occur randomly and with a certain probability and dynamic models where errors occur primarily in bursts. Consequently, error-detecting and correcting codes can be generally distinguished between random-error-detecting/correcting and burst-error-detecting/correcting. Some codes can also be suitable for a mixture of random errors and burst errors.

If the channel capacity cannot be determined or is highly varying, an error-detection scheme may be combined with a system for retransmissions of erroneous data. This is known as automatic repeat request (ARQ) and is most notably used in the Internet. An alternate approach for error control is hybrid automatic repeat request (HARQ), which is a combination of ARQ and error-correction coding.

2.0 Objectives

At the end of this unit, you should be able to

(i) explain the concept of error detection
(ii) list the transmission error detection codes
(iii) describe a parity checking error detection
(iv) discuss on cyclic redundancy check

3.0 Main Content

3.1 Error detection codes

An error detection code is used to detect the presence of an error. Error detection is a method that allows some communication errors to be detected. The data is encoded so that the encoded data contains additional redundant information about the data. The data is decoded so that the additional redundant information must match the original information. This allows some errors to be detected. Unfortunately, some error bursts may cause incorrectly received blocks which pass the error detection test. Therefore, good error detection schemes are designed.

Error detection is most commonly realized using a suitable hash function (or checksum algorithm). A hash function adds a fixed-length tag to a message, which enables receivers to verify the delivered message by re-computing the tag and comparing it with the one provided.

There exists a vast variety of different hash function designs. However, some are of particularly widespread use because of either their simplicity or their suitability for detecting certain kinds of errors (e.g., the cyclic redundancy check's performance in detecting burst errors).
Random-error-correcting codes based on minimum distance coding can provide a suitable alternative to hash functions when a strict guarantee on the minimum number of errors to be detected is desired. Repetition codes, described below are special cases of error-correcting codes: although rather inefficient, they find applications for both error correction and detection due to their simplicity.

3.2 Types of Error Detection

(a) Repetition codes

A repetition code is a coding scheme that repeats the bits across a channel to achieve error-free communication. Given a stream of data to be transmitted, the data is divided into blocks of bits. Each block is transmitted some predetermined number of times. For example, to send the bit pattern "1011", the four-bit block can be repeated three times, thus producing "1011 1011 1011". However, if this twelve-bit pattern was received as "1010 1011 1011" – where the first block is unlike the other two – it can be determined that an error has occurred.
Repetition codes are not very efficient and can be susceptible to problems if the error occurs in exactly the same place for each group (e.g., "1010 1010 1010" in the previous example would be detected as correct). The advantage of repetition codes is that they are extremely simple and are in fact used in some transmissions of numbers stations.

(b) Parity Checking

A **parity bit** is a bit that is added to a group of source bits to ensure that the number of set bits (i.e., bits with value 1) in the outcome is even or odd. It is a very simple scheme that can be used to detect single or any other odd number (i.e., three, five, etc.) of errors in the output. An even number of flipped bits will make the parity bit appear correct even though the data is erroneous.
There are two variants of parity bits: even parity bit and odd parity bit. When using even parity, the parity bit is set to 1 if the number of ones in a given set of bits (not including the parity bit) is odd, making the entire set of bits (including the parity bit) even. When using odd parity, the parity bit is set to 1 if the number of ones in a given set of bits (not including the parity bit) is even, keeping the entire set of bits (including the parity bit) odd. In other words, an even parity bit will be set to "1" if the number of 1's + 1 is even and an odd parity bit will be set to "1" if the number of 1's +1 is odd.

7 bits of data (number of 1s)	**8 bits including parity**	
	even	**odd**
0000000 (0)	**0**0000000	**1**0000000
1010001 (3)	**1**1010001	**0**1010001
1101001 (4)	**0**1101001	**1**1101001
1111111 (7)	**1**1111111	**0**1111111

If an odd number of bits (including the parity bit) are transmitted incorrectly, the parity bit will be incorrect and thus indicates that an error occurred in transmission. The parity bit is only suitable for detecting errors; it cannot correct any errors, as there is no way to determine which

particular bit is corrupted. The data must be discarded entirely and re-transmitted from scratch. On a noisy transmission medium, successful transmission can therefore take a long time or even never occur. However, parity has the advantage that it uses only a single bit and requires only a number of XOR gates to generate.
Parity bit checking is used occasionally for transmitting ASCII characters, which have 7 bits, leaving the 8th bit as a parity bit.
For example, the parity bit can be computed as follows, assuming we are sending a simple 4-bit value 1001 with the parity bit following on the right, and with ^ denoting an XOR gate:

Transmission sent using **even** parity:

A wants to transmit: 1001
A computes parity bit value: 1^0^0^1 = 0
A adds parity bit and sends: 10010
B receives: 10010
B computes parity: 1^0^0^1^0 = 0
B reports correct transmission after observing expected even result.

Transmission sent using **odd** parity:

A wants to transmit: 1001
A computes parity bit value: ~(1^0^0^1) = 1
A adds parity bit and sends: 10011
B receives: 10011
B computes overall parity: 1^0^0^1^1 = 1
B reports correct transmission after observing expected odd result.

This mechanism enables the detection of single bit errors because if one bit gets flipped due to line noise, there will be an incorrect number of ones in the received data. In the two examples above, B's calculated parity value matches the parity bit in its received value, indicating there are no single bit errors. Consider the following example with a transmission error in the second bit:

Transmission sent using even parity:

A wants to transmit: 1001
A computes parity bit value: 1^0^0^1 = 0
A adds parity bit and sends: 10010
*** TRANSMISSION ERROR ***
B receives: 1**1**010
B computes overall parity: 1^1^0^1^0 = 1
B reports incorrect transmission after observing unexpected odd result.

B calculates an odd overall parity indicating the bit error. Here's the same example but now the parity bit itself gets corrupted:

A wants to transmit: 1001

A computes even parity value: 1^0^0^1 = 0
A sends: 10010
*** TRANSMISSION ERROR ***
B receives: 1001**1**
B computes overall parity: 1^0^0^1^1 = 1
B reports incorrect transmission after observing unexpected odd result.

Once again, B computes an odd overall parity, indicating the bit error.

There is a limitation to parity schemes. A parity bit is only guaranteed to detect an odd number of bit errors. If an even number of bits has errors, the parity bit records the correct number of ones, even though the data is corrupt.

Consider the same example as before with an even number of corrupted bits:

A wants to transmit: 1001
A computes even parity value: 1^0^0^1 = 0
A sends: 10010
*** TRANSMISSION ERROR ***
B receives: 1**1**01**1**
B computes overall parity: 1^1^0^1^1 = 0
B reports correct transmission though actually incorrect.

B observes even parity, as expected, thereby failing to catch the two bit errors.

(c) Block check

Block check is a block-based error detection method. The data is divided in blocks in the encoding process. An additional block check is added to each block of data. The check is calculated from the current block. The receiver also performs the same calculation on the block and compares the calculated result with the received result. If these checks are equal the blocks are likely to be valid. Unfortunately, the problem with all block checks is that the block check is shorter than the block. Therefore, there are several different blocks that all have the same checksum. It is possible that the date is corrupted by a random error burst that modifies the block contents so that the block check in the corrupted frame also matches the corrupted data. In this case the error is not detected. Even best block checks cannot detect all error bursts but good block checks minimize this probability. However, the reliability increases as the length of the block check increases.

(i) Block Check Sum

A **checksum** or **hash sum** is a fixed-size datum computed from an arbitrary block of digital data for the purpose of detecting accidental errors that may have been introduced during its transmission or storage. The integrity of the data can be checked at any later time by recomputing the checksum and comparing it with the stored one. If the checksums match, the data were almost certainly not altered (either intentionally or unintentionally).
A checksum of a message is a modular arithmetic sum of message code words of a fixed word length (e.g., byte values). The sum may be negated by means of a one's-complement prior to

transmission to detect errors resulting in all-zero messages. Block check sum is a primitive block check sum that is the sum of all characters in the block. The result is a character that is equally long as the characters in the block. Therefore, the result is sometimes referred as the block check character (BCC). Unfortunately, even a long BCC may allow relatively simple errors. In other words, it is easy to find different blocks that generate the same block check sum. Calculating check sum is certainly fast and easy but the reliability of the check sum is not adequate for today's reliable communications. However, due to its speed it is used in some applications which require that the calculation is done by the software.

The procedure that yields the checksum from the data is called a checksum function or checksum algorithm. A good checksum algorithm will yield a different result with high probability when the data is accidentally corrupted; if the checksums match, the data is very likely to be free of accidental errors.

Checksum schemes include parity bits check digits and longitudinal redundancy checks. Some checksum schemes such as the Luhn algorithm and the Verhoeff algorithm are specifically designed to detect errors commonly introduced by humans in writing down or remembering identification numbers.

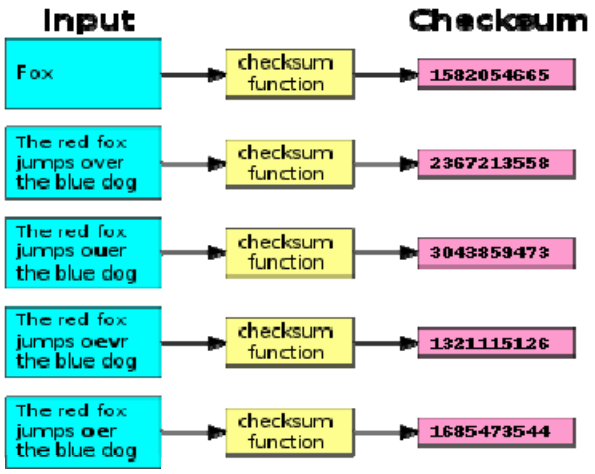


Figure 1.1: Effect of a typical checksum function (the Unix cksum utility).

(ii) Cyclic Redundancy Check (CRC)

A cyclic redundancy check (CRC) is a single-burst-error-detecting cyclic code and non-secure hash function designed to detect accidental changes to digital data in computer networks. The Cyclic redundancy check (CRC) is an intelligent alternative for block check sum. It is calculated by dividing the bit string of the block by a generator polynomial. The value of the cyclic redundancy check is the reminder of the calculation which is one bit shorter than the generator polynomial. This value is also sometimes referred as the frame check sequence (FCS). However, the generator polynomial must be chosen carefully. CRC is a stronger check than the block check sum and it is being used in today's reliable communication. Calculating the CRC requires slightly more processing than the check sum.

Cyclic codes have favorable properties in that they are well suited for detecting burst errors. CRCs are particularly easy to implement in hardware and are therefore commonly used in digital networks and storage devices such as hard disk drives.
Even parity is a special case of a cyclic redundancy check, where the single-bit CRC is generated by the divisor x+1.

Computation of CRC
To compute an *n*-bit binary CRC, let the bits representing the input be in a row and position the (*n*+1)-bit pattern representing the CRC's divisor (called a "polynomial") underneath the left-hand end of the row.

Start with the message to be encoded:
11010011101100

This is first padded with zeroes corresponding to the bit length *n* of the CRC. Here is the first calculation for computing a 3-bit CRC:

```
11010011101100 000 <--- input left shifted by 3 bits
1011               <--- divisor (4 bits)
------------------
01100011101100 000 <--- result
```

If the input bit above the leftmost divisor bit is 0, do nothing and move the divisor to the right by one bit. If the input bit above the leftmost divisor bit is 1, the divisor is XORed into the input (in other words, the input bit above each 1-bit in the divisor is toggled). The divisor is then shifted one bit to the right and the process is repeated until the divisor reaches the right-hand end of the input row. Here is the entire calculation:

```
11010011101100 000 <--- input left shifted by 3 bits
1011               <--- divisor
01100011101100 000 <--- result
 1011              <--- divisor ...
00111011101100 000
  1011
00010111101100 000
   1011
00000001101100 000
      1011
00000000110100 000
       1011
00000000011000 000
        1011
00000000001110 000
```

```
        1011
00000000000101 000
         101 1
----------------
00000000000000 100 <---remainder (3 bits)
```

Since the leftmost divisor bit zeroed every input bit it touched, when this process ends the only bits in the input row that can be nonzero are the n bits at the right-hand end of the row. These *n* bits are the remainder of the division step and will also be the value of the CRC function (unless the chosen CRC specification calls for some post processing).
The validity of a received message can easily be verified by performing the above calculation again, this time with the check value added instead of zeroes. The remainder should equal zero if there are no detectable errors.

```
11010011101100 100 <--- input with check value
1011            <--- divisor
01100011101100 100 <--- result
 1011           <--- divisor ...
00111011101100 100
and so on until:

00000000001110 100
        1011
00000000000101 100
         101 1
-----------------
             0 <--- remainder
```

(d) Cryptographic hash functions

A cryptographic hash function can provide strong assurances about data integrity, provided that changes of the data are only accidental (i.e., due to transmission errors). Any modification to the data will likely be detected through a mismatching hash value. Furthermore, given some hash value, it is infeasible to find some input data (other than the one given) that will yield the same hash value. Message authentication codes, also called keyed cryptographic hash functions, provide additional protection against intentional modification by an attacker.

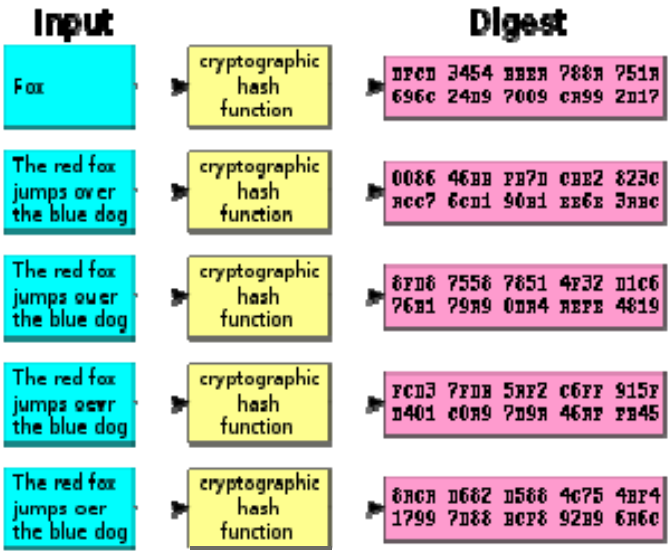


Figure 1.2: A cryptographic hash function (specifically, SHA-1) at work.
Note that even small changes in the source input (here in the word "over") drastically change the resulting output, by the so-called avalanche effect.

(e) Error-correcting codes

Any error-correcting code can be used for error detection. A code with minimum Hamming distance, d, can detect up to d-1 errors in a code word. Using minimum-distance-based error-correcting codes for error detection can be suitable if a strict limit on the minimum number of errors to be detected is desired. Codes with minimum Hamming distance d=2 are degenerate cases of error-correcting codes and can be used to detect single errors. The parity bit is an example of a single-error-detecting code. The Berger code is an early example of a unidirectional error(-correcting) code that can detect any number of errors on an asymmetric channel provided that only transitions of cleared bits to set bits or set bits to cleared bits can occur.

4.0 Conclusion

In this unit, you have learnt about how transmission error can be detected using different detection scheme.

5.0 Summary

The main points in this unit are

- Error Detection Process
  Transmitter
  - For a given frame, an error-detecting code (check bits) is calculated from data bits
  - Check bits are appended to data bits

  Receiver
  - Separates incoming frame into data bits and check bits
  - Calculates check bits from received data bits
  - Compares calculated check bits against received check bits

- Detected error occurs if mismatch

- Parity Check
  Parity bit appended to a block of data
  - Even parity
    - Added bit ensures an even number of 1s
  - Odd parity
    - Added bit ensures an odd number of 1s
  - Example, 7-bit character [1110001]
    - Even parity [11100010]
    - Odd parity [11100011]
- Cyclic Redundancy Check (CRC)
  - Transmitter
    - For a *k*-bit block, transmitter generates an (*n*-*k*)-bit frame check sequence (FCS)
    - Resulting frame of *n* bits is exactly divisible by predetermined number
  - Receiver
    - Divides incoming frame by predetermined number
    - If no remainder, assumes no error

6.0 Tutor-Marked Assignment

(i) What is error detection?
(ii) Mention at least five error detection codes
(iii) Describe how transmission error can be detected with parity checking.
(iv) Discuss on cyclic redundancy check error detection

7.0 References/Further Readings

Clark G. C. and Cain J. B (1981). *Error-Correction Coding for Digital Communications*. New York: Plenum Press. ISBN 0-306-40615-2.

Dana M. (2005). "Communication speed nears terminal velocity". *New Scientist* 187 (2507): 38–41. ISSN 0262-4079.

EDN Magazine (1998). Communication-systems error simulation resolves trade-offs, Available online at: http://www.ednmag.com/reg/1998/061898/13df_03.cfm

Fred H. (1998). Data Communication, Computer Networks and Open Systems, Fourth Edition, Addison-Wesley Publishhing Company Inc., United States of America

Glaise, René J. (1997). "A two-step computation of cyclic redundancy code CRC-32 for ATM networks". *IBM Journal of Research and Development* (Armonk, NY: IBM) 41 (6): 705. Available online at http://www.research.ibm.com/journal/rd/416/glaise.html.

Lemmon, J.J. (2002). Wireless link statistical bit error model. US National Telecommunications and Information Administration (NTIA) Report 02-394

Lin S and Costello D. J. (1983). *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-283796-X.

Richard B. (1994). Web site www.cl.cam.ac.uk/Research/SRG/bluebook/21/crc/crc.html. University of Cambridge Computer Laboratory Systems Research Group.

Ryan W.E. and Lin S. (2009). *Channel Codes: Classical and Modern*. Cambridge University Press. ISBN 978-0-521-84868-8.

Tanenbaum, A. S. (1988). *Computer Networks*, Second Edition. Prentice Hall.

Wicker, Stephen B. (1995). *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-200809-2.

Williams, Ross N. (1996). "A Painless Guide to CRC Error Detection Algorithms V3.00" Available online at: http://www.repairfaq.org/filipg/LINK/F_crc_v3.html.

Wilson, Stephen G. (1996). *Digital Modulation and Coding*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-210071-1.

UNIT 2: ERROR CONTROL

1.0 Introduction

Error control is a method that can be used to recover the corrupted data whenever possible. There are two basic types of error control which are backward error control and forward error control. In backward error control, the data is encoded so that the encoded data contains additional redundant information which is used to detect the corrupted blocks of data that must be resent. On the contrary, in forward error control (FEQ), the data is encoded so that it contains enough redundant information to recover from some communications errors.

2.0 Objectives

At the end of this unit, you should be able to:

(i) explain the concept of backward error control
(ii) mention the types of ARQ
(iii) describe block code
(iv) state the practical application of Golay codes
(v) describe the principle of convolutional encoding
(vi) outline the basic approach to Hybrid scheme

3.0 Main Content

3.1 Types of Error Control/Correction

3.1.1 Backward error control

Using backward error correction requires a two-way communication channel. The sender divides the data in blocks, encodes the data with redundant additional information that is used to detect communications errors. The receiver applies error detection and if the receiver detects errors in the incoming blocks it requests the sender to resend the block. This mechanism is also called the automatic repeat request (ARQ) or Automatic Repeat Query. The ARQ can always repair any errors it can detect but it causes a variable delay on the data transfer.

Automatic Repeat reQuest (ARQ) is an error control method for data transmission that makes use of error-detection codes, acknowledgment (messages sent by the receiver indicating that it has correctly received a data frame or packet) and/or negative acknowledgment messages and timeouts (specified periods of time allowed to elapse before an acknowledgment is to be received) to achieve reliable data transmission over an unreliable service. Usually, when the transmitter does not receive the acknowledgment before the timeout occurs (i.e., within a reasonable amount of time after sending the data frame), it retransmits the frame until it is either correctly received or the error persists beyond a predetermined number of retransmissions. The basic types of ARQ are idle RQ and continuous RQ. The backward error control is used in many data transfer protocols. In addition, in order to use data compression, the communications errors must always be corrected.

(a) Idle RQ

Idle RQ is a fundamental backward correction scheme used in many protocols. The data is transferred in packets by using error detection. The receiver checks the incoming packets and sends an acknowledgement (ACK) to the sender if the packet was valid. If the sender receives an acknowledgement in the specified time it sends the next packet to the receiver. Otherwise, the sender must resend the packet. Idle RQ is very simple but it is often too inefficient. It can only

send data to one direction at a time. In addition, the delay on the data transfer may result in situation where only a small fraction of the capacity of the communications link is used.

(b) Continuous RQ

Continuous RQ is an improvement over Idle RQ when there is a delay on the data transfer. It allows several packets to be sent continuously. Therefore, the sender must use packet numbering. The receiver also receives packets continuously and sends an acknowledgement containing a packet number after receiving a valid packet. In case the sender cannot get an acknowledgement it starts resending packets in either of the two ways. Three types of ARQ protocols are Stop-and-wait ARQ, Go-Back-N ARQ and Selective Repeat ARQ.

In selective repeat, only each block that was corrupted is resent. Selective repeat is complex but it is useful when errors are common. In go-back-n, once a corrupted block is detected, the transmission continues from the corrupted block and all blocks after the corrupted blocks are discarded. Go-back-n is less effective than selective repeat but it is also very simple and it is almost equally effective if the errors are infrequent. Stop-and-wait ARQ is a method used in telecommunications to send information between two connected devices. It ensures that information is not lost due to dropped packets and that packets are received in the correct order. It is the simplest kind of automatic repeat-request (ARQ) method. A stop-and-wait ARQ sender sends one frame at a time, after sending each frame; the sender doesn't send any further frames until it receives an acknowledgement (ACK) signal. After receiving a good frame, the receiver sends an ACK. If the ACK does not reach the sender before a certain time, known as the timeout, the sender sends the same frame again.

3.1.2 Forward error control (FEQ)

Error correction codes or forward correction codes (FEC) or forward error control (FEQ) is designed to detect and correct errors. Using FEQ requires a one-directional channel only. The data is encoded to contain enough additional redundant information to recover from some communication errors. Unfortunately, the FEQ cannot recover from errors only when enough information has been successfully received. There is no way to recover from errors when this is not the case. However, the FEQ operates continuously without any interrupts and it ensures constant delay on the data transfer which is useful for real-time applications.

(a) **Block Code**

In coding theory, block codes are one of the two common types of channel codes (the other one being convolutional codes), which enable reliable transmission of digital data over unreliable communication channels subject to channel noise.

A block code transforms a message m consisting of a sequence of information symbols over an alphabet Σ into a fixed-length sequence c of n encoding symbols, called a code word. In a linear block code, each input message has a fixed length of $k < n$ input symbols. The redundancy added to a message by transforming it into a larger code word enables a receiver to detect and correct errors in a transmitted code word and – using a suitable decoding algorithm – to recover the original message. The redundancy is described in terms of its information rate or more simply – for a linear block code – in terms of its code rate, k/n.

The error correction performance of a block code is described by the minimum Hamming distance d between each pair of code words and is called the distance of the code.

Definition

The encoder for a block code divides the information sequence into message blocks, each message block contains k information symbols over an alphabet set Σ, i.e. a message could be represented as a k-tuple $m = (m_1, m_2, ..., m_k) \in \sum^k$. The total number of possible different message is therefore $|\Sigma|^k$. The encoder transforms message m independently onto an n-tuple codeword $C = (c_1, c_2, ..., c_n) \in \sum^n$. The code of block length n over Σ is a subset of Σ^n : the total number of possible different code words is the same as the total number of messages $|\Sigma|^k$ and k is called dimension. Rate of the block code is defined as $R = \frac{k}{n}$.

The Hamming weight wt(c) of a codeword $\boldsymbol{c}$ is defined as the number of non-zero positions in $\boldsymbol{c}$. The Hamming distance $\Delta(c_1,c_2)$ between two code words c_1 and c_2 is defined as the number of different positions between the code words. The (minimum) distance d of a block code $C \subseteq \sum^n$ is defined as the minimum distance between any two different code words: $d = \min_{c_1 \neq c_2 \in C} \Delta(c_1, c_2)$.

The notation $(n, k, d)_{\sum}$ is used to represent a block code of dimension k, block length n over alphabet set Σ, with distance d. In particular, if alphabet set Σ has size q, the block code is denoted as $(n, k, d)_q$.

(i) Hamming single-bit code

Hamming single-bit is a block code in which each block is separate from each other. The input block size can be made as small as possible. The number of bit errors to be corrected can be specified by adding enough redundant information in the encoding process. The minimum number of bits that differ on all possible code words is called the Hamming distance. This means that the error bursts which is shorter than the Hamming distance can be detected. Therefore, to detect communications errors, the Hamming distance of the line code must be longer than the length of the error bursts. A N-bit error requires an encoding with a Hamming distance of N+1 for detection and 2*N+1 for recovery.

(ii) Golay forward error correction

Golay codes are block codes that allow short code words. The perfect Golay code is an encoding that encodes 12 bits into 23 bits, denoted by (23, 12). It allows the correction of three or fewer single bit errors. The extended Golay code contains an additional parity bit which allows up to four errors to be detected. The resulting code is (24,12) which is also known as half-rate Golay code. The decoding can be performed by using either soft or hard decisions. The soft decisions provide better error correction but require more processing. Golay codes are useful in applications that require low latency and short codeword length. Therefore, Golay codes are used in real-time applications and radio communications.

Practical applications of Golay codes

(i) NASA Deep Space Missions: The Voyager 1 & 2 spacecraft needed to transmit hundreds of color pictures of Jupiter and Saturn in their 1979, 1980, and 1981 fly-bys within a constrained telecommunications bandwidth.

- Color image transmission required three times the amount of data, so the Golay (24,12,8) code was used.
- This Golay code is only triple-error correcting, but it could be transmitted at a much higher data rate than the Hadamard code that was used during the Mariner mission.

(ii) ALE HF data communications: The new American government standards for automatic link establishment (ALE) in High Frequency (HF) radio systems specifies the use of an extended (24,12) Golay block code for forward error correction (FEC).

- The Extended (24,12) Golay Code specified is a (24,12) block code.
- This code encodes 12 data bits to produce 24-bit code words.
- It is furthermore a systematic code, meaning that the 12 data bits are present in unchanged form in the code word.

(iii) BCH Codes

A BCH code is a polynomial code over a finite field with a particularly chosen generator polynomial. It is also a cyclic code.

In technical terms a BCH code is a multilevel cyclic variable-length digital error-correcting code used to correct multiple random error patterns. BCH codes may also be used with multilevel phase-shift keying whenever the number of levels is a prime number or a power of a prime number. A BCH code in 11 levels has been used to represent the 10 decimal digits plus a sign digit.

Simplified BCH Code

Definition: Fix a finite field $GF(q^m)$, where q is a prime. Also fix positive integers n and d such that $n = q^m - 1$ and $2 \le d \le n$. A polynomial code is constructed over $GF(q)$ with code length n, whose minimum Hamming distance is at least d. What remains to be specified is the generator polynomial of this code.

Let α be a primitive nth root of unity in $GF(q^m)$. For all i, let $m_i(x)$ be the minimal polynomial of α^i with coefficients in $GF(q)$. The generator polynomial of the BCH code is defined as the least common multiple $g(x) = lcm(m_1(x), \ldots, m_{d-1}(x))$.

Example

Let $q = 2$ and $m = 4$ (therefore $n = 15$). Different values of d *are considered*. There is a primitive root $\alpha \in GF(16)$ satisfying

$$\alpha^4 + \alpha + 1 = 0 \quad (1)$$

its minimal polynomial over $GF(2)$ is : $m_1(x) = x^4 + x + 1$. Note that in $GF(2^4)$, the equation $(a + b)^2 = a^2 + 2ab + b^2 = a^2 + b^2$ holds and therefore $m_1(\alpha^2) = m_1(\alpha)^2 = 0$. Thus α^2 is a root of $m_1(x)$ and therefore

$$m_2(x) = m_1(x) = x^4 + x + 1.$$

To compute $m_3(x)$, notice that, by repeated application of (1), we have the following linear relations:

- $1 = 0\alpha^3 + 0\alpha^2 + 0\alpha + 1$

- $\alpha^3 = 1\alpha^3 + 0\alpha^2 + 0\alpha + 0$
- $\alpha^6 = 1\alpha^3 + 1\alpha^2 + 0\alpha + 0$
- $\alpha^9 = 1\alpha^3 + 0\alpha^2 + 1\alpha + 0$
- $\alpha^{12} = 1\alpha^3 + 1\alpha^2 + 1\alpha + 1$

Five right-hand-sides of length four must be linearly dependent and indeed we find a linear dependency $\alpha^{12} + \alpha^9 + \alpha^6 + \alpha^3 + 1 = 0$. Since there is no smaller degree dependency, the minimal polynomial of α^3 is: $m_3(x) = x^4 + x^3 + x^2 + x + 1$. Continuing in a similar manner, we find

$$m_4(x) = m_2(x) = m_1(x) = x^4 + x + 1,$$
$$m_5(x) = x^2 + x + 1,$$
$$m_6(x) = m_3(x) = x^4 + x^3 + x^2 + x + 1,$$
$$m_7(x) = x^4 + x^3 + 1.$$

The BCH code with $d = 1,2,3$ has generator polynomial

$$g(x) = m_1(x) = x^4 + x + 1.$$

It has minimal Hamming distance at least 3 and corrects up to 1 error. Since the generator polynomial is of degree 4, this code has 11 data bits and 4 checksum bits.

The BCH code with $d = 4,5$ has generator polynomial

$$g(x) = lcm(m_1(x), m_3(x)) = (x^4 + x + 1)(x^4 + x^3 + x^2 + x + 1) = x^8 + x^7 + x^6 + x^4 + 1$$

It has minimal Hamming distance at least 5 and corrects up to 2 errors. Since the generator polynomial is of degree 8, this code has 7 data bits and 8 checksum bits.

The BCH code with $d = 6, 7$ has generator polynomial

$$g(x) = lcm(m_1(x), m_3(x), m_5(x)) =$$
$$= (x^4 + x + 1)(x^4 + x^3 + x^2 + x + 1)(x^2 + x + 1)$$
$$= x^{10} + x^8 + x^5 + x^4 + x^2 + x + 1$$

It has minimal Hamming distance at least 7 and corrects up to 3 errors. This code has 5 data bits and 10 checksum bits.

The BCH code with $d = 8$ and higher have generator polynomial

$$g(x) = lcm(m_1(x), m_3(x), m_5(x), m_7(x)) =$$
$$= (x^4 + x + 1)(x^4 + x^3 + x^2 + x + 1)(x^2 + x + 1)(x^4 + x^3 + 1)$$
$$= x^{14} + x^{13} + x^{12} + \ldots + x^2 + x + 1$$

This code has minimal Hamming distance 8 and corrects up to 3 errors. It has 1 data bit and 14 checksum bits. In fact, this code has only two codewords: 000000000000000 and 111111111111111.

General BCH codes

General BCH codes differ from the simplified case discussed above in two respects. First, one replaces the requirement $n = q^m - 1$ by a more general condition. Second, the consecutive roots of the generator polynomial may run from $\alpha^c,...,\alpha^{c+d-2}$ instead of $\alpha\ ,...,\alpha^{d-1}$.

Definition: Fix a finite field $GF(q)$, where q is a prime power. Choose positive integers m, n, d, c such that $2 \le d \le n$, $\gcd(n, q) = 1$, and m is the multiplicative order of q modulo n.
As before, let α be a primitive nth root of unity in $GF(q^m)$, and let $m_i(x)$ be the minimal polynomial over $GF(q)$ of α^i for all i. The generator polynomial of the BCH code is defined as the least common multiple $g(x) = lcm(m_c(x),...,m_{c+d-2}(x))$.

Note: if $n = q^m - 1$ as in the simplified definition, then $\gcd(n, q)$ is automatically 1 and the order of q modulo n is automatically m. Therefore, the simplified definition is indeed a special case of the general one.

(iv) Reed-Solomon forward error correction with interleaving

The Reed-Solomon forward error correction with interleaving is a forward error correction scheme that is intended to be used with high-quality video communications. The encoding is performed by filling a two dimensional array of 128*47 octets, that is, 128 octets in each row containing 124 octets of data and 4 octets of redundant check data. The encoding is done by filling the buffer column wise 47 octets at a time. After this has been repeated 124 times the buffer is full and it is encoded by writing each row at a time. This encoding allows two cells to be corrected or 4 cells to be reconstructed. Two interleave buffers are required because a single buffer can only be either read or written at a time. The decoder also needs two buffers for the same reason. The encoder writes a row at a time, then performs the possible recovery and reconstruction of defective cells and reads the array column wise. Unfortunately, the encoding and decoding both cause an additional delay on the data transfer that is equal to the transmission time of a single buffer. This type of encoding does not completely repair all error but it ensures high-quality throughput in real time.

(b) Convolutional forward error correction

In telecommunication, a convolutional code is a type of error-correcting code in which

- each *m*-bit information symbol (each *m*-bit string) to be encoded is transformed into an *n*-bit symbol, where *m*/*n* is the code *rate* ($n \ge m$) and
- the transformation is a function of the last *k* information symbols, where *k* is the constraint length of the code.

Convolutional codes are used extensively in numerous applications in order to achieve reliable data transfer including digital video, radio, mobile communication and satellite communication.
In block codes, each block is independent of other blocks. On the contrary, in the convolutional forward error correction, the encoded data depends on both the current data and the previous data. The convolutional encoder contains a shift register that is shifted each time a new bit is added. The length of the shift register is called the constraint length and it contains the memory of the encoder. Each new input bit is then encoded with each bit in the shift register by using modulo-2 adders.

The decoding is more difficult than the encoding. The data is decoded by using the Viterbi algorithm that tries to find the best solution for the decoding. Of course, all errors cannot be corrected but the error rate can be decreased. The convolutional error correction has an advantage of using all previous correctly received bits for error correction.

Convolutional encoding

To convolutionally encode data, start with *k* memory registers, each holding 1 input bit. Unless otherwise specified, all memory registers start with a value of 0. The encoder has *n* modulo-2 adders (a modulo 2 adder can be implemented with a single Boolean XOR gate, where the logic is: 0+0 = 0, 0+1 = 1, 1+0 = 1, 1+1 = 0) and *n* generator polynomials — one for each adder (figure 2.1). An input bit m_1 is fed into the leftmost register. Using the generator polynomials and the existing values in the remaining registers, the encoder outputs *n* bits. Now bit shift all register values to the right (m_1 moves to m_0, m_0 moves to m_{-1}) and wait for the next input bit. If there are no remaining input bits, the encoder continues output until all registers have returned to the zero state.

Figure 2.1 is a rate 1/3 (*m*/*n*) encoder with constraint length (*k*) of 3. Generator polynomials are G_1 = (1,1,1), G_2 = (0,1,1), and G_3 = (1,0,1). Therefore, output bits are calculated (modulo 2) as follows:

$n_1 = m_1 + m_0 + m_{-1}$

$n_2 = m_0 + m_{-1}$

$n_3 = m_1 + m_{-1}$.

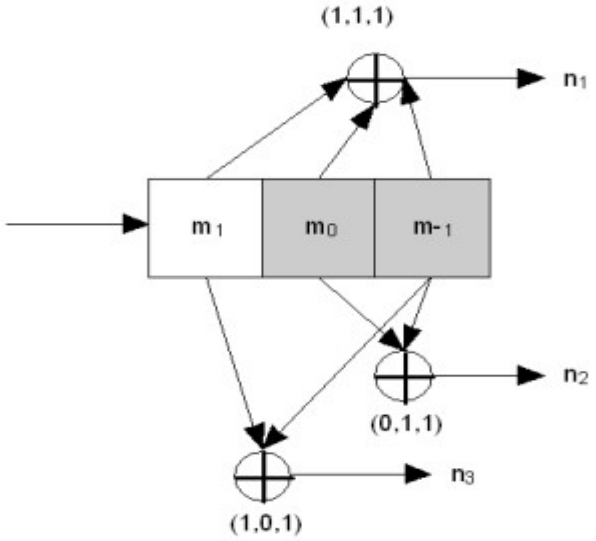


Figure 2.1: Rate 1/3 non-recursive, non-systematic convolutional encoder with constraint length 3

Recursive and non-recursive codes

The encoder on figure 2.1 above is a *non-recursive* encoder. Here's an example of a recursive one:

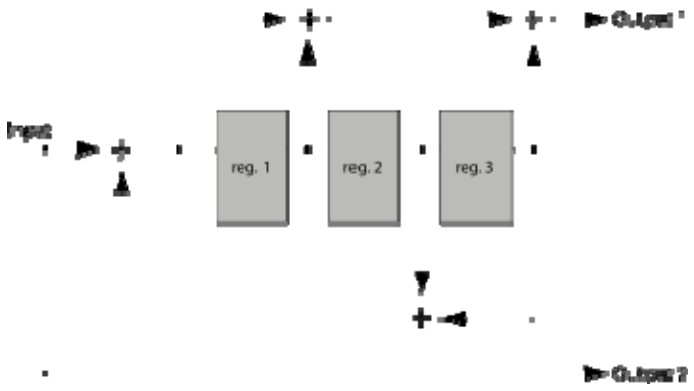


Figure 2.2: Rate 1/2 recursive, systematic convolutional encoder with constraint length 4

From the figure 2.2, the input being encoded is included in the output sequence too (output 2). Such codes are referred to as *systematic*; otherwise the code is called *non-systematic*.
Recursive codes are almost always systematic and conversely, non-recursive codes are non-systematic. It isn't a strict requirement, but a common practice.

Impulse response, transfer function, and constraint length
A convolutional encoder is called, so because it performs a convolution of the input stream with the encoder's impulse responses:

$$y_i^j = \sum_{k=0}^{\infty} h_k^j x_{i-k_j}$$

where x is an input sequence, y^j is a sequence from output *j* and h^j is an impulse response for output *j*.
A convolutional encoder is a discrete linear time-invariant system. Every output of an encoder can be described by its own transfer function, which is closely related to a generator polynomial. An impulse response is connected with a transfer function through Z-transform.
Transfer functions for the first (non-recursive) encoder are:

$$H_1(z) = 1 + z^{-1} + z^{-2},$$
$$H_2(z) = z^{-1} + z^{-2},$$
$$H_3(z) = 1 + z^{-2}$$

Transfer functions for the second (recursive) encoder are:

$$H_1(z) = \frac{1 + z^{-1} + z^{-3}}{1 - z^{-2} - z^{-3}},$$
$$H_2(z) = 1.$$

Define m by

$$m = \max_i poly\deg\left(H_i\left(\frac{1}{z}\right)\right)$$

where, for any rational function $f(z) = P(z)/Q(z)$,

$$poly\ \deg(f) = \max(\deg(P), \deg(Q)).$$

.

Then m is the maximum of the polynomial degrees of the $H_i\left(\frac{1}{z}\right)$ and the constraint length is defined as $K = m + 1$. For instance, in the first example the constraint length is 3, and in the second the constraint length is 4.

Trellis diagram

A convolutional encoder is a finite state machine. An encoder with n binary cells will have 2^n states.

Imagine that the encoder (shown on figure 2.1 above) has '1' in the left memory cell (m_0) and '0' in the right one (m_{-1}). (m_1 is not really a memory cell because it represents a current value). We will designate such a state as "10". According to an input bit the encoder at the next turn can convert either to the "01" state or the "11" state. One can see that not all transitions are possible (e.g., a decoder can't convert from "10" state to "00" or even stay in "10" state).

All possible transitions are shown below:

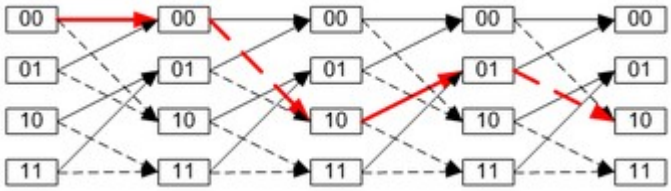


Figure 2.3: A trellis diagram for the encoder on figure 2.1.

A path through the trellis is shown as a red line. The solid lines indicate transitions where a "0" is input and the dashed lines where a "1" is input.

An actual encoded sequence can be represented as a path on this graph. One valid path is shown in red as an example.

This diagram gives us an idea about decoding: if a received sequence doesn't fit this graph, then it was received with errors and we must choose the nearest correct (fitting the graph) sequence. The real decoding algorithms exploit this idea.

Free distance and error distribution

The free distance (d) is the minimal Hamming distance between different encoded sequences. The correcting capability (t) of a convolutional code is the number of errors that can be corrected by the code. It can be calculated as:

$$t = \frac{d-1}{2}$$

Since a convolutional code doesn't use blocks, processing instead a continuous bitstream, the value of t applies to a quantity of errors located relatively near to each other. That is, multiple groups of t errors can usually be fixed when they are relatively far apart.

Free distance can be interpreted as the minimal length of an erroneous "burst" at the output of a convolutional decoder. The fact that errors appear as "bursts" should be accounted for when

designing a concatenated code with an inner convolutional code. The popular solution for this problem is to interleave data before convolutional encoding, so that the outer block (usually Reed-Solomon) code can correct most of the errors.

Decoding convolutional codes

Several algorithms exist for decoding convolutional codes. For relatively small values of *k*, the Viterbi algorithm is universally used as it provides maximum likelihood performance and is highly parallelizable. Viterbi decoders are thus easy to implement in VLSI hardware and in software on CPUs with SIMD instruction sets.

3.1.3 Hybrid schemes

Hybrid ARQ is a combination of ARQ and forward error correction. There are two basic approaches:

- Messages are always transmitted with FEC parity data (and error-detection redundancy). A receiver decodes a message using the parity information, and requests retransmission using ARQ only if the parity data was not sufficient for successful decoding (identified through a failed integrity check).
- Messages are transmitted without parity data (only with error-detection information). If a receiver detects an error, it requests FEC information from the transmitter using ARQ, and uses it to reconstruct the original message.

The latter approach is particularly attractive on an erasure channel when using a rateless erasure code.

4.0 Conclusion

In this unit, you have learnt on how transmission error can be control or corrected using different error control codes such as ARQ, hamming code, Golay codes, BCH codes, Reed Solomon codes, etc.

5.0 Summary

The main points in this unit are

- Automatic Repeat reQuest (ARQ) is an error control method for data transmission that makes use of error-detection codes, acknowledgment and/or negative acknowledgment messages and timeouts to achieve reliable data transmission over an unreliable service.
- Block codes is a common types of channel codes which enable reliable transmission of digital data over unreliable communication channels subject to channel noise.
- Hamming single-bit is a block code in which each block is separate from each other.
- Golay codes are block codes that allow short code words.
- A BCH code is a polynomial code over a finite field with a particularly chosen generator polynomial. It is also a cyclic code.
- Reed–Solomon (RS) codes are non-binary cyclic error-correcting codes, which described a systematic way of building codes that could detect and correct multiple random symbol errors.

- A convolutional code is a type of error-correcting code in which each *m*-bit information symbol (each *m*-bit string) to be encoded is transformed into an *n*-bit symbol, where *m*/*n* is the code *rate* ($n \geq m$) and also the transformation is a function of the last *k* information symbols, where *k* is the constraint length of the code.
- Hybrid ARQ is a combination of ARQ and forward error correction.

6.0 Tutor-Marked Assignment

(i) Discuss on backward error control
(ii) List the types of ARQ
(iii) Describe block code
(iv) Mention the application areas of Golay codes
(v) How is convolutional encoding performed?
(vi) Outline the two basic approaches to Hybrid scheme

7.0 References/Further Readings

Alhoniemi E. (1998). Error Detection and Control in Data Transfer

Clark G. C. and Cain J. B. (1981). *Error-Correction Coding for Digital Communications*. New York: Plenum Press. ISBN 0-306-40615-2.

Dana M. (2005). "Communication speed nears terminal velocity". *New Scientist* 187 (2507): 38–41. ISSN 0262-4079.

Feldman J., Abou-Faycal I. and Frigo M. (2002). "A Fast Maximum-Likelihood Decoder for Convolutional Codes". *Vehicular Technology Conference* 1: 371–375.

Gilbert W. J. and Nicholson W. K. (2004), *Modern Algebra with Applications* (2nd ed.), John Wiley

Hong J. and Vetterli M. (1995), "Simple Algorithms for BCH Decoding", *IEEE Transactions on Communications* 43 (8): 2324–2333

Huffman W. and Pless V. (2003). *Fundamentals of error-correcting codes*. Cambridge University Press. ISBN 13: 9780521782807.

Lin S. and Costello D. J.(1983). *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-283796-X.

Lin S.and Costello D. J.(1983). *Error Control Coding: Fundamentals and Applications*. Prentice-Hall. ISBN 0-13-283796-X.

Lint J. H. (1992). *Introduction to Coding Theory*. GTM. (2nd ed.). Springer-Verlag. ISBN 3-540-54894-7.

M.S. Ryan and G.R. Nudd. (1993). The Viterbi Algorithm. Technical Report University of Warwick RR-238.

Peterson and Davie (2003). *Computer Networks: A Systems Approach*, Third Edition

Reed I. S. and Chen X. (1999), *Error-Control Coding for Data Networks*, Boston, MA: Kluwer Academic Publishers

Reed I.S. and Chen X. (1999), *Error-Control Coding for Data Networks*, Boston, MA: Kluwer Academic Publishers

Rudra A. (2010).*CSE 545, Error Correcting Codes: Combinatorics, Algorithms and Applications*, University at Buffalo, Available online at: http://www.cse.buffalo.edu/~atri/courses/coding-theory/

Ryan, W. E. and Lin S. (2009). *Channel Codes: Classical and Modern*. Cambridge University Press. ISBN 978-0-521-84868-8.
Wicker S. B. (1995). *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-200809-2.
Wicker S.B. (1994). *Error Control Systems for Digital Communication and Storage,* Englewood Cliffs, N.J.: Prentice-Hall
Wilson S. G. (1996). *Digital Modulation and Coding*. Englewood Cliffs NJ: Prentice-Hall. ISBN 0-13-210071-1.
Wireless Networks Spring (2005). Coding and Error Control

UNIT3 : APPLICATIONS OF ERROR CONTROL CODES

1.0 Introduction

In the previous units, you have learnt about how error can be detected in a transmission and how this same error can be control or corrected. In this unit, you will learn the different application areas of error control codes.

2.0 Objectives

At the end of this unit, you should be able to discuss on the application areas of error control codes

3.0 Main Content

3.1 Application areas of Error Control/Correction Codes

Applications that require low latency (such as telephone conversations) cannot use Automatic Repeat reQuest (ARQ); they must use Forward Error Correction (FEC). By the time an ARQ system discovers an error and re-transmits it, the re-sent data will arrive too late.

Applications where the transmitter immediately forgets the information as soon as it is sent (such as most television cameras) cannot use ARQ; they must use FEC because when an error occurs, the original data is no longer available. This is the reason why FEC is used in data storage systems such as RAID and distributed data store. Applications that use ARQ must have a return channel. Applications that have no return channel cannot use ARQ. Applications that require extremely low error rates (such as digital money transfers) must use ARQ.

(i) The Internet

In a typical TCP/IP stack, error control is performed at multiple levels:

- Each Ethernet frame carries a CRC-32 checksum. Frames received with incorrect checksums are discarded by the receiver hardware.
- The IPv4 header contains a checksum protecting the contents of the header. Packets with mismatching checksums are dropped within the network or at the receiver.
- The checksum was omitted from the IPv6 header in order to minimize processing costs in network routing and because current link layer technology is assumed to provide sufficient error detection (e.g. RFC 3819).
- UDP has an optional checksum covering the payload and addressing information from the UDP and IP headers. Packets with incorrect checksums are discarded by the operating system network stack. The checksum is optional under IPv4, only, because the IP layer checksum may already provide the desired level of error protection.
- TCP provides a checksum for protecting the payload and addressing information from the TCP and IP headers. Packets with incorrect checksums are discarded within the network stack, and eventually get retransmitted using ARQ, either explicitly (such as through triple-ack) or implicitly due to a timeout.

(ii) Deep-space telecommunications

Development of error-correction codes was tightly coupled with the history of deep-space missions due to the extreme dilution of signal power over interplanetary distances and the limited power availability aboard space probes. Whereas early missions sent their data uncoded, starting from 1968 digital error correction was implemented in the form of (sub-optimally decoded)

convolutional codes and Reed-Muller codes. The Reed-Muller code was well suited to the noise the spacecraft was subject to (approximately matching a bell curve), and was implemented at the Mariner spacecraft for missions between 1969 and 1977.
The Voyager 1 and Voyager 2 missions, which started in 1977, were designed to deliver color imaging amongst scientific information of Jupiter and Saturn. This resulted in increased coding requirements and thus the spacecraft were supported by (optimally Viterbi-decoded) convolutional codes that could be concatenated with an outer Golay (24,12,8) code. The Voyager 2 probe additionally supported an implementation of a Reed-Solomon code: the concatenated Reed-Solomon-Viterbi (RSV) code allowed for very powerful error correction and enabled the spacecraft's extended journey to Uranus and Neptune.

(iii) Satellite broadcasting (DVB)

The demand for satellite transponder bandwidth continues to grow, fueled by the desire to deliver television (including new channels and High Definition TV) and IP data. Transponder availability and bandwidth constraints have limited this growth because transponder capacity is determined by the selected modulation scheme and Forward error correction (FEC) rate.

(iv) Data storage

Error detection and correction codes are often used to improve the reliability of data storage media.

- A "parity track" was present on the first magnetic tape data storage in 1951. The "Optimal Rectangular Code" used in group code recording tapes not only detects but also corrects single-bit errors.
- Some file formats, particularly archive formats, include a checksum (most often CRC32) to detect corruption and truncation and can employ redundancy and/or parity files to recover portions of corrupted data.
- Reed Solomon codes are used in compact discs to correct errors caused by scratches.
- Modern hard drives use CRC codes to detect and Reed-Solomon codes to correct minor errors in sector reads and to recover data from sectors that have "gone bad" and store that data in the spare sectors.
- RAID systems use a variety of error correction techniques to correct errors when a hard drive completely fails.

(v) Error-correcting memory

Dynamic random access memory (DRAM) may provide increased protection against soft errors by relying on error correcting codes. Such error-correcting memory, known as error correcting code (ECC) or EDAC-protected memory is particularly desirable for high fault-tolerant applications, such as servers, as well as deep-space applications due to increased radiation.
Error-correcting memory controllers traditionally use Hamming codes, although some use triple modular redundancy.
Interleaving allows distributing the effect of a single cosmic ray potentially upsetting multiple physically neighboring bits across multiple words by associating neighboring bits to different words. As long as a single event upset (SEU) does not exceed the error threshold (e.g., a single error) in any particular word between accesses, it can be corrected (e.g., by a single-bit error correcting code) and the illusion of an error-free memory system may be maintained.

4.0 Conclusion

In this unit, you have learnt about the application areas of error control codes.

5.0 Summary

The application areas of error control codes are internet, deep-space telecommunication, satellite broadcasting, data storage and Error-correcting memory.

6.0 Tutor-Marked Assignment

(i) Outline the application areas where ARQ cannot be used.
(ii) Discuss on the application areas of error control codes

7.0 References/Further Readings

Andrews K (2007). The Development of Turbo and LDPC Codes for Deep-Space Applications, Proceedings of the IEEE, 95(11).

Huffman W, Pless V. (2003). , Fundamentals of error-correcting codes, Cambridge University Press, ISBN 9780521782807.

McAuley A. J. (1990) Reliable Broadband Communication Using a Burst Erasure Correcting Code, ACM SIGCOMM

Shu Lin, Daniel J. Costello, Jr. (1983). Error Control Coding: Fundamentals and Applications. Prentice Hall. ISBN 0-13-283796-X.

MODULE 4 : IEEE 802.11 WIRELESS

UNIT 1: OVERVIEW OF IEEE 802.11

1.0 Introduction

The purpose of this unit is to give you a basic overview of the new 802.11Standard, enabling you to understand the basic concepts, principle of operations and some of the reasons behind some of the features and/or components of the Standard.

IEEE 802.11 is a set of standards for implementing wireless local area network (WLAN) computer communication in the 2.4, 3.6 and 5 GHz frequency bands. They are created and maintained by the IEEE LAN/MAN Standards Committee (IEEE 802). The base current version of the standard is IEEE 802.11-2007.

2.0 Objectives

At the end of this unit, you should be able to

- List the IEEE 802.11 standards
- Discuss on the IEEE 802.11 standards

3.0 Main Content

3.1 Overview of IEEE 802.11

The 802.11 family consists of a series of over-the-air modulation techniques that use the same basic protocol. The most popular are those defined by the 802.11b and 802.11g protocols, which are amendments to the original standard. 802.11-1997 was the first wireless networking standard, but 802.11b was the first widely accepted one, followed by 802.11g and 802.11n. Security was originally purposefully weak due to export requirements of some governments and was later enhanced via the 802.11i amendment after governmental and legislative changes. 802.11n is a new multi-streaming modulation technique. Other standards in the family (c–f, h, j) are service amendments and extensions or corrections to the previous specifications.

802.11-1997 (802.11 legacy)

The original version of the standard IEEE 802.11 was released in 1997 and clarified in 1999 but is today obsolete. It specified two net bit rates of 1 or 2 megabits per second (Mbit/s), plus forward error correction code. It specified three alternative physical layer technologies: diffuse infrared operating at 1 Mbit/s; frequency-hopping spread spectrum operating at 1 Mbit/s or 2 Mbit/s; and direct-sequence spread spectrum operating at 1 Mbit/s or 2 Mbit/s. The latter two radio technologies used microwave transmission over the Industrial Scientific Medical frequency band at 2.4 GHz. Some earlier WLAN technologies used lower frequencies, such as the U.S. 900 MHz ISM band. Legacy 802.11 with direct-sequence spread spectrum was rapidly supplanted and popularized by 802.11b.

802.11a

The 802.11a standard uses the same data link layer protocol and frame format as the original standard but an orthogonal frequency division multiplexing (OFDM) based air interface (physical layer). It operates in the 5 GHz band with a maximum net data rate of 54 Mbit/s plus error correction code, which yields realistic net achievable throughput in the mid-20 Mbit/s.

Since the 2.4 GHz band is heavily used to the point of being crowded, using the relatively unused 5 GHz band gives 802.11a a significant advantage. However, this high carrier frequency also brings a disadvantage: the effective overall range of 802.11a is less than that of 802.11b/g. In theory, 802.11a signals are absorbed more readily by walls and other solid objects in their path due to their smaller wavelength and as a result cannot penetrate as far as those of 802.11b. In practice, 802.11b typically has a higher range at low speeds (802.11b will reduce speed to 5 Mbit/s or even 1 Mbit/s at low signal strengths). However, at higher speeds, 802.11a often has the same or greater range due to less interference.

802.11b

802.11b has a maximum raw data rate of 11 Mbit/s and uses the same media access method defined in the original standard. 802.11b products appeared on the market in early 2000, since 802.11b is a direct extension of the modulation technique defined in the original standard. The dramatic increase in throughput of 802.11b (compared to the original standard) along with simultaneous substantial price reductions led to the rapid acceptance of 802.11b as the definitive wireless LAN technology.

802.11b devices suffer interference from other products operating in the 2.4 GHz band. Devices operating in the 2.4 GHz range include: microwave ovens, Bluetooth devices, baby monitors and cordless telephones.

802.11g

In June 2003, a third modulation standard was ratified: 802.11g. This works in the 2.4 GHz band (like 802.11b) but uses the same OFDM based transmission scheme as 802.11a. It operates at a maximum physical layer bit rate of 54 Mbit/s exclusive of forward error correction codes or about 22 Mbit/s average throughput. 802.11g hardware is fully backwards compatible with 802.11b hardware and therefore is encumbered with legacy issues that reduce throughput when compared to 802.11a by ~21%.

The then-proposed 802.11g standard was rapidly adopted by consumers starting in January 2003, well before ratification due to the desire for higher data rates as well as reductions in manufacturing costs. By summer 2003, most dual-band 802.11a/b products became dual-band/tri-mode, supporting a and b/g in a single mobile adapter card or access point.

Like 802.11b, 802.11g devices suffer interference from other products operating in the 2.4 GHz band, for example wireless keyboards.

802.11-2007

In 2003, task group TGma was authorized to "roll up" many of the amendments to the 1999 version of the 802.11 standard. REVma or 802.11ma, as it was called, created a single document that merged 8 amendments (802.11a, b, d, e, g, h, i, j) with the base standard. Upon approval on March 8, 2007, 802.11REVma was renamed to the then-current base standard IEEE 802.11-2007.

802.11n

802.11n is an amendment which improves upon the previous 802.11 standards by adding multiple-input multiple-output antennas (MIMO). 802.11n operates on both the 2.4GHz and the lesser used 5 GHz bands. The IEEE has approved the amendment and it was published in October 2009. Prior to the final ratification, enterprise was already migrating to 802.11n networks based on the Wi-Fi Alliance's certification of products conforming to a 2007 draft of the 802.11n proposal.

4.0 Conclusion

In this unit, you have learnt about different IEEE 802.11 standard.

5.0 Summary

The different IEEE 802.11 standards are: 802.11, 802.11a, 802.11b, 802.11g, 802.11-2007 and 802.11n

6.0 Tutor-Marked Assignment

Mention and explain briefly the IEEE 802.11 standards

7.0 References/Further Readings

Brenner P. (1996). A Technical tutorial on the IEEE802.11protocol. Breezecom Wireless Communication

Intelligraphic (2010). Introduction to IEEE 802.11. Available online at http\www.intelligraphics.com\introduction-ieee-80211

Tutorial_Reports.com (2007). IEEE802.11 Architecture. Available online at http\www.tutorial-reports.com\wireless\WlanWiFi\WiFi_architecure.php

Wikipedia (2011). IEEE 802.11. Available online at http://en.wikipedia.org/wiki/IEEE_802.11

UNIT 2: IEEE 802.11 ARCHITECTURE

1.0 Introduction

The 802.11 architecture is comprised of several components and services that interact to provide station mobility transparent to the higher layers of the network stack.
The station (STA) is the most basic component of the wireless network. A station is any device that contains the functionality of the 802.11 protocol, that being MAC, PHY and a connection to the wireless media. Typically the 802.11 functions are implemented in the hardware and software of a network interface card (NIC).
A station could be a laptop PC, handheld device, or an Access Point. Stations may be mobile, portable, or stationary and all stations support the 802.11 station services of authentication, de-authentication, privacy, and data delivery.

2.0 Objectives

At the end of this unit, you should be able to
(i) define a station
(ii) mention two categories service
(iii) discuss on each type of service

3.0 Main Content

3.1 IEEE 802.11 Component

A station in 802.11 (Wireless Local Area Networks) is referred to as each computer; mobile, portable or fixed. The difference between a portable and mobile station is that a portable station moves from point to point but is only used at a fixed point while a mobile stations access the LAN during movement.

When two or more stations come together to communicate with each other, they form a Basic Service Set (BSS). 802.11 LANs use the BSS as the standard building block. A BSS that stands alone and is not connected to a base is called an Independent Basic Service Set (IBSS) or is referred to as an Ad-Hoc Network. An ad-hoc network is a network where stations communicate only in peer to peer. There is no base and no one gives permission to talk.

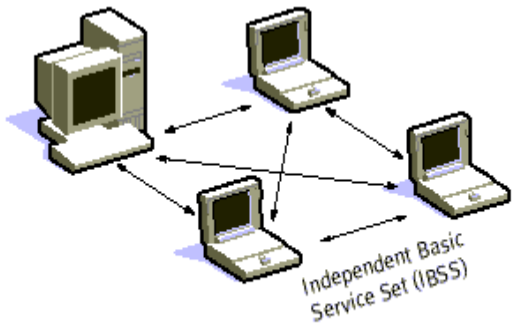


Fig 2.1: "Adhoc Mode"

When two or more BSS's interconnected using a Distribution System or DS, the network becomes one with infrastructure. Each BSS becomes a component of an extended, larger network. Entry to the DS is accomplished with the use of Access Points (AP) which is a station. So, data moves between the BSS and the DS with the help of these access points. The creation of large and complex networks using BSS's and DS's leads to the next level of hierarchy called the Extended Service Set or ESS. The entire network of ESS looks like an independent basic service set to the Logical Link Control layer (LLC). This means that stations within the ESS can communicate or even move between BSS′s transparently to the LLC.

The following diagram shows a typical 802.11 LAN including the components described above:

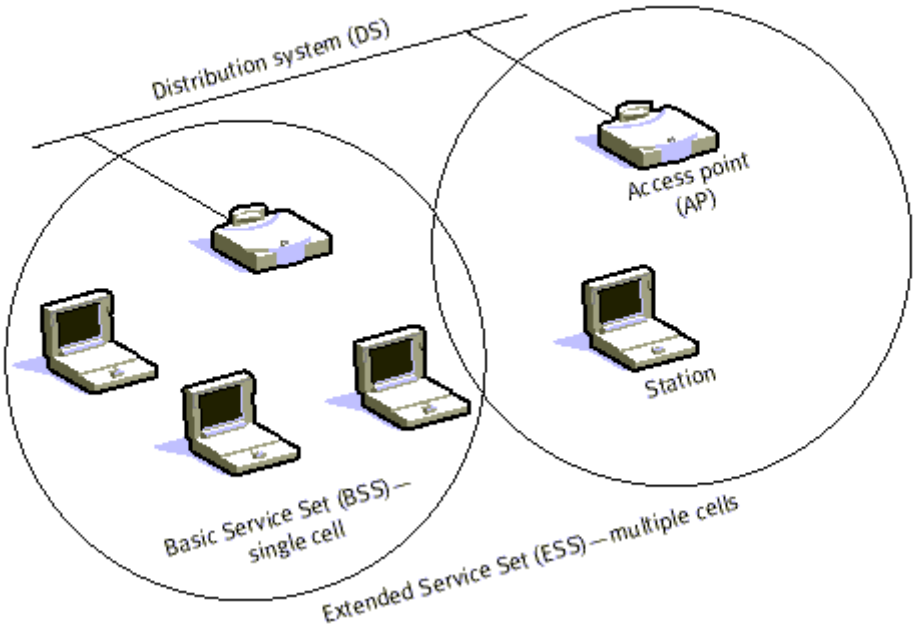


Figure 2.2: A Typical 802.11 LAN in Infrastructure Mode

One of the requirements of IEEE 802.11 is that it can be used with existing wired networks. 802.11 solved this challenge with the use of a Portal. A portal is a device that interconnects between an 802.11 and another 802 LAN. It can also serve as the access point to the DS. All data going to an 802.11 LAN from an 802.X LAN must pass through a portal. It thus functions as bridge between wired and wireless. The implementation of the DS is not specified by 802.11. Therefore, a distribution system may be created from existing or new technologies. A point-to-point bridge connecting LANs in two separate buildings could become a DS.

While the implementation for the DS is not specified, 802.11 do specify the services, which the DS must support. Services are divided into two sections

(a) Station Services (SS)
(b) Distribution System Services (DSS).

(a) Station Services

The 802.11 standard defines services for providing functions among stations. Station services are implemented within all stations on an 802.11 WLAN (including access points). The main thrust behind station services is to provide security and data delivery services for the WLAN.

(i) Authentication

With a wireless system, the medium is not exactly bounded as with a wired system. In order to control access to the network, stations must first establish their identity. This is much like trying to enter a radio net in the military. Before you are acknowledged and allowed to converse, you must first pass a series of tests to ensure that you are who you say you are. That is really all authentication is. Once a station has been authenticated, it may then associate itself. The authentication relationship may be between two stations inside an IBSS or to the AP of the BSS. Authentication outside of the BSS does not take place.

There are two types of authentication services offered by 802.11.

- **Open system authentication:** This is the default authentication method, which is a very simple, two-step process. First the station wanting to authenticate with another station sends an authentication management frame containing the sending station's identity. The receiving station then sends back a frame alerting whether it recognizes the identity of the authenticating station.
- **Shared key authentication:** This type of authentication assumes that each station has received a secret shared key through a secure channel independent of the 802.11 network. Stations authenticate through shared knowledge of the secret key. Use of shared key authentication requires implementation of encryption via the Wired Equivalent Privacy or WEP algorithm.

(ii) De-authentication

The de-authentication service is used to eliminate a previously authorized user from any further use of the network. Once a station is de-authenticated, that station is no longer able to access the WLAN without performing the authentication function again. De-authentication is a notification and cannot be refused. For example, when a station wishes to be removed from a BSS, it can send a de-authentication management frame to the associated access point to notify the access point of the removal from the network. An access point could also de-authenticate a station by sending a de-authentication frame to the station.

(iii) Privacy

Privacy is an encryption algorithm, which is used so that other 802.11 users cannot eavesdrop on your LAN traffic. IEEE 802.11 specifies Wired Equivalent Privacy (WEP) as an optional algorithm to satisfy privacy. If WEP is not used then stations are "in the clear" or "in the red", meaning that their traffic is not encrypted. Data transmitted in the clear are called plaintext. Data transmissions, which are encrypted, are called ciphertext. All stations start "in the red" until they are authenticated.

(iv) MAC Service Data Unit (MSDU) Delivery.

MSDU delivery ensures that the information in the MAC service data unit is delivered between the medium access control service access points (i.e. the data delivery service provides reliable delivery of data frames from the MAC in one station to the MAC in one or more other stations, with minimal duplication and reordering of frames).

(b) Distribution System Services (DSS).

Distribution services provide functionality across a distribution system. Typically, access points provide distribution services. The five distribution services and functions detailed below include: association, disassociation, re-association, distribution, and integration.

(i) Association

The association service is used to make a logical connection between a mobile station and an access point. Each station must become associated with an access point before it is allowed to send data through the access point onto the distribution system. The connection is necessary in order for the distribution system to know where and how to deliver data to the mobile station. The mobile station invokes the association service once and only once, typically when the station enters the BSS. Each station can associate with one access point though an access point can associate with multiple stations.

(ii) Disassociation

The disassociation service is used either to force a mobile station to eliminate an association with an access point or for a mobile station to inform an access point that it no longer requires the services of the distribution system. When a station becomes disassociated, it must begin a new association to communicate with an access point again. An access point may force a station or stations to disassociate because of resource restraints, the access point is shutting down or being removed from the network for a variety of reasons. When a mobile station is aware that it will no longer require the services of an access point, it may invoke the disassociation service to notify the access point that the logical connection to the services of the access point from this mobile station is no longer required.

(iii) Re-association

Re-Association enables a station to change its current association with an access point. The re-association service is similar to the association service, with the exception that it includes information about the access point with which a mobile station has been previously associated. A mobile station will use the re-association service repeatedly as it moves through out the ESS loses contact with the access point with which it is associated, and needs to become associated with a new access point.
The mobile station always initiates re-association.

(iv) Distribution

Distribution is the primary service used by an 802.11 station. A station uses the distribution service every time it sends MAC frames across the distribution system. The distribution service provides the distribution with only enough information to determine the proper destination BSS for the MAC frame. Distribution is simply getting the data from the sender to the intended receiver. The three association services (association, re-association, and disassociation) provide the necessary information for the distribution service to operate. Distribution within the

distribution system does not necessarily involve any additional features outside of the association services, though a station must be associated with an access point for the distribution service to forward frames properly.

(v) Integration

The integration service connects the 802.11 WLAN to other LANs including one or more wired LANs or 802.11 WLANs. A *portal* performs the integration service. The portal is an abstract architectural concept that typically resides in an access point though it could be part of a separate network component entirely.
The integration service translates 802.11 frames to frames that may traverse another network and vice versa as well as translates frames from other networks to frames that may be delivered by an 802.11 WLAN.

4.0 Conclusion

In this unit, you have learnt that the 802.11 architecture comprised of several components and services that interact to provide station mobility transparent to the higher layers of the network stack.

5.0 Summary

An 802.11 LAN is based on a cellular architecture where the system is subdivided into cells. Each cell (called Basic Service Set or BSS, in the 802.11 nomenclature) is controlled by a Base Station (called Access Point or in short, AP). Although a wireless LAN may be formed by a single cell with a single Access Point (optional), most installations will be formed by several cells, where the Access Points are connected through some kind of backbone called Distribution System or DS. This backbone is typically Ethernet and in some cases is wireless itself. The whole interconnected Wireless LAN including the different cells, their respective Access Points and the Distribution System, is seen as a single 802 network to the upper layers of the OSI model and is known in the Standard as Extended Service Set (ESS).

6.0 Tutor-Marked Assignment

(i) What is a station?
(ii) List two types of service
(iii) Discuss briefly on the type service
(iv) What do you understand by the concept: BSS, IBSS and ESS

7.0 References/Further Readings

Brenner P. (1996). A Technical tutorial on the IEEE802.11protocol. Breezecom Wireless Communication

Intelligraphic (2010). Introduction to IEEE 802.11. Available online at http\www.intelligraphics.com\introduction-ieee-80211

Tutorial_Reports.com (2007). IEEE802.11 Architecture. Available online at http\www.tutorial-reports.com\wireless\WlanWiFi\WiFi_architecure.php

Wikipedia (2011). IEEE 802.11. Available online at http://en.wikipedia.org/wiki/IEEE_802.11

UNIT 3: IEEE 802.11 ARCHITECTURE

1.0 Introduction

In this unit, you will learn about IEEE 802.11 layer description as well as the 802.11 standard defines frame type.

2.0 Objectives

At the end of this unit, you should be able to

(i) describe the IEEE 802.11 layer
(ii) state the function of PHY layer
(iii) mention the types of management frame
(iv) explain the three types of control frame

3.0 Main Content

3.1 IEEE 802.11 Layers Description

As any 802.x protocol, the 802.11 protocol covers the MAC and Physical Layer. The Standard currently defines a single MAC which interacts with three PHYs (all of them running at 1 and 2 Mbit/s) as follows:

- Frequency Hopping Spread Spectrum in the 2.4 GHz Band
- Direct Sequence Spread Spectrum in the 2.4 GHz Band, and
- InfraRed

<table>
<tr><td colspan="3">802.2</td><td rowspan="2">Data Link Layer</td></tr>
<tr><td colspan="3">802.11 MAC</td></tr>
<tr><td>FH</td><td>DS</td><td>IR</td><td>PHY Layer</td></tr>
</table>

Beside the standard functionality performed by MAC Layers, the 802.11 MAC also performs other functions that are typically related to upper layer protocols, such as Fragmentation, Packet Retransmissions and Acknowledges.

3.1.1 Data Link Layer

The data link layer within 802.11 consists of two sublayers: Logical Link Control (LLC) and Media Access Control (MAC).
802.11 uses the same 802.2 LLC and 48-bit addressing as other 802 LANs, allowing for very simple bridging from wireless to IEEE wired networks but the MAC is unique to WLANs.

The MAC Layer

The 802.11 MAC is very similar in concept to 802.3, in that it is designed to support multiple users on a shared medium by having the sender sense the medium before accessing it.
For 802.3 Ethernet LANs, the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol regulates how Ethernet stations establish access to the wire and how they detect and handle collisions that occur when two or more devices try to simultaneously

communicate over the LAN. In an 802.11 WLAN, collision detection is not possible due to what is known as the ―near/far‖ problem: to detect a collision, a station must be able to transmit and listen at the same time, but in radio systems the transmission drowns out the ability of the station to ―hear‖ a collision.

To account for this difference, 802.11 uses a slightly modified protocol known as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) or the Distributed Coordination Function (DCF). CSMA/CA attempts to avoid collisions by using explicit packet acknowledgment (ACK), which means an ACK packet is sent by the receiving station to confirm that the data packet arrived intact.

The MAC Layer defines two different access methods; the Distributed Coordination Function and the Point Coordination Function:

(a) Distributed Coordination Function: CSMA/CA

CSMA/CA works as follows. A station wishing to transmit senses the air, and, if no activity is detected, the station waits an additional, randomly selected period of time and then transmits if the medium is still free. If the packet is received intact, the receiving station issues an ACK frame that, once successfully received by the sender, completes the process. If the ACK frame is not detected by the sending station, either because the original data packet was not received intact or the ACK was not received intact, a collision is assumed to have occurred and the data packet is transmitted again after waiting another random amount of time.

CSMA/CA thus provides a way of sharing access over the air. This explicit ACK mechanism also handles interference and other radio related problems very effectively. However, it does add some overhead to 802.11 that 802.3 does not have, so that an 802.11 LAN will always have slower performance than an equivalent Ethernet LAN.

Another MAC-layer problem specific to wireless is the ―hidden node‖ issue, in which two stations on opposite sides of an access point can both ―hear‖ activity from an access point, but not from each other, usually due to distance or an obstruction.

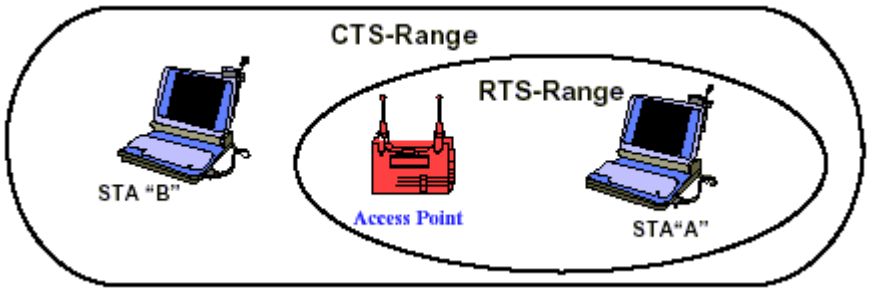


Fig 3.1: RTS/CTS Procedure eliminates the ―Hidden Node‖ Problem

To solve this problem, 802.11 specify an optional Request to Send/Clear to Send (RTS/CTS) protocol at the MAC layer. When this feature is in use, a sending station transmits an RTS and waits for the access point to reply with CTS. Since all stations in the network can hear the access

point, the CTS causes them to delay any intended transmissions, allowing the sending station to transmit and receive a packet acknowledgment without any chance of collision.

Since RTS/CTS adds additional overhead to the network by temporarily reserving the medium, it is typically used only on the largest-sized packets, for which retransmission would be expensive from a bandwidth standpoint.

Finally, the 802.11 MAC layer provides for two other robustness features: CRC checksum and packet fragmentation. Each packet has a CRC checksum calculated and attached to ensure that the data was not corrupted in transit. This is different from Ethernet, where higher-level protocols such as TCP handle error checking. Packet fragmentation allows large packets to be broken into smaller units when sent over the air, which is useful in very congested environments or when interference is a factor, since larger packets have a better chance of being corrupted. This technique reduces the need for retransmission in many cases and thus improves overall wireless network performance. The MAC layer is responsible for reassembling fragments received, rendering the process transparent to higher level protocols.

(b) Point Coordination Function (PCF)

Time-bounded data such as voice and video is supported in the 802.11 MAC specification through the Point Coordination Function (PCF). As opposed to the DCF, where control is distributed to all stations, in PCF mode a single access point controls access to the media. If a BSS is set up with PCF enabled, time is spliced between the system being in PCF mode and in DCF (CSMA/CA) mode. During the periods when the system is in PCF mode, the access point will poll each station for data, and after a given time move on to the next station. No station is allowed to transmit unless it is polled, and stations receive data from the access point only when they are polled. Since PCF gives every station a turn to transmit in a predetermined fashion, a maximum latency is guaranteed. A downside to PCF is that it is not particularly scalable, in that a single point needs to have control of media access and must poll all stations, which can be ineffective in large networks.

3.1.2 802.11 Physical Layer (PHY)

The 802.11 physical layer (PHY) is the interface between the MAC and the wireless media where frames are transmitted and received.

The PHY provides three functions:

(i) First, the PHY provides an interface to exchange frames with the upper MAC layer for transmission and reception of data.

(ii) Secondly, the PHY uses signal carrier and spread spectrum modulation to transmit data frames over the media.

(iii) Thirdly, the PHY provides a carrier sense indication back to the MAC to verify activity on the media.

802.11 provide three different PHY definitions:

- Both Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS) support 1 and 2 Mbps data rates.
- An extension to the 802.11 architecture (802.11a) defines different multiplexing techniques that can achieve data rates up to 54 Mbps.

- Another extension to the standard (802.11b) defines 11 Mbps and 5.5 Mbps data rates (in addition to the 1 and 2Mbps rates) utilizing an extension to DSSS called High Rate DSSS (HR/DSSS).
- 802.11b also defines a rate shifting technique where 11 Mbps networks may fall back to 5.5 Mbps, 2 Mbps, or 1 Mps under noisy conditions or to inter-operate with legacy 802.11 PHY layers.

The three physical layers originally defined in 802.11 included two spread-spectrum radio techniques (Frequency Hopping Spread Spectrum and Direct Sequence Spread Spectrum) and a diffuse infrared specification.

Spread Spectrum

Spread spectrum is a technique trading bandwidth for reliability. The goal is to use more bandwidth that the system really needs for transmission to reduce the impact of localized interference on the media. Spread spectrum spreads the transmitted bandwidth of the resulting signal, reducing the peak power but keeping total power the same. Spread-spectrum techniques, in addition to satisfying regulatory requirements, increase reliability, boost throughput and allow many unrelated products to share the spectrum without explicit cooperation and with minimal interference.

- **Frequency Hopping Spread Spectrum (FHSS)**

Using the frequency hopping technique, the 2.4 GHz band is divided into 75 channels of 1 MHz subchannels each. The sender and receiver agree on a hopping pattern, and data is sent over a sequence of the subchannels. Each conversation within the 802.11 network occurs over a different hopping pattern, and the patterns are designed to minimize the chance of two senders using the same subchannel simultaneously.

FHSS techniques allow for a relatively simple radio design but are limited to speeds of not higher than 2 Mbps. This limitation is driven primarily by FCC (Federal Communications Commission USA) regulations that restrict subchannel bandwidth to 1 MHz. These regulations force FHSS systems to spread their usage across the entire 2.4 GHz band, meaning they must hop often, which leads to a high amount of hopping overhead.

- **Direct Sequence Spread Spectrum (DSSS)**

In contrast, the direct sequence signaling technique divides the 2.4 GHz band into 14 22-MHz channels. Adjacent channels overlap one another partially, with three of the 14 being completely non-overlapping. Data is sent across one of these 22 MHz channels without hopping to other channels.

To compensate for noise on a given channel, a technique called ―chipping‖ is used. Each bit of user data is converted into a series of redundant bit patterns called ―chips.‖ The inherent redundancy of each chip combined with spreading the signal across the 22 MHz channel provides for a form of error checking and correction; even if part of the signal is damaged, it can still be recovered in many cases, minimizing the need for retransmissions.

- **Infrared (IR)**

The Infrared PHY utilizes infrared light to transmit binary data either at 1 Mbps (basic access rate) or 2 Mbps (enhanced access rate) using a specific modulation technique for each. For 1 Mbps, the infrared PHY uses a 16-pulse position modulation (PPM). The concept of PPM is to

vary the position of a pulse to represent different binary symbols. Infrared transmission at 2 Mbps utilizes a 4 PPM modulation technique.

3.2 Frames

Current 802.11 standards define "frame" types for use in transmission of data as well as management and control of wireless links.

Frames are divided into very specific and standardized sections. Each frame has a MAC header, payload and frame check sequence (FCS). Some frames may not have the payload portion. First 2 bytes of MAC header is a frame control field that provides detailed information about the frame. The sub fields of the frame control field are presented in order.

- **Protocol Version:** It is two bits in size and represents the protocol version. Currently used protocol version is zero. Other values are reserved for future use.

- **Type:** It is two bits in size and helps to identify the type of WLAN frame. Control, Data and Management are various frame types defined in IEEE 802.11.

- **Sub Type:** It is four bits in size. Type and Sub type are combined together to identify the exact frame.

- **ToDS and FromDS:** Each is one bit in size. They indicate whether a data frame is headed for a distributed system. Control and management frames set these values to zero. All the data frames will have one of these bits set. However communication within an Independent Basic Service Set (IBSS) network always set these bits to zero.

- **More Fragment:** The More Fragmentation bit is set most notably when higher level packets have been partitioned and will be set for all non-final sections. Some management frames may require partitioning as well.

- **Retry:** Sometimes frames require retransmission and for this there is a Retry bit which is set to one when a frame is resent. This aids in the elimination of duplicate frames.

- **Power Management:** The Power Management bit indicates the power management state of the sender after the completion of a frame exchange. Access points are required to manage the connection and will never set the power saver bit.

- **More Data:** The More Data bit is used to buffer frames received in a distributed system. The access point uses this bit to facilitate stations in power saver mode. It indicates that at least one frame is available and addresses all stations connected.

- **WEP:** The WEP bit is modified after processing a frame. It is toggled to one after a frame has been decrypted or if no encryption is set it has already one.

- **Order:** This bit is only set when the "strict ordering" delivery method is employed. Frames and fragments are not always sent in order as it causes a transmission performance penalty.

The next two bytes are reserved for the Duration ID field. This field can take one of three forms: Duration, Contention-Free Period (CFP) and Association ID (AID).

An 802.11 frame can have up to four address fields. Each field can carry a MAC address. Address 1 is the receiver, Address 2 is the transmitter and Address 3 is used for filtering purposes by the receiver.

- The Sequence Control field is a two-byte section used for identifying message order as well as eliminating duplicate frames. The first 4 bits are used for the fragmentation number and the last 12 bits are the sequence number.
- An optional two-byte Quality of Service control field which was added with 802.11e.
- The Frame Body field is variable in size from 0 to 2304 bytes plus any overhead from security encapsulation and contains information from higher layers.
- The Frame Check Sequence (FCS) is the last four bytes in the standard 802.11 frame. Often referred to as the Cyclic Redundancy Check (CRC), it allows for integrity check of retrieved frames. As frames are about to be sent the FCS is calculated and appended. When a station receives a frame it can calculate the FCS of the frame and compare it to the one received. If they match, it is assumed that the frame was not distorted during transmission.

Management Frames allow for the maintenance of communication. Some common 802.11 subtypes include:

- Authentication frame: 802.11 authentications begin with the wireless network interface control (WNIC) sending an authentication frame to the access point containing its identity. With an open system authentication the WNIC only sends a single authentication frame and the access point responds with an authentication frame of its own indicating acceptance or rejection. With shared key authentication, after the WNIC sends its initial authentication request it will receive an authentication frame from the access point containing challenge text. The WNIC sends an authentication frame containing the encrypted version of the challenge text to the access point. The access point ensures the text was encrypted with the correct key by decrypting it with its own key. The result of this process determines the WNIC's authentication status.
- Association request frame: sent from a station it enables the access point to allocate resources and synchronize. The frame carries information about the WNIC including supported data rates and the service set identifier (SSID) of the network the station wishes to associate with. If the request is accepted, the access point reserves memory and establishes an association ID for the WNIC.
- Association response frame: sent from an access point to a station containing the acceptance or rejection to an association request. If it is an acceptance, the frame will contain information such an association ID and supported data rates.
- Beacon frame: Sent periodically from an access point to announce its presence and provide the SSID and other parameters for WNICs within range.

- Deauthentication frame: Sent from a station wishing to terminate connection from another station.
- Disassociation frame: Sent from a station wishing to terminate connection. It's an elegant way to allow the access point to relinquish memory allocation and remove the WNIC from the association table.
- Probe request frame: Sent from a station when it requires information from another station.
- Probe response frame: Sent from an access point containing capability information, supported data rates, etc., after receiving a probe request frame.
- Reassociation request frame: A WNIC sends a reassociation request when it drops from range of the currently associated access point and finds another access point with a stronger signal. The new access point coordinates the forwarding of any information that may still be contained in the buffer of the previous access point.
- Reassociation response frame: Sent from an access point containing the acceptance or rejection to a WNIC reassociation request frame. The frame includes information required for association such as the association ID and supported data rates.

Control frames facilitate in the exchange of data frames between stations. Some common 802.11 control frames include:

- Acknowledgement (ACK) frame: After receiving a data frame, the receiving station will send an ACK frame to the sending station if no errors are found. If the sending station doesn't receive an ACK frame within a predetermined period of time, the sending station will resend the frame.
- Request to Send (RTS) frame: The RTS and CTS frames provide an optional collision reduction scheme for access point with hidden stations. A station sends a RTS frame to as the first step in a two-way handshake required before sending data frames.
- Clear to Send (CTS) frame: A station responds to an RTS frame with a CTS frame. It provides clearance for the requesting station to send a data frame. The CTS provides collision control management by including a time value for which all other stations are to hold off transmission while the requesting stations transmits.

Data frames carry packets from web pages, files, etc. within the body.

4.0 Conclusion

In this unit, you have learnt about IEEE 802.11 layer which is made up of the data link layer and the physical layer. The data link layer within 802.11 consists of two sublayers: Logical Link Control (LLC) and Media Access Control (MAC).

5.0 Summary

The CSMA/CA scheme implements a minimum time gap between frames from a given user. Once a frame has been sent from a given transmitting station, that station must wait until the time gap is up to try to transmit again. Once the time has passed, the station selects a random amount of time (the backoff interval) to wait before "listening" again to verify a clear channel on which to transmit. If the channel is still busy, another backoff interval is selected that is less than the first. This process is repeated until the waiting time approaches zero and the station is allowed to

transmit. This type of multiple access ensures judicious channel sharing while avoiding collisions.

6.0 Tutor-Marked Assignment

(i) describe the IEEE 802.11 layer
(ii) state the three function of PHY layer
(iii) List the types of management frame that allow for the maintenance of communication.
(iv) discuss on the three types of control frame used to facilitate the exchange of data frames between stations.

7.0 References/Further Readings

Brenner P. (1996). A Technical tutorial on the IEEE802.11protocol. Breezecom Wireless Communication

Cisco Systems, Inc. (2007) "Channel Deployment Issues for 2.4 GHz 802.11 WLANs". http://www.cisco.com/en/US/docs/wireless/technology/channel/deployment/guide/Channel.html.

Garcia Villegas (2007). "Effect of adjacent-channel interference in IEEE 802.11 WLANs". *CrownCom 2007*. ICST & IEEE. Available online at https://upcommons.upc.edu/eprints/bitstream/2117/1234/1/CrownCom07_CReady.pdf.

IEEE (2008) "802.11 Technical Section". Available online at http://wifi.cs.st-andrews.ac.uk/wififrame.html.

IEEE (2008). "Understanding 802.11 Frame Types". Available online at http://www.wi-fiplanet.com/tutorials/article.php/1447501.

IEEE (2009). "IEEE P802.11 - TASK GROUP AC". Available online at http://www.ieee802.org/11/Reports/tgac_update.htm.

Intelligraphic (2010). Introduction to IEEE 802.11. Available online at http\www.intelligraphics.com\introduction-ieee-80211

Tutorial_Reports.com (2007). IEEE802.11 Architecture. Available online at http\www.tutorial-reports.com\wireless\WlanWiFi\WiFi_architecure.php

Wikipedia (2011). IEEE 802.11. Available online at http://en.wikipedia.org/wiki/IEEE_802.11

MODULE 5: BLUETOOTH
UNIT 1: BASIC CONCEPT OF BLUETOOTH

1.0 Introduction

Bluetooth is a proprietary open wireless technology standard for exchanging data over short distances (using short wavelength radio transmissions) from fixed and mobile devices, creating personal area networks (PANs) with high levels of security.

Figure 1.1The Bluetooth logo

2.0 Objectives

At the end of this unit, you should be able to:

(i) define a Bluetooth profile
(ii) list the types of Bluetooth profile
(iii) state areas of application of Bluetooth
(v) differentiate between a Bluetooth and Wi-Fi

3.0 Main Content

3.1 Bluetooth Profile

A Bluetooth profile is a wireless interface specification for Bluetooth-based communication between devices. To use Bluetooth wireless technology, a device must be able to interpret certain Bluetooth profiles, which are definitions of possible applications and specify general behaviors that Bluetooth enabled devices use to communicate with other Bluetooth devices. A Bluetooth profile resides on top of the Bluetooth Core Specification and (optionally) additional protocols. While the profile may use certain features of the core specification, specific versions of profiles are rarely tied to specific versions of the core specification. For example, there are HFP 1.5 implementations using both Bluetooth 2.0 and Bluetooth 1.2 core specifications.
There are a wide range of Bluetooth profiles that describe many different types of applications or use cases for devices.

3.1.1 List of profiles

The following profiles are defined and adopted by the Bluetooth SIG:

(i) Advanced Audio Distribution Profile (A2DP): This profile defines how high quality audio (stereo or mono) can be streamed from one device to another over a Bluetooth connection. For example, music can be streamed from a mobile phone to a wireless headset or car audio or from a laptop/desktop to a wireless headset.
(ii) Audio/Video Remote Control Profile (AVRCP): This profile is designed to provide a standard interface to control TVs, Hi-fi equipment, etc. to allow a single remote control (or other device) to control all of the A/V equipment to which a user has access. It may be used in concert with A2DP or VDP.

(iii) Basic Imaging Profile (BIP): This profile is designed for sending images between devices and includes the ability to resize, and convert images to make them suitable for the receiving device.

(iv) **Basic Printing Profile (BPP):** This allows devices to send text, e-mails, vCards, or other items to printers based on print jobs. It differs from HCRP in that it needs no printer-specific drivers. This makes it more suitable for embedded devices such as mobile phones and digital cameras which cannot easily be updated with drivers dependent upon printer vendors.

(v) **Common ISDN Access Profile (CIP):** This provides unrestricted access to the services, data and signals that ISDN offers.

(vi) Cordless Telephony Profile (CTP): This is designed for cordless phones to work using Bluetooth. It is hoped that mobile phones could use a Bluetooth CTP gateway connected to a landline within the home and the mobile phone network when out of range. It is central to the Bluetooth SIG's '3-in-1 phone' use case.

(vii) Device ID Profile (DIP): This profile allows a device to be identified above and beyond the limitations of the Device Class already available in Bluetooth. It enables identification of the manufacturer, product id, product version and the version of the Device ID specification being met. It is useful in allowing a PC to identify a connecting device and download appropriate drivers. It enables similar applications to those the Plug-and-play specification allows.

(viii) **Dial-up Networking Profile (DUN):** This profile provides a standard to access the Internet and other dial-up services over Bluetooth. The most common scenario is accessing the Internet from a laptop by dialing up on a mobile phone, wirelessly. It is based on Serial Port Profile (SPP) and provides for relatively easy conversion of existing products through many features that it has in common with the existing wired serial protocols for the same task.

(ix) Fax Profile (FAX): This profile is intended to provide a well defined interface between a mobile phone or fixed-line phone and a PC with Fax software installed. Data and voice calls are not covered by this profile.

(x) File Transfer Profile (FTP): Provides the capability to browse, manipulate and transfer objects (files and folders) in an object store (file system) of another system.

(xi) Generic Audio/Video Distribution Profile (GAVDP): provides the basis for A2DP, and VDP.

(xii) **Generic Access Profile (GAP):** provides the basis for all other profiles. GAP defines how two Bluetooth units discover and establish a connection with each other.

(xiii) Generic Object Exchange Profile (GOEP): provides a basis for other data profiles.

(xiv) Hard Copy Cable Replacement Profile (HCRP): This provides a simple wireless alternative to a cable connection between a device and a printer. Unfortunately it does not set a standard regarding the actual communications to the printer, so drivers are required specific to the printer model or range. This makes this profile less useful for embedded devices such as digital cameras and palmtops, as updating drivers can be problematic.

(xv) Health Device Profile (HDP): Profile designed to facilitate transmission and reception of Medical Device data. The API´s of this layer interact with the lower level Multi-Channel Adaptation Protocol (MCAP layer), but also perform SDP behavior to connect to remote HDP devices. Also makes use of the Device ID Profile (DIP).

(xvi) Hands-Free Profile (HFP): Currently in version 1.5, this is commonly used to allow car hands-free kits to communicate with mobile phones in the car. The Bluetooth car kits allow users with Bluetooth-equipped cell phones to make use of some of the phone's features, such as making calls, while the phone itself can be left in the user's pocket or hand bag.

(xvii) Human Interface Device Profile (HID): provides support for devices such as mice, joysticks, keyboards, as well as sometimes providing support for simple buttons and indicators on other types of devices. It is designed to provide a low latency link with low power requirements. PlayStation 3 controllers and Wii Remotes also use Bluetooth HID.

(xviii) Headset Profile (HSP): This is the most commonly used profile providing support for the popular Bluetooth Headsets to be used with mobile phones. It relies on SCO for audio encoded in 64 kbit/s CVSD or PCM and a subset of AT commands from GSM 07.07 for minimal controls including the ability to ring, answer a call, hang up and adjust the volume.

(xix) Intercom Profile (ICP): This is often referred to as the walkie-talkie profile. It is another TCS (Telephone Control protocol Specification) based profile, relying on SCO to carry the audio. It is proposed to allow voice calls between two Bluetooth capable handsets, over Bluetooth.

(xx) LAN Access Profile (LAP): LAN Access profile makes it possible for a Bluetooth device to access LAN, WAN or Internet via another device that has a physical connection to the network. It uses PPP over RFCOMM to establish connections. LAP also allows the device to join an ad-hoc Bluetooth network. The LAN Access Profile has been replaced by the PAN profile in the Bluetooth specification.

(xxi) Object Push Profile (OPP): A basic profile for sending "objects" such as pictures, virtual business cards, or appointment details. It is called push because the transfers are always instigated by the sender (client), not the receiver (server).

(xxii) Personal Area Networking Profile (PAN): This profile is intended to allow the use of Bluetooth Network Encapsulation Protocol on Layer 3 protocols for transport over a Bluetooth link.

(xxii) Phone Book Access Profile (PBAP, PBA): Phone Book Access (PBA) or Phone Book Access Profile (PBAP) is a profile that allows exchange of Phone Book Objects between devices. It is likely to be used between a car kit and a mobile phone to: allow the car kit to display the name of the incoming caller; and also allow the car kit to download the phone book so the user can initiate a call from the car display.

(xxiii) Serial Port Profile (SPP): This profile is based on ETSI 07,10 and the RFCOMM protocol. It emulates a serial cable to provide a simple substitute for existing RS-232, including the familiar control signals. It is the basis for DUN, FAX, HSP and AVRCP.

(xxiv) Service Discovery Application Profile (SDAP): describes how an application should use SDP to discover services on a remote device. SDAP requires that any application should be able to find out which services are available on any Bluetooth enabled device it connects to.

(xxv) SIM Access Profile (SAP, SIM, rSAP): This allows devices such as car phones with built in GSM transceivers to connect to a SIM card in a phone with Bluetooth, thus the car phone itself doesn't require a separate SIM card. This profile is also known as rSAP (remote-SIM-Access-Profile).

(xxvi) Synchronisation Profile (SYNCH): This profile allows synchronization of Personal Information Manager (PIM) items. As this profile originated as part of the infrared specifications but has been adopted by the Bluetooth SIG to form part of the main Bluetooth specification, it is also commonly referred to as IrMC Synchronization.

(xxvii) Video Distribution Profile (VDP): This profile allows the transport of a video stream. It could be used for streaming a recorded video from a PC media center to a portable player, or a live video from a digital video camera to a TV. Support for the H.263 baseline is mandatory. The

MPEG-4 Visual Simple Profile and H.263 profiles 3 and 8 are optionally supported and covered in the specification.1

(xxviii)Wireless Application Protocol Bearer (WAPB): This is a profile for carrying Wireless Application Protocol (WAP) over Point-to-Point Protocol over Bluetooth.

3.2 List of Bluetooth Application

- Wireless control of and communication between a mobile phone and a hands free headset. This was one of the earliest applications to become popular.
- Wireless networking between PCs in a confined space and where little bandwidth is required.
- Wireless communication with PC input and output devices, the most common being the mouse, keyboard and printer.
- Transfer of files, contact details, calendar appointments, and reminders between devices with OBEX.
- Replacement of traditional wired serial communications in test equipment, GPS receivers, medical equipment, bar code scanners, and traffic control devices.
- For controls where infrared was traditionally used.
- For low bandwidth applications where higher USB bandwidth is not required and cable-free connection desired.
- Sending small advertisements from Bluetooth-enabled advertising hoardings to other, discoverable, Bluetooth devices.
- Wireless bridge between two Industrial Ethernet (e.g., PROFINET) networks.
- Dial-up internet access on personal computers or PDAs using a data-capable mobile phone as a wireless modem like Novatel mifi.
- Short range transmission of health sensor data from medical devices to mobile phone, set-top box or dedicated tele-health devices.
- Allowing a DECT phone to ring and answer calls on behalf of a nearby cell phone
- Real-time location systems (RTLS), are used to track and identify the location of objects in real-time using ―Nodes‖ or ―tags‖ attached to, or embedded in the objects tracked and ―Readers‖ that receive and process the wireless signals from these tags to determine their locations
- Tracking livestock and detainees. According to a leaked diplomatic cable, King Abdullah of Saudi Arabia suggested "implanting detainees with an electronic chip containing information about them and allowing their movements to be tracked with Bluetooth. This was done with horses and falcons, the King said."
- Personal security application on mobile phones for theft prevention. The protected item has a Bluetooth marker (e.g. a headset) that is monitored continuously by the security application. If connection is lost (the marker is out of range) then an alarm is raised. The first known implementation of this security application of Bluetooth is BluCop, which is published in December 2010.

Figure 1.2: A typical Bluetooth mobile phone headset.

3.3 Bluetooth vs. Wi-Fi IEEE 802.11 in networking

- Bluetooth and Wi-Fi have many applications: setting up networks, printing, or transferring files.
- Wi-Fi is intended for resident equipment and its applications. The category of applications is outlined as WLAN, the wireless local area networks. Wi-Fi is intended as a replacement for cabling for general local area network access in work areas.
- Bluetooth is intended for non-resident equipment and its applications. The category of applications is outlined as the wireless personal area network (WPAN). Bluetooth is a replacement for cabling in a variety of personally carried applications in any ambiance and can also support fixed location applications such as smart energy functionality in the home (thermostats, etc.).
- Wi-Fi is a wireless version of a traditional Ethernet network, and requires configuration to set up shared resources, transmit files, and to set up audio links (for example, headsets and hands-free devices). Wi-Fi uses the same radio frequencies as Bluetooth but with higher power, resulting in a faster connection and better range from the base station. The nearest equivalents in Bluetooth are the DUN profile, which allows devices to act as modem interfaces, and the PAN profile, which allows for ad-hoc networking.

3.4 Bluetooth devices

Figure 1.3: A Bluetooth USB dongle with a 100 m range. The MacBook Pro shown also has a built in Bluetooth adaptor.

Bluetooth exists in many products such as the iPod Touch, Lego Mindstorms NXT, PlayStation 3, PSP Go, telephones, the Nintendo Wii, some high definition headsets, modems and watches. The technology is useful when transferring information between two or more devices that are near each other in low-bandwidth situations. Bluetooth is commonly used to transfer sound data with telephones (i.e., with a Bluetooth headset) or byte data with hand-held computers (transferring files).

Bluetooth protocols simplify the discovery and setup of services between devices. Bluetooth devices can advertise all of the services they provide. This make the services easier because more of the security, network address and permission configuration can be automated than with many other network types.

4.0 Conclusion

This unit has introduced you to the basic concepts of Bluetooth; profile, applications and device. You have also learnt about the differences between Bluetooth and Wi-Fi.

5.0 Summary

The main points in this unit include the following:

- a Bluetooth profile is a wireless interface specification for Bluetooth-based communication between devices.
- Bluetooth exists in many products such as the iPod Touch, Lego Mindstorms NXT, PlayStation 3, PSP Go, telephones, the Nintendo Wii, some high definition headsets, modems and watches.
- Bluetooth is intended for non-resident equipment and its applications while Wi-Fi is intended for resident equipment and its applications.

6.0 Tutor-Marked Assignment

(i) What are Bluetooth and Bluetooth Profile?
(ii) List out the Bluetooth profile
(iii) State the areas of application of Bluetooth
(vi) What are the differences between a Bluetooth and Wi-Fi

7.0 References/Further Readings

Bluetooth.com. (2010). "Profiles Overview". Available Online at http://www.bluetooth.com/English/Technology/Works/Pages/Profiles_Overview.aspx

Chomienne D. and Eftimakis M. (2010). "Bluetooth Tutorial" (PDF). Available Online at: http://www.newlogic.com/products/Bluetooth-Tutorial-2001.pdf.

Jim K. (2008). "How Bluetooth got its name". Available online at http://www.eetimes.eu/scandinavia/206902019?cid=RSSfeed_eetimesEU_scandinavia

Harold N. (2007). Newton‘s telecom dictionary. New York: Flatiron Publishing.

Wikipedia (2011). Bluetooth. Available online at: http://en.wikipedia.org/wiki/Bluetooth

UNIT 2: BLUETOOTH SPECIFICATION

1.0 Introduction

The Bluetooth specification was developed in 1994 by Jaap Haartsen and Sven Mattisson, who were working for Ericsson in Lund, Sweden. The specification is based on frequency-hopping spread spectrum technology.

The specifications were formalized by the Bluetooth Special Interest Group (SIG). The SIG was formally announced on May 20, 1998. Today it has a membership of over 13,000 companies worldwide. It was established by Ericsson, IBM, Intel, Toshiba and Nokia and later joined by many other companies.

2.0 Objectives

At the end of this unit, you should be able to:

(i) list the types of Bluetooth Specification

(ii) describe the Bluetooth specification

3.0 Main Content

3.1 Bluetooth Specification

(a) Bluetooth v1.0 and v1.0B

Versions 1.0 and 1.0B had many problems and manufacturers had difficulty making their products interoperable. Versions 1.0 and 1.0B also included mandatory Bluetooth hardware device address (BD_ADDR) transmission in the Connecting process (rendering anonymity impossible at the protocol level) which was a major setback for certain services planned for use in Bluetooth environments.

(b) Bluetooth v1.1

- Ratified as IEEE Standard 802.15.1-2002
- Many errors found in the 1.0B specifications were fixed.
- Added support for non-encrypted channels.
- Received Signal Strength Indicator (RSSI).

(c) Bluetooth v1.2

This version is backward compatible with 1.1 and the major enhancements include the following:

- Faster Connection and Discovery
- Adaptive frequency-hopping spread spectrum (AFH), which improves resistance to radio frequency interference by avoiding the use of crowded frequencies in the hopping sequence.
- Higher transmission speeds in practice up to 721 kbit/s than in v1.1.
- Extended Synchronous Connections (eSCO) which improve voice quality of audio links by allowing retransmissions of corrupted packets and may optionally increase audio latency to provide better support for concurrent data transfer.
- Host Controller Interface (HCI) support for three-wire UART.
- Ratified as IEEE Standard 802.15.1-2005
- Introduced Flow Control and Retransmission Modes for L2CAP.

(d) Bluetooth v2.0 + EDR

This version of the Bluetooth Core Specification was released in 2004 and is backward compatible with the previous version 1.2. The main difference is the introduction of an Enhanced Data Rate (EDR) for faster data transfer. The nominal rate of EDR is about 3 Mbit/s, although the practical data transfer rate is 2.1 Mbit/s. EDR uses a combination of GFSK and Phase Shift Keying modulation (PSK) with two variants, π/4-DQPSK and 8DPSK. EDR can provide a lower power consumption through a reduced duty cycle. The specification is published as "Bluetooth v2.0 + EDR" which implies that EDR is an optional feature. Aside from EDR, there are other minor improvements to the 2.0 specification and products may claim compliance to "Bluetooth v2.0" without supporting the higher data rate. At least one commercial device states "Bluetooth v2.0 without EDR" on its data sheet.

(e) Bluetooth v2.1 + EDR

Bluetooth Core Specification Version 2.1 + EDR is fully backward compatible with 1.2 and was adopted by the Bluetooth SIG on July 26, 2007.

The headline feature of 2.1 is secure simple pairing (SSP): this improves the pairing experience for Bluetooth devices while increasing the use and strength of security.

2.1 allows various other improvements including "Extended inquiry response" (EIR) which provides more information during the inquiry procedure to allow better filtering of devices before connection; sniff subtracting which reduces the power consumption in low-power mode

(f) Bluetooth v3.0 + HS

Version 3.0 + HS of the Bluetooth Core Specification was adopted by the Bluetooth SIG on April 21, 2009. Bluetooth 3.0+HS supports theoretical data transfer speeds of up to 24 Mbit/s though not over the Bluetooth link itself. Instead, the Bluetooth link is used for negotiation and establishment and the high data rate traffic is carried over a collocated 802.11 link. Its main new feature is AMP (Alternate MAC/PHY), the addition of 802.11 as a high speed transport. Two technologies had been anticipated for AMP: 802.11 and UWB, but UWB is missing from the specification.

The High-Speed part of the specification is not mandatory and hence only devices sporting the "+HS" will actually support the Bluetooth over Wifi high-speed data transfer. A Bluetooth 3.0 device without the HS suffix will not support High Speed and will only support Unicast Connectionless Data (UCD), as shown in the Bluetooth 3.0+HS specification.

- Alternate MAC/PHY

It enables the use of alternative MAC and PHYs for transporting Bluetooth profile data. The Bluetooth radio is still used for device discovery, initial connection and profile configuration, however when large quantities of data need to be sent, the high speed alternate MAC PHY 802.11 (typically associated with Wi-Fi) will be used to transport the data. This means that the proven low power connection models of Bluetooth are used when the system is idle and the faster radio is used when large quantities of data need to be sent.

- Unicast connectionless data

Permits service data to be sent without establishing an explicit L2CAP channel. It is intended for use by applications that require low latency between user action and reconnection/transmission of data. This is only appropriate for small amounts of data.

- Enhanced Power Control

Updates the power control feature to remove the open loop power control and also to clarify ambiguities in power control introduced by the new modulation schemes added for EDR. Enhanced power control removes the ambiguities by specifying the behaviour that is expected. The feature also adds closed loop power control, meaning RSSI filtering can start as the response is received. Additionally, a "go straight to maximum power" request has been introduced. This is expected to deal with the headset link loss issue typically observed when a user puts their phone into a pocket on the opposite side to the headset.

(g) Bluetooth v4.0

On June 12, 2007, Nokia and Bluetooth SIG announced that Wibree will be a part of the Bluetooth specification, as an ultra-low power Bluetooth technology.

On December 17, 2009, the Bluetooth SIG adopted Bluetooth low energy technology as the hallmark feature of the version 4.0. The provisional names Wibree and Bluetooth ULP (Ultra Low Power) are abandoned.

On April 21, 2010, the Bluetooth SIG completed the Bluetooth Core Specification version 4.0, which includes Classic Bluetooth, Bluetooth high speed and Bluetooth low energy protocols. Bluetooth high speed is based on Wi-Fi and Classic Bluetooth consists of legacy Bluetooth protocols.

- **Bluetooth low energy**

Bluetooth low energy is an alternative to the Bluetooth standard that was introduced in Bluetooth v4.0 and is aimed at very low power applications running off a coin cell. It allows two types of implementation, dual-mode and single-mode. In a dual-mode implementation, Bluetooth low energy functionality is integrated into an existing Classic Bluetooth controller. The resulting architecture shares much of Classic Bluetooth‘s existing radio and functionality resulting in a negligible cost increase compared to Classic Bluetooth. Additionally, manufacturers can use current Classic Bluetooth (Bluetooth v2.1 + EDR or Bluetooth v3.0 + HS) chips with the new low energy stack, enhancing the development of Classic Bluetooth enabled devices with new capabilities.

Cost-reduced single-mode chips, which will enable highly integrated and compact devices, will feature a lightweight Link Layer providing ultra-low power idle mode operation, simple device discovery, and reliable point-to-multipoint data transfer with advanced power-save and secure encrypted connections at the lowest possible cost. The Link Layer in these controllers will enable Internet connected sensors to schedule Bluetooth low energy traffic between Bluetooth transmissions.

4.0 Conclusion

In this unit, you learnt about the various Bluetooth specification.

5.0 Summary

The main points in this unit include the following:

- the Bluetooth specification is based on frequency-hopping spread spectrum technology.
- the various Bluetooth specification are:
  - Bluetooth v1.0 and v1.0B
  - Bluetooth v1.1
  - Bluetooth v1.2

- Bluetooth v2.0 + EDR
- Bluetooth v2.1 + EDR
- Bluetooth v3.0 + HS
- Bluetooth v4.0

6.0 Tutor-Marked Assignment

List and discuss on the Bluetooth Specification

7.0 References/Further Readings

Bluetooth SIG (2008). "How Bluetooth Technology Works". Available Online at http://web.archive.org/web/20080117000828/ http://bluetooth.com/Bluetooth/Technology/Works/

Bluetooth SIG. (2008). "About the Bluetooth SIG". Available online at http://www.bluetooth.com/Bluetooth/SIG/.

Bluetooth SIG (2008). "Specification Documents". Available Online at: http://www.bluetooth.com/Bluetooth/Technology/Building/Specifications/.

Bluetooth SIG (2008)."Wii Controller". Available Online at http://web.archive.org/web/20080220080315/ http://bluetooth.com/Bluetooth/Products/Products/Product_Details.htm?ProductID=2951.

Bluetooth.com. (2010). "Specification Documents". Available Online at http://www.bluetooth.com/Specification%20Documents/AssignedNumbersServiceDiscovery.pdf

Chomienne D. and Eftimakis M. (2010). "Bluetooth Tutorial" (PDF). Available Online at: http://www.newlogic.com/products/Bluetooth-Tutorial-2001.pdf.

Heidi M. (1999). "Bluetooth Technology and Implications". SysOpt.com. http://www.sysopt.com/features/network/article.php/3532506.

Wikipedia (2011). Bluetooth. Available online at: http://en.wikipedia.org/wiki/Bluetooth

UNIT 3: TECHNICAL INFORMATION ON BLUETOOTH

1.0 Introduction

In previous units, you've learnt on Bluetooth profile and specification but in this unit you will acquire knowledge on the technical information on Bluetooth such as the Bluetooth protocol stack, baseband error correction in Bluetooth system as well as Bluetooth setting up connection.

2.0 Objectives

At the end of this unit, you should be able to:

(i) discuss on the Bluetooth protocol stack
(ii) state the baseband error correction in Bluetooth
(iii) list the information transmitted by Bluetooth device

3.0 Main Content

3.1 Bluetooth Protocol Stack

Bluetooth is defined as a layer protocol architecture consisting of core protocols, cable replacement protocols, telephony control protocols, and adopted protocols." Mandatory protocols for all Bluetooth stacks are: Link Management Protocol (LMP), Logical Link Control and Adaptation Protocol (L2CAP) and Service Discovery Protocol (SDP). Additionally, these protocols are almost universally supported: Host/Controller Interface (HCI) and Radio frequency communications (RFCOMM).

(a) Link Management Protocol (LMP)

Link Management Protocol was used for control of the radio link between two devices. It was implemented on the controller.

(b) Logical Link Control & Adaptation Protocol (L2CAP)

It was used to multiplex multiple logical connections between two devices using different higher level protocols. It provides segmentation and reassembly of on-air packets.

In Basic mode, L2CAP provides packets with a payload configurable up to 64kB, with 672 bytes as the default MTU and 48 bytes as the minimum mandatory supported MTU.

In Retransmission and Flow Control modes, L2CAP can be configured for reliable or isochronous data per channel by performing retransmissions and CRC checks.

Bluetooth Core Specification Addendum 1 adds two additional L2CAP modes to the core specification. These modes effectively deprecate original Retransmission and Flow Control modes:

- **Enhanced Retransmission Mode** (ERTM): This mode is an improved version of the original retransmission mode. This mode provides a reliable L2CAP channel.
- **Streaming Mode** (SM): This is a very simple mode with no retransmission or flow control. This mode provides an unreliable L2CAP channel.

Reliability in any of these modes is optionally and/or additionally guaranteed by the lower layer Bluetooth BDR/EDR air interface by configuring the number of retransmissions and flush timeout (time after which the radio will flush packets). In-order sequencing is guaranteed by the lower layer.

Only L2CAP channels configured in ERTM or SM may be operated over AMP logical links.

(c) Service Discovery Protocol (SDP)

Service Discovery Protocol (SDP) allows a device to discover services supported by other devices and their associated parameters. For example, when connecting a mobile phone to a Bluetooth headset, SDP will be used for determining which Bluetooth profiles are supported by the headset (Headset Profile, Hands Free Profile, Advanced Audio Distribution Profile (A2DP) etc.) and the protocol multiplexer settings needed to connect to each of them. Each service is identified by a Universally Unique Identifier (UUID), with official services (Bluetooth profiles) assigned a short form UUID (16 bits rather than the full 128).

(d) Host/Controller Interface (HCI)

Standardised communication between the host stack (e.g., a PC or mobile phone OS) and the controller (the Bluetooth IC). This standard allows the host stack or controller IC to be swapped with minimal adaptation.

There are several HCI transport layer standards, each using a different hardware interface to transfer the same command, event and data packets. The most commonly used are USB (in PCs) and UART (in mobile phones and PDAs).

In Bluetooth devices with simple functionality (e.g., headsets) the host stack and controller can be implemented on the same microprocessor. In this case the HCI is optional, although often implemented as an internal software interface.

(e) Radio Frequency Communications (RFCOMM) (Serial Port Emulation)

Radio frequency communications (RFCOMM) is a cable replacement protocol used to create a virtual serial data stream. RFCOMM provides for binary data transport and emulates EIA-232 (formerly RS-232) control signals over the Bluetooth baseband layer.

RFCOMM provides a simple reliable data stream to the user, similar to TCP. It is used directly by many telephony related profiles as a carrier for AT commands, as well as being a transport layer for OBEX over Bluetooth.

Many Bluetooth applications use RFCOMM because of its widespread support and publicly available API on most operating systems. Additionally, applications that used a serial port to communicate can be quickly ported to use RFCOMM.

(f) Bluetooth Network Encapsulation Protocol (BNEP)

BNEP is used for transferring another protocol stack's data via an L2CAP channel. Its main purpose is the transmission of IP packets in the Personal Area Networking Profile. BNEP performs a similar function to SNAP in Wireless LAN.

(g) Audio/Video Control Transport Protocol (AVCTP)

It is used by the remote control profile to transfer AV/C commands over an L2CAP channel. The music control buttons on a stereo headset use this protocol to control the music player.

(h) Audio/Video Distribution Transport Protocol (AVDTP)

It is used by the advanced audio distribution profile to stream music to stereo headsets over an L2CAP channel. It is intended to be used by video distribution profile.

(i) Telephony control protocol

Telephony control protocol-binary (TCS BIN) is the bit-oriented protocol that defines the call control signaling for the establishment of voice and data calls between Bluetooth devices. Additionally, "TCS BIN defines mobility management procedures for handling groups of Bluetooth TCS devices." TCS-BIN is only used by the cordless telephony profile which failed to attract implementers. As such it is only of historical interest.

(j) Adopted protocols

Adopted protocols are defined by other standards-making organizations and incorporated into Bluetooth's protocol stack, allowing Bluetooth to create protocols only when necessary. The adopted protocols include:

- Point-to-Point Protocol (PPP): Internet standard protocol for transporting IP datagrams over a point-to-point link.
- TCP/IP/UDP: Foundation Protocols for TCP/IP protocol suite
- Object Exchange Protocol (OBEX): Session-layer protocol for the exchange of objects, providing a model for object and operation representation
- Wireless Application Environment/Wireless Application Protocol (WAE/WAP):WAE specifies an application framework for wireless devices and WAP is an open standard to provide mobile users access to telephony and information services.

3.2 Baseband Error Correction

Three types of error correction are implemented in Bluetooth systems:

- 1/3 rate forward error correction (FEC)
- 2/3 rate FEC
- Automatic repeat-request (ARQ): also known as Automatic Repeat Query, is an error-control method for data transmission that uses acknowledgements (messages sent by the receiver indicating that it has correctly received a data frame or packet) and timeouts (specified periods of time allowed to elapse before an acknowledgment is to be received) to achieve reliable data transmission over an unreliable service. If the sender does not receive an acknowledgment before the timeout, it usually re-transmits the frame/packet until the sender receives an acknowledgment or exceeds a predefined number of re-transmissions.

3.3 Setting up connections

Any Bluetooth device in discoverable mode will transmit the following information on demand:

- Device name
- Device class
- List of services
- Technical information (for example: device features, manufacturer, Bluetooth specification used, clock offset)

Any device may perform an inquiry to find other devices to connect to and any device can be configured to respond to such inquiries. However, if the device trying to connect knows the address of the device, it always responds to direct connection requests and transmits the

information shown in the list above if requested. Use of a device's services may require pairing or acceptance by its owner but the connection itself can be initiated by any device and held until it goes out of range. Some devices can be connected to only one device at a time and connecting to them prevents them from connecting to other devices and appearing in inquiries until they disconnect from the other device.
Every device has a unique 48-bit address. However, these addresses are generally not shown in inquiries. Instead, friendly Bluetooth names are used, which can be set by the user. This name appears when another user scans for devices and in lists of paired devices.
Most phones have the Bluetooth name set to the manufacturer and model of the phone by default. Most phones and laptops show only the Bluetooth names and special programs are required to get additional information about remote devices. This can be confusing as, for example, there could be several phones in range named T610.

4.0 Conclusion

In this unit, you learnt about Bluetooth protocol stack, the baseband error correction in Bluetooth as well as Bluetooth setting up connection.

5.0 Summary

The main points in this unit include the following:

- the protocols for all Bluetooth stacks are: Link Management Protocol (LMP), Logical Link Control and Adaptation Protocol (L2CAP) and Service Discovery Protocol (SDP).
- the three types of error correction implemented in Bluetooth systems are:
  - 1/3 rate forward error correction (FEC)
  - 2/3 rate FEC
  - automatic repeat-request (ARQ)
- any Bluetooth device in discoverable mode will transmit the following information on demand: device name, device class, list of services and technical information

6.0 Tutor-Marked Assignment

(i) discuss on the Bluetooth protocol stack
(ii) outline the baseband error correction in Bluetooth
(iii) mention the information transmitted by Bluetooth device when in discoverable mode

7.0 References/Further Readings

BlueZ (2010). "Official Linux Bluetooth protocol stack". Available Online at: http://www.bluez.org/.

Chomienne D. and Eftimakis M. (2010). "Bluetooth Tutorial" (PDF). Available Online at: http://www.newlogic.com/products/Bluetooth-Tutorial-2001.pdf.

Information Age (2007). "The Bluetooth Blues". Available Online at: http://web.archive.org/web/20071222231740/http://www.information-age.com/article/2001/may/the_bluetooth_blues.

Kewney G. (2004). "High speed Bluetooth comes a step closer: enhanced data rate approved". Newswireless.net. http://www.newswireless.net/index.cfm/article/629.

Meyer D. (2009). "Bluetooth 3.0 released without ultrawideband". zdnet.co.uk. Available Online at: http://news.zdnet.co.uk/communications/0,1000000085,39643174,00.htm.

Nokia (2007). "Wibree forum merges with Bluetooth SIG" (PDF). Press release. Available Online at: http://www.wibree.com/press/Wibree_pressrelease_final_1206.pdf.
Peterson and Davie (2003). Computer Networks: A Systems Approach.
William S. (2005). Wireless communications and networks. Upper Saddle River, NJ: Pearson Prentice Hall.
Wikipedia (2011). Bluetooth. Available online at: http://en.wikipedia.org/wiki/Bluetooth

MODULE 6: RADIO FREQUENCY IDENTIFICATION

UNIT 1: RADIO FREQUENCY IDENTIFICATION: OPERATING PRINCIPLE

1.0 Introduction

Radio frequency identification (RFID) is a rapidly growing technology that has the potential to make great economic impacts on many industries. While RFID is a relatively old technology, more recent advancements in chip manufacturing technology are making RFID practical for new applications and settings, particularly consumer item level tagging. These advancements have the potential to revolutionize supply-chain management, inventory control and logistics.

Radio Frequency Identification (RFID) is an automatic identification technology where information is carried by radio waves. It is a technology that uses communication via radio waves to exchange data between a reader and an electronic tag attached to an object, for the purpose of identification and tracking.

RFID makes it possible to give each product in a grocery store its own unique identifying number, to provide assets, people, work in process, medical devices etc. all with individual unique identifiers - like the license plate on a car and for every item in the world. This is a vast improvement over paper and pencil tracking or bar code tracking that has been used since the 1970s. With bar codes, it is only possible to identify the brand and type of package in a grocery store, for instance. Furthermore, passive RFID tags (those without a battery) can be read if passed within close enough proximity to an RFID reader. It is not necessary to "show" them to it, as with a bar code. In other words it does not require line of sight to "see" an RFID tag, the tag can be read inside a case, carton, box or other container and unlike barcodes, RFID tags can be read hundreds at a time. Bar codes can only read one at a time.

Some RFID tags can be read from several meters away and beyond the line of sight of the reader. The application of bulk reading enables an almost-parallel reading of tags.

3.0 Objectives

At the end of this unit, you should be able to:

(i) define a radio frequency identification
(ii) state two types of radio frequency identification reader
(iii) outline two types of radio frequency identification tags
(iv) discuss on the two major operating principles of radio frequency identification

4.0 Main Content

3.1 Overview of RFID

Radio-frequency identification involves the hardware known as interrogators (readers) and tags (labels) as well as RFID software or RFID middleware.

Most RFID tags contain at least two parts: one is an integrated circuit for storing and processing information, modulating and demodulating a radio-frequency (RF) signal and other specialized functions; the other is an antenna for receiving and transmitting the signal.

RFID can be either passive (using no battery), active (with an on-board battery that always broadcasts or beacons its signal) or battery assisted passive "BAP" which has a small battery on board that is activated when in the presence of an RFID reader. Passive tags meant to be mounted on metal or withstand gamma sterilization. Active tags for tracking containers, medical assets or monitoring environmental conditions in data centers.BAP tags have sensor capability like temperature and humidity.

The term RFID refers to the technology while the tags is called "RFID tags" not "RFIDs".

3.1.1 Types of RFID Reader

Fixed RFID and Mobile RFID: Depending on mobility, RFID readers are classified into two different types:

(i) Fixed RFID: If the reader reads tags in a stationary position, it is called fixed RFID. These fixed readers set up specific interrogation zones and create a "bubble" of radio frequency (RF) energy that can be tightly controlled if the physics is well engineered. This allows a very definitive reading area when tags go in and out of the interrogation zone.

(ii) Mobile RFID: if reader reads tags in a mobile position, it is called mobile RFID. Mobile readers include hand held‘s, carts and vehicle mounted RFID readers from manufacturers such as Motorola, Intermec, etc.

3.1.2 Types of RFID Tags

There are three types of RFID tags:

(i) Passive RFID tags: which have no power source and require an external electromagnetic field to initiate a signal transmission

(ii) Active RFID tags: which contain a battery and can transmit signals once an external source ('Interrogator') has been successfully identified

(iii) Battery Assisted Passive (BAP) RFID tags: which require an external source to wake up but have significant higher forward link capability providing greater range.

3.1.3 Elements of an RFID system

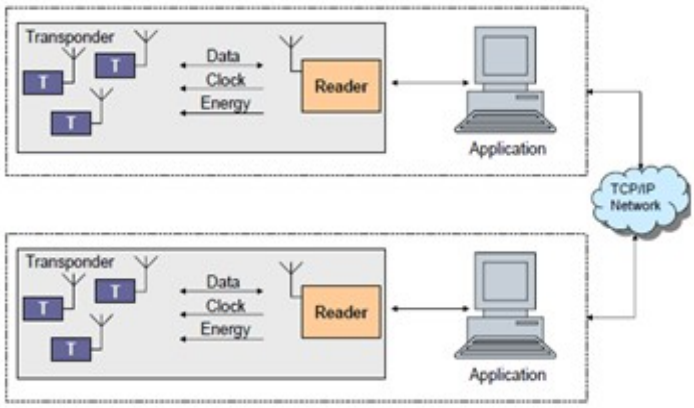


Figure 1.1: Elements of an RFID System
Source: Handy M.(2004)

Transponder Power Supply
Active (Battery-Assisted) Transponders

- Own Energy Source (E.G. Battery)
- Transponder Transmits Radio Signal
- Higher Read Range
- Finite Lifetime

Passive Transponders

- No Power Supply "On Board"
- Transponder Reflects/Modulates Radio Signal From Reader
- Shorter Read Range
- Lifetime Not Limited By Energy Source

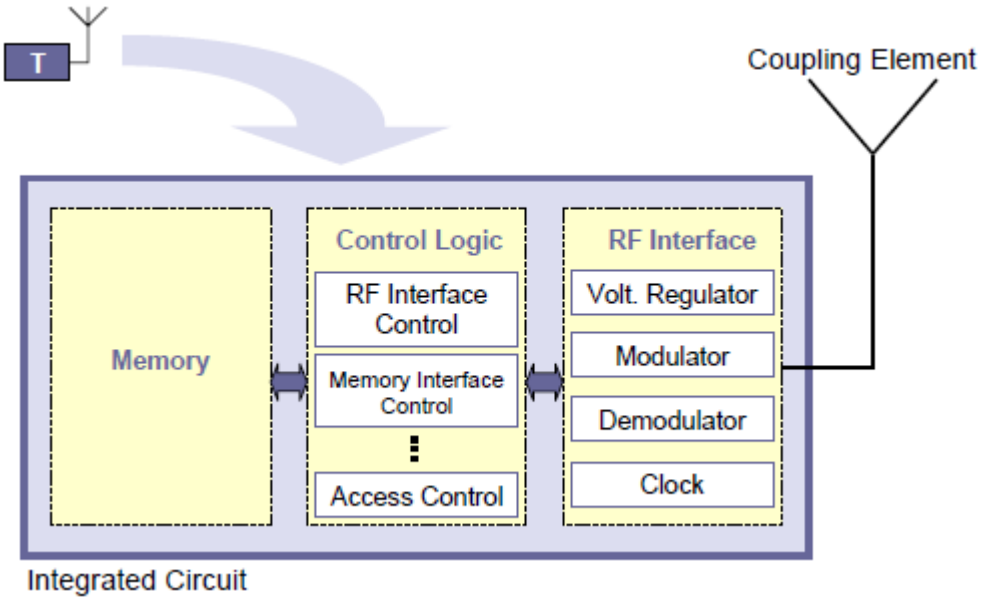


Figure 1.2: A Passive Transponder
Source: Handy M.(2004)

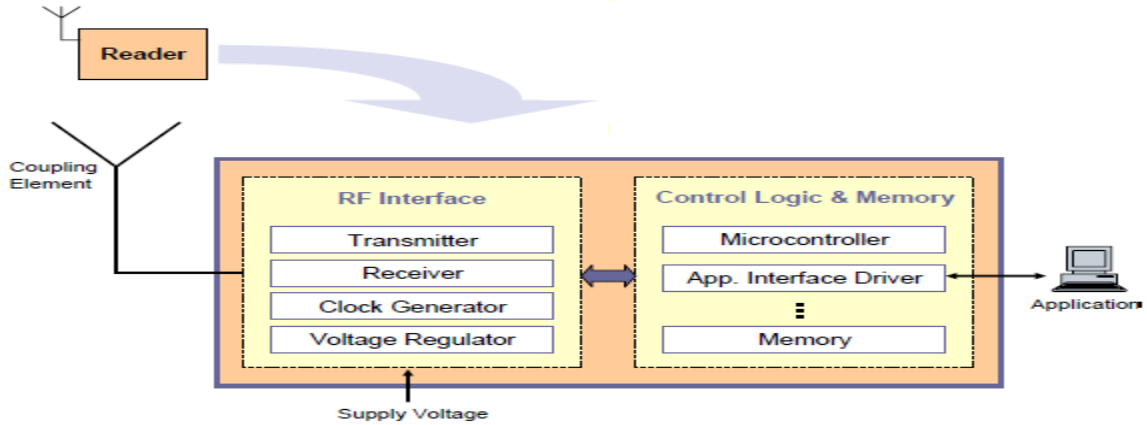


Figure 1.3 : Readers – What‘s inside?
Source: Handy M.(2004)

Special reader design depends on

- Type of coupling
- Communication sequence
- Type of data transmission
  Transponder → reader
- Operating frequency

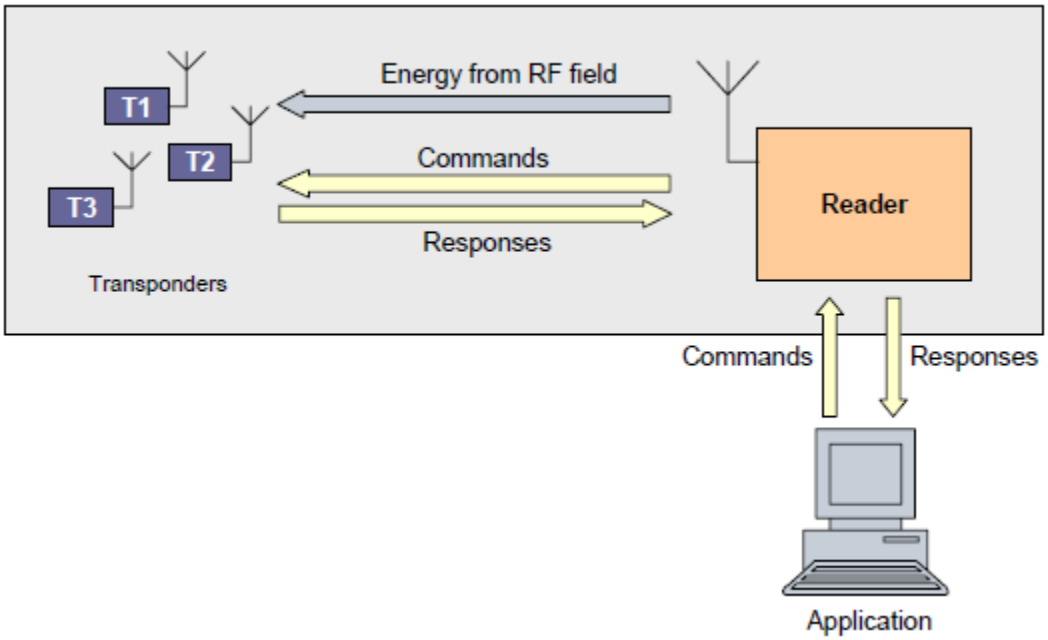


Figure 1.4: Communication Model (Passive)
Source: Handy M..(2004)

Communication Example:

- Application → Reader:
  - ‒Show me IDs of all tags in range!‖
- Reader → Tags:
  - ‒Deliver your ID!‖
- Tags → Reader:
  - ‒T1‖, ‒T2‖, ‒T3‖
- Reader → Application:
  - ‒IDs of tags in range: T1, T2, T3‖

3.2 Operating principles of RFID systems

There is a huge variety of different operating principles for RFID systems. Figure 3.2(a) below provides a short survey of known operation principles. The most important principles - ‗inductive coupling' and ‗backscatter coupling' are described more detailed below.

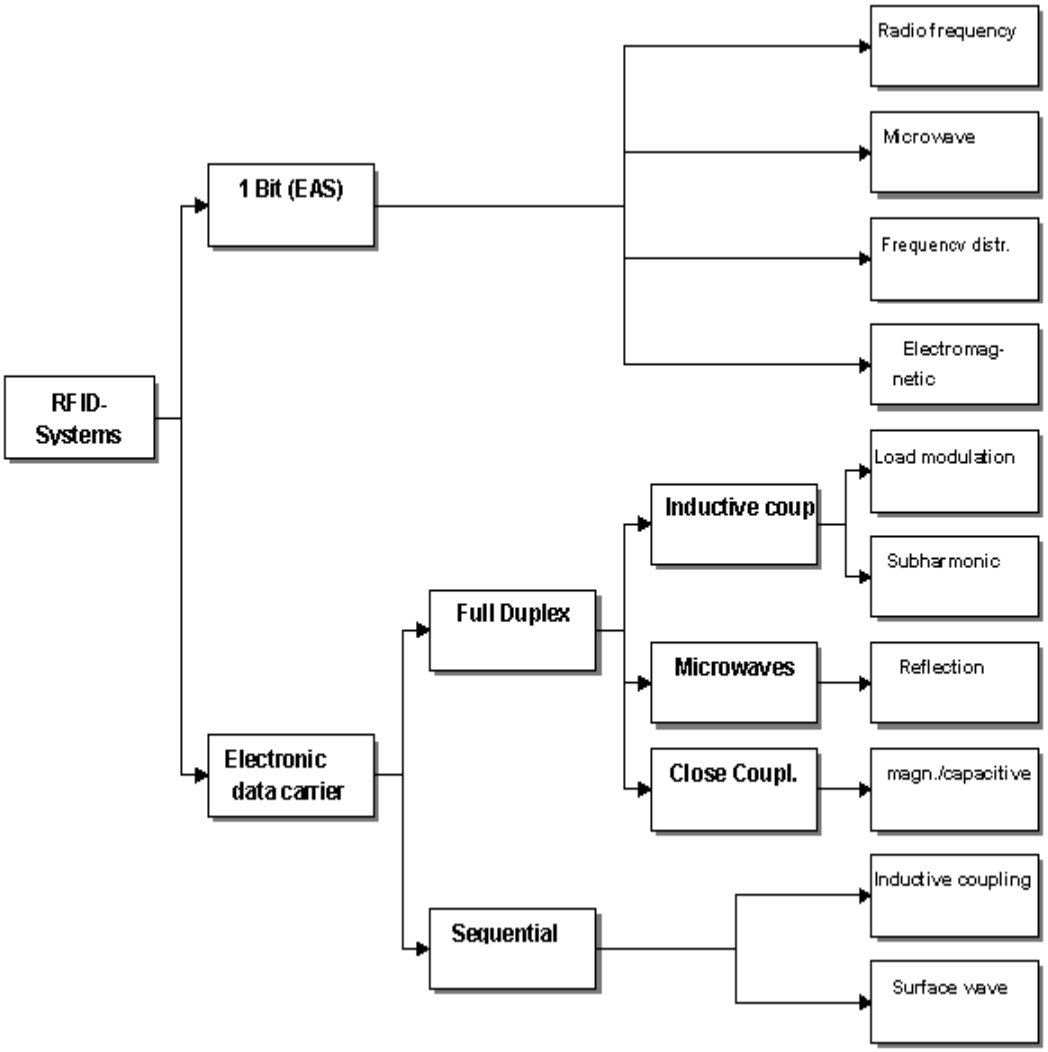


Figure 1.5: The allocation of the different operating principles of RFID systems
Source: http//RFID-handbook.com

Inductive Coupling

An inductively coupled transponder comprises of an electronic data carrying device, usually a single microchip and a large area coil that functions as an antenna.

Inductively coupled transponders are always operated passively. This means that all the energy needed for the operation of the microchip has to be provided by the reader. For this purpose, the reader's antenna coil generates a strong, high frequency electro-magnetic field, which penetrates the cross-section of the coil area and the area around the coil. The electro-magnetic field may be treated as a simple magnetic alternating field with regard to the distance between transponder and antenna because the wavelength of the frequency range used (< 135 kHz: 2400 m, 13.56 MHz: 22.1 m) is several times greater than the distance between the reader's antenna and the transponder.

A small part of the emitted field penetrates the antenna coil of the transponder, which is some distance away from the coil of the reader. By induction, a voltage Ui is generated in the transponder's antenna coil. This voltage is rectified and serves as the power supply for the data

carrying device (microchip). A capacitor C1 is connected in parallel with the reader's antenna coil, the capacitance of which is selected such that it combines with the coil inductance of the antenna coil to form a parallel resonant circuit, with a resonant frequency that corresponds with the transmission frequency of the reader. Very high currents are generated in the antenna coil of the reader by resonance step-up in the parallel resonant circuit, which can be used to generate the required field strengths for the operation of the remote transponder.
The antenna coil of the transponder and the capacitor C1 to form a resonant circuit tuned to the transmission frequency of the reader. The voltage U at the transponder coil reaches a maximum due to resonance step-up in the parallel resonant circuit.

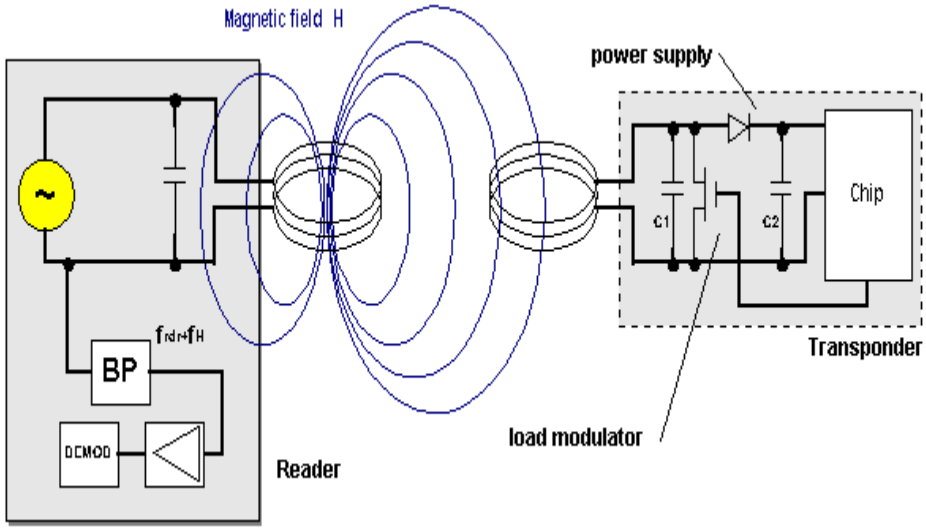


Figure 1.6: Operation principle of Inductive Coupling
Source: http//RFID-handbook.com

As described above, inductively coupled systems are based upon a transformer-type coupling between the primary coil in the reader and the secondary coil in the transponder. This is true when the distance between the coils does not exceed 0.16 l, so that the transponder is located in the near field of the transmitter antenna.
If a resonant transponder (i.e. the self-resonant frequency of the transponder corresponds with the transmission frequency of the reader) is placed within the magnetic alternating field of the reader's antenna, then this draws energy from the magnetic field. This additional power consumption can be measured as voltage drop at the internal resistance in the reader antennae through the supply current to the reader's antenna. The switching on and off of a load resistance at the transponder's antenna therefore effects voltage changes at the reader's antenna and thus has the effect of an amplitude modulation of the antenna voltage by the remote transponder. If the switching on and off of the load resistor is controlled by data, then this data can be transferred from the transponder to the reader. This type of data transfer is called load modulation.
To reclaim the data in the reader, the voltage measured at the reader's antenna is rectified. This represents the demodulation of an amplitude modulated signal. An example of the circuit is shown in figure 1.7.

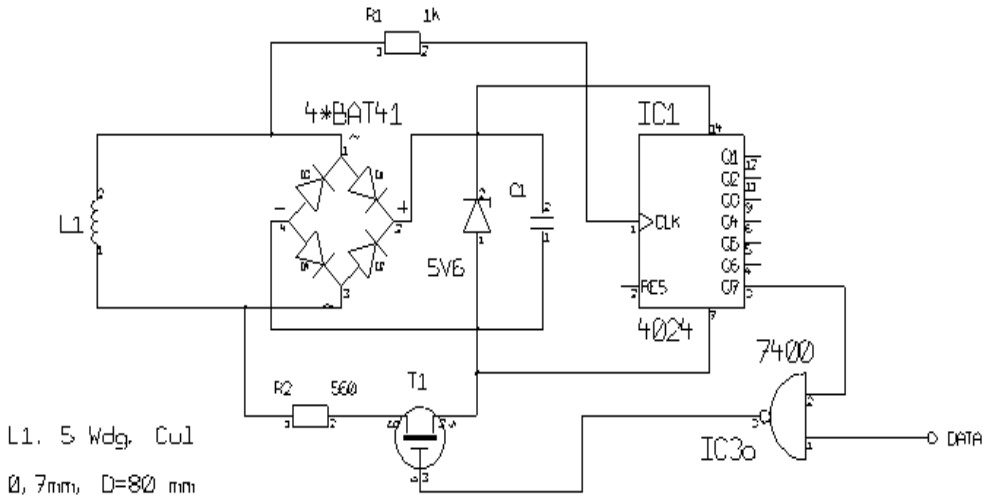


Figure 1.7: Sample circuit of the power supply and load modulator in a transponder
Source: http//RFID-handbook.com

In figure 3.2(c) above, If the additional load resistor in the transponder is switched on and off at a very high elementary frequency f_H, then two spectral lines are created at a distance of $\pm f_H$ around the transmission frequency of the reader and these can be easily detected (however f_H must be less than f_{READER}). In the terminology of radio technology the new elementary frequency is called a subcarrier. Data transfer is by the ASK, FSK or PSK modulation of the subcarrier in time with the data flow. This represents an amplitude modulation of the subcarrier.

Backscatter Coupling (3.2.2)
It is known from the field of RADAR technology that electromagnetic waves are reflected by objects with dimensions greater than around half the wavelength of the wave. The efficiency with which an object reflects electromagnetic waves is described by its reflection cross-section. Objects that are in resonance with the wave front that hits them, as is the case for antenna at the appropriate frequency for example, have a particularly large reflection cross-section.

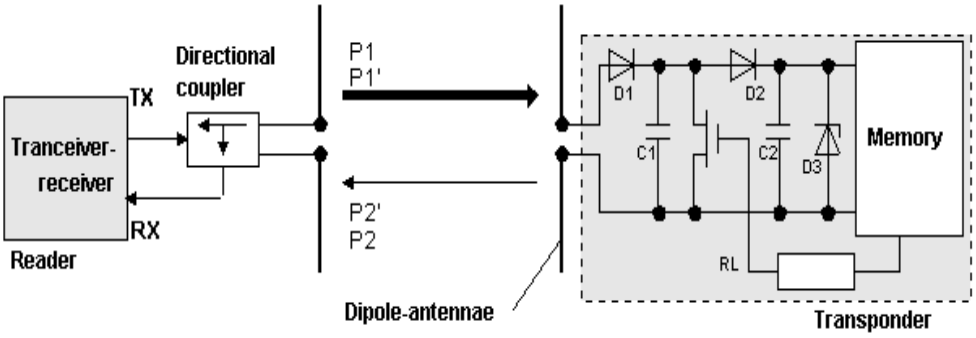


Figure 1.8: Operation principle of a backscatter transponder
Source: http//RFID-handbook.com

Power P1 is emitted from the reader's antenna, a small proportion of which (free space attenuation) reaches the transponder's antenna. The power P1' is supplied to the antenna

connections as HF voltage and after rectification by the diodes D1 and D2 this can be used as turn on voltage for the deactivation or activation of the power saving "power-down" mode. The diodes used here are low barrier Schottky diodes, which have a particularly low threshold voltage. The voltage obtained may also be sufficient to serve as a power supply for short ranges.
A proportion of the incoming power P1' is reflected by the antenna and returned as power P2. The reflection characteristics (= reflection cross-section) of the antenna can be influenced by altering the load connected to the antenna. In order to transmit data from the transponder to the reader, a load resistor RL connected in parallel with the antenna is switched on and off in time with the data stream to be transmitted. The amplitude of the power P2 reflected from the transponder can thus be modulated (à modulated backscatter).
The power P2 reflected from the transponder is radiated into free space. A small proportion of this (free space attenuation) is picked up by the reader's antenna. The reflected signal therefore travels into the antenna connection of the reader in the "backwards direction" and can be decoupled using a directional coupler and transferred to the receiver input of a reader. The "forward" signal of the transmitter, which is stronger by powers of ten, is to a large degree suppressed by the directional coupler.
The ratio of power transmitted by the reader and power returning from the transponder (P1 / P2) can be estimated using the radar equation.

5.0 Conclusion

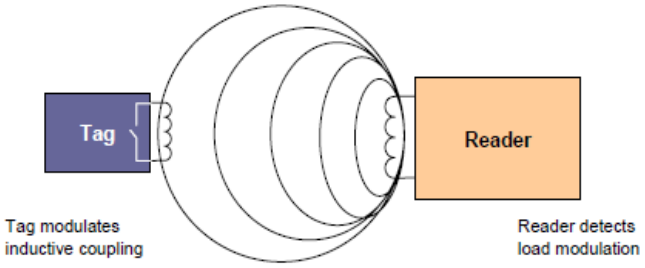
This unit has introduced you to the basic concepts of Radio Frequency Identification. You have also learnt about the operating principles of radio frequency identification.

6.0 Summary

The main points in this unit are:

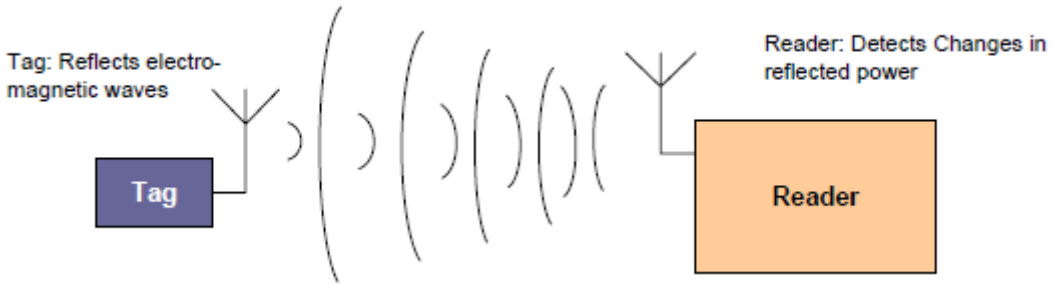
- Radio Frequency Identification (RFID) is an automatic identification technology where information is carried by radio waves.
- Other Auto-ID-Technologies are Bar Code, Smart Cards and Biometrics (e.g. fingerprint)
- Special Characteristics of Radio Communication are: No physical contact, No line-of-sight, Imperceptible
- Inductive Coupling
  Transponder's power supply:
  - Reader generates magnetic alternating field.
  - This field induces a voltage in tag's antenna coil → power supply

  Data transfer from transponder → reader:
  - Resonant transponder draws energy from magnetic alternating field.
  - This can be detected in reader's antenna (voltage drop)
  - Switching a load resistor on and off at the *transponder* → amplitude modulation of voltage at the *reader* antenna (load modulation)



- Electromagnetic Backscatter Coupling
  Transponder's power supply:
  - Tag draws energy from electromagnetic field of the reader.
  - Tag may be battery-assisted.

  Data transfer from transponder → reader:
  - Similar to radar technology
  - Incoming power at the tag antenna is reflected partially.
  - Reflection characteristics depend on load connected to the antenna.
  - Load resistor in parallel with the antenna can modulate the amplitude of the reflected power (modulated backscatter).



6.0 Tutor Marked Assignment

(i) What is radio frequency identification?
(ii) state two types of radio frequency identification reader
(iii) mention two types of radio frequency identification tags
(iv) discuss on the inductive coupling operating principle of radio frequency identification
(v) explain on the backscatter coupling operating principle of radio frequency identification

7.0 References/Further Reading

Albrecht K. and McIntyre L. (2005). Spychips : How Major Corporations and Government Plan to Track Your Every Move with RFID. Nelson Current Publishing.

Avoine G. (2006). Security and Privacy in RFID Systems Bibliography. Available at: http://lasecwww.epfl.ch/~gavoine/rfid/.

Auto-ID Labs. (2006). Webpage. Available at: http://www.autoidlabs.org.

Baird, J.L. (1928). ‒Improvements in or relating to apparatus for transmitting views or images to a distance‖. Patent #GB292,185.

Bridgelall R. (2004). RADAR Technology for Commodity Goods.

Bono S., Green M., Stubblefield A., Rubin A., Juels A. and Szydlo M. (2005). Analysis of the Texas Instruments DST RFID. Available at: http://rfidanalysis.org.

EPCglobal. (2006). Webpage. Available at: http://www.epcglobalinc.org.

ExxonMobile Speedpass. (2006). Webpage. Available at: http://www.speedpass.com.

Finkenzeller K (2003). RFID-Handbook; John Wiley & Sons.

Food and Drug Administration. (2004). Combating Counterfeit Drugs. Technical Report. United States Department of Health and Human Services. Available at: http:// www.fda.gov/oc/initiatives/counterfeit/report02_04.html.
Handy M.. RFID Technology. Institute of Applied Microelectronics & CS University of Rostock, RFID-Workshop, 30.9./1.10.04, Berlin
Harmon C.K. (2003). Basics of RFID Technology, MIT RFID Privacy Workshop, Cambridge, MA..
Heute Technologie Von Morgen Beherrschen (2011). Physical Principles of the RFID-Handbook. Available online at http://RFID-handbook.com
Ibid. (2003). Identification cards -- Contactless integrated circuit(s) cards – Vicinity cards. ISO/IEC 15693.
Ibid. (2003). Wal-Mart Expands RFID Mandate. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/539/1/1/.
Ibid. (2004). RFID for Item Management. ISO/IEC 18000.
Ibid. (2005). AmEx Adds RFID to Blue Credit Cards. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/1646/1/1/.
International Organization for Standardization (ISO). (2003). Identification cards -Contactless integrated circuit(s) cards -- Vicinity cards. ISO/IEC 14443.
Juels A. (2003). RFID Tags: Privacy and Security without Cryptography; MIT RFID Privacy Workshop, Cambridge, MA.
Juels, A. (2006). RFID Security and Privacy: A Research Survey. Selected Areas of Cryptography.
Juels A. and Pappu R. (2003). Squealing Euros: Privacy-Protection in RFID-Enabled Banknotes. Financial Cryptography. Lecture Notes in Computer Science. Volume 2742,103-121.
Juels A., Rivest R. L. and Szydlo M. (2003). The Blocker Tag: Selective Blocking of RFID Tags for Consumer Privacy, ACM Press.
Juels, A. and Weis S.A. (2005). Authenticating Pervasive Devices with Human Protocols. Advances in Cryptology – Crypto '05. Lecture Notes in Computer Science. Volume 3621. 293-308.
Krane J. (2003). Benetton clothing to carry tiny tracking transmitters. Associated Press.
Nokia. (2004). Nokia Mobile RFID Kit. Available at: http://www.nokia.com/nokia/0,,55738,00.html.
Pister K. (2004). Smart Dust: Autonomous Sensing and Communications in a Cubic Millimeter. Available at: http://robotics.eecs.berkeley.edu/~pister/SmartDust/.
Reynolds M.(2003). Physics of RFID, MIT RFID Privacy Workshop, Cambridge, MA
RFID Journal. (2003). Gillette Confirms RFID Purchase. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/258/1/1/.
Rieback, M.R., Crispo, B., Tanenbaum, A.S. (2006). Is your cat infected with a computer virus? Pervasive Computing and Communications. IEEE Press. Pages 169-179.
Royal Air Force. (2006). Royal Air Force History. Available at: http://www.raf.mod.uk/ history/line1940.html.
Stockman, H. (1948). Communication by Means of Reflected Power. Proceedings of the Institute of Radio Engineers. October. Pages 1196-1204.
Subramanian, V., Chang, P., Huang, D., Lee, J., Molesa, S., Redinger, D., and Volkman, S. (2006). Conference on VLSI Design. Pages 709-714. IEEE Press.
Uniform Code Council. (2006). Webpage. Available at: http://www.uc-council.org.

United States Department of Defense. (2006). Radio Frequency Identification. Available at: http://www.acq.osd.mil/log/rfid/index.htm.
University of California, Berkeley Organic Electronics Group. Website. Available at: http://organics.eecs.berkeley.edu/. (Last Accessed: March 11, 2006.)
Wikipedia (2011). Radio Frequency Identification. Available online at http://en.wikipedia.org/wiki/Radio-frequency_identification

UNIT 2: RADIO FREQUENCY IDENTIFICATION: SENSING APPLICATIONS

1.0 Introduction

This unit provides an insight into the nature of the applications that have attracted RFID deployment to date and also provide a window into future uses. Direct sensing of product identity is important in environments in which it is too complex, uncertain or expensive to extract information about product movement via indirect methods – generally these involve computer tracking models and simple proximity sensing devices of some form. Following the contrast between bar coding and RFID systems in previous unit, it is clear that an easily automated, wireless, non-line of sight system is required. These characteristics are reflected in applications of RFID till date such as supply chain management; electronic tolling, facilities management (e.g. libraries); airline baggage handling, asset tracking, inventory system, asset management and retail sales.

2.0 Objectives

At the end of this unit, you should be able to

(i) List the application areas of RFID in transportation
(ii) explain the various application areas of RFID
(iii) outline the RFID technology standards

3.0 Main Contents

Current uses

In 2010 three key factors drove a significant increase in RFID usage: decreased cost of equipment and tags, increased performance to a reliability of 99.9% and a stable international standard around UHF passive. The two areas of significant use are financial services for IT asset tracking and healthcare

(i) Payment by mobile phones

MicroSD cards was developed since 2009. When inserted into a mobile phone, the microSD card can both be a passive tag and an RFID reader. After inserting the microSD, a user's phone can be linked to bank accounts and used in mobile payment. Nokia's 2008 device, the 6212, has RFID capabilities. Credit card information can be stored and bank accounts can be directly accessed using the enabled handset. The phone, if used as a vector for mobile payment, the users would be required to enter a passcode or PIN before payment is authorized.

(ii) Transportation payments

Governments use RFID applications for traffic management, while automotive companies use various RFID tracking solutions for product management.

Figure 2.1: Automobile Security

- **Car-sharing:** The Zipcar car-sharing service uses RFID cards for locking and unlocking cars and for member identification.
- **Season parking tickets:** RFID is used to replace the paper Season Parking Ticket (SPT).
- **Toll roads:** RFID cards have been used as e-toll in the motorways and bridges as a payment system; it is also used in public transportation systems such as buses and trains.
- **Public transit (bus, rail, subway)** The public transport payment is RFID based. The card allows the user to credit money in advance and to be debited according to the distance travelled, as determined by the check-in and check-out stations. The card can also be used to pay taxi drivers and some shops offer card readers as well.

(iii) Asset management and retail sales

RFID combined with mobile computing and Web technologies provide a way for organizations to identify and manage their assets. Mobile computers, with integrated RFID readers, can now deliver a complete set of tools that eliminate paperwork, give proof of identification and attendance. This approach eliminates manual data entry.

Web based management tools allow organizations to monitor their assets and make management decisions from anywhere in the world. Web based applications now mean that third parties, such as manufacturers and contractors can be granted access to update asset data, including for example, inspection history and transfer documentation online ensuring that the end user always has accurate, real-time data. Organizations are already using RFID tags combined with a mobile asset management solution to record and monitor the location of their assets, their current status, and whether they have been maintained.

(iv) Product tracking

RFID use in product tracking applications begins with plant-based production processes and then extends into post-sales configuration management policies for large buyers.

- **Casino chip tracking**

RFID tags were placed on high value chips. These tags gave the casinos the ability to detect counterfeit chips, track betting habits of individual players, speed up chip tallies and determine counting mistakes of dealers.

- **IT asset tracking**
  - High-frequency RFID or HFID/HighFID tags are used in library book or bookstore tracking, jewelry tracking, pallet tracking, building access control, airline baggage tracking, apparel and pharmaceutical items tracking. High-frequency tags are widely used in identification badges, replacing earlier magnetic stripe cards. These badges need only be held within a certain distance of the reader to authenticate the holder.
  - UHF, Ultra-HighFID or UHFID tags are commonly used commercially in case, pallet, shipping container tracking, truck and trailer tracking in shipping yards.
  - Ultrahigh-frequency identification (UHFID) tags is used to help monitor agricultural equipment.

(v) Transportation and logistics

- Logistics and transportation are major areas of implementation for RFID technology. For example, yard management, shipping and freight and distribution centers are some areas where RFID tracking technology is used. Transportation companies around the world value RFID technology due to its impact on the business value and efficiency.
- The railroad industry operates an automatic equipment identification system based on RFID. Locomotives and rolling stock are equipped with two passive RFID tags (one mounted on each side of the equipment); the data encoded on each tag identifies the equipment owner, car number, type of equipment, number of axles, etc. The equipment owner and car number can be used to derive further data about the physical characteristics of the equipment from the Association of American Railroads' car inventory database and the railroad's own database indicating the lading, origin, destination, etc. of the commodities being carried.[32]

(vi) Animal identification

Figure 2.2: Animal management using RFID technology. Santa Gertrudis cattle: The calf has an electronic ear tag and herd management tag (yellow).

RFID tags for animals represent one of the oldest uses of RFID technology. Originally meant for large ranches and rough terrain, since the outbreak of mad-cow disease, RFID has become crucial in animal identification management. An implantable variety of RFID tags or transponders can also be used for animal identification. The transponders are more well-known as passive RFID technology, or simply "chips" on animals.

- **RFID tracking and tracing for meatpackers**

The Canadian Cattle Identification Agency began using RFID tags as a replacement for barcode tags. The tags are required to identify a bovine's herd of origin and this is used for tracing when a packing plant condemns a carcass.

(vii) Inventory systems

An advanced automatic identification technology such as the Auto-ID Labs system based on the Radio Frequency Identification (RFID) technology has significant value for inventory systems. Notably, the technology provides an accurate knowledge of the current inventory. In an academic study performed at Wal-Mart, RFID reduced Out-of-Stocks by 30 percent for products

selling between 0.1 and 15 units a day. Other benefits of using RFID include the reduction of labor costs, the simplification of business processes and the reduction of inventory inaccuracies.

(viii) Hospital operating rooms

The SmartSponge System was the first RFID-based system approved for use in the operating room. The system, consisting of an electronic reader and high frequency RFID-tagged disposable gauze, sponges and towels, is designed to improve patient safety and O.R. efficiency. The system aims to reduce or eliminate the most common and costly surgical "never event", unintentionally retained foreign objects in surgery. The system automatically provides a device-reconciled count by directly matching the unique identifier on each tagged item both entering into and then out of the surgical case. The system also provides a reusable wand which may be used to scan the patient as an additional safety measure or to assist in locating misplaced sponges.

(ix) Promotion tracking

Manufacturers of products sold through retailers promote their products by offering discounts for a limited period on products sold to retailers with the expectation that the retailers will pass on the savings to their customers. However, retailers typically engage in *forward buying*, purchasing more product during the discount period than they intend to sell during the promotion period. Some retailers engage in a form of arbitrage, reselling discounted product to other retailers, a practice known as *diverting*. To combat this practice, manufacturers are exploring the use of RFID tags on promoted merchandise so that they can track exactly which product has being sold through the supply chain at fully discounted prices.

(x) Libraries

Figure 2.3: RFID tags used in libraries: square book tag, round CD/DVD tag and rectangular VHS tag.

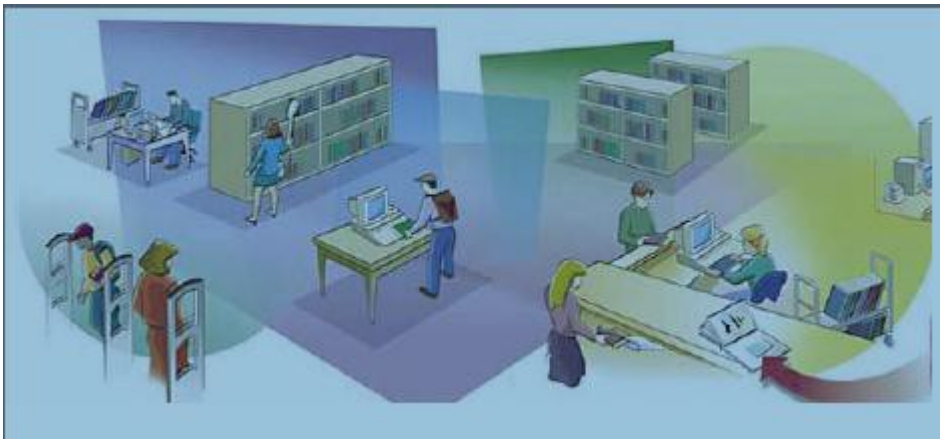


Figure 2.4: RFID tags incorporated into books and other library media
Source: Hodges S and McFarlane D. (2005)

Among the many uses of RFID technology is its deployment in libraries. This technology has slowly begun to replace the traditional barcodes on library items (books, CDs, DVDs, etc.). The RFID tag can contain identifying information, such as a book's title or material type, without having to be pointed to a separate database . The information is read by an RFID reader, which replaces the standard barcode reader commonly found at a library's circulation desk. It may replace or be added to the barcode, offering a different means of inventory management by the staff and self service by the borrowers. It can also act as a security device, taking the place of the more traditional electromagnetic security strip.

(xi) Passports

The RFID passport (**biometric passport or e-passport** or **ePassport)** is a combined paper and electronic passport that contains biometric information that can be used to authenticate the identity of travelers. It uses contactless smart card technology, including a microprocessor chip (computer chip) and antenna (for both power to the chip and communication) embedded in the front or back cover or center page of the passport. The passport's critical information is both printed on the data page of the passport and stored in the chip. The e-passports record the travel history (time, date and place) of entries and exits from the country. Public Key Infrastructure (PKI) is used to authenticate the data stored electronically in the passport chip making it expensive and difficult to forge when all security mechanisms are fully and correctly implemented. The currently standardized biometrics used for this type of identification systems are facial recognition, fingerprint recognition, and iris recognition. These were adopted after assessment of several different kinds of biometrics including retinal scan.

(xii) Schools and universities

RFID card system is used to check or track pupils and staff in and out of the school/university main gate or building via specially designed cards and to both track attendance and prevent unauthorized entrance. Some schools uses RFID in IDs for borrowing books and also gates in those particular schools have RFID ID scanners for buying items at a school shop and canteen, library and also to sign in and sign out for student and teacher's attendance.

(xiii) Museums

RFID technologies are now also implemented in end-user applications in museums. An example was the custom-designed temporary research application, "eXspot. A visitor entering the museum received an RF Tag that could be carried as a card. The eXspot system enabled the visitor to receive information about specific exhibits. Aside from the exhibit information, the visitor could take photographs of themselves at the exhibit. It was also intended to also allow the visitor to take data for later analysis. The collected information could be retrieved at home from a "personalized" website keyed to the RFID tag.

(xiv) Social retailing

When customers enter a dressing room, the mirror reflects their image and also images of the apparel item being worn by celebrities on an interactive display. A webcam also projects an image of the consumer wearing the item on the website for everyone to see. This creates an interaction between the consumers inside the store and their social network outside the store. The technology in this system is an RFID interrogator antenna in the dressing room and Electronic Product Code RFID tags on the apparel item.

(xv) Race timing

Figure 2.5: J-Chip 8-channel receiver next to timing mat.
The athlete wears a chip on a strap around his ankle.

Many forms of RFID race timing have been in use for timing races of different types. It is used for registering race start and end timings for animals or individuals in large running races or multi-sport races where it is impossible to get accurate stopwatch readings for every entrant. In the race, the racers wear passive or active tags that are read by antennae placed alongside the track or on mats across the track. UHF based tags instead of low or high frequency last-generation tags provide accurate readings with specially designed antennas. Rush error, lap count errors and accidents at start time are avoided since anyone can start and finish any time without being in a batch mode.

(xvi) Ski resorts

A number of ski resorts have adopted RFID tags to provide skiers hands-free access to ski lifts. Skiers do not have to take their passes out of their pockets.

(xvii) Human implants

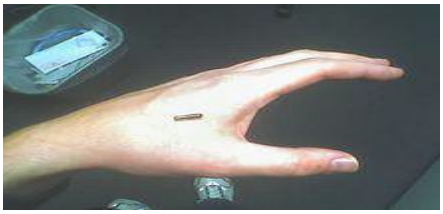


Figure 2.6: Hand with the planned location of the RFID chip.

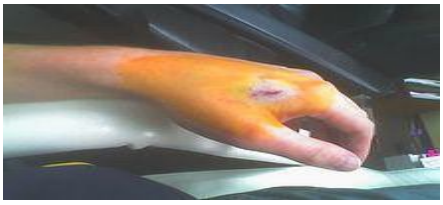


Figure 2.7: Just after the operation to insert the RFID tag was completed.
The yellow is from the iodine disinfection before inserting the chip.

A human microchip implant is an integrated circuit device or RFID transponder encased in silicate glass and implanted in the body of a human being. A subdermal implant typically contains a unique ID number that can be linked to information contained in an external database, such as personal identification, medical history, medications, allergies, and contact information.

Potential uses

RFID can be used in a variety of applications such as:

- Access management
  - Building access with proximity cards
  - Ski-lift passes
  - Concert tickets
  - Automobile ignition systems

- Tracking of goods and RFID in retail

- Tracking of persons and animals
- Toll collection and contactless payment

- Machine readable travel documents
- Smartdust (for massively distributed sensor networks)

- Tracking sports memorabilia to verify authenticity
- Airport baggage tracking logistics
- Anti-Counterfeiting:
  - Casino tokens, e.g. Wynn Casino Las Vegas
  - High-denomination currency notes,
  - Luxury goods, e.g. Prada
  - Prescription drugs

Complement to barcode

RFID tags are often a complement, but not a substitute barcodes. They may not ever completely replace barcodes, due to their higher cost and the advantage of multiple data sources on the same object. Also, unlike RFID labels, barcodes can be generated and distributed electronically, *e.g.* via e-mail or mobile phone, for printing and/or display by the recipient. An example is airline boarding passes.

The unique identity is a mandatory requirement for RFID tags, despite special choice of the numbering scheme. RFID tag data capacity is large enough that each individual tag will have a unique code, while current bar codes are limited to a single type code for a particular product. The uniqueness of RFID tags means that a product may be tracked as it moves from location to location, finally ending up in the consumer's hands. This may help to combat theft and other forms of product loss. The tracing of products is an important feature that gets well supported with RFID tags containing a unique identity of the tag and also the serial number of the object.

This may help companies to cope with quality deficiencies and resulting recall campaigns, but also contributes to concern about tracking and profiling of consumers after the sale.

Telemetry

Active RFID tags also have the potential to function as low-cost remote sensors that broadcast telemetry back to a base station. Applications of tagometry data could include sensing of road conditions by implanted beacons, weather reports, and noise level monitoring. Passive RFID tags can also report sensor data. For example, the Wireless Identification and Sensing Platform is a passive tag that reports temperature, acceleration and capacitance to commercial Gen2 RFID readers. It is possible that active or semi-passive RFID tags used with or in place of barcodes could broadcast a signal to an in-store receiver to determine whether the RFID tag (product) is in the store.

Identification of patients and hospital staff

Hospital has deployed an RFID and barcode based bedside medication verification system that improves patient safety by reducing medication errors. Nurses use a PDA equipped with a portable RFID reader and barcode scanner to check patient ID and medications before administering any drugs, including drugs delivered through IV pumps. Using patient ID cards with magnetic stripes enables a healthcare professional to call up valuable information in seconds by using a card reader. Instead of time-consuming photocopies of each individual's insurance card, this technology cuts the time a patient must take to identify herself or himself to a provider. There are also claim adjudication applications, as well as the ability to generate information on benefits and co-pays with a swipe of a card.

Standards that have been made regarding RFID technology include:

- ISO 14223 – Radiofrequency [*sic*] identification of animals – Advanced transponders
- ISO/IEC 14443: This standard is a popular HF (13.56 MHz) standard for HighFIDs which is being used as the basis of RFID-enabled passports under ICAO 9303. The Near Field Communication standard that lets mobile devices act as RFID readers/transponders is also based on ISO/IEC 14443.
- ISO/IEC 15693: This is also a popular HF (13.56 MHz) standard for HighFIDs widely used for non-contact smart payment and credit cards.
- ISO/IEC 18000: Information technology — Radio frequency identification for item management
- ISO/IEC 18092 Information technology — Telecommunications and information exchange between systems — Near Field Communication — Interface and Protocol (NFCIP-1)
- ISO 18185: This is the industry standard for electronic seals or "e-seals" for tracking cargo containers using the 433 MHz and 2.4 GHz frequencies.
- ISO/IEC 21481 Information technology — Telecommunications and information exchange between systems — Near Field Communication Interface and Protocol -2 (NFCIP-2)
- ASTM D7434, Standard Test Method for Determining the Performance of Passive Radio Frequency Identification (RFID) Transponders on Palletized or Unitized Loads
- ASTM D7435, Standard Test Method for Determining the Performance of Passive Radio Frequency Identification (RFID) Transponders on Loaded Containers

- ASTM D7580 Standard Test Method for Rotary Stretch Wrapper Method for Determining the Readability of Passive RFID Transponders on Homogenous Palletized or Unitized Loads

4.0 Conclusion

In this unit, you learnt about the various application areas and potential uses of radio frequency identification. You have also learnt about standards that have been made regarding RFID technology.

5.0 Summary

RFID has many applications; for example, it is used in enterprise supply chain management to improve the efficiency of inventory tracking and management. The Healthcare industry has used RFID to create tremendous productivity increases by eliminating "parasitic" roles that don't add value to an organization such as counting, looking for things or auditing items. Many financial institutions use RFID to track key assets and automate Sarbanes Oxley SOX compliance. Also with recent advances in social media RFID is being used to tie the physical world with the virtual world. RFID in Social Media first came to light in 2010 with facebook's annual conference (f8).

7.0 Tutor-Marked Assignment

(i) Mention the application areas of RFID in transportation
(ii) Discuss on the various application areas of RFID
(iii) outline the standards that have been made regarding RFID technology.

7.0 References/Further Reading

Albrecht K. and McIntyre L. (2005). Spychips : How Major Corporations and Government Plan to Track Your Every Move with RFID. Nelson Current Publishing.

Avoine G. (2006). Security and Privacy in RFID Systems Bibliography. Available at: http://lasecwww.epfl.ch/~gavoine/rfid/.

Auto-ID Labs. (2006). Webpage. Available at: http://www.autoidlabs.org.

Baird, J.L. (1928). ―Improvements in or relating to apparatus for transmitting views or images to a distance‖. Patent #GB292,185.

Bridgelall R. (2004). RADAR Technology for Commodity Goods.

Bono S., Green M., Stubblefield A., Rubin A., Juels A. and Szydlo M. (2005). Analysis of the Texas Instruments DST RFID. Available at: http://rfidanalysis.org.

EPCglobal. (2006). Webpage. Available at: http://www.epcglobalinc.org.

ExxonMobile Speedpass. (2006). Webpage. Available at: http://www.speedpass.com.

Finkenzeller K (2003). RFID-Handbook; John Wiley & Sons.

Food and Drug Administration. (2004). Combating Counterfeit Drugs. Technical Report. United States Department of Health and Human Services. Available at: http:// www.fda.gov/oc/initiatives/counterfeit/report02_04.html.

Handy M.. RFID Technology. Institute of Applied Microelectronics & CS University of Rostock, RFID-Workshop, 30.9./1.10.04, Berlin

Harmon C.K. (2003). Basics of RFID Technology, MIT RFID Privacy Workshop, Cambridge, MA..

Heute Technologie Von Morgen Beherrschen (2011). Physical Principles of the RFID-Handbook. Available online at http://RFID-handbook.com

Hodges S. and McFarlane D. (2005). Radio Frequency Identification: Technology, Applications and Impact. Uk: Auto-ID Lab, Cambridge University

Ibid. (2003). Identification cards -- Contactless integrated circuit(s) cards – Vicinity cards. ISO/IEC 15693.

Ibid. (2003). Wal-Mart Expands RFID Mandate. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/539/1/1/.

Ibid. (2004). RFID for Item Management. ISO/IEC 18000.

Ibid. (2005). AmEx Adds RFID to Blue Credit Cards. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/1646/1/1/.

International Organization for Standardization (ISO). (2003). Identification cards -Contactless integrated circuit(s) cards -- Vicinity cards. ISO/IEC 14443.

Juels A. (2003). RFID Tags: Privacy and Security without Cryptography; MIT RFID Privacy Workshop, Cambridge, MA.

Juels, A. (2006). RFID Security and Privacy: A Research Survey. Selected Areas of Cryptography.

Juels A. and Pappu R. (2003). Squealing Euros: Privacy-Protection in RFID-Enabled Banknotes. Financial Cryptography. Lecture Notes in Computer Science. Volume 2742,103-121.

Juels A., Rivest R. L. and Szydlo M. (2003). The Blocker Tag: Selective Blocking of RFID Tags for Consumer Privacy, ACM Press.

Juels, A. and Weis S.A. (2005). Authenticating Pervasive Devices with Human Protocols. Advances in Cryptology – Crypto '05. Lecture Notes in Computer Science. Volume 3621. 293-308.

Krane J. (2003). Benetton clothing to carry tiny tracking transmitters. Associated Press.

Nokia. (2004). Nokia Mobile RFID Kit. Available at: http://www.nokia.com/nokia/0,,55738,00.html.

Pister K. (2004). Smart Dust: Autonomous Sensing and Communications in a Cubic Millimeter. Available at: http://robotics.eecs.berkeley.edu/~pister/SmartDust/.

Reynolds M.(2003). Physics of RFID, MIT RFID Privacy Workshop, Cambridge, MA

RFID Journal. (2003). Gillette Confirms RFID Purchase. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/258/1/1/.

Rieback, M.R., Crispo, B., Tanenbaum, A.S. (2006). Is your cat infected with a computer virus? Pervasive Computing and Communications. IEEE Press. Pages 169-179.

Royal Air Force. (2006). Royal Air Force History. Available at: http://www.raf.mod.uk/ history/line1940.html.

Stockman, H. (1948). Communication by Means of Reflected Power. Proceedings of the Institute of Radio Engineers. October. Pages 1196-1204.

Subramanian, V., Chang, P., Huang, D., Lee, J., Molesa, S., Redinger, D., and Volkman, S. (2006). Conference on VLSI Design. Pages 709-714. IEEE Press.

Uniform Code Council. (2006). Webpage. Available at: http://www.uc-council.org.

United States Department of Defense. (2006). Radio Frequency Identification. Available at: http://www.acq.osd.mil/log/rfid/index.htm.

University of California, Berkeley Organic Electronics Group. Website. Available at: http://organics.eecs.berkeley.edu/. (Last Accessed: March 11, 2006.)

Wikipedia (2011). Radio Frequency Identification. Available online at http://en.wikipedia.org/wiki/Radio-frequency_identification

World Health Organization. (2006). Counterfeit Medicines. Available at: http:// www.who.int/mediacentre/factsheets/fs275/en/.

UNIT 3: RADIO FREQUENCY IDENTIFICATION: CHALLENGES

1.0 Introduction

While RFID adoption yields many efficiency benefits, it still faces several hurdles. Besides the typical implementation challenges faced in any information technology system and economic barriers, there are major concerns over security and privacy in RFID systems. Without proper protection, RFID systems could create new threats to both corporate security and personal privacy.

2.0 Objectives

At the end of this unit, you should be able to

(i) mention the challenges of radio frequency identification
(ii) discuss on the problems of radio frequency identification

3.0 Main Content

3.1 Challenges of RFID

(i) Data flooding

Each tag generating a message each time when passing a reader may be a desired outcome. However, event filtering is required to reduce this data inflow to a meaningful depiction of moving goods passing a threshold. Various concepts have been designed, mainly offered as middleware performing the filtering from noisy and redundant raw data to significant processed data.

(ii) Global standardization

The frequencies used for RFID in the USA are currently incompatible with those of Europe or Japan. Furthermore, no emerging standard has yet become as universal as the barcode. To address international trade concerns, it is necessary to utilize a tag that is operational within all of the international frequency domains. An example of such a tag is a Sentry-M WW from RCD Technology.

(iii) Exploitation

A RFID buffer overflow bug that could infect airport terminal RFID databases for baggage and also passport databases to obtain confidential information on the passport holder had been reported.

(iv) Passports

In an effort to make passports more secure, several countries have implemented RFID in passports. However, the encryption on UK chips was broken in less than 48 hours. Since that incident, further efforts have allowed researchers to clone passport data while the passport is being mailed to its owner. Where a criminal used to need to secretly open and then reseal the envelope, now it can be done without detection, adding some degree of insecurity to the passport system.

(v) Shielding

A number of products are available on the market that will allow a concerned carrier of RFID-enabled cards or passports to shield their data. Shielding is again a function of the frequency

being used. Low-frequency LowFID tags, like those used in implantable devices for humans and pets are relatively resistant to shielding, though thick metal foil will prevent most reads. High frequency HighFID tags (13.56 MHz — smart cards and access badges) are sensitive to shielding and are difficult to read when placed within a few centimetres of a metal surface. UHF Ultra-HighFID tags (pallets and cartons) are difficult to read when placed within a few millimetres of a metal surface, although their read range is actually increased when they are spaced 2–4 cm from a metal surface due to positive reinforcement of the reflected wave and the incident wave at the tag. UHFID tags can be successfully shielded from most reads by being placed within an anti-static plastic bag.

(vi) Temperature exposure

Currently, RFID tags are created by gluing an integrated circuit (IC) to an inlay. This poses a problem as vibration and high temperatures will loosen the connection. If the IC loses connection with the inlay, the RFID tag will no longer transmit. A new design was filed for patent (currently pending approval) where the IC is soldered to a circuit board and the circuit board is then soldered to the inlay. This process replaces the adhesive with solder which is much more durable and temperature resistant.

(vii) Security and Privacy

Many concerns have been expressed over the security and privacy of RFID systems. Traditional applications, like large-asset tracking, were typically closed systems where tags did not contain sensitive information. Tags on railway cars contained the same information painted on the side of the cars themselves. However, as more consumer applications are developed, security and especially privacy, will become important issues.

- **Eavesdropping**

Perhaps the biggest security concerns in RFID systems are espionage and privacy threats. As organizations adopt and integrate RFID into their supply chain and inventory control infrastructure, more and more sensitive data will be entrusted on RFID tags. As these tags inevitably end up in consumer hands, they could leak sensitive data or be used for tracking individuals.

An attacker able to eavesdrop from long range could possibly spy on a passive RFID system. Despite the fact that passive tags have a short operating range, the signal broadcast from the reader may be monitored from a long distance. This is because the reader signal actually carries the tag‘s power and thus necessarily must be strong. A consequence is that a reader communicating with a passive tag in, for instance, a UHF setting might be monitored from a range up to 100-1000 meters. While this only reveals one side of a communication protocol, some older protocols actually broadcast sensitive tag data over the forward channel. Newer specifications, like the EPCglobal class-1 generation-2, avoid this.

Although short-range eavesdropping requires nearby physical access, it can still be a threat in many settings. For example, a corporate spy could carry a monitoring device while a retail store conducts its daily inventory. Alternatively, a spy could simply place bugging devices that log protocol transmissions.

Espionage need not be passive. Attackers could actively query tags for their contents. Rather than waiting to eavesdrop on legitimate readers, an active attacker could simply conduct tag read

operations on its own. Active attackers may be easy to detect in a closed retail or warehouse environment, but may be difficult to detect in the open.
Both eavesdropping and active queries pose threats to individual privacy. RFID tags can be embedded in clothes, shoes, books, key cards, prescription bottles and a slew of other products. Many of these tags will be embedded without the consumer ever realizing they are there. Without proper protection, a stranger in public could tell what drugs you are carrying, what books you are reading, perhaps even what brand of underwear you prefer.
Besides leaking sensitive data, individuals might be physically tracked by the tags they carry.
Of course, cellular phones can track individuals. Unlike a cell phone, which is only supposed to be tracked by a cellular provider, RFID tags might be tracked by anyone (granted, within a relatively short read range). Readers will eventually be cheap to acquire and easy to conceal. Clearly, tracking someone is trivial if an attacker is able to actively query unique identifying numbers from tags. Even if unique serial numbers are removed from tags, an individual might be tracked by the ―constellation‖ of brands they carry. A unique fashion sense might let someone physically track you through an area by your set of favorite brands.

- **Forgery**

Rather than simply trying to glean data from legitimate tags, adversaries might try to imitate tags to readers. This is a threat to RFID systems currently being used for access control and payment systems. While an adversary who can obtain a tag can always clone it, the real risk is someone able to ―skim‖ tags wirelessly for information that can be used to produce forgeries. For instance, if tags simply respond with a static identification number, skimming is trivial. Forgery is obviously a major issue in RFID systems used specifically as an anti-counterfeiting device. A cautionary example is the ExxonMobil SpeedPass, which uses an RFID keychain fob that allows customers to make purchase at ExxonMobil gas stations .
A team of researchers from Johns Hopkins University and RSA Security broke the weak security in SpeedPass and produce forgeries that could be used to make purchases at retail locations.
.

- **Denial of Service**

Weaker attackers unable to conduct espionage or forgery attacks may still be able to sabotage RFID systems or conduct denial of service attacks. An adversary may simply jam communication channels and prevent readers from identifying tags. An attacker could also seed a physical space with ―chaff‖ tags intended to confuse legitimate readers or poison databases. Locating and removing chaff tags might be very difficult in a warehouse environment, for instance.
Powerful electromagnetic signals could physically damage or destroy RF systems in a destructive denial of service attack. Fortunately, attempting these attacks from long range would require so much power that it would affect other electronic components and be easily detected. While these denial of service and sabotage attacks may seem to be simply nuisances, they could represent serious risks. This is especially true in defense or medical applications. For example, the United States Department of Defense is moving towards RFID-based logistics control. An attack against the RFID infrastructure could delay crucial shipments of war materiel or slow down troop deployments.

- **Viruses**

In 2006, researchers demonstrated a RFID virus based on an SQL injection attack [21]. The virus payload was an SQL database query that would overwrite existing RFID identifiers in the database with the virus payload. When tags were updated from the infected database, the virus would be propagated. This virus assumes that RFID contents are essentially ―executed‖ without any validation. It also assumes that future reads from an infected system can overwrite tag contents, which is often not the case in practice. In fact, nothing about the virus was particular to RFID systems. Input from any source such as network connection, USB port or keyboard could spread viruses when insecurely executed without validation.

4.0 Conclusion

In this unit, you have learnt about the various problems faced by radio frequency identification.

5.0 Summary

The various problems faced by radio frequency identification are: data flooding, global standardization, exploitation, passports, shielding, temperature exposure and security and privacy.

6.0 Tutor-Marked Assignment

List and discuss in detail the problems faced by radio frequency identification

7.0 References/Further Reading

Weis S. A. RFID (Radio Frequency Identification): Principles and Applications

Wikipedia (2011). Radio Frequency Identification. Available online at http://en.wikipedia.org/wiki/Radio-frequency_identification

.