Resource: https://ebooks.inflibnet.ac.in/itp12/chapter/wireless-communication/

Terms used in Mobile Communication

MS (Mobile Station)

Mobile station is combination of user's all equipment (mobile phone, SIM, card etc.) and software needed for communication with a GSM network.

Mobile station communicates the information with the user and modifies it to the transmission protocols of the air interface to communicate with the Base Station Subsystem (BSS).

The information of the user communicates with the MS through a microphone and speaker for the speech, keyboard and display for short messaging and wire and cable connection for other data terminals.

In GSM, MS consists of four main components:

- Mobile Termination (MT)
- Terminal Equipment (ME)
- Terminal Adapter (MA)
- Subscriber Identity Module (SIM)

Base Station (BS)

- A **Base station** transmits and receives user data in the cellular network to customer phones and cellular devices. It is connected to an antenna (or multiple antennas).
- BS is a fixed point of communication for customer cellular phones on a carrier network.
- BSs (Base stations) are company specific. However one single site may host multiple base stations from competing telecommunication companies.
- Different types of base stations can be setup according to the coverage required, as follows:
    - Macrocells
    - Picocells

Subscriber Identity Module (SIM)

It is a smart card which stores data for GSM cellular telephone subscriber. It is also called portable memory chip. Data stored by the SIM includes user identity, location and phone number, network authorization data, contact lists, personal security keys and stored text messages. Security features contains authentication and encryption to protect data and prevent eavesdropping.

Base Transceiver Station (BTS)

The BTS is used for data transmission between the mobile phone and the base station. It has a equipment for encryption and decryption of communications, spectrum filtering equipment, antenna and transceivers (TRX).

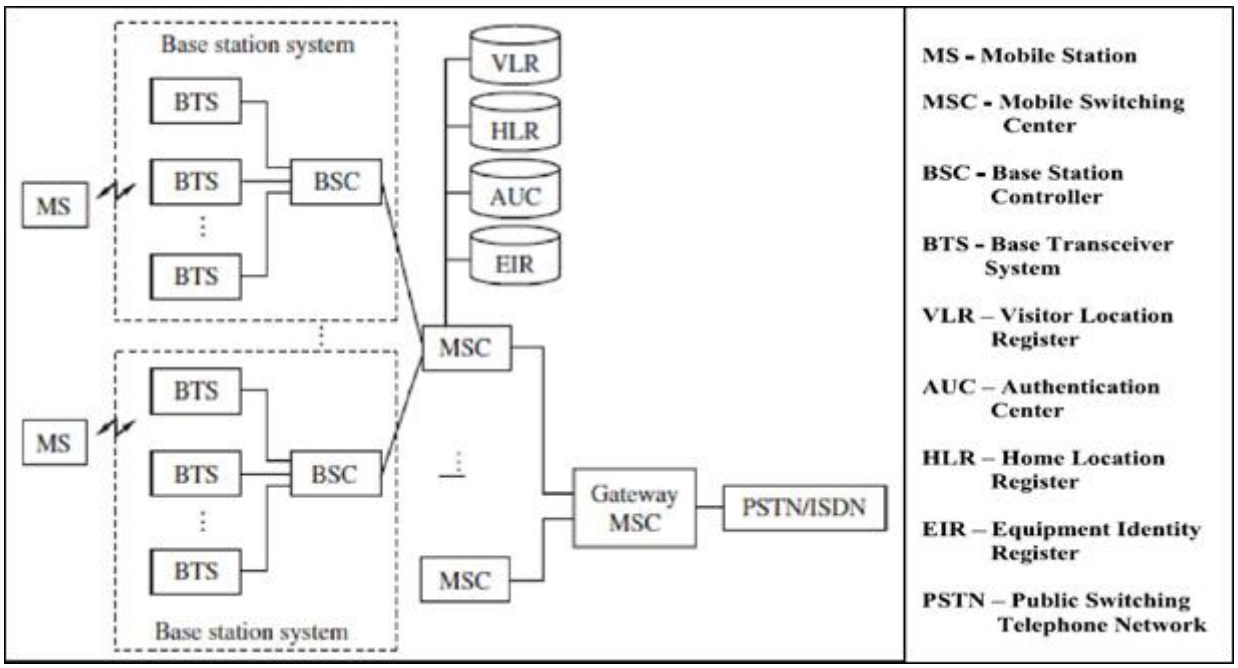
A Base Transceiver Station consists of the following:

- Antennas that relays radio messages
- Transceivers
- Duplexers
- Amplifiers

Mobile Switching Center (MSC)

It is a telephone exchange that is actually used to make the connection between mobile users within the network, from mobile users to the public switched network (PSTN) and from mobile users to other mobile networks.

MSC is the hardware part of wireless switch. It also provides support for registration and maintenance of the connection with the mobile stations.



Base Station Controller (BSC)

BSC is used to control a group of Base Transceiver Stations. It is used for the allocation of radio resources to a mobile call and for the handovers that are made between base stations (BS) under their control. Other handovers are controlled by the MSC.

Channels

Channel is a range of frequency allotted to particular service or systems.

Carrier

Carrier is a company to which your mobile device connects to, such as Idea, Airtel, BSNL, Vodafone etc.

Transceiver

Transceiver is a device capable to perform simultaneously transmitting and receiving radio signals.

Gateway

It a network point that acts as an entrance to another network.

GSM

GSM is called Global System for Mobile Communication. It is a standard to describe protocols for digital cellular networks used by mobile phones.

Mobile Computing: Introduction

Learning Objectives

- Introduction to Mobile Computing
- Characteristics of a mobile computing Environment
- Mobile Computing Entities
- Brief overview of the paper

Introduction

"Nothing is constant in this world other than the change"

Human kind has also undergone several changes since the old stone age. Of many factors contributing to this change, most prominent is advent of computers. It made our life simple. Knowledge got coded into information and intelligence got documented in algorithms. Hence there arised need to share information

amongst different computers. Internet was conceived. It made the world smaller and closer. The computer industry then collaborated with telecommunications industry. The convergence was called Information and Communication technology or ICT. At the same time the communication was becoming wireless. Mobility got redefined. Devices started moving so did the information. Diversified devices synchronized amongst themselves got empowered with computing ability, networking capability and storage capacity.

All these developments have generated an environment which has facilitated us with the provision of **"ANY WHERE ANY TIME"** services and stay connected round the clock. We have transformed rather than changed under the banner of force called "Mobile Computing".

This is the first introductory lecture of this paper. In this module we will understand the meaning of mobile computing, characteristics of mobile computing environment, basic entities and security issues in mobile computing environment.

Mobile Computing: Definition

In a formal way Mobile Computing can be defined as "Set of geographically or temporally distributed computing systems service providers, servers that participate, connect, and synchronize through mobile communication protocols"

The goal of any mobile Computing Environment is to provide decentralized computations on diversified devices, systems, and networks, which are mobile, synchronized, and interconnected via mobile communication standards and protocols

Mobile Computing is an umbrella term used to describe technologies that enable people to access network services anyplace, anytime, and anywhere.

Mobile Computing has different names in different context. Some of them are:

- **Ubiquitous Computing:** Refers to the blending of computing devices with environmental objects empowering them with computing capability
- **Pervasive Computing:** As per Oxford Dictionary, the literal meaning of pervasive "exists in all parts of a place or thing". Pervasive Computing is term used for next generation of computing in which information and communication technology is used everywhere, by everyone, and at all times
- **Nomadic Computing:** Mobile may also, however, refer to access in a fixed location via equipment that users can relocate as required, but is stationary while in operation

Characteristics of a mobile computing environment

Any Computing Environment is called Mobile Computing environment if it supports one or all of the following characteristics:

• **User Mobility:** User accesses the same service while on move from one physical location to another in home network or remote network. For eg. An application can be accessed by a person even when he is at home or in office

• **Network Mobility:** This type of mobility can be seen in two context

o User moves from one network to another accessing the service seamlessly. For eg. Opening an email using Ethernet LAN at office and then going home to access it at Wi-Fi and same service used at 4G connection at a cinema hall.

o Network itself is mobile. A typical case is MANET or Mobile ad-hoc network. . In MANET each node act as a host as well as a "router" to forward the traffic to other specified node in the network. As the nodes move, the routers also move implicating movement of network. For eg. VANET : Vehicular adhoc network in which the vehicles on move form a network with each other and communicate with each other.

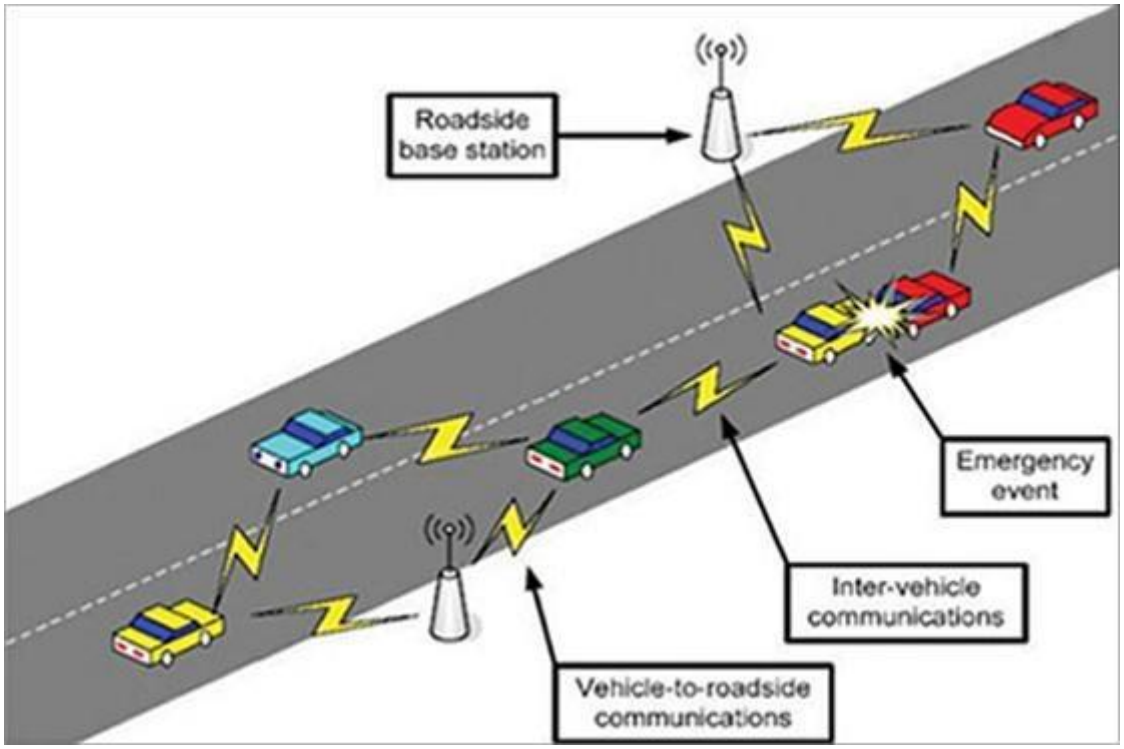


Figure 1 VANET

- **Bearer Mobility:** User uses the same service while switching the bearer. Examples of bearer are HTTP, TCP/TP SMS, WAP, Voice. If a user accesses a service on one bearer at a place and then he moves to another place he can switch to other bearer.
- **Device Mobility:** Use the same service while switching from one device to other. For eg. Opening an online shopping application on desktop and purchasing an item using the application on laptop
- **Session Mobility:** User session should be able to move from one user-agent environment to other. For eg. If a session is interrupted due to some reason on one device/network, it should be able to resume on other device or network
- **Host Mobility:** User device can be either server or host.
- **Agent Mobility:** User-agent or the applications move from node to another. Agents can be browsers, crawlers, aglets etc. They should be able to move when the node moves.

Mobile computing entities

The following entities comprises of Mobile Computing environment:

User with a device

User device is Combination of Hardware device called as User Equipment with software called User Agent embedded inside. A device can be fixed or wireless. It can be computing device or Communicating device but use of them is blended into a task flow that their function is integrated. The examples of devices in their respective category are:

Computing
•Computers(fixed) •Mobile Phones •Laptops •PDA •Palm top computers

Communicating
•Telephone(Fixed) •Mobile Phones •Digital TV •Pagers

Dialogue Control

The dialogues can be long session oriented dialogues or sessionless dialogues that are short term. The type of dialogue depends on type of device used and input output devices available. For eg. Checking bank balance can be a long session oriented dialogue which comprises of entering the URL, authentication via user name and password, checking balance by entering the account number and then logging out. The same thing can be accomplished via sending a message through mobile phone which would be a short term sessionless transaction.

Networks

The networks used for mobile computing are fixed or wireless, infrastructure or adhoc and bearer. Fixed or wirelined networks are installed with the help of fiber optical cables, coaxial cables. These are generally public networks and cover a large area. WIRELESS NETWORKS do not need wires. The nodes interact with each other via radio interface. Infrastructure LANs need a infrastructure in the form of access point which acts as a central hub through which all the nodes communicate. AD-HOC LANS do not have any access point. They are temporary for a particular time and span over a limited range. They can be easily configured and de-configured. Optical, Infrared and radio communication are helpful for formation of Adhoc LANs. Bearers are used for transportation of data. Different type of networks uses different type of bearers. Examples of different categories of networks are summarized as below

Fixed Networks
•Broadband networks over DSL line •Cables •PSTN •Satellite(used as part of infrastructure) •Internet Backbone

Wireless Networks
•Personal Cellular systems(PCS) •AMPS(Advanced Mobile Phone System) •GSM •CDMA •DoCoMo •GPRS •Wireless in Local Loop(WLL) •Public Land Mobile Network

Ad-Hoc Networks
• Bluetooth • Wireless LAN(802.11) • IrDA • VANET • MANET • Sensor Networks

Bearers
• TCP/IP • HTTP • SMS • Unstructured supplementary service data(USSD) • WAP • Voice

Middleware

Middleware is software entity between user application and operating system or between application and device. In context with mobile different middleware are applicable:

Communication middleware	Transaction Processing Middleware	Behaviour Management Middleware
•Connectors for different services like TN3270 for IBM mainframe services, Javamail Connector for IMAP,POP3 services	•Convert session oriented (SoD) to sessionless dialogue(SID) •Management of web components	•Manage device specific rendering at run-time independent of the application •Eg. one application run on WEB, other on WAP and different one for SMS.

Gateways

Gateways act as interface between transport bearers. For eg. IVR gateway to interface voice with computer or a WAP gateway to access internet over phone or a SMS gateway for sending SMS from application.

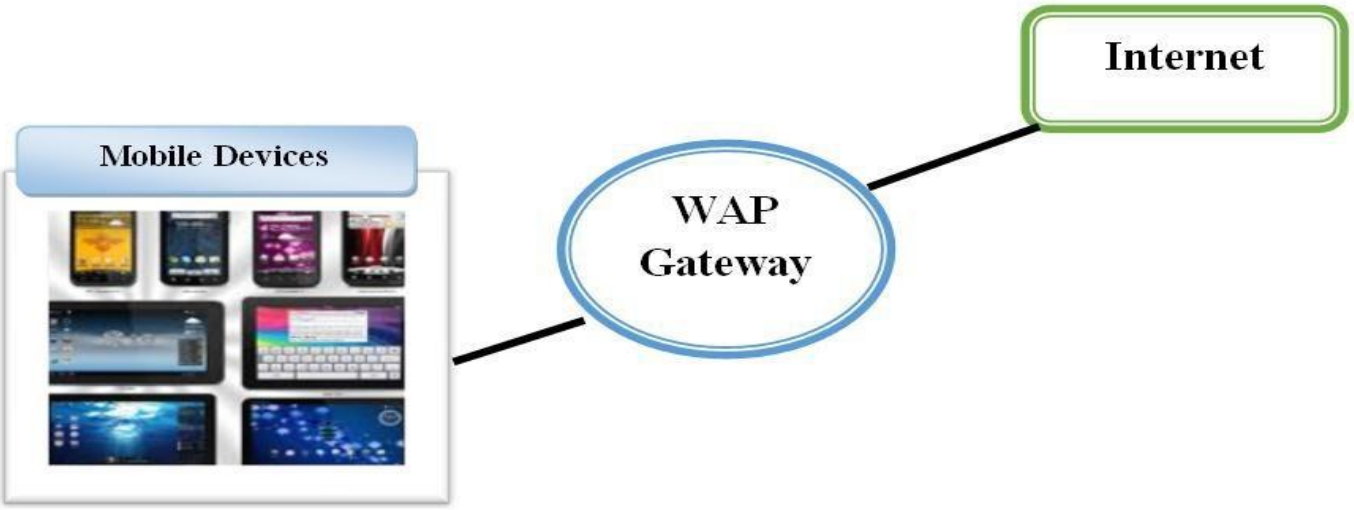


Figure 2 WAP Gateway

Content

Data and information, applications and services all constitute the content. The applications and services run on the origin server known as content servers. Depending on the context, content can be different for different people at different times. For eg. Viewing stock update at one time via one application at the same time using another application to book movie ticket at other time. Looking at current scenario when there are around 2 billion applications on apple store and 2,5 billion applications on android store, there cannot be any comprehensive classification or listing of mobile applications. Yet some of the common categories and examples of applications in those categories are as follows:

Personal	Transaction Oriented	Perishable
•PAYTM •Reminders •Organisers •Remote Monitoring •Tour organisers •Home designs	•Mobile Shopping •Mobile Banking •Bill Payment	•News •Weather forecast •Stock updates •Sports Update

Location Specific	Corporate	Entertainment
•Navigation •Location tracking •Bookings	•Business alerts •Mails •ERP •MIS •Virtual office	•Chatting •Social Networking •Interactive games

Government	Scientific
• Public announcements • Tax payment • Complaint booking • Citizen reporter	• Water quality monitoring • Wind speed measure • Virtual Laboratories • Simulators • Citizen science

Standards

Standards are documented agreements containing technical specifications or criteria to be used as rules, guidelines or definitions of characteristics". Materials, products, processes and services should comply to these guidelines. Standards are necessary for interoperability of goods and services. The bodies who develop and maintain standards work at regional, national and international level. These standard bodies are formed by governments, professional institutes and industry consortiums. For eg. BSI or Bureau of Indian Standard. Some of the standard bodies related to mobile computing technologies are:

1. **The International Organization for Standardization (ISO)** is a worldwide federation of national standards bodies from more than 140 countries, one from each country. ISO is a non-government organization established in 1947. The mission of ISO is to promote the development of standardization and related activities in the world with a view to facilitating the international exchange of goods and services.
2. **Internet Engineering Task Force (IETF)** is the standard – making body for Internet and related technologies. IETF is an open international community of network designers, operators, vendors and researchers concerned with the evolution of Internet architecture and smooth operation of the Internet.
3. **ETSI (the European Telecommunications Standards Institute)** is an organization whose mission is to produce telecommunication standards that will be used for decades to come throughout Europe and possibly beyond. ETSI unites members from countries inside and outside of Europe, and represents regulators, network operators, manufacturers, service provider's research bodies and users.
4. **The Open Mobile Alliance (OMA)** has been established by the consolidation of the WAP Forum and the Open Mobile Architecture initiative. It intends to expand the market for the entire industry by removing barriers to interoperability and supporting a seamless and easy – to –use mobile experience for end users.
5. **ITU (International Telecommunication Union)** is an organization within the United Nations System. It was founded on the principle of cooperation between governments and the private sector. With a membership encompassing telecommunication policy- makers and regulators, network operators, equipment manufacturers, hardware and software developers, regional standards-making organizations and financing institutions, ITU's activities, policies and strategic direction are determined and shaped by the industry it serves.
6. **IEEE Standards Association (IEEE-SA)** is an organization that produces standards, which are developed and used internationally. Standard for Wireless LAN are created, maintained and managed by IEEE. These are defined through different 802.11 standards.
7. **The Electronic Industries Alliance (EIA)** is a national trade organization within the US that includes the full spectrum of its electronics industry. The Alliance is a partnership of electronic and high-tech associations and companies whose mission is promoting the market development and competitiveness of the US high-tech industry through domestic and international policy efforts.
8. **World Wide Web Consortium (W3C)** develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential. W3C is a forum for information, commerce, communication and collective understanding. By promoting interoperability and encouraging an open forum for discussion, W3C is committed to leading the technical evolution of the Web.
9. **3GPP** is to produces globally applicable technical specifications and technical reports for 3rd Generation Mobile System based on evolved GSM core network and radio access technologies that

they support, i.e., Universal Terrestrial Radio Access (UTRA) both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes. The scope was subsequently amended to include maintenance and development of the Global System for Mobile communication (GSM) technical specifications and technical reports including evolved radio access technologies (e.g., General Packet Radio Service (GPRS) and Enhanced Data rates for GSM Evolution (EDGE)).

10. **The American National Standard Institute (ANSI)** is the national standard organization in the United States. ANSI Standard X3.4- 1968 defines the 'American National Standards Code for Information Interchange (ASCII)' character set. ASCII character set is used in almost every modern computer today.
11. **Universal Mobile Telecommunications system (UMTS)** represents an evolution in terms of services and data speeds from today's second-generation mobile networks life GSM. As a key member of the global family of third generation (3G) mobile technologies identified by the ITU, UTMS, is the natural evolutionary choice for operators of GSM networks.
12. **Bluetooth** wireless technology is a worldwide specification for a small –form factor, low-cost radio solution that provides links between mobile computers, mobile phones, other portable handheld devices, and connectivity to the Internet. The standard and specification for Bluetooth are developed, published and promoted by the Bluetooth Special Interest Group.
13. **The CDMA Development Group (CDG)** is an international consortium of companies who have joined together to lead the adoption and evolution of CDMA wireless systems around the world the world. The CDG comprises the world's leading CDMA service providers and manufacturers. By working together, the members will help ensure interoperability among systems, while expediting the availability of CDMA technology to consumers.
14. **The Public-Key Cryptography Standard (PKCS)** are specifications produced by RSA Laboratories in cooperation with secure systems developers worldwide for the purpose of accelerating the deployment of public-key cryptography.
15. **The Presence and Availability Management (PAM) Forum** is an independent consortium with a goal to accelerate the commercial deployment of targeted presence and availability applications and services that respect users'. Preferences, permissions and privacy
16. **The Parlay Group** is a multi-vendor consortium formed to develop open, technology- independent application programming interfaces (APIs). Parlay integrates intelligent network (IN) services with IT application via a secure, measured, and billable interface. By releasing developers from underlying code, networks and environments, Parlay APIs allow for innovation within the enterprise
17. **DECT stands for Digital Enhanced Cordless Communications.** It is an ITSI standard for portable phones. DECT is known in ITU as a 3G system and is commonly referred as IMT- FT (IMT Frequency Time).
18. **WiMAX Forum** is Worldwide Interoperability for Microwave Access Forum dedicated to certifying the operations of interconnecting devices. WiMAX aims to provide wireless data over long distances in different forms ranging from point-to-point links to full scale mobile access networks for wireless broadband communication.
19. **TTA is Telecommunications Technology Association**. TTA is an IT standards organization catering to development of new standards based in Korea. It provides one-stop services for comprehensive IT standards.
20. **Wi-Fi owns trademark to Wi-Fi alliance**. It was previously known as Wireless Ethernet Compatibility Alliance. It is focused on interoperability and compatibility of Wireless LAN devices and committed to continuous improvements in design and better user experience.

21. **Association of Radio Industries and Businesses (ARIB)** is an institution, based in Japan, dedicated to efficient use of radio spectrum and its implications in businesses. China Communications Standards Association (CCSA) is an attempt of Chinese Ministry of IT to reform telecommunications industry and market. It aims to become a nationally unified standards organization in china.
22. **Digital Living Network Alliance (DLNA)** is a cross-industry association of consumer electronics, computing industry and mobile device companies. The objective of DLNA is to establish a conglomeration of wired and wireless interoperable network of personal computers, consumer electronics and mobile devices in the home and outside in order to enable a seamless environment.

Summary

- This is the introductory lecture of the paper
- Mobile computing is a paradigm to describe technology which provides access to services on move. Also known as ubiquitous, pervasive or nomadic computing.
- Any computing environment involving user, device, network, bearer, session, host and agent mobility can be called mobile computing environment

User device, networks, gateways, middle-wares, dialogue control and content servers and applications and standards form components of mobile computing infrastructure.

WIRELESS Communication

Learning Objectives

- Introduction to wireless Communication
- Types of Wireless Communications
- Advantages of Wireless Communications
- Challenges for Efficient Wireless Communications

History of Wireless Communications

1. Introduction

From application of light to generate on/off patterns to highly advance 5G technologies; From the very primitive radio control toys to the highly sophisticated satellites; from 535 kHz AM radio to 2.4 GHz LAN; from very long km waves in submarine communication to ultra wide band short pulses; from global cellular GSM Network to the extremely close body area network, wireless communication has revolutionized the way in which we work, connect or communicate to the world. It has made the world a small close entity and has uncovered concept of pervasive and ubiquitous computing with the ideal of "ANYWHERE ANYTIME".

2. Wireless Communication

"Transmitting or receiving voice & data over a distance without use of electrical conductors or wires"

When one or all of communicating devices are mobile in nature, it is known as Mobile communications. The distance of communication can be short as few meters in case of Wi-Fi, Bluetooth, TV Remote or it can be significantly long as thousands of kms in case of radio, mobile phones, satellite communications, and High Altitude platforms.

Then how does the communication takes place?

Wireless communications require some form of energy to transmit information. Principle source for wireless communication is the electromagnetic waves. Electromagnetic waves are movement of massless particles moving in sinusoidal fashion with the speed of light.They are charactrised by different wavelength and frequency.

- **Frequency:** The literal meaning of frequency is rate of change. In context with E.M. waves, frequency means number of cycles/time. Unit of frequency is Hertz.

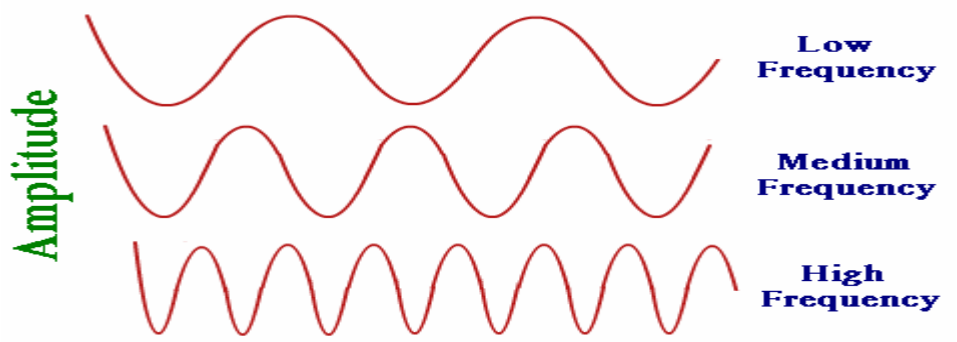


Figure 1: Attributes of E.M. waves

- **Wavelength:** Distance between two peak values of a E.M. wave. Wavelength is inversely proportional to frequency (Fig. 1). Unit is same as of distance.
- **Amplitude:** Strength of E.M. wave. More the amplitude more power an wave as. For ex. For light amplitude specifies brightness of light.

The E.M. waves are arranged according to frequency, wavelength and energy in the form of spectrum known as electromagnetic spectrum. The different bands of E.M. Spectrum range from Gamma rays with highest frequency (order of THz) to radio waves with lowest frequency (of order KHz).

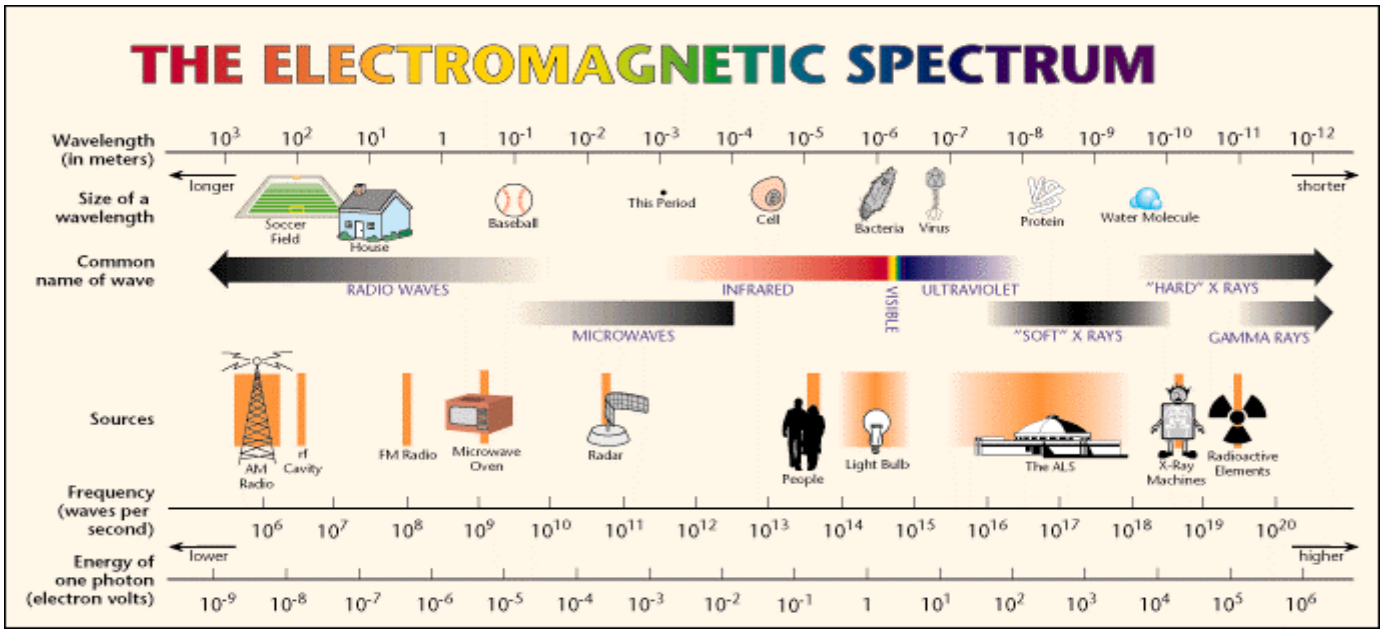


Figure 2: ElectroMagnetic Spectrum

The commonly used bands of electromagnetic spectrum appropriate of wireless communication are:

- Radio waves: Radio, Television, Mobile phones, wireless LAN etc.

    - It can travel through long distances
    - Easily generated
    - Less obstructed by obstacles
    - Can travel along Earth's surface(Ground Waves)
    - Penetrate through layers of atmosphere
    - Can travel across hemisphere of Earth by reflection from ionosphere(Sky Waves/Short Waves)
    - At higher wavelengths, can travel under water, as required in submarine communication
    - Highly licensed
    - Behave differently at different frequency
- Microwaves: Radar, cellular phones
    - Can travel through short distances due to higher frequency
    - Obstructed easily
    - Line-of-sight transmission
- Infrared waves: Remote controllers, Infrared data Association (IrDA)
    - Used for short communication
    - Highly Directional
- Visible: Laser light for wireless Intranet, barcode readers, indoor communication using LEDs
    - Travel in straight line
    - Need Line-of-sight
    - Need to be focused using laser
    - Effected by bad weather

Other forms of energy are :

acoustic used in hydrophones and wireless sensor networks
ultrasonic
electronic where an electron beam is used to transfer the information

1. Current wireless systems

Cellular systems based mobile phones (2G……5G) Cordless Phones
Wireless LAN's
Wimax GPS
Satellite Systems Paging
PAN (Bluetooth)
Vehicular adhoc networks Sensor Networks (RFID) Body area network
UWB – Ultra wide band radios
Low cost low power Bluetooth and Zigbee
…….***and many more to come***

2. Advantages of Wireless Communications

- **Freedom from wire**

When you are not tied, you can fly

Wireless offers a neat and hassle free environment without worrying about messy yet costly wires, jacks and plugs which makes it easy to establish, use and maintain. Apart from wireless headphones, keyboards,

wireless phone chargers, ultra wide band also promises wireless desktops in coming years. Bluetooth technology enabling a sms on mobile phone to be wireless transmitted to a printer have reduced the labor of connecting to jacks, plugs and messy yet costly wires.

- **Mobility**

The waves supporting wireless communications move freely in air hence do not tie a device to a particular location as done by messy wires and cables. Technologies like Mobile IP and Cellular IP provide undisrupted services to the users on move. Roaming facility initiated by GSM now has been accepted by all standards providing flexibility to stay connected anywhere on globe.

- **Global connectivity *Where the wires cannot go, wireless can go***

Wireless communications comes with the tagline "Anywhere anytime". Communications can reach the places which are otherwise infeasible or costly to be connected by wires like rural areas, moving vehicles, hilly regions, underwater submarine. Satellite technologies enables seamless connectivity in remote and inaccessible places like deserts, dense forests, battle fields, calamity affected areas. For ex: NASA'S FINDER is a handheld device which detects heart beats of humans buried under during Earthquakes. Teleteaching applications, portals for farmers to provide them information regarding crops are many other advantages of global connectivity via wireless communications.

- **Convenience *Easy is the best***

Automatic instant communications without physical setup e.g.: Wi-Fi, Bluetooth. RFID technology enabling automatic toll collections at booth, UPS delivery system which uploads the delivery status real time to the server.

- **Reduced cost**

Since wireless setups do not use elaborate infrastructure, the installation and maintenance cost is less as compared to wired which includes per foot charges of cables and other devices. Time and labor to plan wiring is expensive and further more damage to buried cables requires replacements and replanning hence maintenance is also a costly affair. Conversely, wireless setups if done properly require very less maintenance and that too is easy and cheap.

- **Flexibility*You can grow only when you are flexible***

A wireless setup can easily be scaled by configuring more devices to the existing setup which is otherwise difficult in wired one because of limited number of user supports which are physically connected to the device.

3. Challenges

Ray Davis has truly said "Challenges become obstacles only when you bow to it" Wireless communications in any form has to undergo and overcome many challenges some of which are listed down. There are lot of efforts already made to cope up with these challenges and many more are in pipeline.

- **Efficient Hardware**

The devices used in the wireless communications should be functionally good yet should be small, consume low power and should be low cost hence offering a green wireless communication. With the

introduction of 3G and following technologies, multimedia applications are most sought after which is a most power hungry scenario.

For ex. Power is the most critical resource in Wireless Body area Networks which provide ubiquitous healthcare infrastructure like heart monitoring system etc.

At the equipment level, power saving can be achieved by switching it to sleep mode when not in use and using low power transmitters and receivers as one of the essential hardware requirements for every wireless device. Advances in semiconductor technology and transmissions at high frequencies have resulted in small antenna sizes hence making the devices light weight and energy efficient. Energy efficient signal processing components like CMOS offers an energy efficient, cheap yet slow option than its counterpart ECL which is fast but power hungry.

At operating system level, power can be controlled by use of power aware file systems, graphics drivers and other components. Technologies should be designed in such a way that they extract less of power from the device on which they are installed like Bluetooth, Zigbee the most efficient is Ultra wide band in terms of bandwidth, cost , power , physical size and data rates . Nanotechnology ,a boom in increasing the efficiency of devices by using Magnetic nano particles and metaparticles to reduce size of Antenna, Carbon Nano Tube memory to increase memory, less energy consumption and provide better human interfaces

- **Finite Radio Spectrum**

Though the recent wireless technologies use license free 2.4 GHz ISM band but radio frequencies used for long distance communications like satellite communication, radios, televisions, mobile phones still rely on licensed frequency spectrum which is limited and regulated by international standards. So it is required to efficiently use the available spectrum without or very less interference allowing more subscribers bases. One approach is Cellular frequency reuse in which the geographical areas are divided into small cells with neighboring cells assigned non-overlapping frequencies. Frequency reuse is supported by assigned same frequency to those cells which are geographically distant from each other. Cognitive radio is emerging as another alternative as it increases the utilization of radio spectrum to support the traffic. Spread Spectrum technologies like Frequency Hopped Spread Spectrum and CDMA allows using entire bandwidth without interference. Cell sizes are getting small to increase the spectral efficiency supported by sectorized antennas.

- **Quality of service**

As wireless communication is getting more and more ubiquitous, a high quality of service in delivery of voice, video and data is expected. In mobile telephony, QoS is estimated by voice quality, signal strength, low call blocking and dropping, high data rates for multimedia applications, whereas for wireless networks QoS depends on throughput, delay, packet loss-rate, low packet error rate and reliability. Factors effecting QoS are multipath interference, noise, improper handoffs during mobility and many more. The solutions to improve QoS:

- Signal processing to overcome attenuation
- Repeaters
- Wide spread cellular infrastructure
- MIMO technology
- OFDM to reduce multipath interference

- Proper handoff mechanism
- 3G and 4G technologies(99.999 reliability)
- Mobile IP
- Improved TCP algorithms to increase throughput
- Improved MAC algorithms for carrier sense and collision detection

- **Mobility Management**

An ideal wireless technology particularly in Mobile computing environment, should maintain uninterrupted connectivity when a mobile device changes location. Proper infrastructure management for installation and maintenance of cells and their networks is essential. Proper Localization techniques and registration methods should be applied to trace and connect the device to the network. GSM uses HLR and VLR databases which maintain information about the location of the devices. Mobility in Internet to support is supported by Mobile IP technology.

- **Multipath Propagation**

When the signal travels from sender to receiver through the radio channel, it is reflected, refracted, diffracted or scattered by the objects in the environment like buildings, atmosphere etc in mobile telephony and furniture in WLAN. As a result the signal does not follow line-of-sight and follow different propagation paths in the form of multipath components. Each multipath component travel with different amplitude, delay and phase shifts among each other. These components add up constructively or destructively at receiving end affecting total signal amplitude .This effect is called fading. Furthermore each multipath component arrive the receiver as a sequence of pulses with different time, amplitude and phase called Intersymbol interference. MIMO and OFDM are the solution to this affect. OFDM signals can be made orthogonal to each other so that they do not interfere. It is used by 802.11n, Wi-Fi, LTE-advanced, Wimax and many more technologies. MIMO is a technology which uses multiple antennas to setup multiple data streams on the same channel to increase its capacity. It takes advantage of multipath propagation in providing additional capacity to the channel.

- **Security**

Wireless communications use air as the medium hence are highly susceptible to threats like interception, masquerading, denial-of-service, replay attacks and many more. Though wireless communications in all forms should be performed in a secured way but is more crucial in mobile computing applications in the field of banking, business, governance etc. and is therefore area of special concern. Every technology comes with its own security measure. CDMA itself offers built in security as the reciver has to know the code to decode the data. GSM offers authenticity both at device level and network level. At device level, unique PIN is used to authenticate SIM card. At network level authentication center performs authentication using challenge response technique with the help of secret key on the SIM card. It also same key to encrypt the traffic using algorithm stored on SIM.WLAN offers security through SSID, WEP (Wired equivalent privacy) and authentication servers like radius, Kerberos etc. Digital signatures and certificates are used for verification of records. To maintain anonymity, the data is transmitted using temporary identifiers of the users like TMSI number in GSM. Various Intrusion prevention and detection systems are available for adhoc networks like MANET and VANET.

4. History of Wireless Communications

History of wireless communications goes far back in ancient times when light was used to generate on/off patterns in China and smoke signals were used in Greece. The Timeline showing evolution of wireless and mobile communications from its inception by invention of optical telegraph till date.

1794: Claude Chappe invented the optical telegraph using optical telegraph lines. The communication used optical frequencies which could not be focused and was also deviated by rain and fog.

1831: Demonstration of electromagnetic induction by Michael Faraday and Joseph Henry

1864: James C. Maxwell laid foundation for electromagnetic fields with his famous equation

1886: Heinrich Hertz demonstrated the wave character of electrical transmission Nikola Tesla experimented by increasing the distance of electromagnetic transmission

1895: Guglielmo Marconi marked beginning of radio communication. Demonstrated wireless telegraphy

1901: First transatlantic transmission at a distance 3200 kms.

1906: First radio broadcast when Fessenden transmitted voice and music for Christmas.

1907: First commercial transatlantic connections were set up.

1915: First wireless voice transmission was set up between New York and San Francisco.

1920: first commercial radio station started

1920: First commercial radio station started (KDKA from Pittsburgh).

1920: Marconi discovered short waves. They have advantage of being reflected at Ionosphere and hence can be send around the world.

1926: First telephone in a train was available on Berlin Hamburg line with wires parallel to the rail track working as antenna.

1927: First car radio was available

1928: Television broadcasting across Atlantic by John L. Baird

1932: First teleteaching started from CBS Station. Up until then, all wireless communication used amplitude modulation, which offered relatively poor quality due to interference.

1933: Edwin H. Armstrong invented frequency modulation.

1958: First analog wireless network A-NETZ in Germany using carrier frequency of 160 MHz

1971: ALOHANET developed at University of Hawaii based on packet radio (First wireless lamp).

1972: B-NETZ was followed in Germany

1973: Motorola invented first mobile cellular phone

1979: Northern European countries of Denmark, Finland, Norway, and Sweden agreed upon the nordic mobile telephone (NMT) system which use 450 MHz carrier.

1982: The 'Groupe Special Mobile' was founded which used 900 MHz, allow roaming throughout Europe, was fully digital and offer voice and data service. "GSM was formed"

1983: AMPS, an analog mobile phone system working at 850 MHz started.

1984: Standard CT1 Cordless Telephone for cordless phones formed

1991: ETSI adopted the standard Digital European Cordless Telephone (DECT) for digital cordless telephony which works at a spectrum of 1880-1900 MHz with a range of 100-500 m. It used 120 duplex channels and offered data rate of 1.2 Mbit/sec. Fully digital systems were introduced.

1991: GSM was standardized in a document of more than 5,000. This first version of GSM, now called "Global system for Mobile Communication", works at 900 MHz and uses 124 full-duplex channels.

1996: ETSI standardized the 'High Performance Radio Local Area Network' (HIPERLAN) which operated on 5.2 GHz and offered data rate of 23.5 Mbit/s.

1997: IEEE standard 802.11 was founded and was a strong contender in local area networks. It used 2.4 GHz licensed free band offering data rate of 2 Mbps-10Mbps.

1998: Marked the beginning of mobile communication using satellites with the Iridium system (Iridium, 2002). It consists of 66 satellites in lower earth orbit and uses 1.6 GHz for communication. Marked the beginning of portable mobile satellite phones. Universal Mobile Telecommunication System (UMTS) was standardised. It combined GSM networks technology with CDMS Solution.

1999: ITU standardized IMT-2000 which included family of standards like UMTs, CDMA2000, and DECT.

1999: Bluetooth was commercialized

1999: Wireless Application protocol started.

2000: GPRS offering higher data rates and packet oriented transmissions 3G spectrum started.

2001: Third generation of mobile communication started in Japan with the FOMA service, in Europe with several field trails and in, Korea with cdma2000.IEEE released a new WLAN standard; 802.11a, operating at 5 GHz and offering gross data rates of 54 Mbit/s.

2002: Digital terrestrial TV started, first UMTS network was launched allowing high-speed applications such as mobile TV and video calling.

2003: EDGE was deployed by AT&T on Singular network in the USA. IEEE 802.11g was added to the 802.11 standard, allowing transmission speeds up to 54 Mb/s. Bluetooth specification 1.2 was released. This new specification includes Adaptive Frequency-hopping (AFH), which reduces RF interference.

2004: Newest version of IEEE 802.16 is added and it completely changes the WiMAX standard. Bluetooth specification 2.0 is released.

2008: 4G launched

2009: The latest in Wi-Fi standards 802.11n approved

2010: IMT-A ratified by ITU and 4G communications deployed all over. First 4G handset is introduced at International CTIA WIRELESS show.

5. Conclusion

In this module we have introduced wireless communication and its types. We have seen advantages and challenges of wireless communication and have browsed through the history since its inception.

"We are living in an era of five point reliability" and a long way to go towards 5G".

Cellular Systems: Part 1

Learning objectives:

- Understand the motivation behind development of Cellular systems
- Discussion of cellular system architecture and terms associated with it
- Discussion on shape and size of the cell
- Types of the cell
- Frequency reuse in cellular systems

Introduction

Cellular Systems or networks are a technology in which a wide geographical area is broken up into small units called cells. The basic motive of cellular technology was to use many low power transistors strategically placed all over the geographical region. The low power antennas are connected to a central

exchange. Neighboring cells are assigned distinct frequencies or channels which can be reused after a certain distance by the means of clustering thus facilitating frequency reuse. In this module we will discuss the cellular technology and understand the various terms like cell, cluster, frequency reuse etc. We will understand why shape of the cell is hexagonal and classify cells on the basis of their sizes.

Motivation

Let us first understand the rationale behind the development of cellular technology. To understand let us go back as early as in 1946 when first mobile call was made. A single transmitter was used with limited capacity and channels. For eg. In New York, for 2000 subscribers only 12 channels were available. A subscriber had to wait for 30 minutes to make a call and the instrument used was large in size. To improvise the situation, cellular telephone service was conceived at Bell labs. The system consisted of low power transmitter spreaded strategically throughout the city facilitated with frequency reuse and automatic call handoffs. The reason for having low power antennas was that the transmission of antennas would reach only to a certain range beyond which the signal strength decreases. So the same channel can be used once the device is out of range of the transmitter. This is the principle behind cellular technology known as frequency reuse. The goal is to improve the spectral efficiency by providing services to millions of subscribers within limited spectrum.

This is the principle behind the cellular technology. Large geographical area is divided into small geographic regions called **cells** hence the name **cellular technology** and the conventional radio phones got the names **cell phones**.

Cellular System Architecture

Cell : is a basic geographic unit of a cellular system. It comprises of small area which has a low power antenna installed in the form of a unit called the **base station**. The location of base station depends on whether the antenna is omnidirectional or directional.

- **Center excited cells:** Uses omnidirectional antennas and the base station is at the center of the cell
- **Edge excited cells:** Uses directional antennas and the base station is at the periphery of the cell

Footprint of the cell: The transmission of the base station can be detected only if it is within a threshold value. This area is called footprint or coverage area of the cell. This is expected to be spherical in nature owing to the isotropic nature of the radiator.

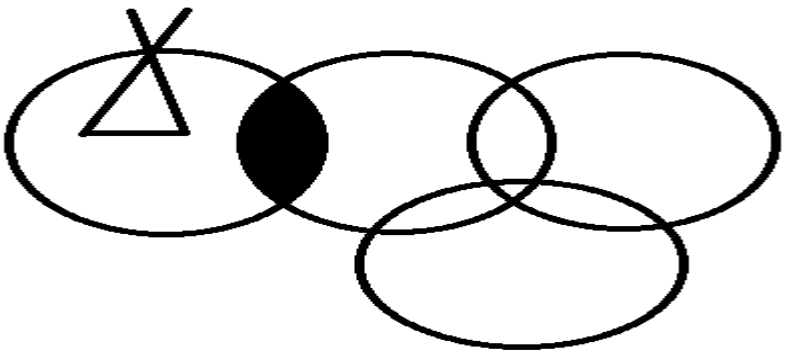


Figure 1 Footprint of the cell

Cluster:

Each cell is assign a frequency channel or group of channels. Adjacent cells are assigned non – overlapping frequency channels. Group of such cells with non- over lapping frequency is called a cluster. Size of the cluster (number of cells) can be 3, 4, 7, 9, 12, 21. The size of the cluster C, is according to shift parameters i and j. i is step along one direction and j is step along another direction. C is given as:

$C = i2 + ij + j2$

Typical values of C are given in the table below:

Cluster Size(C)	Values of i and j
C=1	i=1, j=0
C=3	i=1, j=1
C=4	i=2, j=0
C=7	i=2, j=1
C=9	i=3, j=0
C=12	i=0, j=2

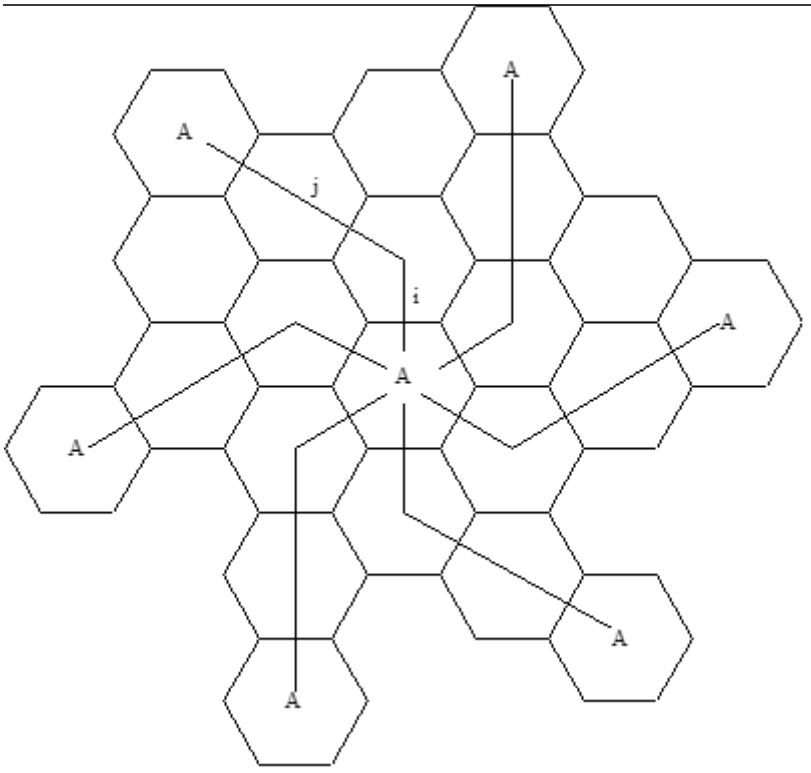


Figure 2: I and j parameters

Common Cluster sizes:

- GSM cluster size C = 3 and 4
- CDMA cluster size C = 1
- FM cluster size C = 7 or 9

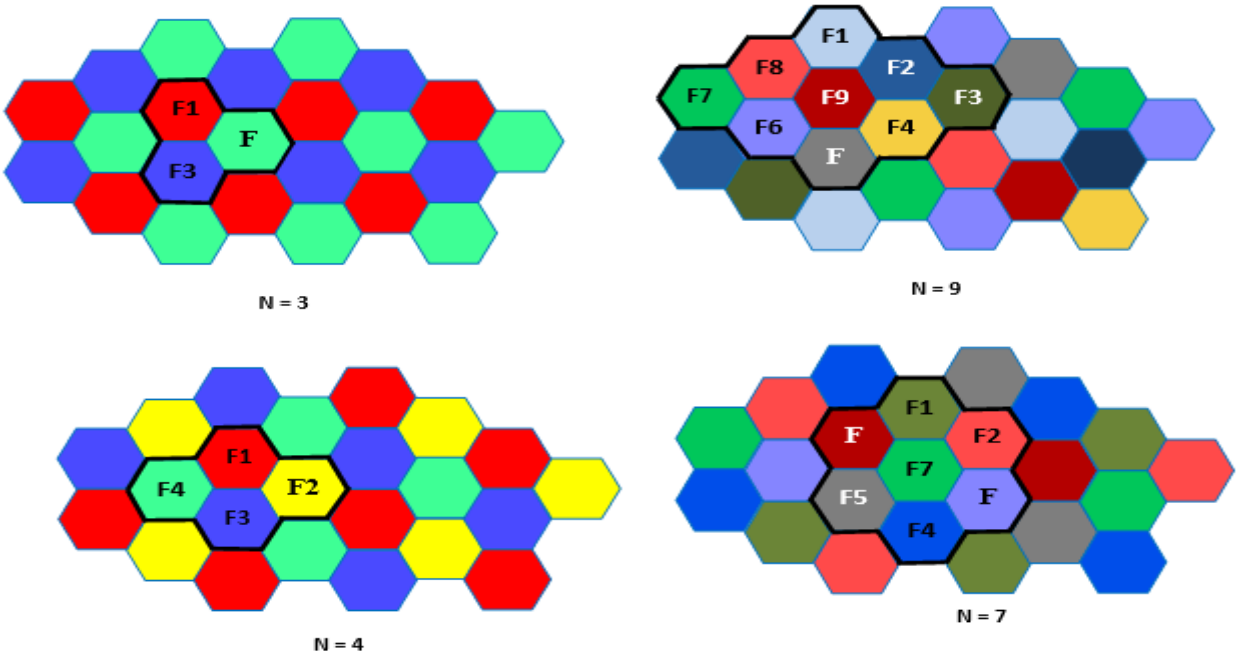


Figure 3 Cluster of different sizes

Same frequency (F1 and F1) can be assign to different cells in such a way that they are at a sufficient distance from each other and communication in this cell do not interfere with each other. Such cells are called **Co-channel cells**. If number of cells in cluster is C then frequency reuse factor is **1/C.** Distance between co-channel cells is called **reuse distance.** Larger the cluster, more the reuse distance, less interference. Smaller the cluster, lesser the reuse distance.

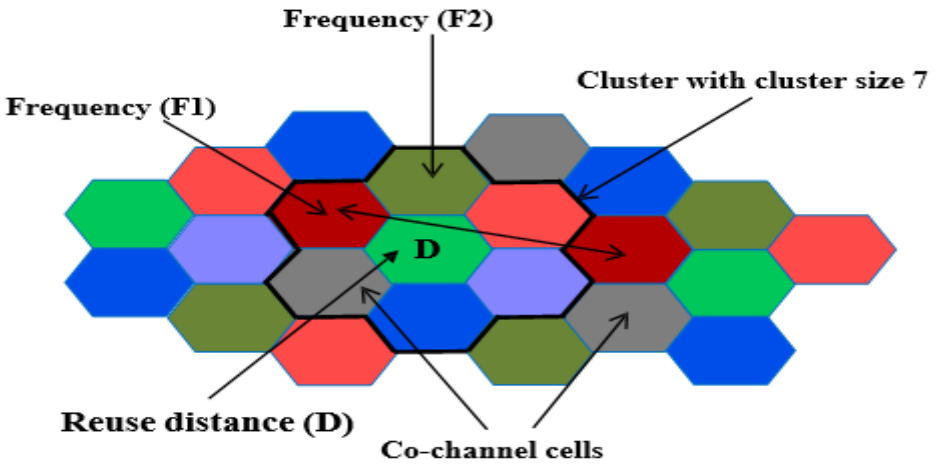


Figure 4 Reuse distance

This cluster can repeat itself and hence the same set of channels can be used again and again. By repeating the cluster in a systematic way, the same frequency can be reuse in different area for a different transmission. As the number of times cluster is repeated increases capacity increase. The closest distance between the co-channel cells (in different clusters) is determined by the choice of the cluster size and the

layout of the cell cluster. The distance between the co-channels cells depends on size of cluster. More the number of cells in the cluster, more the co-channel distance If cluster size is small number of time it will be repeated will be more hence capacity will increase. But if cluster size is small, the distance between co-channels cells will be less hence interference will be more. To tradeoff this situation, smallest cluster size for which interference can be tolerated should be taken. Larger cluster size demands for more number of unique channels hence number of channels allocated to each cells will be reduce decreasing the capacity.

Shape and Size of the Cell

Shape of the cell

Expected shape of the cell owing to the isotropic nature of the base station antenna should be a circle. Gaps results in loss of coverage and overlapping lead to interference of the signal. Therefore a polygon should be taken which can cover entire area without overlap and gaps and also it should rep geometry of circle. An equilateral triangle (all side equal), square or regular hexagon (all side equal) can be taken. They do not overlap or leave gaps. The next criteria is that the area covered should be maximum and also represent geometry of circle. The area of all the three shapes as compared to circle(Fig 5):

Equilateral triangle: 17.77 % of area of circle
Square: 63.7 % of area of circle
Hexagon: 83% of area of circle

Therefore hexagon is the ideal shape for a cell. In hexagon distance between center and the farthest point in the cell, a regular hexagon covers maximum area. Therefore regular hexagon is the ideal choice for shape of cell. This size is only the logical shape. In practice a cell exhibits irregular geometry.

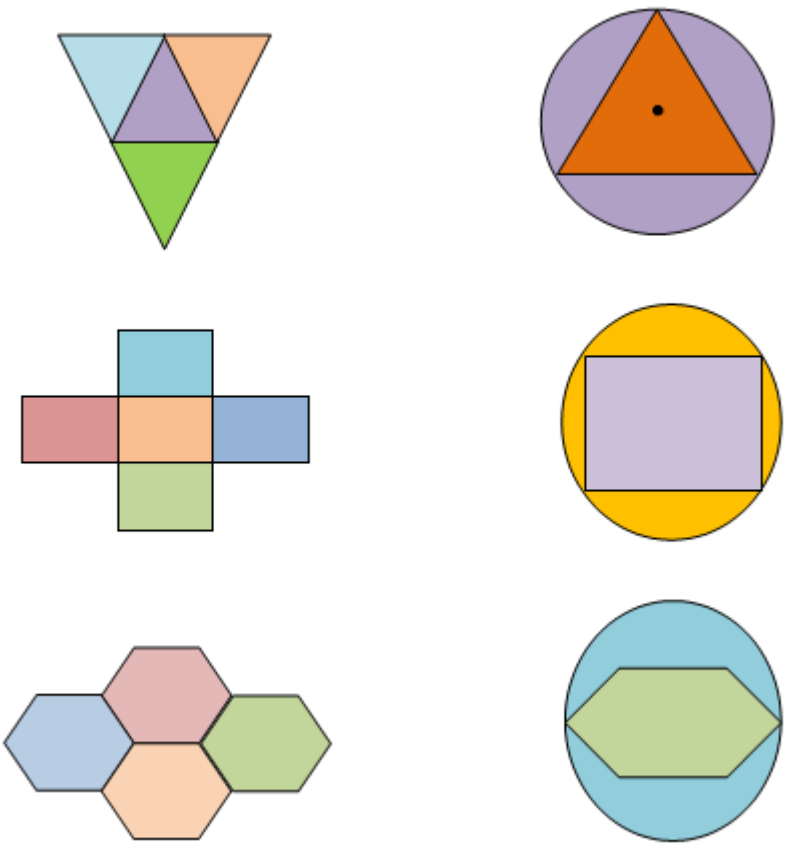


Figure 5 Eligible geometry for cell shape

Size of cell: The size of the cell can be controlled by controlling the power of antenna. The size of the cell depends on tele density and topography within a particular region (Fig 6).

Figure 6 Cell sizes in suburbs are big whereas small in densely populated cities

By the increasing number of cells the overall capacity of cellular system. But if cells are made smaller, more base stations will be requested, increasing the cost.

According to size of cell, they are names as:

1. **Macro cells**: Macro cells used for remote or sparsely populated areas. Areas of the cells are 10km or more in diameter. Used in suburbs of cities.
2. **Micro cells**: – Micro Cells used in densely populate area with diameter around 10km.
3. **Pico cells**: Covers small areas like buildings, a small volley between mountains, a tunnel etc. where the coverage from large cells is not possible. Diameter is of few meters
4. **Selective cells**: Selective cells used when 360 degree coverage is not requested. They are used to fill the gapes/holes in the coverage of a BS.

Advantage of using small cells

1. **Higher capacity**: Huge cells allow limited number of users/km2. Small cells can allow more users with the help of SDM, frequency reuse, FDM and frequency hopping.
2. **Low power Tx/Rx**: Power is a critical resource for mobile stations. Base stations can easily transmit high power so that the coverage area is increase but it has ill effects on the people surrounding it. Secondly mobile devices cannot do the same. Therefore small cells which requires only low transmit power are useful.
3. **Reduced Interference**: Small cells means less distance more the distance covered by antenna, more is the interferenc Small cells covering small distance offer less interference.
4. **Robustness**: If a base station covering large area fails, it influences connection of many people where as if it is divided into small cells, failure of BS of small cell will effect communication within a small area Small cells offer a decentralized system.
5. **Umbrella cel**ls: A heavily used road crosses an area when there are lots of microc This would required lots of handovers as people cross the road. This is handled by umbrella cells which controls all these cells

Disadvantages of using small cells

1. **Large Infrastructure**: Smaller the cells, more the cells more BS are required and a complex infrastructure to connect these base stations are also

2. **Frequent Handover**: To provide uninterrupted services in form of cells or data, proper handovers are required. The frequent of handover would be more when he cell size is small and speed of movement is also more.

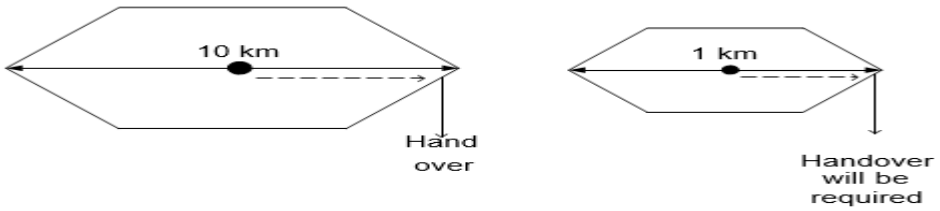


Figure 7 Frequent Handover

3. **Frequency Planning**: Smaller the cells, distribution of frequencies have to be done carefully to avoid the interference at the same time keeping in mind scarcity of frequencies.

Summary

Cellular systems were developed with the motive of replacing large high power antenna with small low power antennas so as to increase the spectral efficiency, reduce interference and allow more subscribers. It facilitates frequency reuse by repeating groups of cells called cluster in a systematic way. The ideal shape of cell should be hexagon as it covers maximum area and is closest to area of circle. Cell sizes depend on teledensity and topography of the area. Small cells increases capacity, interference and offer a decentralized robust system but require frequent handovers and are complex to implement. In the next module you will learn more about clusters, how they increase the capacity, how channel allocation is done and how handovers are handled in cellular systems.

Mobile IP

This is an **IETF (Internet Engineering Task Force)** standard communications protocol designed to allow mobile devices' (such as laptop, PDA, mobile phone, etc.) users to move from one network to another while maintaining their permanent IP (Internet Protocol) address.

Defined in RFC (Request for Comments) 2002, mobile IP is an enhancement of the internet protocol (IP) that adds mechanisms for forwarding internet traffic to mobile devices (known as mobile nodes) when they are connecting through other than their home network.

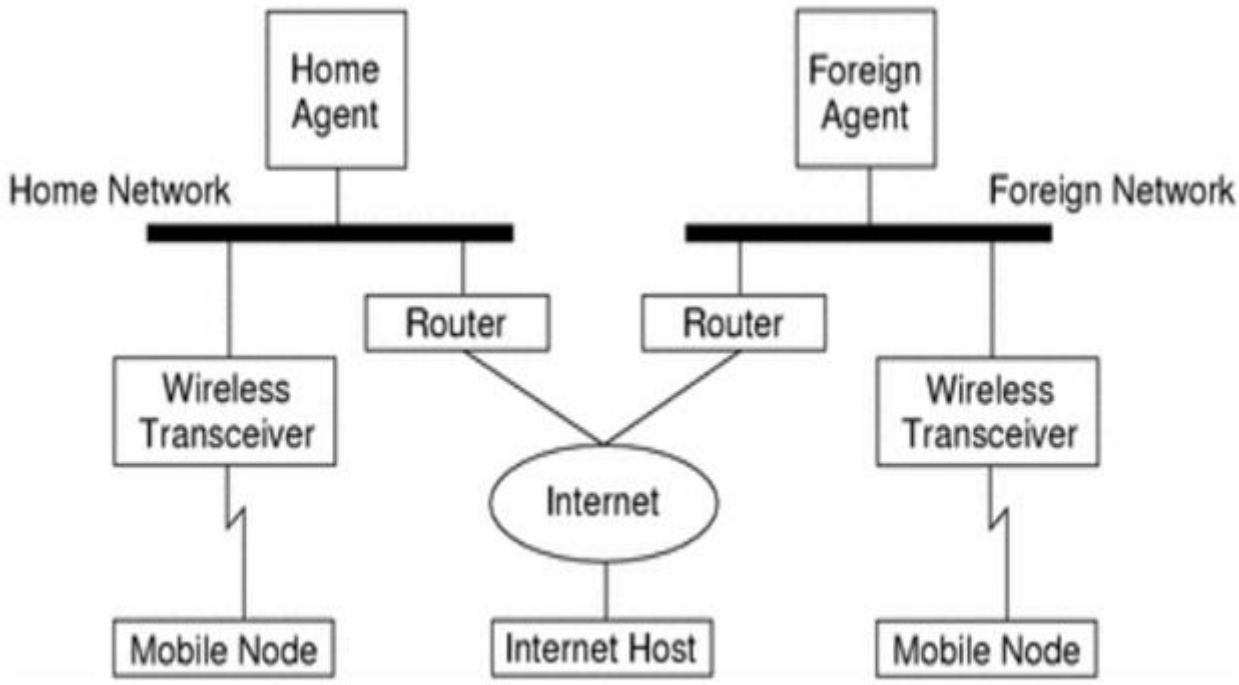


Fig: Mobile IP topology

The following case shows how a datagram moves from one point to another within the Mobile IP framework.

- First of all, the internet host sends a datagram to the mobile node using the mobile node's home address (normal IP routing process).
- If the mobile node (MN) is on its home network, the datagram is delivered through the normal IP (Internet Protocol) process to the mobile node. Otherwise the home agent picks up the datagram.
- If the mobile node (MN) is on foreign network, the home agent (HA) forwards the datagram to the foreign agent.
- The foreign agent (FA) delivers the datagram to the mobile node.
- Datagrams from the MN to the Internet host are sent using normal IP routing procedures. If the mobile node is on a foreign network, the packets are delivered to the foreign agent. The FA forwards the datagram to the Internet host.

In the case of wireless communications, the above illustrations depict the use of wireless transceivers to transmit the datagrams to the mobile node. Also, all datagrams between the Internet host and the MN use the mobile node's home address regardless of whether the mobile node is on a home or foreign network. The care-of address (COA) is used only for communication with mobility agents and is never seen by the Internet host.

Components of Mobile IP

The mobile IP has following three components as follows:

1. Mobile Node (MN)

The mobile node is an end system or device such as a cell phone, PDA (Personal Digital assistant), or laptop whose software enables network roaming capabilities.

2. Home Agent (HA)

The home agent provides several services for the mobile node and is located in the home network. The tunnel for packets towards the mobile node starts at home agent. The home agent maintains a location registry, i.e. it is informed of the mobile node's location by the current COA (care of address). Following alternatives for the implementation of an HA exist.

- Home agent can be implemented on a **router** that is responsible for the home network. This is obviously the best position, because without optimization to mobile IP, all packets for the MN have to go through the router anyway.
- If changing the router's software is not possible, the home agent could also be implemented on an **arbitrary node** in the subset. One biggest disadvantage of this solution is the double crossing of the router by the packet if

the MN is in a foreign network. A packet for the mobile node comes in via the router; the HA sends it through the tunnel which again crosses the router.

3. Foreign Agent (FA)

The foreign agent can provide several services to the mobile node during its visit to the foreign network. The FA can have the COA (care or address) acting as a tunnel endpoint and forwarding packets to the MN. The foreign agent can be the default router for the MN.

Foreign agent can also provide security services because they belong to the foreign network as opposed to the MN which is only visiting.

In short, FA is a router that may function as the point of attachment for the mobile node when it roams to a foreign network delivers packets from the home agent to the mobile node.

4. Care of Address (COA)

The Care- of- address defines the current location of the mobile node from an IP point of view. All IP packets sent to the MN are delivered to the COA, not directly to the IP address of the MN. Packet delivery toward the mobile node is done using a tunnel. To be more precise, the COA marks the endpoint of the tunnel, i.e. the address where packets exit the tunnel. There are two different possibilities for the location of the care of address:

1. **Foreign Agent COA:** The COA could be located at the foreign agent, i.e. the COA is an IP address of the foreign agent. The foreign agent is the tunnel endpoint and forwards packets to the MN. Many MN using the FA can share this COA as common COA.
2. **Co-located COA:** The COA is co-located if the MN temporarily acquired an additional IP address which acts as a COA. This address is now topologically correct, and the tunnel endpoint is at the mobile node. Co-located address can be acquired using services such as DHCP. One problem associated with this approach is need for additional addresses if MNs request a COA. This is not always a good idea considering the scarcity of IPv4 addresses.

5. Correspondent Node (CN)

At least one partner is needed for communication. The correspondent node represents this partner for the MN. The correspondent node can be a fixed or mobile node.

6. Home Network

The home network is the subset the MN belongs to with respect to its IP address. No mobile IP support is needed within this network.

7. Foreign network

The foreign network is the current subset the MN visits and which is not the home network.

Process of Mobile IP

The mobile IP process has following three main phases, which are:

1. Agent Discovery

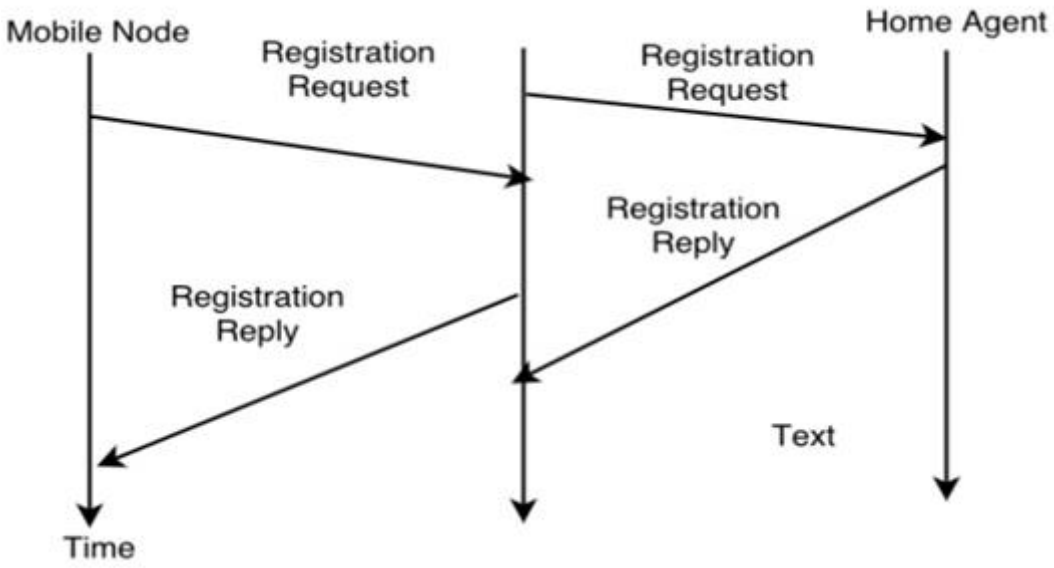
During the agent discovery phase the HA and FA advertise their services on the network by using the ICMP router discovery protocol (IROP).

Mobile IP defines two methods: agent advertisement and agent solicitation which are in fact router discovery methods plus extensions.

- **Agent advertisement:** For the first method, FA and HA advertise their presence periodically using special agent advertisement messages. These messages advertisement can be seen as a beacon broadcast into the subnet. For this advertisement internet control message protocol (ICMP) messages according to RFC 1256, are used with some mobility extensions.
- **Agent solicitation:** If no agent advertisements are present or the inter arrival time is too high, and an MN has not received a COA, the mobile node must send agent solicitations. These solicitations are again bases on RFC 1256 for router solicitations.

2. Registration

The main purpose of the registration is to inform the home agent of the current location for correct forwarding of packets.



Registration can be done in two ways depending on the location of the COA.

- **If the COA is at the FA**, the MN sends its registration request containing the COA to the FA which is forwarding the request to the HA. The HA now set up a **mobility binding** containing the mobile node's home IP address and the current COA.

Additionally, the mobility biding contains the lifetime of the registration which is negotiated during the registration process. Registration expires automatically after the lifetime and is deleted; so a mobile node should register before expiration. After setting up the mobility binding, the HA send a reply message back to the FA which forwards it to the MN.

- **If the COA is co-located**, registration can be very simpler. The mobile node may send the request directly to the HA and vice versa. This by the way is also the registration procedure for MNs returning to their home network.

3. Tunneling

A tunnel is used to establish a virtual pipe for data packets between a tunnel entry and a tunnel endpoint. Packets which are entering in a tunnel are forwarded inside the tunnel and leave the tunnel unchanged. Tunneling, i.e., sending a packet through a tunnel is achieved with the help of encapsulation.

Tunneling is also known as "**port forwarding**" is the transmission and data intended for use only within a private, usually corporate network through a public network.

MACA (Medium access control with collision avoidance)

Learning Objectives

- Understand need of medium access control is necessary in wireless environment
- Understand motivation behind separate MAC protocols for wireless environment
- Understand Hidden and Exposed terminal problem
- Understand Near and far terminal problem
- MACA algorithm to solve the problems due to CSMA/CD in wireless environment
- MAC protocols used in IEEE 802.11

Introduction

Medium Access Control, allow several users to share a common medium of communication simultaneously. An efficient MAC technique should have goal of maximum channel utilization with minimum interference and collisions and provide reliable point-to-point or multipoint connection between different devices on medium. The common MAC algorithms existing in wired networks cannot be

simply replicated in wireless networks due to situations like Hidden and Exposed terminal problem and near and far terminal problem. Due to these problems, the existing MAC algorithms like CSMA/CD fails in wireless scenario. This module discusses these problems and presents the motivation behind need of specialized MAC algorithms in wireless scenario. The module also discusses some of the alternative algorithms to CSMA/CD which works well in the wireless environment like MACA, MACAW and virtual carrier sense.

Medium Access Control

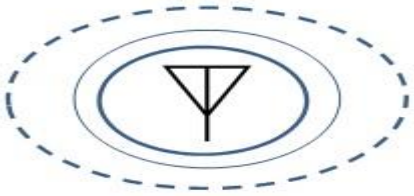
Medium Access Control is protocol of data link layer. It is used to regulate the control of access among different users without or very less collisions. The transmission medium in wireless communication is air or atmosphere which is shared by multiple users or subscribers. In such a situation, simultaneous access by multiple users can lead to collisions. A good MAC algorithm should minimize the number of collisions hence increasing the throughput at the same time maintain fairness among the users.

The perfect analogy to this situation is Highway where more than one vehicle can arrive at same or different points of time. If the traffic on highway is not controlled in an efficient and systematic way, accidents can occur. Therefore different traffic control mechanisms should be applied. Similarly in wireless networks access to the transmission media should be controlled using different modulation and multiplexing techniques. Medium access control is one of the two sub-layers of Data Link layer of ISO/OSI reference model. The biggest challenge of medium access control is that wireless devices should transmit without interfering with the signals of neighboring wireless devices.

Need for Specialized MAC in wireless Communications

Let us now understand whether the standard MAC algorithms used in wired networks, can be replicated in wireless scenario. For this we first understand the basic CSMA/CD **Carrier Sense multiple access with collision detection** used in IEEE 802.3 wired networks. It works as follows:

- Sense the medium
    - Analogy: In a round table conference, different people participate and communicate. They sense through their eyes and ears to find if anyone is talking. If anyone senses someone talking, he remains quite i.e. “Listen before you talk”.
- If free, transmit else wait
    - Analogy: If it is found that no one is talking, press the button to initiate talk and start talking. For that time, others will sense the medium to be busy.
- Continuously listen to the medium for any collisions
    - Analogy: Observe if the speech has coincided with some other person.
- Stop in case of collision detection and sends a jamming signal
    - Analogy: The person stops talking and repeats its previous speech. The scheme works well in wired scenario. In wired communication, all devices are connected through wire and the strength of the signal is uniform throughout the wire hence all the devices can listen to the medium and detect the collision if it exists. But in wireless scenario there are many other issues which will not allow CSMA/CD to function properly. They are:



a) In wireless environment, signal propagates in omni directional way in all directions and the strength of the signal decreases inversely as square of distance from the transmitter.

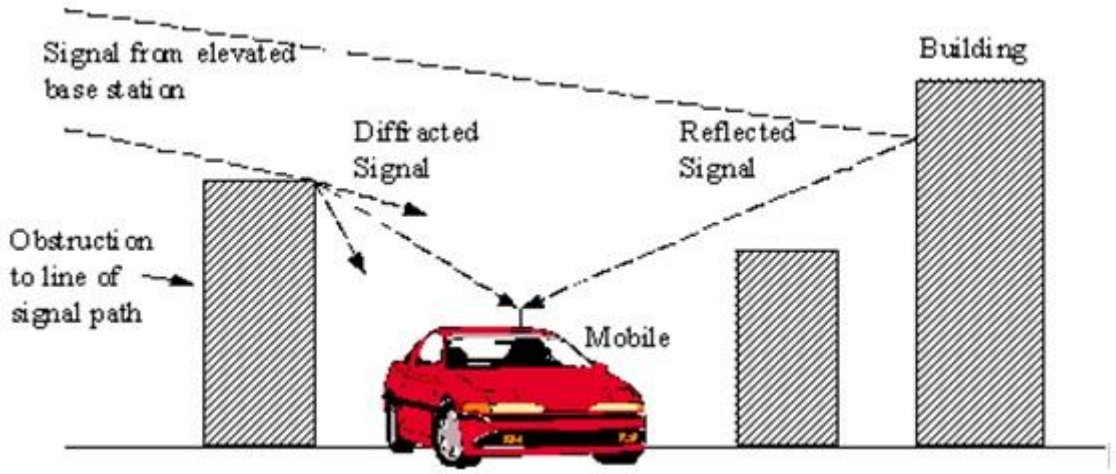


Figure 1: Various effects degrading the quality of the signal

b) The objects in the way from sender to receiver also offers various effects like reflection, scattering, diffraction leading to multipath propagation and many other undesirable effects which degrades the signal as shown in 1

c) The receiving power is very much less than transmitting Wireless transceivers can't send and receive on the same channel at the same time, so they can't detect collisions.

d) Hidden and exposed terminal problem and near and far terminal problem are other situations which fails the use of CSMA/CD in wireless networks.

Let us understand the hidden and exposed terminal problem and near and far terminal problem typical to wireless networks.

Hidden Terminal Problem

As the strength of the signal transmitting from a wireless device, decreases with distance, the transmission is limited to a certain area known as transmission range after which the signal diminishes. A device can listen only to those devices which are in its range others are said to be **hidden** from it.

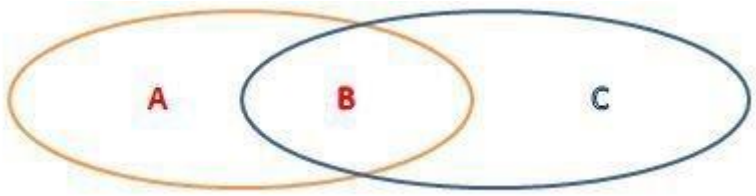


Figure 2: Hidden terminals A and C

In Figure 2, A is in transmission range of B. B is in transmission range of C but A&C are in not in transmission range of each other. A and C are said to be hidden from each other.

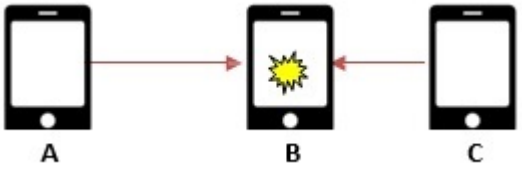


Figure 3: A and C are hidden from each other

Following sequence of events (Fig. 3) illustrate why CSMA/CD fails in hidden terminal situation:

- A wish to transmit, senses the medium, finds it to be idle.
- A transmits. C cannot hear transmission of A
- C wish to transmit, senses the medium, also finds it to be idle.**(Carrier sense fails)**
- Collision occurs at B
- Collision not heard by both A&C **(Collision detection fails)**
- Both Continue transmitting
- A and C are said to hidden from each other

Hidden terminal problem decreases the throughput because of collisions

Analogy: Two people want to talk to a third person but they cannot hear each other. So when one person is talking, other cannot sense it, finds the way free, he also talks and both the conversation coincides. Further they do not even know that there is collision of speech and continue talking.

Exposed terminal problem

- B is transmitting to A, C can hear it
- D is in transmission range of C
- C wish to transmit to D
- C sense medium, finds it busy because of transmission of B
- C waits causing delay(Fig. 4)

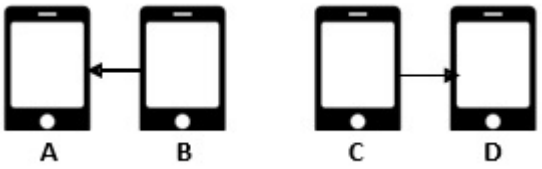


Figure 4: Exposed Terminal Problem

Near and far terminal problem

Figure 5: Near and Far Terminal Problem

To illustrate this problem, let us view a situation when three terminals A, B and C are such that B is farther from A and nearer to C as shown in Fig. 5. In this situation when A and B both transmit with equal

power, but the strength of signal of A decreases as it reaches C due to distance from C. At the same time power of B is strong and it drowns signal of A. This problem is more acute in Code division multiplexing because all stations transmit with same power yet they are at different distance from the receiver. Hence the near terminals drown signal of far terminals. Therefore regular power regulation is required in CDM. E.g. UMTS regulates power 1,500 times in a second.

Analogy: People sitting in the same hall speak at the same time. The receiver gets all the signals. But the people who are sitting near to the receiver drowns the signal of the people sitting far apart.

Multiple Access with Collision Avoidance: A Solution to Hidden and exposed terminal problem

MACA is a scheme proposed by Karn in 1990 which solves the problem of hidden and exposed terminal problem. In this scheme instead of sensing the medium, consent of receiver is taken before transmitting. The receiver if free, signals transmission following which the sender transmit. This is accomplished by the use of two fixed length (32 bytes) additional signaling packets called request to send **(RTS)** and clear to send **(CTS).** They **are also called control packets.**

RTS: A control packet used by the sender to seek permission from the receiver to transmit. It contains name of sender, receiver of user data and duration of transmission.

CTS: A control packet used by the receiver to grant permission to the sender to transmit. It contains name of sender, receiver of user data and duration of transmission.

Whenever any station wants to transmit, it sends an RTS to the receiver. If receiver is free, it signals the transmission by sending CTS. Sender sends the packet and receiver on receiving the packet, sends the acknowledgement.

How MACA solves Hidden terminal problem

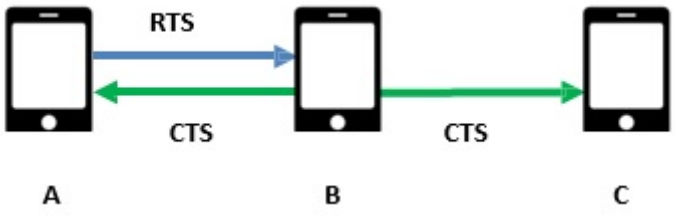


Figure 6: A and C are hidden from each other

A, B, C are 3 terminals where A and C are hidden from each other.

A wants to transmit to B. It broadcasts RTS to B (Figure 4). When broadcasted, the RTS will be heard by all the stations in its range so the RTS is heard by B but not by C (A&C are hidden) B sends CTS. C hears CTS (B&C are in transmission range).C is not allowed to transmit anything for the duration mentioned in RTS. **Hidden terminal problem is solved.**

How MACA solves exposed terminal problem

B wants to transmit to A. B sends RTS to A. (RTS heard by C)

A sends CTS, CTS not heard by C.
C understands it is out of range with A.
C now starts transmission with D.
Exposed terminal problem is solved.

Limitations of MACA

MACA offered a three way handshake only.
MACA did not provide specifications about parameters What are RTS, CTS packet sizes ?
Collisions of RTS may occur when more than one station send RTS at same time. In that case none of the stations gets CTS.

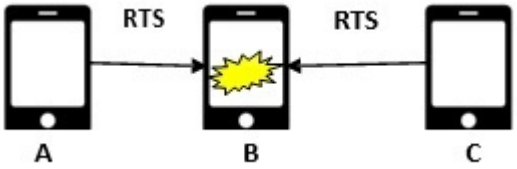


Figure 7: Collision of RTS from A and C at B

The overhead is affordable when data packets are large but in case of short and time-critical data packets this overhead can be quite expensive.

MACAW

It is refined and extended MACA. Used Information sharing to achieve fairness. It supports

- Four-way handshake (reliable, recover at MAC layer)
- Five-way handshake (relieve exposed terminal problem)
- RRTS (unfairness) It works as follows
- Sender sends Ready-to-Send (RTS)
- Receiver responds with Clear-to-Send (CTS)
- Sender sends DATA PACKET
- Receiver acknowledge with ACK
- RTS and CTS announce the duration of the transfer
- Nodes overhearing RTS/CTS keep quiet for that duration
- Sender will retransmit RTS if no ACK is received
- If ACK is sent out, but not received by sender, after receiving new RTS, receiver returns ACK instead of CTS for new RTS

Time State diagram of MACA algorithm

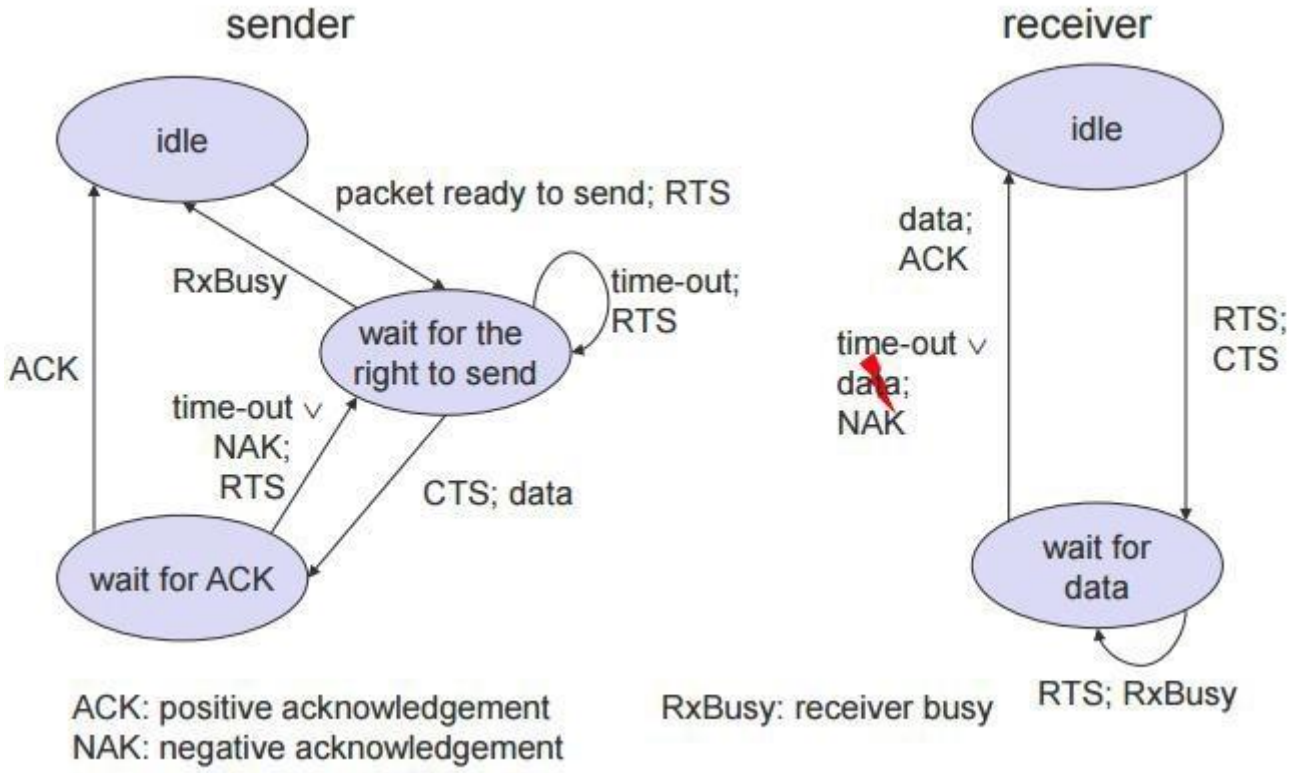


Figure 8: State diagram of MACA Algorithm: Picture courtesy: Mobile communications by Joschen Schiller

Figure 5 shows state diagram for sender in the MACA. The sender can be either in ***idle*** state, waiting for CTS state or ***waiting for acknowledgement*** state. When sender wishes to transmit, it sends RTS and moves in to ***wait for CTS*** state. If the receiver is idle, it will send CTS. Sender will transmit and move to **wait for ACK state** and receiver will go to **wait for data state**. If the receiver is busy or time-out for CTS has occurred, sender again goes to idle state. On receiving the data, receiver gives ACK, sender goes back to idle. In case of NACK, sender again sends initiates transmission and waits for CTS.

MACAW with five-way handshake

- Sender sends Ready-to-Send (RTS)
- Receiver responds with Clear-to-Send (CTS)
- Sender sends DATA SENDING (DS)
- Sender sends DATA PACKET
- Receiver acknowledge with ACK
- RTS and CTS announce the duration of the transfer
- Nodes overhearing RTS/CTS keep quiet for that duration

Multiple Access Techniques in wireless networks

Learning Objectives

- Understand need of medium access control in wireless environment
- Understand multiple access schemes used in cellular systems in four dimensions
- SDMA: Understand the need and implementation of multiple access in space domain
- FDMA: Understand the need and implementation of multiple access in frequency domain
- TDMA: Understand the need and implementation of multiple access in time domain

- CDMA: Understand the need and implementation of multiple access using pseudo- random code

Introduction

Medium Access Control is protocol of data link layer. It is used to regulate the control of access among different users without or very less collisions. The transmission medium in wireless communication is air or atmosphere which is shared by multiple users or subscribers. In such a situation, simultaneous access by multiple users can lead to collisions. A good MAC algorithm should minimize the number of collisions hence increasing the throughput at the same time maintain fairness among the users.

Medium Access Control, allow several users to share a common medium of communication simultaneously. An efficient MAC technique should have goal of maximum channel utilization with minimum interference and collisions and provide reliable point-to-point or multipoint connection between different devices on medium. The common MAC algorithms existing in wired networks cannot be simply replicated in wireless networks due to situations like Hidden and Exposed terminal problem and near and far terminal problem. Due to these problems, the existing MAC algorithms like CSMA/CD fails in wireless scenario. In the previous module these problems were discussed and the motivation behind need of specialized MAC algorithms in wireless scenario was presented. In the previous module we discussed some of the alternative algorithms to CSMA/CD which works well in the wireless environment like MACA, MACAW and virtual carrier sense. In this module we present different multiple access techniques used in cellular systems and other wireless communication systems. The multiple access can be regulated in four dimensions namely SPACE, FREQUENCY, TIME and CODE. This module presents various ways in which these multiple access techniques are implemented and used.

Multiple Access Schemes for Cellular systems

Let us consider two real life situations which demand for an efficient multiple access mechanisms and different ways in which they are handled.

Analogy: Let us assume two analogous situations to wireless communications.

- First is a Highway which is used by many vehicles. Maximum number of vehicles should use the highway without collisions.
- Second is Hall of people whose conversations are to be regulated so that maximum people can talk without interference.

The goal is common: Collisions should be less and utilization should be maximum.

Highway Problem: Maximum vehicles should pass without collision. Some of the solutions are:

- Divide the common highway into number of lanes **(Space division)**. Every vehicle can use a single lane. But the problem is that number of vehicles which can pass without collision is limited to the number of lanes. So its not an efficient utilization of the available space.
- The subsequent solution is that we allow more than one vehicle to pass the same lane provided they move with different speeds **(Frequency division)** hence avoiding collisions. Here the problem is that vehicles moving with same speed in same lane would collide
- To solve above stated problem, vehicles moving with same speed in same lane should arrive at different time **(Time division)**.

Hall of people Problem: Second situation is conversation between people sitting in a hall. Some of the ways in which more and more people can talk without interference are:

- One way to reduce interference amongst their conversation to divide people among separate groups and make them sit at a substantially far distance from each other **(Space Division)**.
- To avoid interference amongst the group, they can be restricted to talk with different pace and at different time **(Frequency and time division)**
- Another way which allows all the people to speak at the same place with same pace at same time is conversation in different languages. That is all people speak in different languages with the constraint that languages should be separable from each other; there should be very less correlation between the languages. For example Hindi and Gujarati are very much similar to each other therefore signals received from two people speaking in Hindi and Gujarati are not separable from each other. At the same time care should be taken that people close to the receiver should not drain the signals of people sitting far away.

Motivated from the above scenario, Multiple Access for Wireless Communications is regulated in four dimensions i.e. Space, Frequency, Time and code. Accordingly, there are four schemes:

- **SDMA: Space Division Multiple Access**
- **FDMA: Frequency Division Multiple Access**
- **TDMA: Time Division Multiple Access**
- **CDMA: Code Division Multiple Access**

In next four sections, we will discuss these schemes one by one

Space Division Multiple Access

In this scheme the different users sharing the bandwidth are physically separated from each other by allocating them a separate space. The users should be sufficiently separated from each other so as to avoid interference. For eg Cell formations in cellular systems, FM radio transmission implemented using Space division multiplexing. In cellular systems the geographical area is divided into different spaces called cells. Each cell is associated with a unique frequency/set of frequencies. Same frequency is allocated to different cells if they are at a sufficient distance from each other to avoid interference. Hence facilitating frequency reuse (Fig 1).

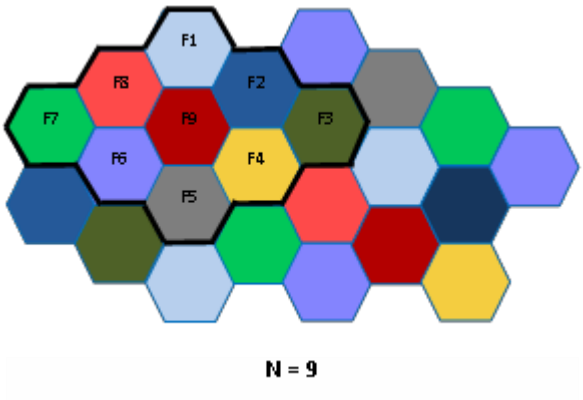


Figure 1: Space division multiple access by Cell formations

Additional separation is provided by the use of directional antennas. The cells are further divided into 3 sectors of 120 0 each by the use of sectorized antennas (Fig. 2). New technologies like smart antennas or

adaptive arrays use dynamic beam forming to shrink signals into narrow beams that can be focused on specific users, excluding all others. This technology further decreases interference and increases the capacity. Satellite dish antennas are highly directional that allow transmit signals to many zones on the earth's surface using duplicate frequencies for different zones (Fig. 3).

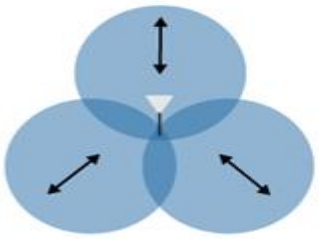


Figure 2: Sector formation

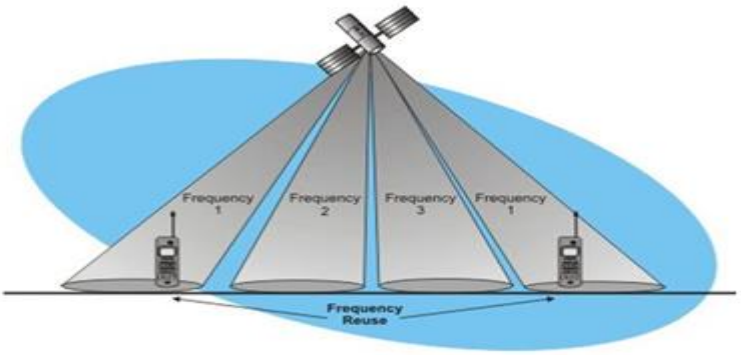


Figure 3: Directional Antennas

Polarization division multiple access (PDMA), is a variant of SDMA in which signals are separated by using different polarizations of the antennas. Two different signals then can use the same frequency, one transmitting a vertically polarized signal and the other transmitting a horizontally polarized signal. The signals won't interfere with one another even if they're on the same frequency because they're orthogonal and the antennas won't respond to the oppositely polarized signal. Separate vertical and horizontal receiver antennas are used to recover the two orthogonal signals. This technique is widely used in satellite systems

Limitation of SDMA

This scheme clearly offers wastage of bandwidth because the channel remains with the user whether it uses it or not as in the case of radio stations. Another problem arises when more than one user wish to transmit in the same space. Like more than one station wants to transmit in the same city with same frequency. Therefore SDMA is never used alone instead it is always used in conjunction with FDMA and CDMA.

Frequency Division Multiple access

In this scheme, the bandwidth is divided into number of non-overlapping frequency bands with guard spaces between them. Each band can be used by single user. Guard spaces are needed to avoid adjacent channel interference (Fig 4) .

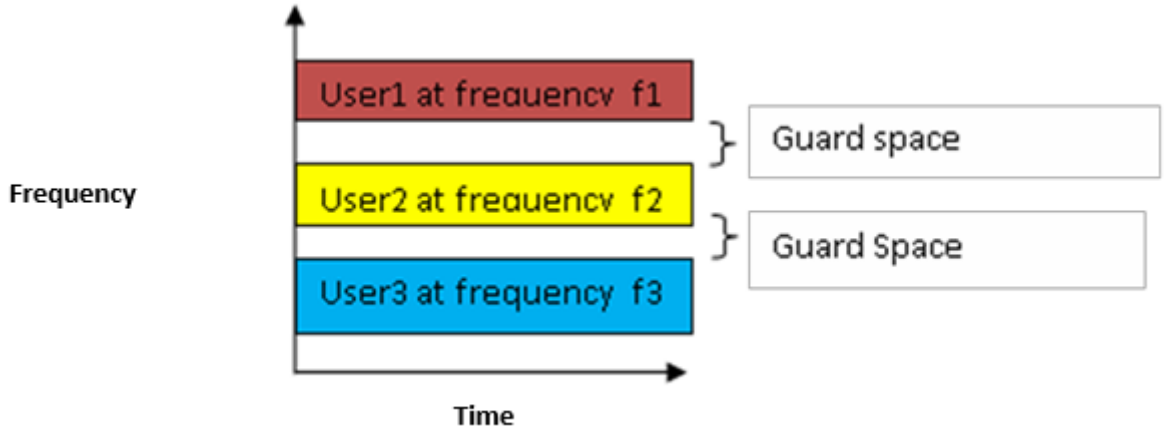


Figure 4: Pure FDMA

The allocation of frequencies in a FDMA system can be fixed or dynamic.

Fixed FDMA

In this scheme, the frequencies allocated to the channels are fixed and remain unchanged. E.g. radio stations. The limitation of this scheme is bandwidth wastage when the user is not using the frequencies allocated to it.

Demand Driven FDMA

The channels are allocated according to the load and demand. The frequencies assigned to the channels can be same for the whole time **(Pure FDMA)** or it can vary at different interval of times (FDMA + TDMA). FDMA + TDMA allows transmission to be on one of the frequency channels for some time and then jump to other frequency channel after some time (Fig. 5). This scheme is called **frequency hopping** and is used in Bluetooth, IEEE 802.11, GPS and other many other technologies. The pattern of channel usage is known as hopping sequence which both sender and receiver has to agree upon.

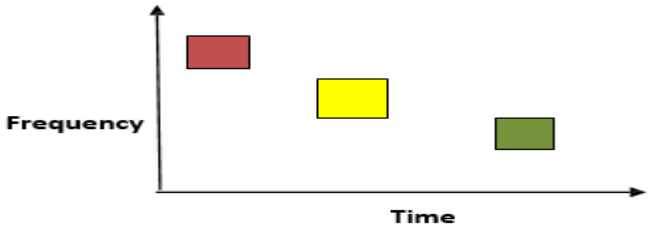


Figure 5: Frequency Hopping

If B= Total Bandwidth; G=width of Guard bands and ch is bandwidhth of single radio channel Then Number of channels in FDMA system is given by

N=B-2G/ch

FDMA in GSM-900

GSM uses FDMA with an additional Frequency division duplexing scheme. Duplexing is required because GSM being a phone network, requires the transmission in both the directions simultaneously. The two

directions are from mobile station to base station and base station to mobile station. To facilitate full duplexing the available bandwidth is divided in such a way that are two separate channels for transmission in both directions. The frequencies allocated for transmission from mobile station to base station are known as **UPLINK** and from base station to mobile stations are known as **DOWNLINK.** The **UPLINK frequencies range from 890.2 MHz to 915 MHz** and **DOWNLINK frequencies** range from 935.2 **MHz** to 960 **MHz** . The uplink and downlink bands are separated by a guard band of 20 MHz.

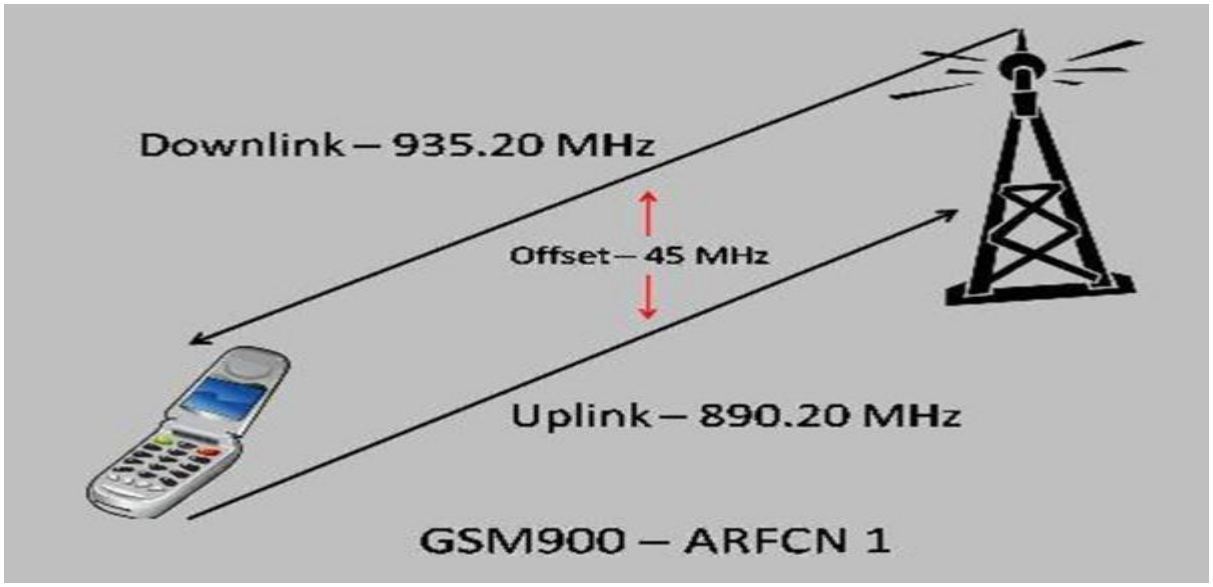


Figure 6: Full duplexing

Fu= 890 MHz + 0.2 * n MHz; where n is the number of uplink channel.

Both uplink and downlink bandwidth is divided into 124 channels of 200 kHz each with guard bands of 100 kHz(Fig. 7). The uplink and downlink frequencies are calculated as :

Fd = 935 MHz + 0.2 * n MHz; where n is the number of downlink channel Fd = Fu + 45 MHz

In practice GSM uses combination of FDMA and TDMA.

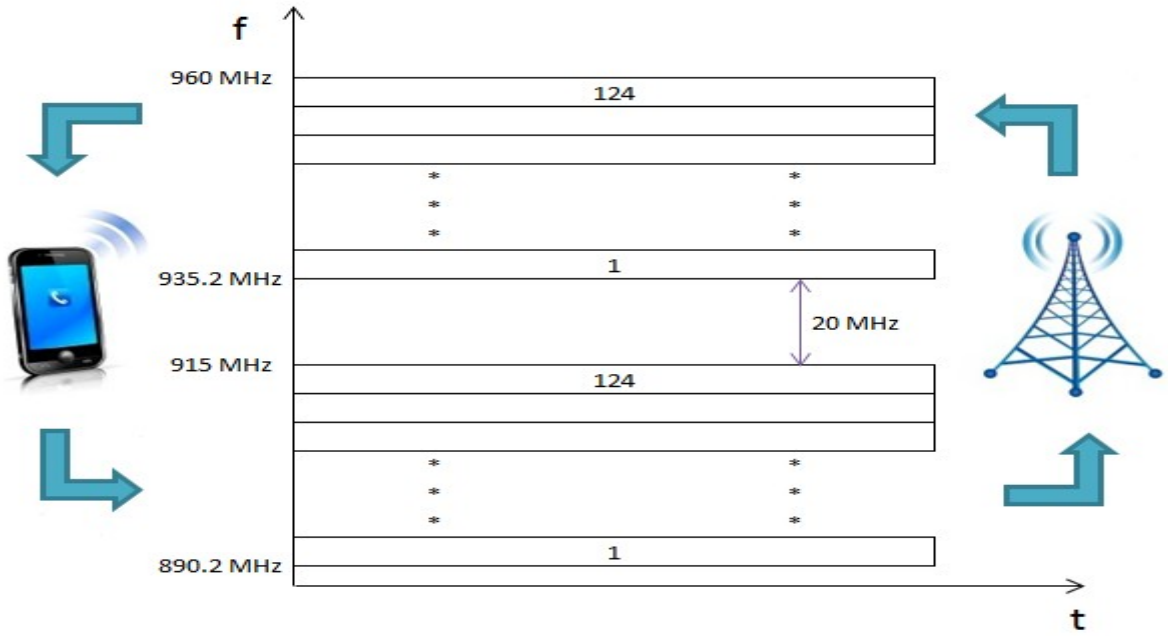


Figure 7: FDMA with FDD in GSM

Features of FDMA

- It can be used both for digital as well as analog
- Receiver should tune in to same frequency as the sender
- In digital transmission, the channel is not needed for the whole time. During that time the frequency remains unused
- FDMA demands highly efficient filters in the radio hardware, contrary to CDMA and TDMAFDMA is devoid of timing issues that exist in TDMA
- As a result of the frequency filtering, FDMA is not prone to the near-far problem that exists in CDMA.
- All users transmit and receive at different frequencies because every user receives an individual frequency slot.

Time Division Multiple Access

In this scheme, multiple users can transmit using same frequency but at different time. Each user transmit using common frequency band but at different non-overlapping time slots. The entire bandwidth is available to the user but only for a finite time.

Frequency

Time

In the figure above user 1 gets the complete frequency for time t1, user 2 at time t2 and so on. Data transmission in TDMA is not continuous but occurs in bursts. This results in low battery consumption since the subscriber transmitter can be turned OFF when not in use. Bandwidth can be supplied on demand to different users by concatenating or reassigning time slot based on priority. TDMA is used in GSM, IS-136, SS7,satellite systems etc. The two variants of TDMA are:

Fixed TDMA

The slot sizes are fixed. If T is the time in seconds and N is the number of users, then slot size=T/N. In this scheme if a user does not transmit during the allocated time slot, the corresponding bandwidth is wasted.

Dynamic TDMA

A scheduling algorithm dynamically allocates time slots according to the traffic demand of each.

Code Division Multiple Access

In this scheme, all the users use the complete bandwidth for whole time. Separation is achieved by assigning each channel its own code. The situation can be compared when people at a common place speak at same time with same frequency. Each pair speaks its own language and for others who might overhear the conversation, it may appear as noise. Thus CDMA also comes with built-in security. The languages should be clearly separated. For eg. Hindi and Gujarati won't work but Hindi and French will serve the purpose. In code division multiple access the code separates the users as well as spreads the signal converting a narrow band signal to a broad band signal (Fig. 9).

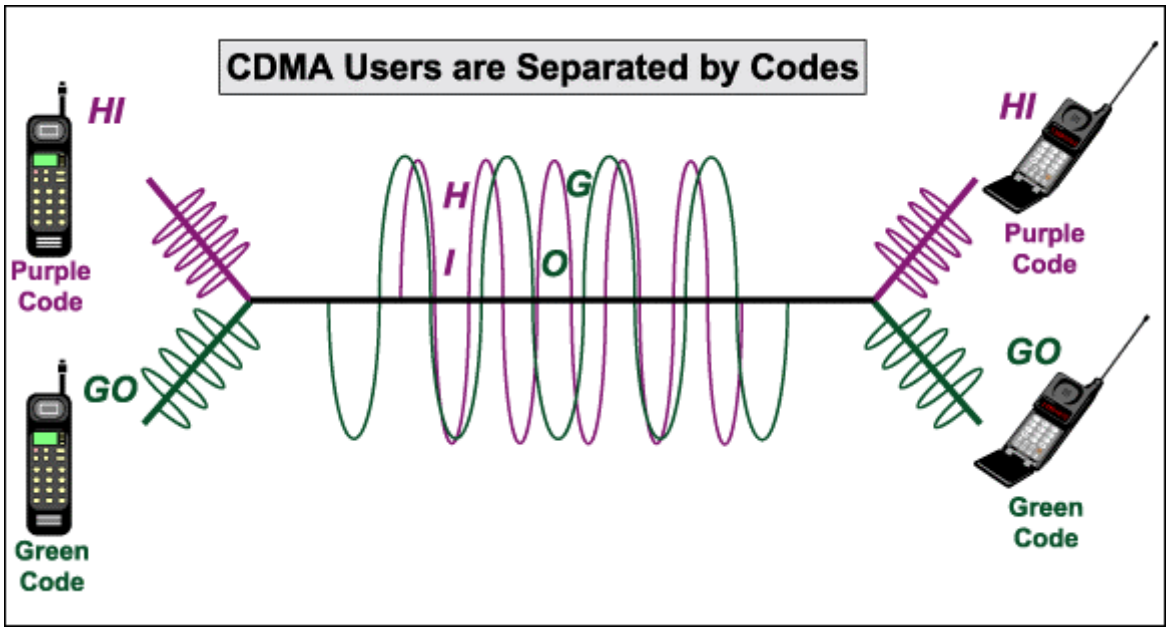


Figure 9: Example of CDMA

A good code should have these two properties:

A code for a user should have a good auto-correlation ie the absolute value of inner product of the code with itself should be a big value.

For eg. The Barker code (+1,-1,+1, +1, -1, +1, +1, +1, -1, -1, -1) has a good autocorrelation=11. This code is used in IEEE 802.11

The code should be orthogonal with other codes ie the absolute value of inner product of the code with other code should be zero. Eg. (2,5,0) and (0,0,7) are orthogonal because (2,5,0) .(0,0,7)=0+0+0=0.

Advantage of CDMA

- Built in Security because only the receiver who knows the code can interpret it for others it is noise. That is why the code is also called pseudo random noise

- Protection against interference and tapping

Disadvantage of CDMA

- Huge Code Space is required
- Complexity at receiver side because the receiver has to know the code and separate it with user data
- Receiver should be precisely synchronized with the sender
- The strength of all the signals should be same otherwise the nearby signals will drain the far signals. Hence regular power regulations are required.

Summary

- Efficient MAC algorithms are a necessity for wireless communications
- Multiple access in wireless can be regulated in four domains
- SDMA separates users spatially like in cellular systems. Simple to implement but not efficient utilization of bandwidth
- FDMA provides different channels to different users statically or dynamically
- TDMA provides different time slots to different users. It can be static allocation or demand driven
- CDMA utilizes whole bandwidth whole time where separation is achieved using different codes

DHCP and Cellular IP

- Introduction
- Need of Dynamic Configuration
- DHCP message format
- Address acquisition state
- Cellular IP
- Network of CIP
- Summary

Introduction

Internet service provider (ISP) provides a block of Internet Protocol (IP) addresses to the organization. There are many ways to do address assignment in the network of any organization. IP address assignment is done in the organization by network administrator. Each computer connected with the TCP/IP network required to get an IP address of the network. Also, other information such as address of router, address of name server and address mask for the network are required for communication within a network.

There was a one general protocol Bootstrap Protocol (BOOTP) which allows a host machine to obtain all the information automatically during the startup process. When a computer is booting on the network, the operating system send BOOTP messages to the network to get an IP address. BOOTP was implemented using User Datagram Protocol (UDP). It was initially used for UNIX disk-less workstation to get the location of network to assign IP address. Some network card manufacturers have embedded this protocol in the Basic Input Output System (BIOS) of the network interface card which allow direct network booting.

Later on the standardized network protocols Dynamic Host Configuration Protocol (DHCP) was developed which provide functionality of leasing of an IP address. DHCP is used to assign dynamic IP addresses to the host of network.

Need of Dynamic configuration

In the current era of Internet most of the people are using laptops or mobile phone for wireless internet connections. ISPs have a continually changing set of customers, and portable laptop computers with wireless connections make it possible to move a computer from one place to another place very easily and quickly. Whenever any new computer is connected to the network it sends the request to the server for an IP address. The server selects one of the addresses from the range of IP addresses available and assigned the address to the computer based on the list provided by ISP.

DHCP allows three types of address assignment manual configuration, automatic configuration and dynamic configuration.

- In manual IP configuration technique an administrator or manager configures a specific address for each and every computer connected to the network.
- While in automatic configuration an administrator or manager allows a DHCP server to assign a fixed address when a computer is attached to the network first time.
- In dynamic configuration method an IP address is assigned to the machine for a short period of time e. based on the loan period a server will assign an IP address. Dynamic IP addressing is most powerful aspects of DHCP. When a client contacts a DHCP server, the client sends an identifier, usually the client's hardware address. The server uses the client's identifier and the network to which the client has connected to determine how to assign the client and IP address.

DHCP message format

DHCP client and server both are using the same message format which is shown in the figure

1.1. Each and every field has their own usage and requirements. Table 1.1 describes the value which is available in the message format.

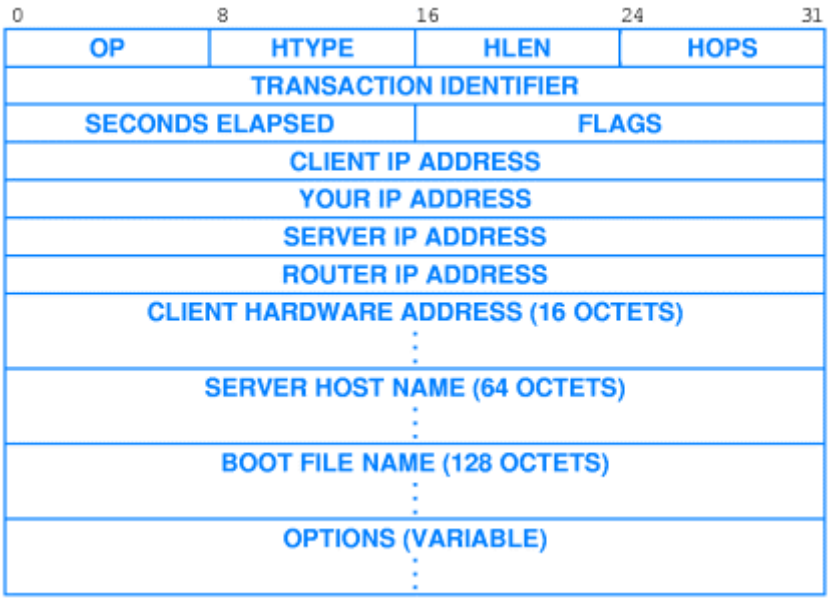


Figure 1.1 Format of DHCP message [Reference: Internetworking with TCP/IP Principles, Protocols and Architecture, 5th Edition Douglas E. Comer]

Field Name	Description
OP	Specifies which operation is performed (1) Request or (2) Reply

HTYPE	Hardware Type (1 for Ethernet)
HLEN	Hardware Length (6 for Ethernet)
HOPS	Initially Zero (0), Server will increment Hops count value.
TRANSACTION ID	Unique integer ID used to match reply with request.
SECONDS	Specifies the number of seconds before the client computer started.
FLAGS	It controls the server response via unicast or broadcast.
CLIENT IP ADDRES	Client's IP address if it is known by client.
YOUR IP ADDRESS	Server return the client's IP address if previous field contain ZERO
SERVER IP ADDRESS	Client specifies the server IP address if server is specific.
ROUTER IP ADDRESS	IP address of router if it is known by client machine.
CLIENT HARDWARE ADDRESS	MAC address of client machine specify by client
SERVER HOST NAME	Name of the specific server from where client want response.
BOOT FILE NAME	Client specifies the name of booting file for small diskless machine.
OPTIONS	Uses Type-Length-Value style encoding.

Table 1.1 DHCP message field and description

DHCP Address Acquisition State

During the dynamic IP address assignment process client machine can be in one of six states. The transitions of states are shown in figure 1.2. First of all a client machine is booted and it enters into the INITIALIZE state. From this state a client is discovering about the server from where it can get an IP address. So, client will broadcast the message using DHCPDISCOVER message type. Now all the servers available on the local network will receive the message and only one or more server will give reply depending on the availability of IP addresses. Here, server will send DHCPOFFER message to the client machine. If all the servers are busy than client will not get any response. Client will be in the SELECT state so all the offers of servers can be view by client and if it finds feasibility than send request to server using DHCPREQUEST message type. Here, a client will be in the REQUEST state for the waiting of a response from the server. Server will give DHCPACK message for confirming the request and client will be in a BOUND state for the given period of time.

After 50% of lease period expire client can send request to the server for extending the lease period. Depending on the availability of IP addresses and pre-request for an IP address of any client at the server side server will give DHCPACK or DHCPNACK message. If client receive DHCPACK that indicate server has extended the time period of lease else DHCPNACK indicate server has not extended the lease period. Whenever a lease period is extended after 50% the client will be in the RENEW state.

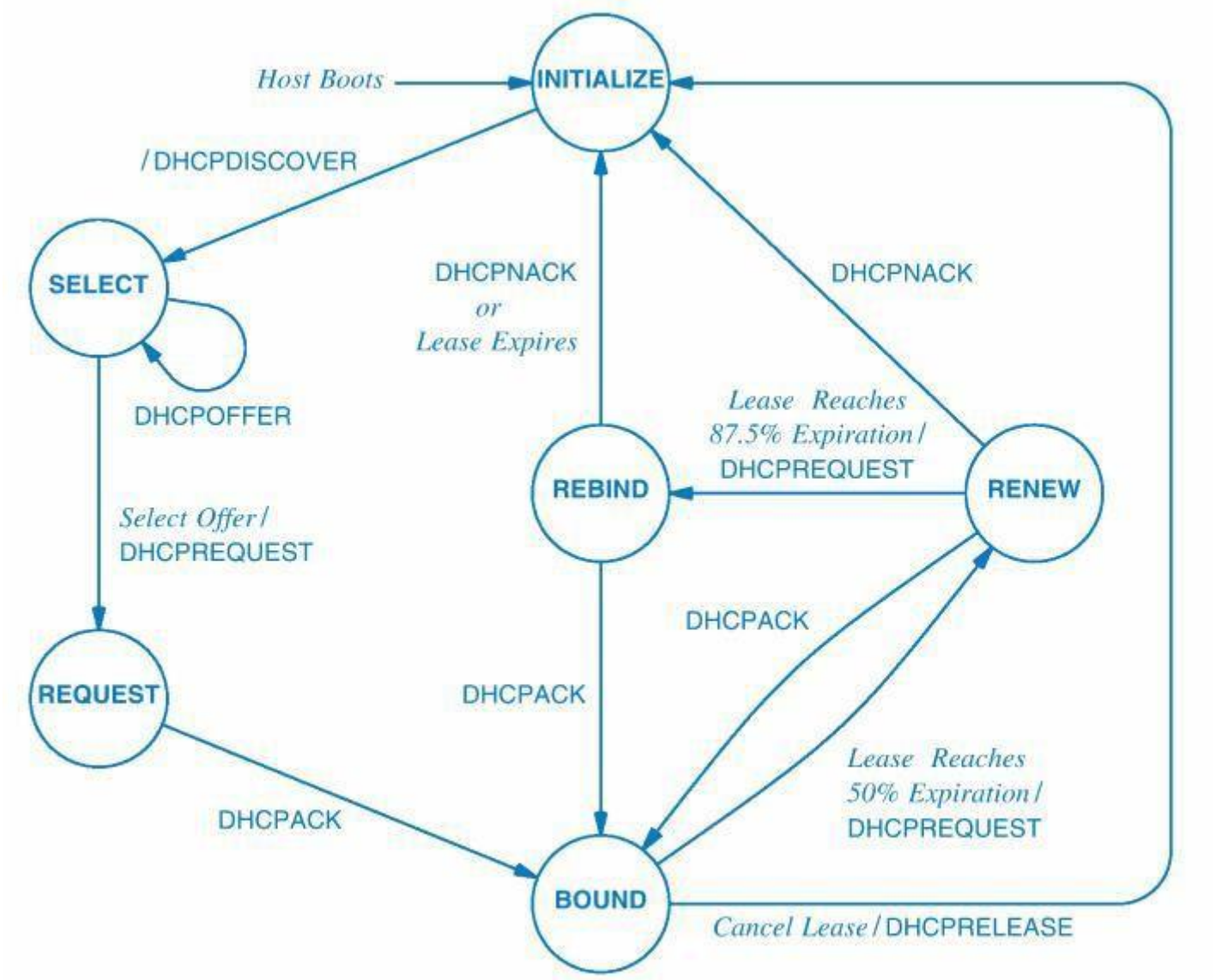


Figure 1.2 DHCP address acquisition state diagram [Reference: Internetworking with TCP/IP Principles, Protocols and Architecture, 5th Edition Douglas E. Comer]

After completion of 87.5% of lease period the client can send DHCPREQUEST for extending the lease period. Again the server will check the availability of IP addresses and pre-request for an IP address of any client at the server side than server will give DHCPACK or DHCPNACK message. If DHCPACK message comes than client will be in REBIND state else for DHCPNACK message client machine has to use an IP address for the limited period only.

After completing of 100% lease period client will be in INITIALIZE state again. In between if client want to release the lease period or cancel the lease than he/she will send DHCPRELEASE message to the server so, the state of BOUND will be change into INITIALIZE state.

So, after the first timer expires, the client can move into renew state. If a second timer expires before renewal state than client can move into rebind state. If the final timer expires before a lease has been renewed, the client stops using the IP address and returns to the initial state to obtain a new IP address.

Cellular IP

IP technology is improving day by day and significant innovations are happening in the world for Internet Protocol such as Voice over Internet Protocol (VOIP), Mobile IP, and Cellular IP.

VOIP is used for sending communication and multimedia sessions over Internet network. Mobile IP allows any mobile device user to move from one place to another by maintaining their permanent IP address. It is also suited for client/server mechanism where security is must. Cellular IP (CIP) is a standard provided by Internet Engineering Task Force (IETF). It was first proposed in the year 2000.

CIP is used to work with wireless network where internet work communication is possible. CIP can accommodate large no of users by maintaining distributed Paging and Routing caches. There is no need for new packet formats, address allocation or even for encapsulations.

Four fundamental design principles of the cellular internet protocol are:

1. location information is stored in distributed data bases,
2. location information referring to a mobile host is created and updated by regular IP datagram originated by the said mobile host
3. Location information is stored as soft state
4. Location management for idle mobile hosts is separated from location management of hosts that are actively transmitting or receiving data

Network of Cellular IP

A client machine wants wireless interface connection which is changing frequently. There is a mechanism which indicates the delivery of packet from one machine to another machine with high portability. There are two mechanisms for handover in cellular IP hard handover and semi-soft handover. During the handover mechanism quality of signal matters so to avoid the degradation of service semi-soft handoff technique is used. Cellular IP can interwork with Mobile IP to support wide area mobility.

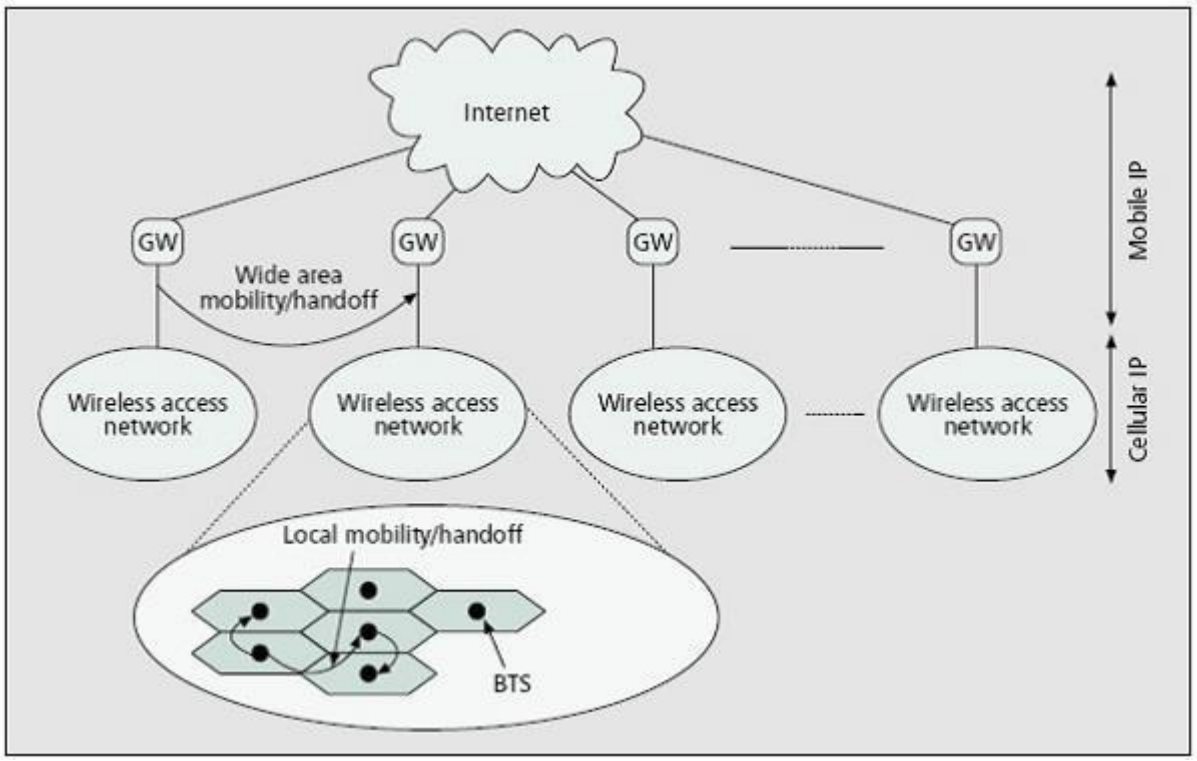


Figure 1.3 Cellular IP and Mobile IP Network

Summary

- The dynamic host configuration protocol is always beneficial to those networks where number of IP addresses is limited and host machines are scale in and scale out.
- DHCP allows a server to assign an IP addresses automatically or dynamically.
- To use DHCP, a host machine becomes client in the network. Client starts three timers after obtaining an IP address from the server.
- DHCP can be used in multiple subnets.
- Cellular IP can work with all Mobile IP networks that support wireless communication.
- CIP can accommodate large no of users by maintaining distributed Paging and Routing caches.

Indirect TCP

Learning Objectives

- Understand the problems associated with wireless or mobile networks
- How TCP handles packet loss
- Why standard TCP cannot be applied to wireless networks
- Need of specialized protocols for wireless networks
- Goals of specialized protocols for wireless networks
- I-TCP architecture and functioning
- Advantages and limitations of I-TCP

Introduction

Upto now, we have seen implication of mobility on lower layer protocols like data link and network layer. For example for data link layer we have seen that hidden and exposed terminal problems fail the application of CSMA\CD which works well in wired networks. We discussed some of the protocols specially designed for MAC layer of all layer like MACA, MAAW, DAMA, PRMA and so on. Similarly for network layer, there is mobile IP technology to support the mobility yet providing seamless connectivity but support for mobility only at lower layers is not enough. The application required mobility support at transport layer also. Mobile computing is a paradigm that provide services to the users on "anywhere" "anytime" basis. In such a scenario mobile hosts expects same services as offered to the fixed hosts hence a transport layer protocol is required for mobile domain also.

Mobile network itself is a hostile environment therefore of TCP and UDP, TCP is best choice as it is connected oriented transport layer oriented transport layer protocol that provides in order delivery of packets to the receiver. But TCP undergoes the congestion control mechanisms in the form of "Slow start" during packet loss assuming that there cannot be any other for packet loss other than congestion. But in wireless or mobile environment there are many reasons for the packet loss like higher bit error rate, effect of disconnections, improper handoff's may also lead to packet loss therefore. Replication of TCP there would cause serious drop in through put.

Therefore specialized TCP is requirement for wireless environment. In this module we will understand functioning of standard TCP and how it handles packet losses. Then we will see why we cannot replicate standard TCP in wireless networks. We will learn about need and goals of specialized protocols for

transport layer in mobile computing environment. There have been several protocols proposed in this respect.

In this module we will learn about protocol I-TCP and in subsequent modules we will learn about more protocols.

Regular TCP

It is a connection oriented protocol. It has been tuned to work for networks composed of wired links and stationary hosts. It is a reliable protocol. Reliability is maintained by observing acknowledging and resending lost packets. The number of packet sent is regulated by increase or decrease size of the window. TCP sender uses the cumulative acknowledgements sent by the receiver. To adapt with the problems of wired link the window size is changed. The wired data transmission is through fiber optics copper wires etc. and works without introducing transmission errors. If a software is mature enough it will not drop packets or flip bits so a packet drop cannot be due to hardware or software error. Therefore the reason for packet loss is some overload in transmission path i.e. congestion. It assumes congestion to be the main reason of packet loss.

Congestion is a condition when the buffers of the router are filled, the sum of input rates of packet is greater than output link. The router cannot forward the packets fast enough and drop the packets. It is lost. The receiver notices a gap in stream of incoming packets. It continues to acknowledge all the in-sequence packets up to the missing one. Sender notices missing acknowledgement and assumes packet loss. Therefore TCP perform following congestion control scheme. The scheme has 3 parts:

- Slow start
- Congestion Avoidance
- Fast Retransmit

Slow Start

It works as follows:

- Sender calculates congestion window for the receiver
- The transmission starts with congestion window (cwnd=1)
- Sender sends one packet; If acknowledgement received increase window size by 1 i.e. 2
- Sender now sends two packets; If acknowledgement received for two packets, window is increased by 2 one for each acknowledgement i.e. 4
- For each received acknowledge TCP increase the window exponentially only upto a threshold. Once congestion threshold is achieved, congestion phase starts.

Congestion Avoidance

- As soon as congestion threshold is received, window is increased linearly by adding 1 to the congestion window each time acknowledgement is received
- it continues to increase its window size linearly until
- it reaches the receiver's maximum advertised window or
- packet loss is observed due to time-out
- duplicate acknowledgement is received for same packet

How TCP understands condition for packet loss

TCP measures how long acknowledgements take to return back and use it to determine which packets have reached the receives and provides reliability by retransmitting lost packets. For this purpose, it maintains a running average of round trip delay and an estimate of expected deviation from this average. If current delay is longer then average by more than 4 times expected deviation. TCP assumes packet loss and retransmits.

It also assumes packet loss if duplicate acknowledgements received. This is because receiver acknowledges higher in order sequence number. If it receives out-of-order packets, it generates acknowledgements for same highest in order sequence number. If it receives out-of-order sequence number and that result in duplicate acknowledgement. In this case also it assumes packet loss but not due to congestion due to error on wireless link.

Fast Retransmit

If sender receives 3 duplicate acknowledgements then it activates fast retransmit algorithm. It assumes data segment indicated by duplicate acknowledgements is lost and retransmits lost segment. No time is wasted waiting for timeout in order for retransmission. Fast retransmission is documented in RFC 2001 & RFC 2581

Illustration

- Sequence#1→Sent : ACK for 1 received
- Sequence#2→Sent : ACK for 2 received
- Sequence#3→Sent : ACK not received
- Sequence#4→Sent: received at receiver but ACK for 3 sent by receiver
- Sequence#5→Sent: received at receiver but ACK for 3 sent by receiver
- Sequence#6→Sent: received at receiver but ACK for 3 sent by receiver

Number of duplicate acknowledgement is 3!!!

Packet 3 is lost!!!

- Only sequence number 3 is retransmitted

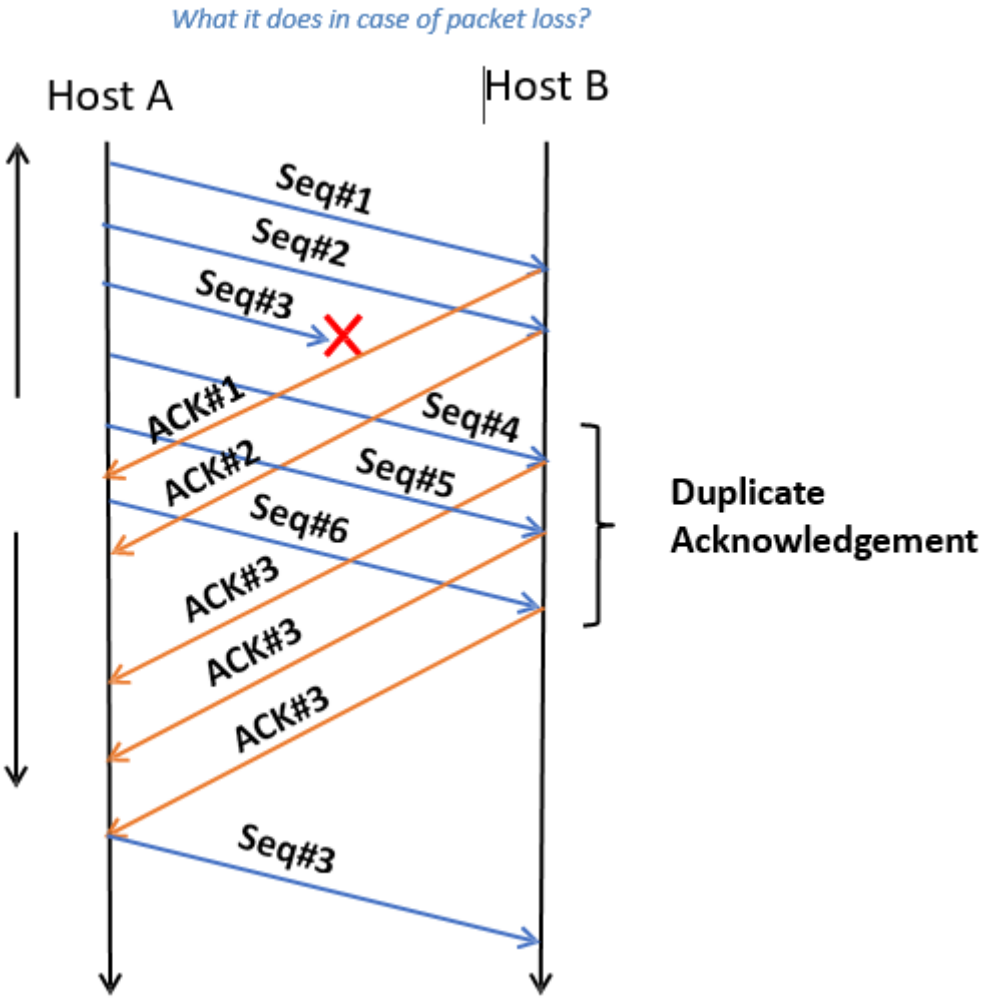


Figure 1 Fast retransmit

TCP reacts any packet loss by

1. Dropping threshold value into half the current window or 2 whichever is larger
2. Resetting the transmitting window size to 1 activating slow start
3. Reset retransmission timer to a back off interval that doubles with each consecutive timeout

Can TCP be applied for mobile hosts as well?

TCP was not designed for mobile host so it is not expected to perform well in wireless/mobile network when a packet is loss, regular TCP assumes that it is due to congestion. Whereas in mobile networks there can be packet losses unrelated to congestion. Some of them are:

- **High Bit Error rate**: Wireless links suffers from a high bit error rate high (can be of the order of 10-5).BER causes packets to be corrupted resulting in loss of TCP data segments or acknowledgements. Missing acknowledgements triggers time out at sender causes re-transmission with window size to be one which drastically reduces the throughput. FEC at data link may handle high BER but it would be wastage of wireless bandwidth when correction is not necessary
- **Black outs:** When the mobile move from one cell to another call or service is transferred from old base station called Hand off. Hard Handover- there is brief disconnection known as black out period. Any packet sent during this period might be lost.
- **Call Blocking:** Disconnections due to call blocking; a condition when mobile station do not get any channel due to unavailability of the channels. The disconnection period can be of few seconds and can also last for 1 minute. These disconnections also result in loss of packets and

acknowledgements. Since these disconnections are very lengthy the FEC scheme is ineffective cell sizes

- **Disconnection due to small cell latencies:** When the cell sizes are small to accommodate more and more users and increase the capacity results in small cell latencies and hence frequency disconnection.
- **Signal Blocking:** When radio signals are blocked due to buildings and other objects in the environment
- **Power Scarcity:** Mobile computer are battery operated hence power required is more

In such a scenario if TCP triggers congestion control procedures, it will significantly reduce the throughput. There has been lot of research in the methods to improve TCP in wireless environments. The goal of protocols:

- It is not possible to change the entire TCP because it is the base on which the entire internet is based. Therefore proposed protocol should not change standard TCP
- The protocol should not go into slow start when it is not genuine to go
- Isolate the mobility related problems from existing network protocols when fixed host and wireless station to communicate.

Indirect TCP(I-TCP)

It is a split connection protocol proposed by Bakre and Badrinath in 1995. It splits an End-to End TCP connection between fixed host and mobile host into two separate connections at an intermediate called MSR (Mobility support router).MSR can be an access point, foreign agent in case mobile IP, entry point of network like IWF in GSM, GGSN in case of GPRS.

- The link between fixed computer and AP is supported by regular TCP and between MSR and mobile node optimized TCP is used

The MSR acts as proxy between fixed computer and wireless node. It is seen as wireless node to fixed computer as fixed computer to wireless node.

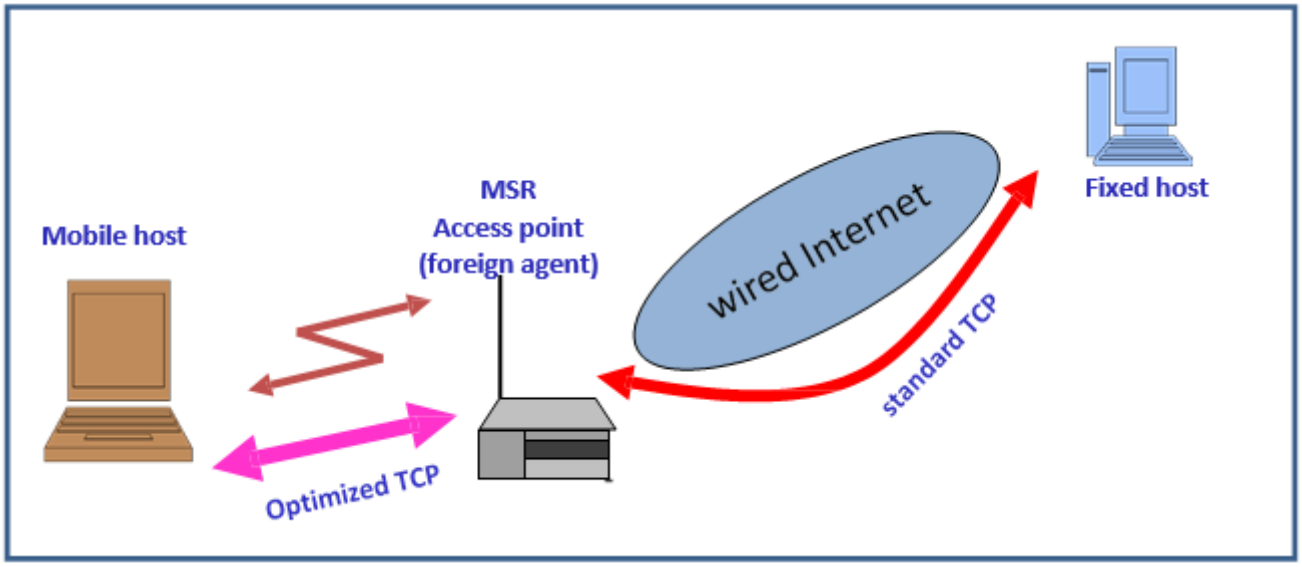


Figure 2 I-TCP

Packetgoesfromfixednodetowirelessnode

- Sender sends packet to MSR
- It acts as proxy; buffers the packets ; gives acknowledgement to the receiver
- MSR then forwards packet to mobile host

- If packet is lost on wired link, standard TCP handles it
- If packet lost on wireless link, MSR re-transmits with low round trip time

Packet goes from wireless node to fixed node

- Mobile node sends the packet
- MSR acknowledges it and forwards to the fixed computer
- If packet is lost, it is retransmitted on wireless with lower round trip time

Handovers

- At the time of handover old proxy forwards the packets to the new proxy because acknowledgements are already sent for those packets
- New MSR then forwards those packets to the mobile host Current state of TCP (sequence numbers, address, ports)reflected by socket should

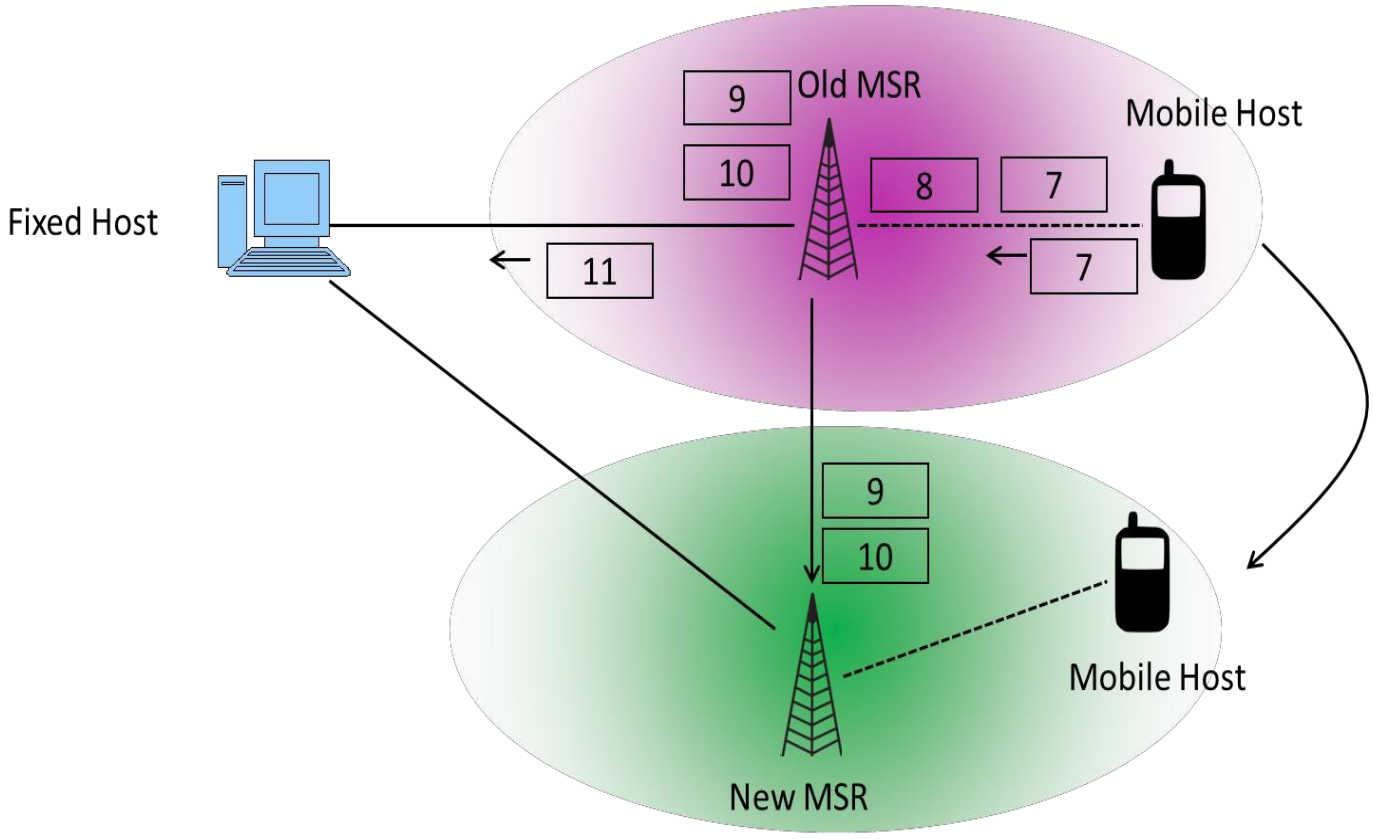


Figure 3 Packets forwarded during handover

Software Components The components for the software to support I-TCP

1. **MH Side:** I-TCP can be accessed as a transport protocol by application running on mobile host using special library calls. They are similar in interface and function to the socket calls made by an application using register end – to – end TCP
2. **MSR Side:** In the proposed protocol, most of the functionality for supporting I-TCP connection lies with the MSR. If acts as actual bridge connecting the wired and wireless parts of the connection. It consists of a user level UNIX process pumping data from one part of connection to the other. Hand off support for I-TCP connection is implemented in MSR kernels

Advantages

1. No need to make changes in TCP protocol used by computers in fixed network or any other device which do not need optimization
2. Transmission error on wireless links do not propagate in fixed network
3. The optimized TCP is used only on wireless link without affecting stability of internet

Disadvantages

1. Loss of end – to – end semantics packets already acknowledgements before being delivered
2. Whole scheme fails if MSR crashes
3. Increased handover tendency
4. MSR should be a trusted entity. Serious security threat if MSR is compromised

Snooping TCP

Learning Objective

- Understand need for optimized TCP in wireless networks
- Understand limitations of I-TCP
- Snooping protocol and its design in order to remove limitations of I-TCP
- Advantages and limitations of Snooping TCP

Introduction

The wireless communication is characterized by following properties:

1. Limited bandwidth
2. High latency rate
3. High bit error rate
4. Temporary disconnections
5. User Mobility and handoffs

All these factors affect the protocols for wireless communication. There have been attempts to modify the existing protocols in wired networks to comply with wireless environment. Like MACA for data link layer to replace CSMA/CD and Mobile IP as a substitute for IP in network layer. To support mobility, protocols in higher layers like TCP in transport layer also needs to be modified or some alternate protocols need to be proposed. In this module we will understand the performance of transport layer protocols in wireless communication scenario and study a protocol called Snooping TCP which is adapted as per characteristics of wireless communications.

Need for optimized transport layer protocols

Standard TCP is a well-established transport layer protocol for wired links and fixed hosts. TCP provides reliable transmission by re-transmission on time-out and handles end to end delays and packet losses efficiently. Assuming that the bit error rate over wired links is low, TCP assumes congestion to be the only cause of packet loss and reacts by reducing the window size before re-transmission of packets. This mechanism is known as **slow start**. The scheme works well in wired networks but in wireless networks, high error rate of links, intermittent connectivity, improper hand offs are the other reasons for packet loss. In such a situation, if TCP goes into slow-start, it will result in reduction of bandwidth utilization, poor throughput and high delay hence the performance will be degraded. To handle this, lots of research has been going on to improve the performance of TCP over wireless links. All of the researches believe that TCP is the only appropriate model for wireless networks since many network applications are built on top of TCP therefore it is not possible to change the entire protocol. Hence it is necessary to propose optimized versions of standard TCP maintaining its performance. The optimized versions should not tend to make changes on fixed hosts which mean that it should not be aware of the errors on the wireless link. Many TCP protocols for wireless networks have been proposed, one of them is I-TCP. In this module, the design and functioning of snoop protocol and how it overcomes the limitations of I-TCP has been described.

Limitation of I-TCP

In I-TCP, the foreign agent or router or the base station splits the TCP connection into 2 parts. Between fixed computer and base station, standard TCP is used and between base station and wireless host, optimized protocol specific to wireless links is used. The limitations of this scheme are:

- **Loss of end-to-end semantics of TCP**: The acknowledgements are given by the foreign agent as soon as it receives the packets .The acknowledgements are received by the sender which makes it understand that the packet has been received even before the receiver gets it.
- **Centralized Proxy**: The base station or the foreign agent acts as a proxy forwarding the
- Packets in both the directions therefore whole scheme fails if foreign agent is crashed
- **Security Threat:** The foreign agent acting as a proxy receives all the packets so it should be a trusted entity

Snooping TCP

The Snooping TCP was proposed by Balakrishnaet. al. in 1995.This approach is designed in such a way so as to overcome the end-to-end semantics loss in I-TCP. Snooping TCP offers a transparent design which leave the end-end connection of TCP intact. Basic idea is to buffer packets close to mobile host and perform local re-transmission in case of packet loss.The scheme works as follows:

- Foreign agent buffers the packet until it receives acknowledgement from the mobile host.
- Foreign agent snoops the packet flow and acknowledgement in both the directions.
- If the foreign agent does not receive acknowledgement from the mobile host or receives duplicate acknowledgements, it assumes either the packet or the acknowledgement is lost
- Foreign agent directly retransmits the packet from its buffer.
- Foreign agent also maintains its own timer for retransmission of buffered packet in case it is lost on wireless link
- If the foreign agent crashes, a timeout at fixed host will work and cause
- re-transmission and the scheme falls to standard TCP. The foreign agent in contrast to forwarding the packets as in case of I-TCP, just buffers the packets intended for mobile host
- To maintain transparency foreign agent does not acknowledge the packet to the fixed host(END-TO-END Semantics is maintained)

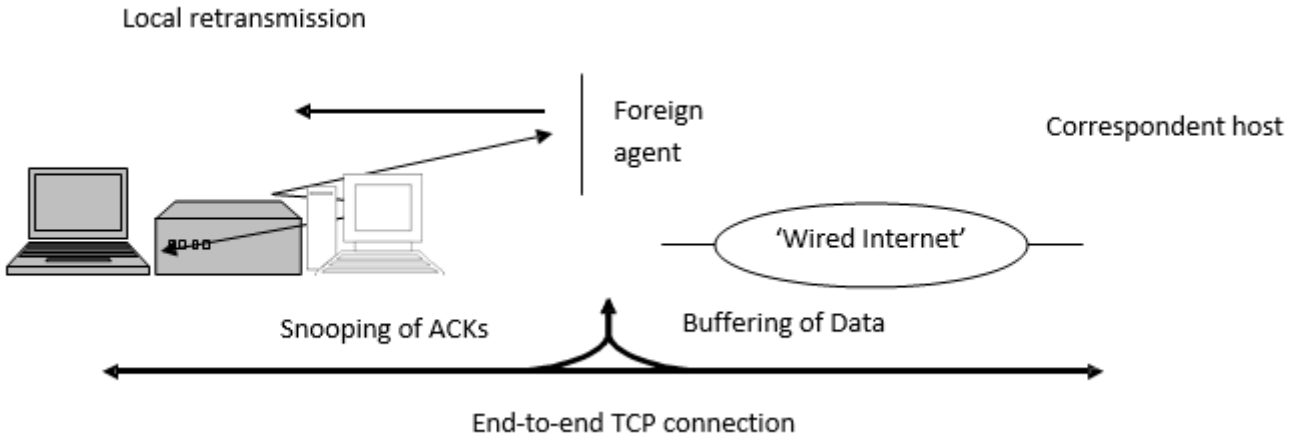


Figure 1: Architecture of Snoop Protocol

Design of Snoop Protocol

The snoop protocol is implemented as follows:

- The TCP connection on the wired link between foreign agent and fixed host is standard TCP

- The TCP connection on the wireless link between foreign agent and mobile host is
- optimized TCP
- The foreign agent snoops the flow of packets and acknowledgements and caches the packets towards the mobile host
- The scheme suggests some changes on the routing code of the foreign agent when there is packet flow from fixed host to the mobile host
- For data transfer from mobile host to fixed host additional mechanisms at mobile hosts are required
- No changes are required at the fixed host

Additional mechanisms required to implement Snoop Protocol

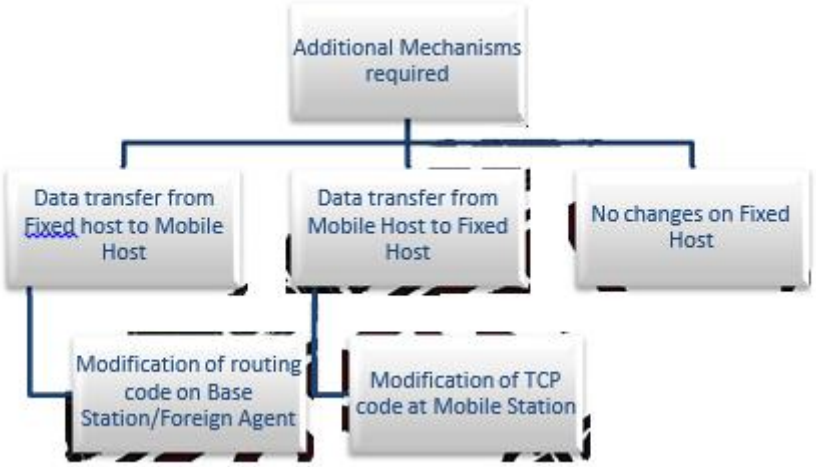


Figure 2: Additional mechanisms required by Snoop Module

Data transfer from fixed host to mobile host

The change in the routing code of (BS)is addition of snooping-module. The function of module is to monitor the packets passing through the connection in both the direction.The snoop module caches TCP packets sent from fixed host (FH) that are not acknowledged by mobile host (MH)i.e. the packets are cached until they are acknowledged by mobile host. The snooping module keeps track of the entire acknowledgement sent from the mobile host. If a packet loss is detected by the Base Station (BS) either by arrivals of duplicate acknowledgement or by local timeout,snoop module re-transmits the lost packets to mobile host(MH) which has been cached.

The snoop module has two procedures Snoop Data and Snoop acknowledge. Snoop data process and cache packets towards mobile host and Snoop acknowledge process acknowledgements coming from mobile host (MH) and drive local re-transmission.

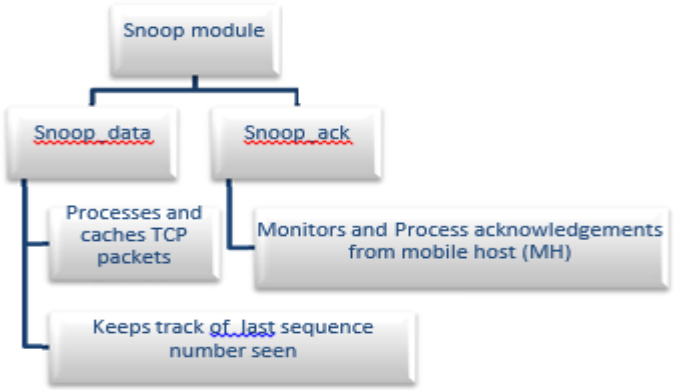


Figure 3: the procedures of Snoop Module

Snoop data (): This module processes packets arriving from the fixed host. A TCP packet is uniquely identified by the sequence number of its first byte of data and size. The BS/AP keeps track of past sequence number seen for connection. Appropriate actions are performed by this module depending on sequence number of the packet and status of the cache. The packets arriving at the Base Station (BS) from the fixed computer are one of the three types mentioned below:

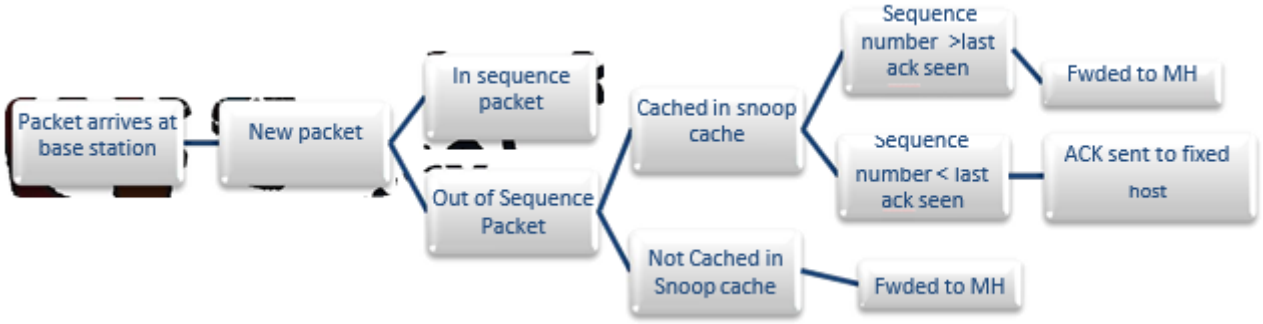


Figure 4: Types of packets arriving at base station and action taken accordingly

- **In sequence Packet:**The packet arrives in the normal in sequence. The packet is added to snoop cache and forwarded to the mobile host (MH). A time stamp is placed on one packet per transmitted window in order to estimate the round trip time of wireless link.
- **Out of sequence Packet:** A packet whose sequence number is out-of-order. There is possibility that this packet has been cached or not been cached.
- **Out of sequence Packet that has been cached:**This happens when dropped packet causes timeouts at the sender. Packet after TCP sender fast re-transmission arrives at the sender. The base Station (BS) now sees whether this packet is greater or less than the last acknowledged packet and different actions are taken accordingly.
- If the sequence number is greater than last acknowledged packet it would mean that the packet didn't reach the mobile host (MH) earlier and hence it is forwarded.
- If the sequence number is less than the previously acknowledge, it is assumed that packet is received by mobile host (MH).One thing to do is to discard the packet but it is not a wise thing to do because the acknowledgement from mobile host (MH) to fixed host (FH) might have lost due to congestion a TCP acknowledgement correspondence to last acknowledgement seen at base station (BS) is generated with source address and port number of the mobile and sent to fixed host (FH).
- **Out of sequence Packet that has NOT been cached:**This case happens when packet was lost due to congestion or it is delivered out of order by the network. This packet is forwarded to the mobile host and on additional information regarding packet re-transmitted by sender is associated at base station (BS).
- **Snoop Acknowledge ():** This module monitors and processes the acknowledgement sent by mobile host (MH) and different types of actions are performed depending on type and number of acknowledgements received.

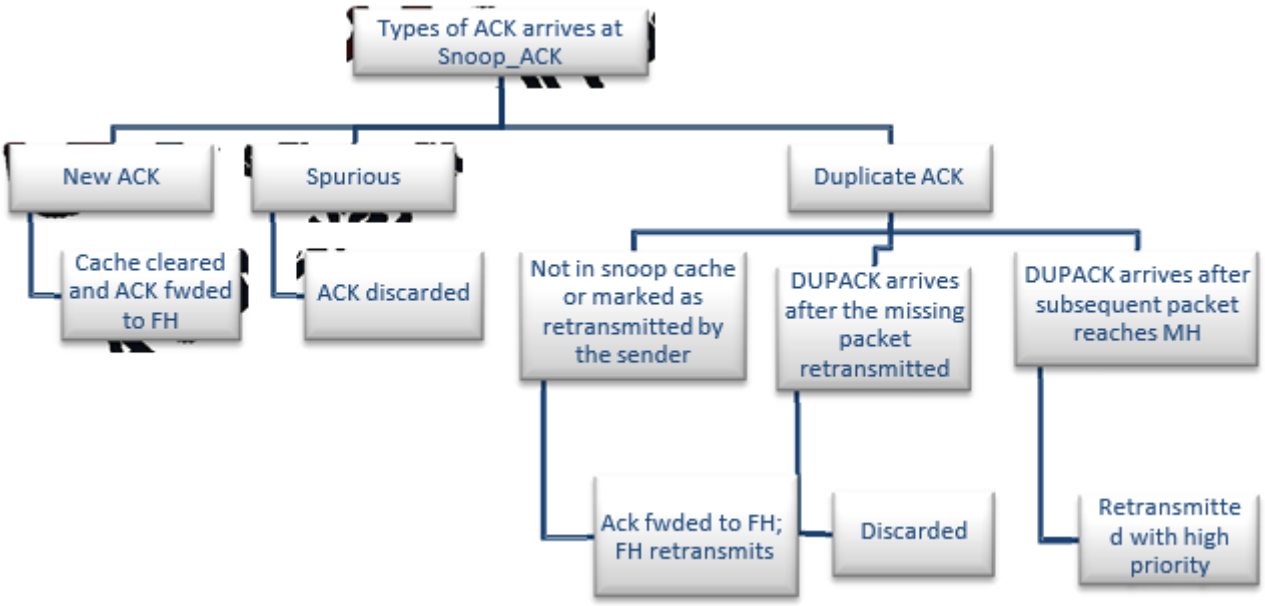


Figure 5: Types of acknowledgements received and action taken

- **New acknowledgement**: Indicates that the packet is received at the mobiles host indicating the link is error free. The packet sequence number arriving at the receiver will be increased. On receiving this acknowledgement, the snoop cache is cleaned and all the packets which are acknowledged are freed. The acknowledge packet is forwarded to the fixed host.
- **Spurious acknowledgement**:This is an acknowledgement for the packet with sequence number less than the last seen acknowledgement. This acknowledgement is discarded.
- **Duplicate acknowledge:**The acknowledgement is same as the previous one receipt packet indicates that the next packet in sequence with the duplicate acknowledge has not seen received by the mobile host (HS). But some subsequent packets might have been received because a duplicate acknowledgement is generated for every TCP segment received out of sequence. When the base host (BH) of foreign agent (FA) notices duplicate acknowledge it takes action depending on the snoop cache state and type of acknowledgement.
    - If the packet is not in cache the duplicate acknowledgement is forwarded to the fixed host (FH) because now it should be resent from the fixed computers.
    - If the packet is marked as sender re-transmitted packet, then also it is forwarded to the fixed host (FH) because on number of duplicate acknowledgements it receives when it re-transmits a packet.
    - When the missing packet is already being retransmitted when the first DUPLICATE ACKNOWLEDGE arrived so this is discarded.
    - When the first duplicate acknowledgement of the packet arrives after the subsequent packet in the stream reaches mobile host (MH). Now for every successive packet in the window, a duplicate acknowledgement should be generated so to minimize the number of duplicate acknowledgements the lost packet should be transmitted as soon as possible. This packet should be re-transmitted with a higher priority than other packets. To implement this, two queues of re-transmitting packets are maintained one for high priority packets and other for normal packets Re-transmission using fast queue also improves the performance. The approach of maintaining two queues is helpful when there are low or medium error rates because when the error rate is high; all the packets are required to be re-transmitted again therefore there is no need to maintain 2 queues. But when bit rates are lower or medium,

fast queue enables the lost packets to be re-transmitted soon hence increasing the performance.

Data Transfer from a mobile host

The proposed protocol suggests a modification at TCP code at the mobile host also. This is necessary because the modification made at the base because when the transfer is from mobile host to the fixed host (FH) and a packet is lost, it cannot be found whether it is on wireless link or due to congestion.In the proposed protocol, track of lost packets is kept as well as NACKS are generated for these lost packets back to the mobile. This is helpful when there are several packet losses in a single transmission window which may be due to high interference or when strength or quality of signal is low. On receiving the NACK, the mobile host retransmits the missing packet immediately. This retransmitted packet would arrive out of sequence at the fixed host hence they should be rerecorded at the fixed host.

Handoffs

In case of handoffs, the new base station (BS) should perform the task of snooping so the new base station (BS) should prepare their snoop cache for the mobile host. This is the transition state called "buffering" state and the base station (BS) cannot snoop onto the acknowledgements. As soon as hand off occurs, snoop cache is synchronized to the new base station (BS) and the process continues.

Advantages

1. **End-to-end semantics is preserved**: The FA does not acknowledge the packet. Even if the foreign agent (FA) or base station (BS) crashes, the approach automatically fall into standard TCP
2. **No Modifications at Fixed Host**:The fixed computer TCP does not require any modifications. Most of the modifications are at foreign agent (FA) / base station (BS) and some on mobile host.
3. **No packet loss during handovers**: In case of handovers, if there is some data not
4. transferred to the new foreign agent, there will be a time-out at fixed host triggering retransmission of packet, following mobile IP, to a new COA.If the new base station does not comply with scheme, approach will fall back to standard TCP.

Disadvantages

1. If the packet is lost or delayed during the re-transmission from buffer of the foreign agent, due to error on wireless link, time-out will occur at fixed host (FH).Therefore problems on wireless link are not isolated.
2. The wireless links offers very high delay as compared to wired link almost by a factor of 10. In this case the timers in foreign agent (FA) and fixed host are almost equal and approach is almost ineffective.
3. Use of NAK between foreign agent and mobile host assumes additional mechanism on the mobile host.
4. Snooping and buffering won't be applicable if there is end-end encryption between fixed host and mobile host. As per RFC 2406 in IP encapsulation security payload TCP protocol header are encrypted and the snooping on sequence numbers won't be possible.
5. Re-transmission from foreign agent (FA)/ base station (BS) may not work because many schemes prevent replay attacks and re-transmission may be interpreted as replay. Therefore snooping protocol is used only when encryption is used above transport layers.
6. Architecture of snooping protocol to overcome limitations of I- TCP.
7. Functioning Snoop-module at base station (BS)/ foreign agent (FA).

M-TCP

Learning Objectives

- Limitations of I-TCP and Snooping TCP
- Introduction to M-TCP
- Architecture and TCP layer
- Functioning of M-TCP
- Advantages and limitations of M-TCP

Introduction

We have seen in previous modules that standard TCP used with fixed hosts and wired networks pose serious drop in throughput if used without modification in wireless environment. The reason is that standard TCP understands only congestion as the reason of packet loss and triggers congestion control mechanism whereas in wireless networks there can be many other reasons of packet loss like high bit error rate, frequency disconnection, improper handovers. In such a situation, applying congestion control mechanism would lead to serious drop in the throughput. Therefore there is need optimized algorithms for TCP in wireless environment. But fixed networks are already based on standard TCP and it is not possible to entirely change the entire TCP. Therefore the algorithms proposed should have the following goals.

1. It should not aim to modify the entire TCP because it is the based on which the whole of internet rely
2. The protocol should not trigger congestion control mechanism when it is not required
3. The network related problems on wireless link should be isolated from the fixed hosts

The TCP variants devised for wireless environment should adhere to these goals. In the previous module we discussed two protocols I-TCP and snooping TCP. Both are split connection protocols and divide the end-to-end TCP connection into two parts via an intermediate host. The connection between fixed host and intermediate host has standard TCP whereas optimized TCP resides on link between intermediately host and mobile host. In I-TCP the intermediately hosts act as proxy which takes packets from fixed host, gives acknowledgement and then forwards the packet to mobile host and same thing is repeated on other side. This completely loses end-to-end semantics of TCP because the packets are acknowledged even before they are delivered. The situation is worst when the mobile moves to new network. All the packets buffered for the mobile device needs to be forwarded moreover whole scheme fails when the intermediator crashes. As a revision to this scheme, snooping TCP was proposed in which the intermediator just caches the packets from fixed host towards the mobile host at the same time it snoops the flow of acknowledgements and packets header. In case of missing acknowledgements it forwards the cached packet to the mobile host. The scheme retains end- end semantics. Both I-TCP and Snooping TCP handles packet losses due to bit error rate and do not work well when packets are lost due to frequent disconnections.

In 1997, M-TCP was proposed by Brown & Singh. The protocol deals with the problem of periodic disconnection as compared to existing solutions of that time which deals with problem of high bit error rate. The proposed solution is to send the sender into persist mode when the mobile device goes into disconnection state. This is done by shrinking the window of the sender. In this module we will discuss the M-TCP protocol in detail and how it performs in the situation of frequent disconnections.

M-TCP

Let us see what are the reasons of frequent disconnections in mobile environment and what are the problems associated with it

Reasons of frequent disconnections

- Brief blackout period (or disconnection) during handovers

- Disconnections may also be caused by physical obstacles in the environment that block radio signals, such as buildings.
- Due to load on cells, the mobile device may not receive any bandwidth for large time periods (call blocking)

Problems with existing solutions during disconnections

- If the mobile is disconnected for a long period of time, the sender will invoke congestion control.
- For Snooping TCP, in the case of handoff's the new snoop will start with empty cache and till the cache is built ,there would be no retransmission or filtering and there would be decrease in throughput. The problem of degradation would be more serious when cell sizes are small (pico or micro cell) environment
- Serial Timeout: A serial time out is a condition wherein multiple consecutive retransmissions of same segment are transmitted to mobile while it is disconnected and these retransmissions are also loss. The retransmission timer also gets doubled with every retransmission(Fig 1).

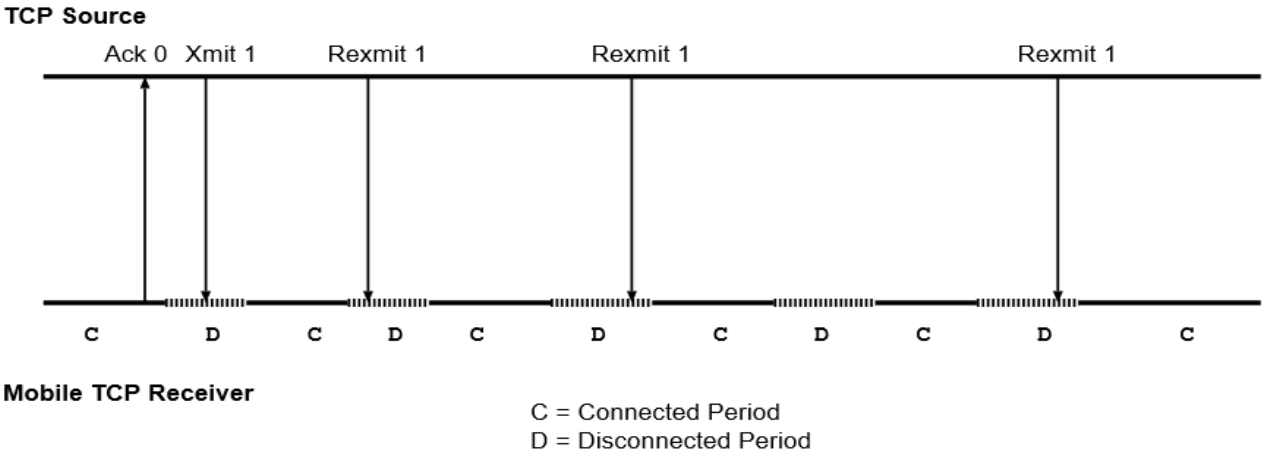


Figure 1: Serial Timeouts

In Fig. 1 , C & D indications connection & disconnection states It can be seen that retransmission are taking place in disconnection states which again triggers doubling of timer. Hence there is no activity till 64 sec even if the device regains connection. The split connection approached using TCP on both sides of connection respond poorly to lengthy disconnections

M-TCP is aims to solve the problem of frequent disconnection, preserving end-end semantics and throughput

M-T P Architecture

- The authors of M-TCP proposed it as three-layer architecture (Fig. 2)
- Lowest level- Mobile host
- Next level- MSS (Mobile support station) is a node which controls mobile host
- Highest level- SH or a supervisory host controlling many MSS

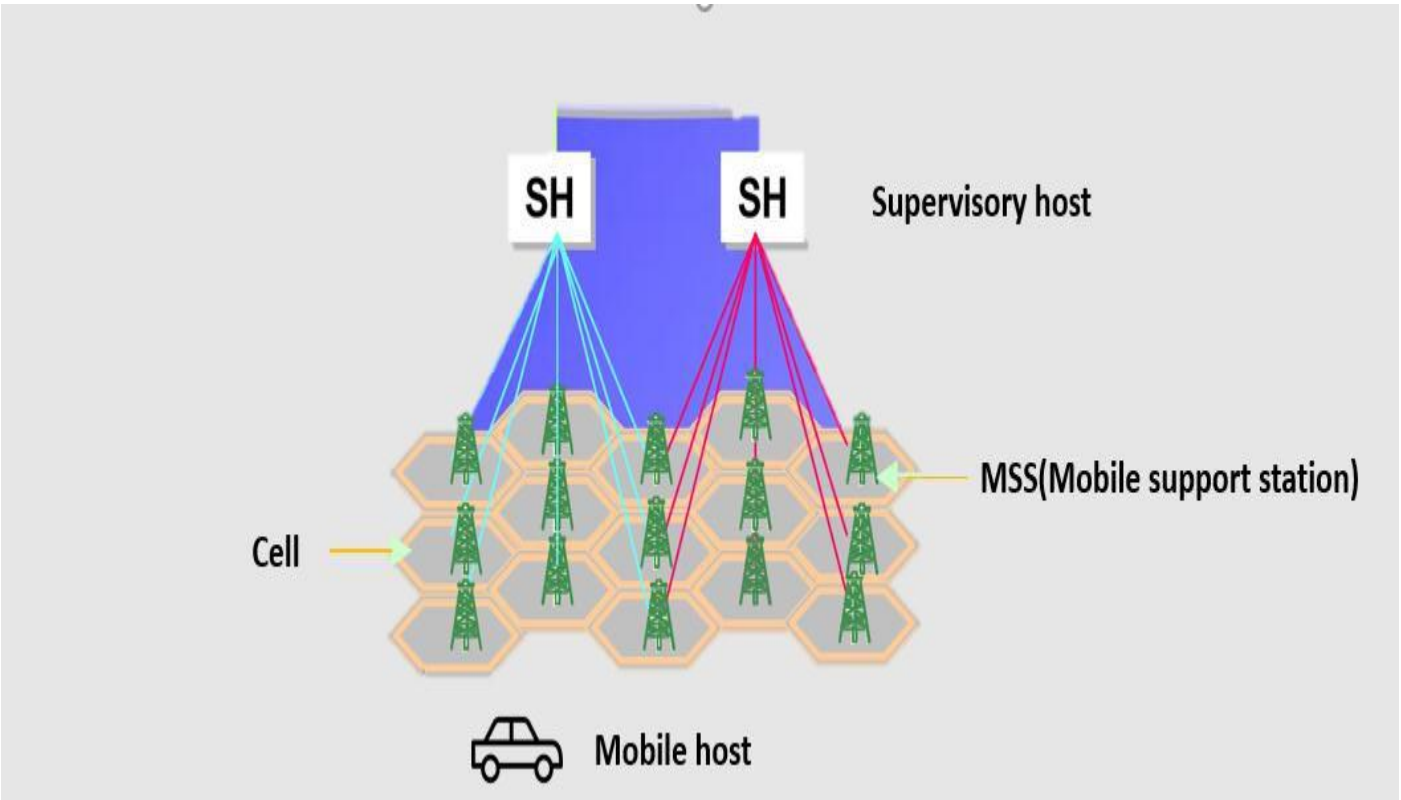


Figure 2: Layered M-TCP architecture

Advantage of having MSS

1. MSS are cheaper devices as they serve as only connection points whereas in I-TCP & snoop protocols, sophisticated MSS are used. Hence cost of infrastructure is increases particularly in case of picocelluar networks
2. No need to transfer the states when MH roams from one cell into another controlled by same MH

M-T P: Transport layer Design

The proposed protocol divides the connection into two parts at SH supervisory host (Fig. 3).

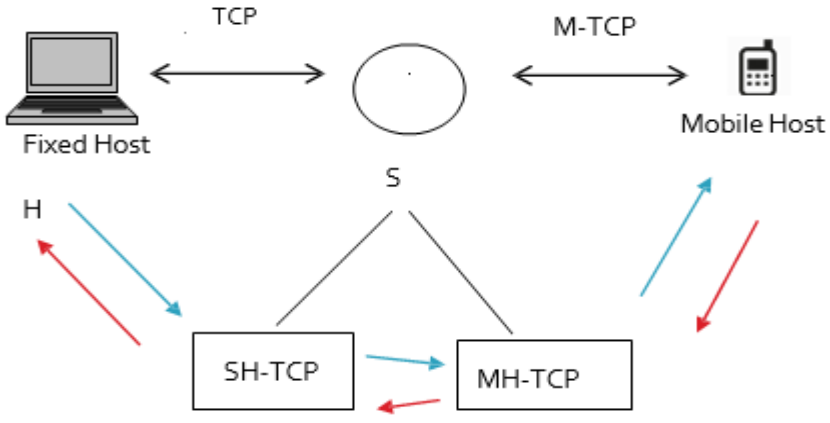


Figure 3: Layered M-TCP architecture

- The TCP sender on fixed network uses unmodified TCP to send data to the SH while SH uses modified TCP to send data to the MH.
- The TCP client at SH receiving data from fixed host is called SH-TCP

- TCP client at SH towards MH is called M-TCP .It receives packets from SH-TCP and passes to mobile host.
- The acknowledgements are received by M-TCP from mobile host and forwarded to SH- TCP which delivers it to the fixed host

SH-TCP client

- SH-TCP receives a packet from fixed host
- It passes it to M-TCP client.
- It does not acknowledge maintaining end-end semantics.
- M-TCP at SH passes the packet to MH
- It receives ACK's from Mobile host via M-TCP
- Sends the ACK's to Fixed Host

In a situation when disconnection is there, the approach is to choke the TCP sender when mobile device is disconnected & allow the sender to transmit at full speed when the connection is regained. The mode with choked window is called persist mode. According to RFC 1112, TCP clients go to persistent mode when window size is shrunk to zero. For this purpose, a new Ack is required. SH-TCP follows a strategy to send the sender into persistent mode. It keeps one Ack from mobile host unacknowledged . The strategy is as follows:

- Suppose the window size is 100.
- 70 bytes are sent to mobile host.
- Mobile host Acks all 70 bytes
- The SH-TCP sends Ack for only 69 bytes to the sender
- When the MH goes into disconnection state after acknowledging 70 bytes, it stops sending Acks.
- When M-TCP does not receive Ack's it assumes the mobile to be disconnected and informs SH TCP
- It uses the last Ack for 70th byte.
- Sends it to TCP sender along with window size update
- When TCP sender receives window size update message, it goes into persistent mode
- As long as the sender is in the persistent mode, no time out occur, no exponential back off of retransmission timers occur and neither it goes into slow start When connection is regained,
- When the MH regains the connection, it sends a greeting packet to SH.
- M-TCP receives it, passes it to SH-TCP.
- SH-TCP sends Ack to the sender and reopens it's receive window.
- The sender leaves persist mode and begin sending data again at full speed from bytes number w+1 where w is the last acknowledged byte.

The only issue with this strategy is when there only w' bytes to be sent to MH. According to protocol, SH-TCP would send Ack's for only W'-1. This will cause timeout at sender & trigger retransmission and after 12 retrains sender will quit. The solution to this is that when SH-TCP understands that there are no more Acks from the MH it should not allow the sender to go in persist mode. It should send the Ack for last byte.

Design of M-TCP The goal of M-TCP is to avoid serial timeouts & recover from losses due to disconnection. The modification made in this protocol is notification of wireless connectivity.

- M-TCP observes flow of Acks from MH
- If it does not receive Ack for some time, it understands mobile host is disconnected.
- M-TCP freezes all its timers. This is done to avoid congestion control mechanism at M- TCP

- It informs to SH-TCP about disconnection
- When the mobile host regains the connection, it sends a last Ack to mark the reconnection.
- M-TCP informs this to SH-TCP who opens senders transmit window
- When connection is gained, M-TCP at MH sends a specially marked Ack to M-TCP at SH, which contains Seq. no of highest byte received so far
- M-TCP unfreezes timers to resume operation

Advantage

- End-End semantics is maintained because the SH does not Ack message it just forwards Ack from MH
- No buffering of packets at SH as in I-TCP so it is not necessary to forward buffers to SH
- The efficiency of the TCP connection is not degraded due to disconnections since the sender is prevented from going into exponential backoff and slow-start
- If the MH moves from the domain of one SH into the domain of another SH, the old SH does not need to forward the socket buffers (as in I-TCP).

Limitations

- Errors on wireless link are also propagated to the fixed sender
- It assumes bit error rate to be very low which is not a valid assumption

Summary

- M-TCP is proposed protocol which deals with packet losses due to frequent disconnections
- Idea is to choke the senders window during the period of disconnection and re-transmit with full speed when connection is gained hence avoiding slow start
- The protocol divides TCP connection into two parts at supervisory host. It forwards packets but do not acknowledges them maintaining end-to-end semantics
- It has limitation of not isolating the wired and wireless connections

Variants of TCP Optimized for Wireless Networks

Learning Objectives

- Recap of previous module
- Limitations of I-TCP and Snooping TCP and M-TCP
- Fast re-transmit protocol
- Freezing protocol
- Selective retransmission
- Transaction oriented TCP
- Comparison of all the variants of TCP

Introduction

We have seen in previous modules that standard TCP cannot be replicated in wireless environment because standard TCP assumes only congestion as reason for packet loss and goes into slow start by lowering the rate at which the packets are sent. This decreases the throughput. Whereas in wireless and mobile environments there are many reasons of packet loss other than congestion like call blocking improper handover, high bit error rate , disconnections and many more. Applying standard TCP here would degrade the performance. Many protocols have been optimized as per the wireless environment. Each protocol has its strengths and weaknesses and caters a particular problem of wireless scenarios. In the previous modules, we have

learnt about three protocols namely I-TCP, Snooping TCP and M-TCP. All are split connection protocols and divide the end-to-end TCP connection into two parts via an intermediate host. The connection between fixed host and intermediate host has standard TCP whereas optimized TCP resides on link between intermediately host and mobile host. In I-TCP the intermediate hosts act as proxy which takes packets from fixed host, gives acknowledgement and then forwards the packet to mobile host and same thing is repeated on other side. This completely loses end-to-end semantics of TCP because the packets are acknowledged even before they are delivered. The situation is worst when the mobile moves to new network. All the packets buffered for the mobile device needs to be forwarded. Moreover, whole scheme fails when the intermediate crashes. As a revision to this scheme, snooping TCP was proposed in which the inter-mediator just caches the packets from fixed host towards the mobile host at the same time it snoops the flow of acknowledgements and packets header. In case of missing acknowledgements, it forwards the cached packet to the mobile host. The scheme retains end-end semantics.

Both I-TCP and Snooping TCP handles packet losses due to bit error rate but do not work well when packets are lost due to frequent disconnections. A revision to these schemes which deals with the problem of periodic disconnection was proposed as M-TCP. The proposed solution is to send the sender into persist mode when the mobile device goes into disconnection state. This is done by shrinking the window of the sender.

All these schemes do not take into account the problems such as packet loss due to improper handover or due to lengthy disconnections. In this module we will discuss four more protocols devised to handle these issues in mobile computing environment. These protocols are:

- Fast re-transmit/ Fast recovery
- Timeout/Freezing
- Selective retransmission
- Transaction oriented TCP

Next sections consists of discussion about these protocols

Fast re-transmit/ Fast recovery

This protocol is designed for situations when mobile phone moves from one network to another network. As with Mobile IP technology, it comes under new foreign agent. Packets needs to hand offed to new foreign agent. Improper Hand-offs may result in temporary loss of packets to MH because with non-overlapping cells there may be a small disconnection before mobile connects to new base station MH and old BS may attempt to send packets to each other, during this period; it may result in loss of packets. The existing approaches cannot handle this situation because of reasons given below:

Behavior of existing protocol in case of improper handoff

- In I-TCP approach hard state at base station must be moved to new base station or foreign agent.
- In Snoop protocol soft state need not be moved but while the new base station/foreign agent builds new state, packet losses may not be recovered locally.
- M-TCP avoids reduction of congestion window due to handoff. When host moves, route changes, and new route may be more congested. Therefore it is not wise to start transmitting at full window after handoff.

Issues during handoffs

During the long delay for a handoff to complete a whole window worth of data may be lost. After handoff is complete, ACKs are not received by the TCP sender. As a result, sender eventually times out, and retransmits; If handoff still not complete, another timeout will occur. This is a serious performance penalty as time is wasted until timeout occurs and Window gets shrunk after timeout.

Fast retransmit behavior of TCP

Duplicate ACKs for the same packet are received implies that all the packets upto the acknowledged packet has been received. This implies that receiver is receiving something and packet loss is not due to congestion, it is due to transmission error. Hence missing packet can be sent before timer expires (fast retransmit) and No initiation of slow start is required. Sender transmits with the current window size (fast recovery). This behavior is utilized in the protocol discussed in next section.

Fast retransmit/fast recovery protocol takes advantage of this behavior of TCP

Idea was presented by Caceres in 1995. The idea is to forcefully impose fast re-transmit behavior when it moves from one foreign agent to another. Artificially force fast re-transmit/fast recovery.

When mobile host enters new foreign agent area, it sends duplicate ACKs .The algorithm proposed three duplicate ACKs. When fixed host receives duplicate ACKs, it goes into fast re-transmit mode. It continues transmitting with same speed.

Advantage

- Simple to implement
- Minor changes in mobile host software required
- No need to change foreign agent or fixed host

Disadvantage

- More cooperation between mobile IP and TCP layer is required.
- Change in one influences the other insufficient isolation of packet losses.
- Re-transmitted packet crosses whole network between fixed host and mobile host
- Packet losses only due to improper handovers are considered

Transmission/timeout freezing

This algorithm is used in a situation when Mobile hosts can be disconnected for a longer time. For e.g. when a mobile phone enters a tunnel or a disconnection occurs due to overloaded cells etc. In such a situation TCP disconnects after time-out completely. The proposed protocol uses the fact that MAC layer is often able to detect interruption in advance .MAC layer knows reason for disruption. MAC can inform TCP layer of upcoming loss of connection. If interruption is anticipated, early both mobile host and fixed host can be informed. Reason of disruption is also informed to prevent triggering of slow-start When TCP is informed of congestion, all the timers and current window states are freezed..TCP sender stops sending. When MAC detects connectivity again, it signals TCP about it. Transmission resumes from where it stopped. Timers do not advance because of freezing. No timeout occurs and TCP does not go into slow start.

Advantages

- Resume connection after long disconnection

- Can be used with encrypted data

Disadvantages

- Both fixed host and mobile host needs to be changed
- Dependence on MAC layer for information of interruption
- If any encryption mechanism uses time dependent encryption schemes, freezing would not help.

Selective retransmission

TCP acknowledgements are often cumulative. TCP acknowledges in-sequence receipt of packets up to certain packet. If a single packet is lost, whole packet sequence beginning at the gap has to be re-transmitted (go-back-n), thus wasting bandwidth.

To solve this issue, selective retransmission was presented as solution in RFC 2018. It allows for acknowledgements of single packets. Sender can now re-transmit only the missing packets.

Advantage

- Much higher efficiency
- Requires less bandwidth
- Helpful in slow wireless links

Disadvantage

- more complex software in a receiver, more buffer needed at the receiver to re-sequence the packets

Transaction oriented TCP

A TCP connection consists of three phases: setup, data transmission and connection release using 3-way-handshake. It needs 3 packets for setup and release, respectively. This overhead is minimal and acceptable when connections are with large traffic or long duration connection. But if the connections are short, it is overhead. Because for even short messages a minimum of 7 packets are needed (Fig. 1). To solve this issue, T-TCP was proposed by Barden in 1994. It was documented in RFC 1644. The proposed proposal combines connections establish and release phases with data packets. Hence number of packets reduces to two instead of seven.

Advantages

- Reduces overhead

Disadvantages

- Needs changes at mobile host and fixed host
- Mobility is not hidden
- Security issues

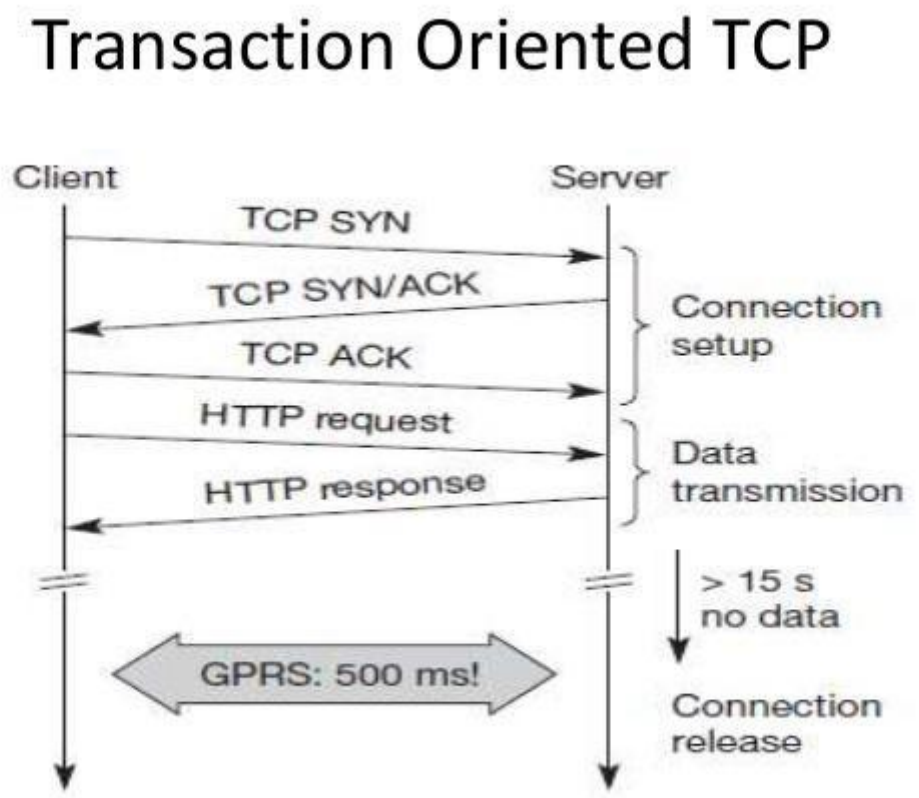


Figure 1: 7 packets for sending data

Summary:

The below table summarizes the mechanisms, strengths and limitations of all the variants of TCP studied so far.

Approach	Mechanism	Advantages	Disadvantages
Indirect TCP	splits TCP connection into two connections	isolation of wireless link, no change in TCP or fixed host	loss of TCP semantics, higher latency at handover, security threat, lack of robustness
Snooping TCP	"snoops" data and acknowledgements, local retransmission	transparent for end- to-end connection, MAC integration possible, TCP semantics maintained	problematic with encryption, bad isolation of wireless link, no transmission during period of disconnection while handovers
M-TCP	splits TCP connection, chokes sender via window size	Maintains end-to- end semantics, handles long term and frequent disconnections	Bad isolation of wireless link, processing overhead due to bandwidth management

Approach	Mechanism	Advantages	Disadvantages
Fast retransmit/ fast recovery	avoids slow-start after roaming by forced retransmission	simple and efficient	mixed layers, not transparent
Transmission/ time-out freezing	Mac layer notifies of the disruption, freezes TCP state at disconnect, resumes, after reconnection	independent of content or encryption, works for longer interrupts	changes in TCP required, MAC dependent
Selective retransmission	retransmit only lost packets	Lower bandwidth requirement	slightly more complex receiver software, more buffer needed
Transaction oriented TCP	combine connection setup/release and data transmission	Efficient for certain applications	changes in TCP required, not transparent

Dynamic Host Configuration Protocol

Dynamic Host Configuration Protocol (DHCP) is a network management protocol used to dynamically assign an IP address to nay device, or node, on a network so they can communicate using IP (Internet Protocol). DHCP automates and centrally manages these configurations. There is no need to manually assign IP addresses to new devices. Therefore, there is no requirement for any user configuration to connect to a DHCP based network.

DHCP can be implemented on local networks as well as large enterprise networks. DHCP is the default protocol used by the most routers and networking equipment. DHCP is also called RFC (Request for comments) 2131.

DHCP does the following:

- DHCP manages the provision of all the nodes or devices added or dropped from the network.
- DHCP maintains the unique IP address of the host using a DHCP server.
- It sends a request to the DHCP server whenever a client/node/device, which is configured to work with DHCP, connects to a network. The server acknowledges by providing an IP address to the client/node/device.

DHCP is also used to configure the proper subnet mask, default gateway and DNS server information on the node or device. There are many versions of DCHP are available for use in IPV4 (Internet Protocol Version 4) and IPV6 (Internet Protocol Version 6).

How DHCP works

DHCP runs at the application layer of the TCP/IP protocol stack to dynamically assign IP addresses to DHCP clients/nodes and to allocate TCP/IP configuration information to the DHCP clients. Information includes subnet mask information, default gateway, IP addresses and domain name system addresses.

DHCP is based on client-server protocol in which servers manage a pool of unique IP addresses, as well as information about client configuration parameters, and assign addresses out of those address pools.

The DHCP lease process works as follows:

- First of all, a client (network device) must be connected to the internet.
- DHCP clients request an IP address. Typically, client broadcasts a query for this information.
- DHCP server responds to the client request by providing IP server address and other configuration information. This configuration information also includes time period, called a lease, for which the allocation is valid.
- When refreshing an assignment, a DHCP clients request the same parameters, but the DHCP server may assign a new IP address. This is based on the policies set by the administrator.

Components of DHCP

When working with DHCP, it is important to understand all of the components. Following are the list of components:

- **DHCP Server:** DHCP server is a networked device running the DCHP service that holds IP addresses and related configuration information. This is typically a server or a router but could be anything that acts as a host, such as an SD-WAN appliance.
- **DHCP client:** DHCP client is the endpoint that receives configuration information from a DHCP server. This can be any device like computer, laptop, IoT endpoint or anything else that requires connectivity to the network. Most of the devices are configured to receive DHCP information by default.
- **IP address pool:** IP address pool is the range of addresses that are available to DHCP clients. IP addresses are typically handed out sequentially from lowest to the highest.
- **Subnet:** Subnet is the partitioned segments of the IP networks. Subnet is used to keep networks manageable.
- **Lease:** Lease is the length of time for which a DHCP client holds the IP address information. When a lease expires, the client has to renew it.
- **DHCP relay:** A host or router that listens for client messages being broadcast on that network and then forwards them to a configured server. The server then sends responses back to the relay agent that passes them along to the client. DHCP relay can be used to centralize DHCP servers instead of having a server on each subnet.

Benefits of DHCP

There are following benefits of DHCP:

Centralized administration of IP configuration: DHCP IP configuration information can be stored in a single location and enables that administrator to centrally manage all IP address configuration information.

Dynamic host configuration: DHCP automates the host configuration process and eliminates the need to manually configure individual host. When TCP/IP (Transmission control protocol/Internet protocol) is first deployed or when IP infrastructure changes are required.

Seamless IP host configuration: The use of DHCP ensures that DHCP clients get accurate and timely IP configuration IP configuration parameter such as IP address, subnet mask, default gateway, IP address of DND server and so on without user intervention.

Flexibility and scalability: Using DHCP gives the administrator increased flexibility, allowing the administrator to move easily change IP configuration when the infrastructure changes.

Introduction to Wireless LAN

Wireless LAN stands for **Wireless Local Area Network**. It is also called LAWN (**Local Area Wireless Network**). WLAN is one in which a mobile user can connect to a Local Area Network (LAN) through a wireless connection.

The IEEE 802.11 group of standards defines the technologies for wireless LANs. For path sharing, 802.11 standard uses the Ethernet protocol and CSMA/CA (carrier sense multiple access with collision avoidance). It also uses an encryption method i.e. wired equivalent privacy algorithm.

Wireless LANs provide high speed data communication in small areas such as building or an office. WLANs allow users to move around in a confined area while they are still connected to the network.

In some instance wireless LAN technology is used to save costs and avoid laying cable, while in other cases, it is the only option for providing high-speed internet access to the public. Whatever the reason, wireless solutions are popping up everywhere.

Examples of WLANs that are available today are NCR's waveLAN and Motorola's ALTAIR.

Advantages of WLANs

- **Flexibility:** Within radio coverage, nodes can communicate without further restriction. Radio waves can penetrate walls, senders and receivers can be placed anywhere (also non-visible, e.g., within devices, in walls etc.).
- **Planning:** Only wireless ad-hoc networks allow for communication without previous planning, any wired network needs wiring plans.
- **Design:** Wireless networks allow for the design of independent, small devices which can for example be put into a pocket. Cables not only restrict users but also designers of small notepads, PDAs, etc.

- **Robustness:** Wireless networks can handle disasters, e.g., earthquakes, flood etc. whereas, networks requiring a wired infrastructure will usually break down completely in disasters.
- **Cost:** The cost of installing and maintaining a wireless LAN is on average lower than the cost of installing and maintaining a traditional wired LAN, for two reasons. First, after providing wireless access to the wireless network via an access point for the first user, adding additional users to a network will not increase the cost. And second, wireless LAN eliminates the direct costs of cabling and the labor associated with installing and repairing it.
- **Ease of Use:** Wireless LAN is easy to use and the users need very little new information to take advantage of WLANs.

Disadvantages of WLANs

- **Quality of Services:** Quality of wireless LAN is typically lower than wired networks. The main reason for this is the lower bandwidth due to limitations is radio transmission, higher error rates due to interference and higher delay/delay variation due to extensive error correction and detection mechanisms.
- **Proprietary Solutions:** Due to slow standardization procedures, many companies have come up with proprietary solutions offering standardization functionality plus many enhanced features. Most components today adhere to the basic standards IEEE 802.11a or 802.11b.
- **Restrictions:** Several govt. and non-govt. institutions world-wide regulate the operation and restrict frequencies to minimize interference.
- **Global operation:** Wireless LAN products are sold in all countries so, national and international frequency regulations have to be considered.
- **Low Power:** Devices communicating via a wireless LAN are typically power consuming, also wireless devices running on battery power. Whereas the LAN design should take this into account and implement special power saving modes and power management functions.
- **License free operation:** LAN operators don't want to apply for a special license to be able to use the product. The equipment must operate in a license free band, such as the 2.4 GHz ISM band.
- **Robust transmission technology:** If wireless LAN uses radio transmission, many other electrical devices can interfere with them (such as vacuum cleaner, train engines, hair dryers, etc.).Wireless LAN transceivers cannot be adjusted for perfect transmission is a standard office or production environment.

Fundamentals of WLANs

1. HiperLAN

- HiperLAN stands for High performance LAN. While all of the previous technologies have been designed specifically for an adhoc environment, HiperLAN is derived from traditional LAN environments and can support multimedia data and asynchronous data effectively at high rates (23.5 Mbps).
- A LAN extension via access points can be implemented using standard features of the HiperLAN/1 specification. However, HiperLAN does not necessarily require any type of access point infrastructure for its operation.
- HiperLAN was started in 1992, and standards were published in 1995. It employs the 5.15GHz and 17.1 GHz frequency bands and has a data rate of 23.5 Mbps with coverage of 50m and mobility< 10 m/s.
- It supports a packet-oriented structure, which can be used for networks with or without a central control (BS-MS and ad-hoc). It supports 25 audio connections at 32kbps with a maximum latency of 10 ms, one video connection of 2 Mbps with 100 ms latency, and a data rate of 13.4 Mbps.
- HiperLAN/1 is specifically designed to support adhoc computing for multimedia systems, where there is no requirement to deploy a centralized infrastructure. It effectively supports MPEG or other state of the art real time digital audio and video standards.
- The HiperLAN/1 MAC is compatible with the standard MAC service interface, enabling support for existing applications to remain unchanged.
- HiperLAN 2 has been specifically developed to have a wired infrastructure, providing short-range wireless access to wired networks such as IP and ATM.

The two main differences between HiperLAN types 1 and 2 are as follows:

- Type 1 has a distributed MAC with QoS provisions, whereas type 2 has a centralized schedule MAC.
- Type 1 is based on Gaussian minimum shift keying (GMSK), whereas type 2 is based on OFDM.

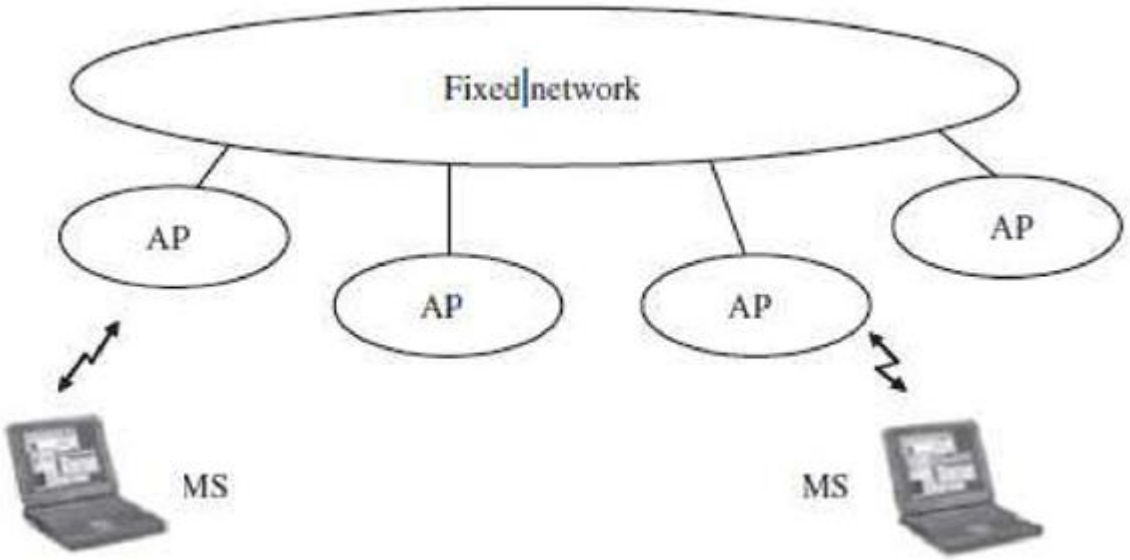
- HiperLAN/2 automatically performs handoff to the nearest access point. The access point is basically a radio BS that covers an area of about 30 to 150 meters, depending on the environment. MANETs can also be created easily.

The goals of HiperLAN are as follows:

- QoS (to build multiservice network)
- Strong security
- Handoff when moving between local area and wide areas
- Increased throughput
- Ease of use, deployment, and maintenance
- Affordability
- Scalability

One of the primary features of HiperLAN/2 is its high speed transmission rates (up to 54 Mbps). It uses a modulation method called OFDM to transmit analog signals. It is connection oriented, and traffic is transmitted on bidirectional links for unicast traffic and unidirectional links toward the MSs for multicast and broadcast traff

This connection oriented approach makes support for QoS easy, which in turn depends on how the HiperLAN/2 network incorporates with the fixed network using Ethernet, ATM, or IP.

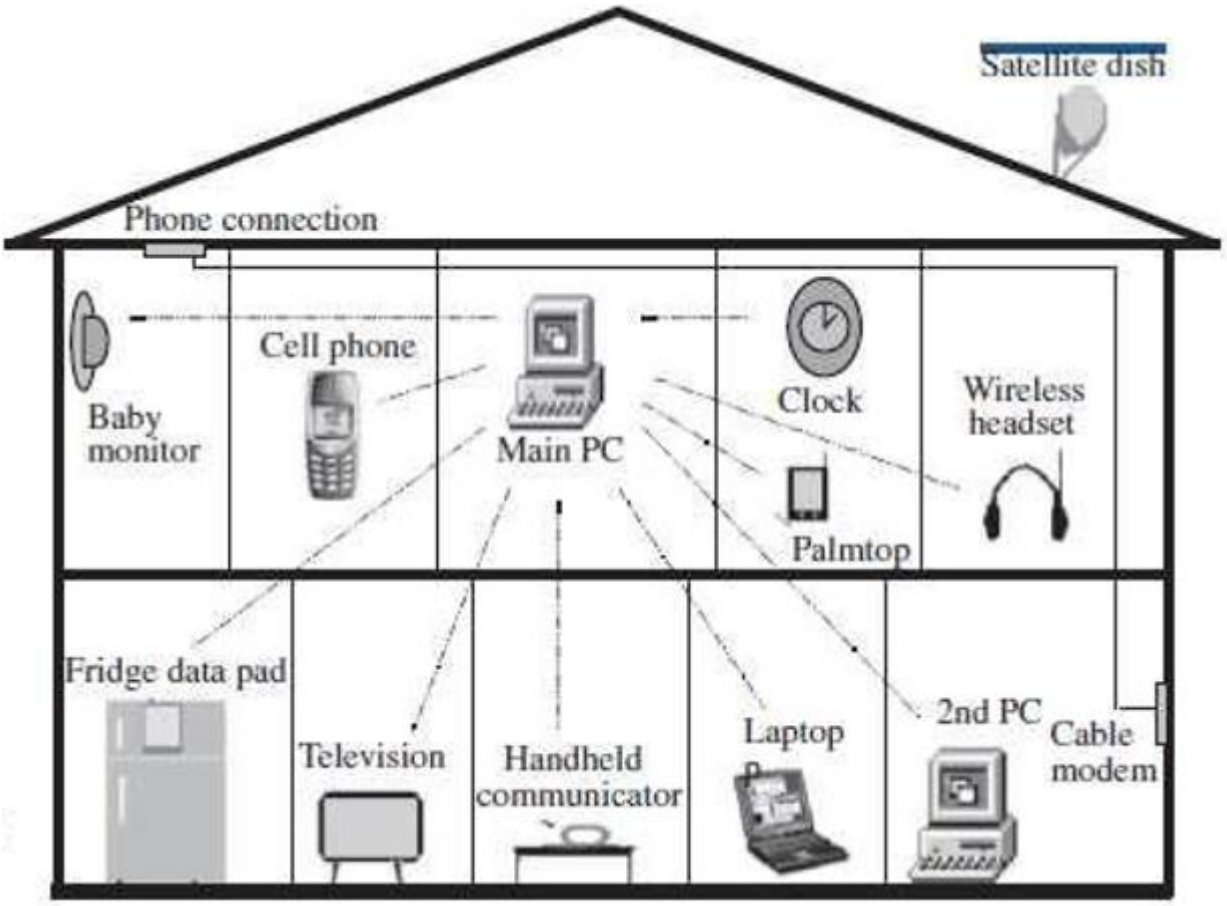


The HiperLAN/2 architecture shown in the figure allows for interoperation with virtually any type of fixed network, making the technology both network and application independent.

HiperLAN/2 networks can be deployed at "hot spot" areas such as airports and hotels, as an easy way of offering remote access and internet services.

2. Home RF Technology

- A typical home needs a network inside the house for access to a public network telephone and internet, entertainment networks (cable television, digital audio and video with the IEEE 1394), transfer and sharing of data and resources (printer, internet connection), and home control and automation.
- The device should be able to self-configure and maintain connectivity with the network. The devices need to be plug and play enabled so that they are available to all other clients on the network as soon as they are switched on, which requires automatic device discovery and identification in the system.
- Home networking technology should also be able to accommodate any and all lookup services, such as Jini. Home RF products allow you to simultaneously share a single internet connection with all of your computers - without the hassle of new wires, cables or jacks.
- Home RF visualizes a home network as shown in the figure:



- A network consists of resource providers, which are gateways to different resources like phone lines, cable modem, satellite dish, and so on, and the devices connected to them such as cordless phone, printers and fileservers, and TV.
- The goal of Home RF is to integrate all of these into a single network suitable for all applications and to remove all wires and utilize RF links in the network suitable for all applications.
- This includes sharing PC, printer, fileserver, phone, internet connection, and so on, enabling multiplayer gaming using different PCs and consoles inside the home, and providing complete control on all devices from a single mobile controller.
- With Home RF, a cordless phone can connect to PSTN but also connect through a PC for enhanced services. Home RF makes an assumption that simultaneous support for both voice and data is needed.

Advantages of Home RF

- In Home RF all devices can share the same connection, for voice or data at the same time.
- Home RF provides the foundation for a broad range of interoperable consumer devices for wireless digital communication between PCs and consumer electronic devices anywhere in and around the home.
- The working group includes Compaq computer corp. Ericson enterprise network, IBM Intel corp., Motorola corp. and other.
- A specification for wireless communication in the home called the shared wireless access protocol (SWAP) has been developed.

3. IEEE 802.11 Standard

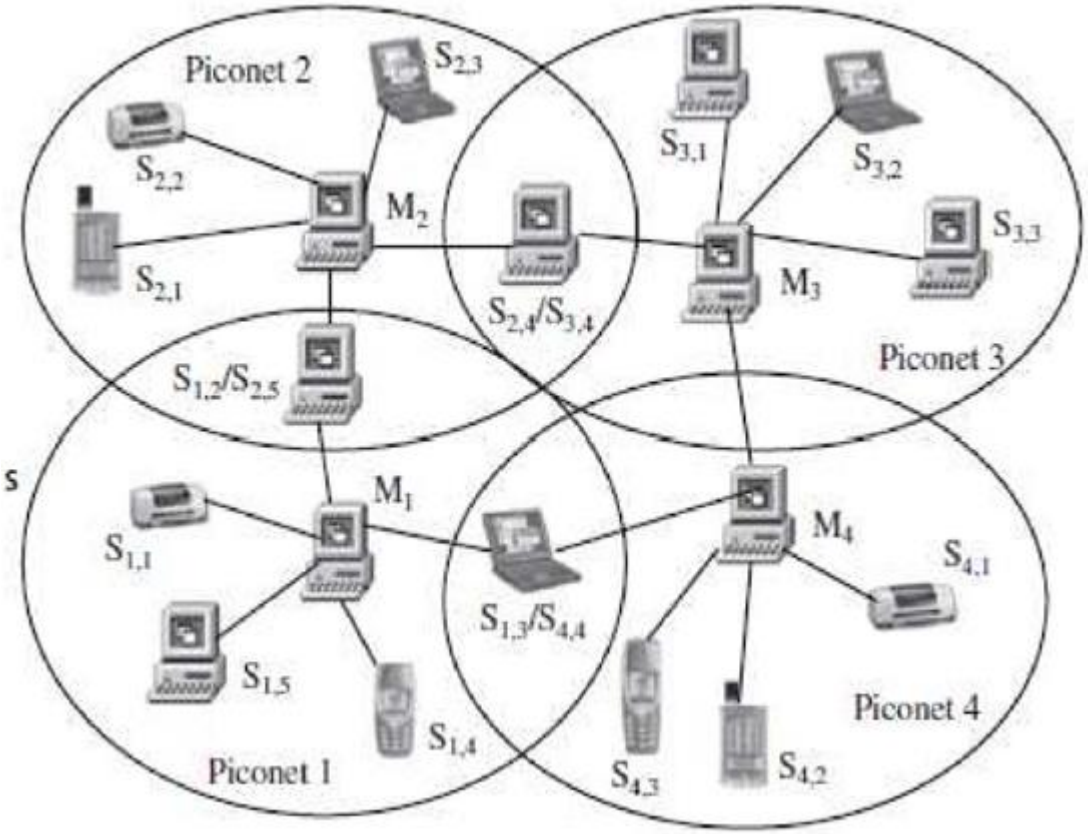
IEEE 802.11 is a set of standards for the wireless area network (WLAN), which was implemented in 1997 and was used in the industrial, scientific, and medical (ISM) band. IEEE 802.11 was quickly implemented throughout a wide region, but under its standards the network occasionally receives interference from devices such as cordless phones and microwave ovens. The aim of IEEE 802.11 is to provide wireless network connection for fixed, portable, and moving stations within ten to hundreds of meters with one medium access control (MAC) and several physical layer specifications. This was later called 802.11a. The major protocols include IEEE 802.11n; their most significant differences lie in the specification of the PHY layer.

4. Bluetooth

Bluetooth is one of the major wireless technologies developed to achieve WPAN (wireless personal area network). It is used to connect devices of different functions such as telephones, computers (laptop or desktop), notebooks, cameras, printers, and so on.

Architecture of Bluetooth

- Bluetooth devices can interact with other Bluetooth devices in several ways in the figure. In the simplest scheme, one of the devices acts as the master and (up to) seven other slaves.
- A network with a master and one or more slaves associated with it is known as a piconet. A single channel (and bandwidth) is shared among all devices in the piconet.



- Each of the active slaves has an assigned 3-bit active member address. many other slaves can remain synchronized to the master though remaining inactive slaves, referred to as parked nodes.
- The master regulates channel access for all active nodes and parked nodes. Of two piconets are close to each other, they have overlapping coverage areas.
- This scenario, in which nodes of two piconets intermingle, is called a scatternet. Slaves in one piconet can participate in another piconet as either a master or slave through time division multiplexing.
- In a scatternet, the two (or more) piconets are not synchronized in either time or frequency. Each of the piconets operates in its own frequency hopping channel, and any devices in multiple piconets participate at the appropriate time via time division multiplexing.
- The Bluetooth baseband technology supports two link types. Synchronous connection oriented (SCO) types, used primarily for voice, and asynchronous connectionless (ACL) type, essentially for packet data.

Infrared vs Radio Transmission

Infrared Transmission

- Infrared technology uses diffuse light reflected at walls, furniture etc. or a directed light if a line of sight (LOS) exists between sender and receiver.
- Infrared light is the part of the electromagnetic spectrum, and is an electromagnetic form of radiation. It comes from the heat and thermal radiation, and it is not visible to the naked eyes.

- In infrared transmission, senders can be simple light emitting diodes (LEDs) or laser diodes. Photodiodes act as receivers.
- Infrared is used in wireless technology devices or systems that convey data through infrared radiation. Infrared is electromagnetic energy at a wave length or wave lengths somewhat longer than those of red light.
- Infrared wireless is used for medium and short range communications and control. Infrared technology is used in instruction detectors; robot control system, medium range line of sight laser communication, cordless microphone, headsets, modems, and other peripheral devices.
- Infrared radiation is used in scientific, industrial, and medical application. Night vision devices using active near infrared illumination allow people and animals to be observed without the observer being detected.
- Infrared transmission technology refers to energy in the region of the electromagnetic radiation spectrum at wavelength longer than those of visible light but shorter than those of radio waves.
- Infrared technology allows computing devices to communicate via short range wireless signals. With infrared transmission, computers can transfer files and other digital data bidirectional.

Advantages of infrared

- The main advantage of infrared technology is its simple and extremely cheap senders and receivers which are integrated into nearly all mobile devices available today.
- No licenses are required for infrared and shielding is very simple.
- PDAs, laptops, notebooks, mobile phones etc. have an infrared data association (IrDA) interface.
- Electrical devices cannot interfere with infrared transmission.

Disadvantages of Infrared

- Disadvantages of infrared transmission are its low bandwidth compared to other LAN technologies.
- Limited transfer rates to 115 Kbit/s and we know that even 4 Mbit/s is not a particular high data rate.
- Their main disadvantage is that infrared is quite easily shielded.
- Infrared transmission cannot penetrate walls or other obstacles.
- Typically, for good transmission quality and high data rates a LOS (Line of site), i.e. direct connection is needed.

Radio Transmission

- Almost all networks use radio waves for data transmission, e.g., GSM at 900, 1800, and 1900 MHz, DECT at 1880 MHz etc. Radio transmission technologies can be used to set up ad-hoc connections for work groups, to connect, e.g., a desktop with a printer without a wire, or to support mobility within a small area.
- The two main types of radio transmission are AM (Amplitude Modulation) and (FM) Frequency Modulation.
- FM minimizes noise and provides greater reliability. Both AM and FM process sounds in patterns that are always varying of electrical signals.
- In an AM transmission the carrier wave has a constant frequency, but the strength of the wave varies. The FM transmission is just the opposite; the wave has constant amplitude but a varying frequency.
- Usually the radio transmission is used in the transmission of sounds and pictures. Such as, voice, music and television.
- The images and sounds are converted into electrical signals by a microphone or video camera. The signals are amplified, and transmitted. If the carrier is amplified it can be applied to an antenna.
- The antenna converts the electrical signals into electromagnetic waves and sends them out or they can be received. The antenna consists commonly of a wire or set of wires.

Advantages of Radio Transmission

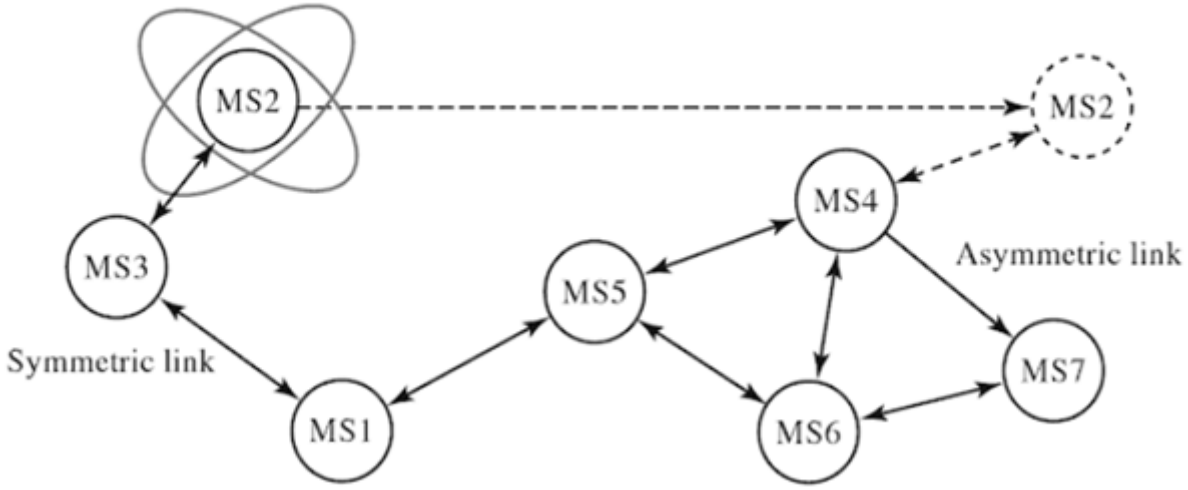
- Advantages of radio transmission include the long-term experiences made with radio transmission for wide area networks (e.g. microwave links) and mobile cellular phones.
- Radio transmission can cover larger areas and can penetrate (thinner) walls, plants, furniture etc.
- Additional coverage is gained by reflection.
- Radio typically does not need a LOS (Line of Site) if the frequencies are not too high.
- Higher transmission rates (e.g. 54 Mbit/s) than infrared (directed laser links, which offer data rates well above 100 Mbit/s).

Disadvantages of Radio Transmission

- Radio transmission can be interfered with other senders, or electrical devices can destroy data transmitted via radio.
- Bluetooth is simple than infrared.
- Radio is only permitted in certain frequency bands.
- Shielding is not so simple.
- Very limited ranges of license free bands are available worldwide and those that are available are not the same in all countries.
- A lot harmonization is going on due to market pressure.

Mobile Adhoc Network (MANET)

- A MANET consists of a number of mobile devices that come together to form a network as needed, without any support from any existing internet infrastructure or any other kind of fixed stations.
- A MANET can be defined as an autonomous system of nodes or MSs(also serving as routers) connected by wireless links, the union of which forms a communication network modeled in the form of an arbitrary communication graph.
- This is in contrast to the well-known single hop cellular network model that supports the needs of wireless communication between two mobile nodes relies on the wired backbone and fixed base stations.
- In a MANET, no such infrastructure exists and network topology may be changed dynamically in an unpredictable manner since nodes are free to move and each node has limiting transmitting power, restricting access to the node only in the neighboring range.
- MANETs are basically peer-to-peer, multi-hop wireless networks in which information packets are transmitted in a store and forward manner from a source to an arbitrary destination, via intermediate nodes as given in the figure:



- As nodes move, the connectivity may change based on relative locations of other nodes. The resulting change in the network topology known at the local level must be passed on to other nodes so that old topology information can be updated.
- For example, as MS2 in the figure changes its point of attachment from MS3 to MS4, other nodes that are part of the network should use this new route to forward packets to MS2. In the figure, we assume that it is not possible to have all nodes within each other's radio range. In case all nodes are closed by within each other's radio range, there are no routing issues to be addressed.
- In figures raise another issue, that of symmetric and asymmetric (bidirectional) and asymmetric (unidirectional) links. Consider symmetric links with associative radio range; for example, if MS1 is within radio range of MS3, then MS3 is also within radio range of MS1. The communication links are symmetric. This assumption is not always valid because of differences in transmitting power levels and the terrain. Routing in asymmetric networks is relatively hard task. In certain cases, it is possible to find routes that exclude asymmetric links, since it is cumbersome to find the return path. The issue of efficient is one of the several challenges encountered in a MANET.

- The other issue is varying the mobility patterns of different nodes. Some other nodes are highly mobile, while others are primarily stationary. It is difficult to predict a node's movement and direction of movement and numerous studies have been performed to evaluate their performance using different simulators.

Characteristics of MANET

Some characteristics of adhoc network are as follows:

- **Dynamic topologies:** nodes are free to move arbitrarily; thus the network topology may be changed randomly and unpredictably and primarily consists of bidirectional links. In some cases where the transmission power of two nodes is different, a unidirectional link may exist.
- **Bandwidth-constrained and variable capacity links:** wireless links continue to have significantly lower capacity than infrastructure networks.
- **Energy-constrained operation:** some or all of the MSs in a MANET may rely on batteries or other exhaustible means for their energy. For these nodes or devices, the most important system design optimization criteria may be energy conservation.
- **Limited physical security:** MANETs are generally more prone to physical security threats than wire line networks. The increased possibility of eavesdropping, spoofing, and denial of services (DoS) attacks should be considered carefully. To reduce security threats, many existing link security techniques are often applied within wireless networks.

Applications of MANET

Some specific applications of ad hoc networks include industrial and commercial applications involving cooperative mobile data exchange. There are many existing and future military networking requirements for robust, IP-compliant data services within mobile wireless communication networks, with many of these networks consist of highly dynamic autonomous topology segments. Advanced features of Mobile ad hoc networks, including data rates compatible with multimedia applications global roaming capability, and coordination with other network structures are enabling new applications.

- **Defense applications:** Many defense applications require on the fly communications set-up, and ad hoc/sensor networks are excellent candidates for use in battlefield management.
- **Crisis management applications:** These arise, for example, as a result of natural disasters in which the entire communication infrastructure is in disarray. Restoring communications quickly is essential.
- **Telemedicine:** The paramedic assisting the victim of a traffic accident in a remote location must access medical records (e.g. X-rays) and may need video conference assistance from a surgeon for an emergency intervention. In fact, the paramedic may need to instantaneously relay back to the hospital the victim's X-rays and other diagnostic tests from the site of the accident.
- **Tele-geoprocessing application:** The combination of GPS, GIS (Geographical Information Systems), and high-capacity wireless mobile systems enables a new type of application referred to as tele- geo processing.
- **Virtual Navigation:** A remote database contains the graphical representation of building, streets, and physical characteristics of a large metropolis. They may also "virtually" see the internal layout of buildings, including an emergency rescue plan, or find possible points of interest.
- **Education via the internet:** Educational opportunities available on the internet or remote areas because of the economic infeasibility of providing expensive last-mile wire line internet access in these areas to all subscribers.
- **Vehicular area network:** This a growing and very useful application of adhoc network in providing emergency services and other information. This is equally effective in both urban and rural setup. The basic and exchange necessary data that is beneficial in a given situation.

Routing

Routing is the process of finding the best path for traffic in a network, or across multiple networks. The role of routing is similar to the road map for a hotel. In both cases, we need to deliver messages at proper location and in an appropriate way.

Routing in a mobile ad-hoc network depends on many factors such as:

- Modeling of the topology,

- Selection of routers,
- Initiation of a route request,
- And specific underlying characteristics that could serve as heuristics in finding the path effectively.

In a MANET, each node or device is expected to serve as a router, and each router is indistinguishable from another in the sense that all routers execute the same routing algorithm to compute paths through the entire network.

Need for Routing

There are following needs for routing:

- Since centralized routing in a dynamic and even for small networks is impossible therefore routing computation must be distributed.
- Route computation should not add many more nodes.
- If any host demands for the route, they must have quick access.
- Maintenance of a global state should not involve in the route computation.
- Each node should care about their destination node to its route and should not be involved in frequent topology updates for those portions of the network that have no traffic.
- Since broadcast can be time consuming for MANETs, it must be avoided as much as possible.
- In routing there must have a backup route when the primary route has become stale.

Routing Classification

Routing protocol can be classified as:

1. Proactive Protocol
2. Reactive Protocol
3. Hybrid Protocol

1. Proactive Protocol

Proactive protocols attempt to evaluate continuously the routes within the network. It means proactive protocol continuously maintain the routing information, so that when a packet needs to be forwarded, the path is known already and can be immediately used. The family of distance vector protocols is an example of proactive scheme.

The advantage of the proactive schemes is that whenever a route is needed, there is negligible delay in determining the route.

Unfortunately, it is a big overhead to maintain routing tables in the MANET environment. Therefore, this type of protocol has following common disadvantages:

- Requires more amounts of data for maintaining routing information.
- Low reaction on re-structuring network and failures of individual nodes.

2. Reactive Protocols

Reactive protocols do not maintain routes but invoke a route determination procedure only on demand or we can say reactive protocols build the routes only on demand. Thus, when a route is required, some sort of global search procedure is initiated. The family of classical flooding algorithms belongs to the reactive protocol group. Examples of reactive ad-hoc network routing protocols include ad hoc on demand distance vector (AODV) and temporally ordered routing algorithm (TORA).

These protocols have the following advantages:

- No large overhead for global routing table maintenance as in proactive protocols.
- Reaction is quick for network restructure and node failure.
  Even though reactive protocols have become the main stream for MANET routing, they still have the following disadvantages:
- Latency time is high in route finding
- Excessive flooding can lead to network clogging.

3. Hybrid Protocols

Hybrid protocols attempt to take advantage of best of reactive and proactive schemes. The basic idea behind such protocols is to initiate route discovery on demand but at a limited search cost. One of the popular hybrid protocols is zone routing protocol (ZRP).

Routing protocols may also be categorized as follows:

1. Table-driven protocols
2. Source initiated on -demand protocols

1. Table-driven routing protocol

- These protocols are called table-driven because each node is required to maintain one or more tables containing routing information on every other node in the network.
- They are **proactive** in nature so that the routing information is always consistent and up to date.
- The protocols respond to changes in network topology by propagating the updates throughput the network so that every node has a consistent view of the network.

The table driven routing protocols are categorized as follows:

Destination - sequenced distance vector routing

- Destination sequenced distance vector routing (DSDV) is a table driven routing protocol for MANET based on Bellman-Ford algorithm.
- DSDV was developed by **C. Perkins and P. Bhagwat in 1994**. The main contribution of the algorithm was that the algorithm works correctly, even in the presence of the loops in the routing table.
- As we know, each mobile node maintains a routing table with a route to every possible destination in the network and the number of hops to the destination.
- Each entry in the table contains a sequence number assigned by the destination node.
- The sequence numbers allow the node to distinguish stale routes from new ones, and help avoid formation of routing loops.
- **A new route broadcast contains:**
  - The destination address.
  - The number of hops required to reach the destination.
  - The sequence number of the information received about the destination and a new sequence number unique to the broadcast.
- If there multiple routes are available for the same destination, the route with the most recent sequence number is used. If two updates have the same sequence number, the route with smaller metric is used to optimize the routing.

For example the routing table of Node A from the above network is:

Destination	Next Hop	No. of Hops	Sequence no.	Install time
A	A	0	A46	001000
B	B	1	B36	001200
C	B	2	C28	001500

Basically the table stores description of all possible paths reachable by node A, along with the hop, number of hops, sequence number and install time.

Advantages

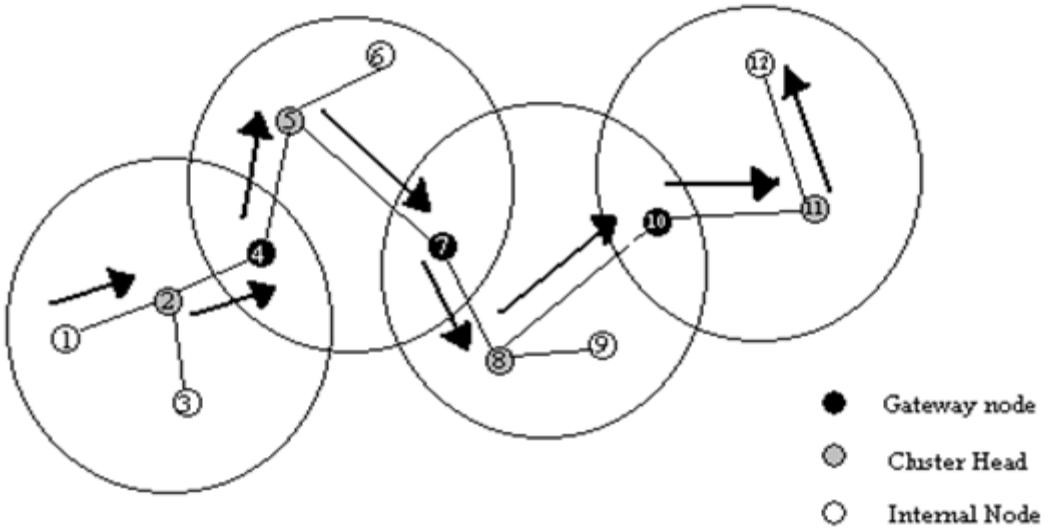
- Destination sequenced distance vector routing was one of the early algorithms available. It is suitable for creating ad-hoc networks with small no. of nodes.

Disadvantage

- Destination sequenced distance vector routing requires a regular update of its routing tables, which uses more battery power and a small amount of bandwidth even when the network is idle.
- This algorithm is not suitable for highly dynamic networks.

Cluster Head gateway switch Routing

- The cluster head (CH) gateway switch routing (CGSR) protocol is different from the destination sequenced distance vector routing in the type of addressing and the network organization scheme employed.
- Instead of a flat network, CGSR uses cluster heads, which control a group of ad hoc nodes and hence achieve a hierarchical framework for code separation among clusters, routing, channel access, and bandwidth allocation.
- Identification of appropriate clusters and selection of cluster heads is quite complex. Once clusters have been defined, it is desirable to use a distributed algorithm within the cluster to elect a node as the cluster head.
- The disadvantage of using a cluster head scheme is that frequent changes adversely affect performance as nodes spend more time selecting a cluster head rather than relaying packets. Hence, the least cluster change (LCC) clustering algorithm is used rather than CH selection every time the cluster membership changes. Using LCC, CHs change only when two CHs come into contact, or when a node moves out of contact with all other CHs.



- In this scheme, each node must maintain a cluster member table (CMT), which stores the destination CH for each node in the network. The cluster member tables are broadcast periodically by the nodes using the DSDV algorithm.
- When a node receives such a table from a neighbor, it can update its own information. As expected, each node also maintains a routing table to determine the next hop required to reach any destination.

Wireless routing protocol (WRP)

The wireless routing protocol is a proactive unicast routing protocol for MANETs. It uses an enhanced version of the distance vector routing protocol, which uses the Bellman - Ford algorithm to calculate paths.

For the wireless routing protocol (WRP) each node maintains 4 tables:

- Distance table
- Routing table
- Link cost table
- Message retransmission list (MRL) table

Each entry in the message retransmission list has a sequence number of the update message, a retransmission counter, an acknowledgment required flag vector with one entry per neighbor, and a list of updates sent in the update message. When

any node receives a hello message from a new node, it adds the new node to its routing table and sends the new node a copy of its routing table. A node must send a message to its neighbors within a certain time to ensure connectivity.

Advantages

- The advantage of WRP is similar to DSDV. In addition, it has faster convergence and adds fewer table updates.

Disadvantage

- The complexity of maintenance of multiple tables demands a large amount of memory and greater processing power from nodes in the MANET.
- Since it suffers from limited scalability therefore WRP is not suitable for highly dynamic and for a very large ad hoc wireless network.

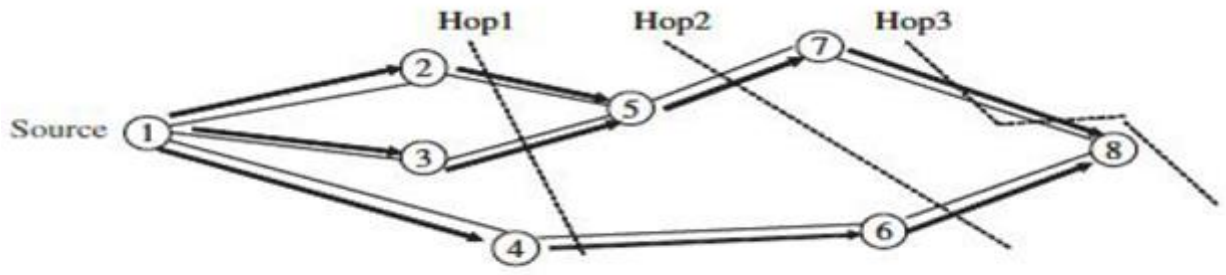
2. Source initiated on -demand protocols

- Source - initiated on demand routing is **reactive** in nature, unlike table driven routing. This type of protocols generates routes only when a source demands it.
- In other words, when a source node requires a route to a destination, the source initiates a route discovery process in the network. This process finishes when a route to the destination has been discovered or all possible routes have been examined without any success.
- The discovered route is maintained by a route maintenance procedure, until it is no longer desired or the destination becomes inaccessible.

The source initiated on demand routing is categorized as follows:

Ad hoc on demand distance vector routing (AODV)

- AODV is a routing protocol for MANETs (mobile ad hoc networks) and other wireless ad hoc networks.
- It is a reactive routing protocol; it means it establishes a route to a destination only on demand.
- AODV routing is built over the DSDV algorithm. It is a significant improvement over DSDV.
- The devices that are not on a particular path do not maintain routing information, nor do they participate in the routing table exchanges.
- When a source requires sending a message to a destination and does not have a valid route to the latter, the source initiates a route discovery process.
- Source sends a route request (RREQ) packet to all its neighbors, the latter forward the request to all their neighbors, and so on, until either the destination or an intermediate mobile (node) with a "fresh enough" route to the destination is reached.

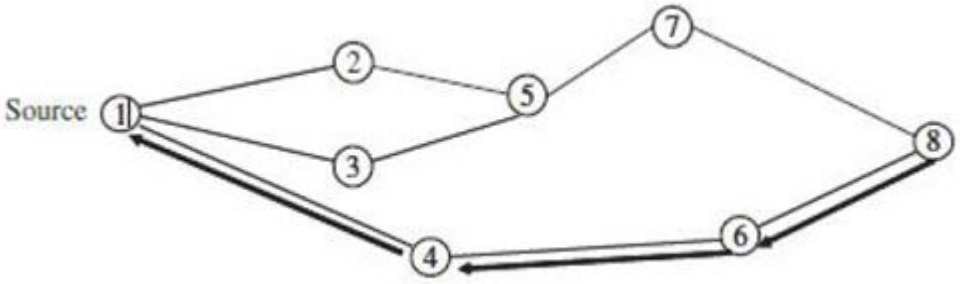


(a) Propagation of route request (RREQ) packet

The above figure illustrates the propagation of the broadcast request (RREQs) across the network. Since in DSDV, destination sequence numbers are used to ensure that all routes are loop free and contain the most recent route information. Each node has a unique sequence number and a broadcast ID, which is incremented each time the node, initiates RREQ.

The broadcast ID, together with the IP address of node, uniquely identifies every RREQ.

Intermediate mobile reply only if they have a route to the destination with a sequence number greater than or at least equal to that contained in the RREQ. To optimize the route performance, intermediate nodes record the address.



(b) Path taken by the route reply (RREP) packet

From the above figure, since RREP (route reply packet) travels back on the reverse path, the nodes on this path set up their forward route entries to point to the node from which RREP had just been received. These forward route records indicate the active forward route. The RREP continues traveling back along the reverse path till it reaches the initiator of the route discovery. Thus, AODV can support only the use of symmetric links.

Dynamic Source Routing (DSR)

- Dynamic source routing is an on-demand routing protocol which is based on source routing.
- It is very similar to AODV in that it forms a route on demand when a transmitting computer requests one. But, it uses source routing instead of relying on the routing table at each intermediate device. Many successive refinements have been made to dynamic source routing.
- This protocol works in two main phases:
    - Route discovery
    - Route maintenance
- When a node has a message to send, it contacts to the route cache to determine whether is it has a route to the destination. If an active route to the destination exists, it is used to send a message.
- Otherwise a node initiates a route discovery by broadcasting a route request packet. The route request stores the destination address, the source address, and a unique identification number.
- Each device that receives the route request checks whether it has a route to the destination. If it does not, it adds its own address to the route record of the packet and then rebroadcasts the packet on its outgoing links.
- To minimize the no. of broadcasts, a mobile rebroadcasts a packet only if it has not seen the packet before and its own address was not already in the route record.

Temporally Ordered Routing Algorithm (TORA)

- **TORA (Temporally Ordered Routing Algorithm)** is a source initiated on demand routing protocol.
- It was invented by **Vincent Park and M. Scott Corson** from university of Maryland in 1997 for wireless ad hoc network.
- TORA is a highly adaptive, efficient, loop-free and scalable routing protocol based on link reversal algorithm.
- The main objective of TORA is to limit message propagation in the highly dynamic mobile computing environment. It means, it is designed to reduce communication overhead by adapting local topological changes in ad hoc network. Another main feature of TORA routing protocol is the localization of control packets to a small region (set of nodes) near the occurrence of a topological changes due to route break. Hence, each node of the network required to contain its local routing and topology information about adjacent nodes.
- TORA supports multiple routes to transmit data packet between source and destination nodes of mobile ad hoc network. In short, TORA exhibits multipath routing capability.
- The TORA's operation can be compared to that of water flowing downhill toward a sink node through a grid of tubes that model the routes in the real world network. The tube junctions represent the nodes, the tube themselves represent the route links between the nodes, the tube's water represents the packets flowing between nodes through the route links toward the destination, as shown in the figure:

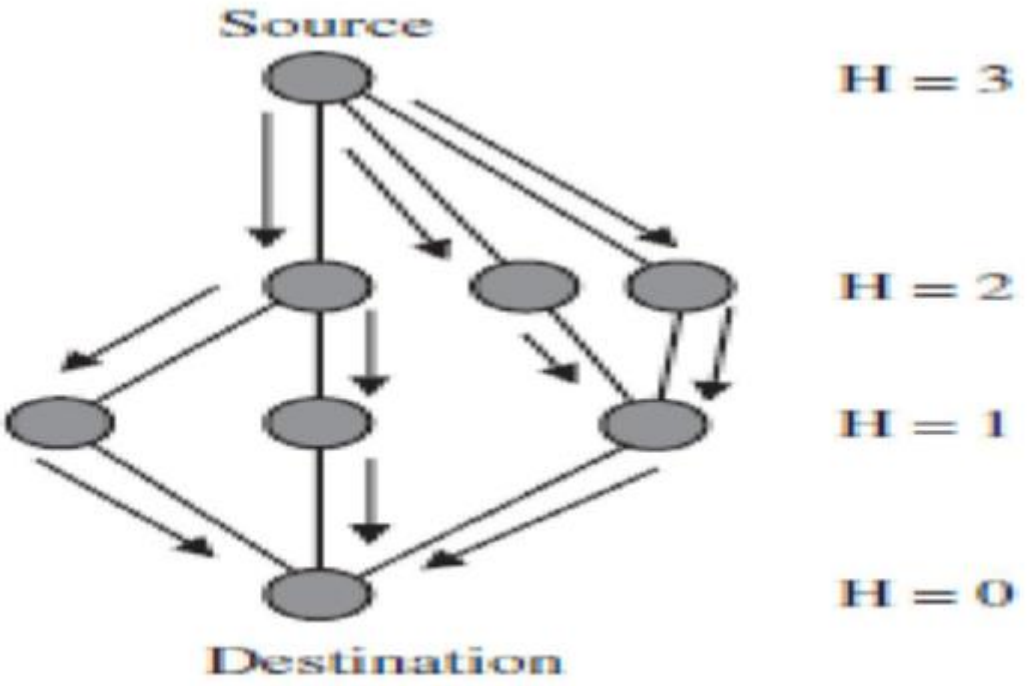


Fig: TORA height metric

- Considering the data flow to be downhill, each node has a height with respect to the destination node. The analogy also makes it easy to correct routes in case of link failure or error.
- One of the biggest advantages of TORA is that it can operate smoothly in a highly dynamic mobile environment. It provides multiple paths for any source-destination pair. For this purpose, teach node must maintain routing information about their one-hop neighbors.
- **TORA works in three main phases:**
  - **Route creation:** Route creation from source to destination.
  - **Route maintenance:** Maintenance of the route.
  - **Route erasure:** Erasing of the route when the route is no longer valid.
- TORA attempts to build a separate directed acyclic graph (DAG) by each node to every destination. When a route to a particular destination is required, the source node broadcasts a QUERY packet containing the address of the destination. The route query propagates via the network till it reaches either the destination or an intermediate node containing the route to the destination.
- **TORA contains a quintuple metric which consists of:**
  - Logical time of link failure.
  - Unique ID of the node that defines the new reference level.
  - A reflection indicator bit.
  - A propagation ordering parameter.
  - Unique ID of the node.

Hybrid Protocol - Zone Routing Protocols

Hybrid protocols attempt to take advantage of best of reactive and proactive schemes. The basic idea behind such protocols is to initiate route discovery on demand but at a limited search cost. One of the popular hybrid protocols is zone routing protocol (ZRP).

Zone routing protocol (ZRP)

- Zone routing protocol is a hybrid of reactive and proactive protocols. It combines the advantage of both reactive and proactive schemes.
- ZRP was invented by Zygmunt Haas of Cornell University. Zone routing protocol finds loop free routes to the destination.

- ZRP divides the network into zones of variable size; size of the zone is determined radius of length ?, where the ? is the number of hops or nodes to the perimeter of the zone and not the physical distance.
- In other words we can say that, the neighborhood of the local node is called a routing zone. Specifically, a routing zone of the node is defined as the set of nodes whose minimum distance in hops from the node is no greater than the zone radius.
- A node maintains routes to all the destinations proactively in the routing zone. It also maintains its zone radius, and the overlap from the neighboring routing zones.
- To create a routing zone, the node must identify all its neighbors first which are one hop away and can be reached directly.
- The Process of neighbor discovery is governed by the NDP (Neighbor Discovery Protocol), a MAC level scheme. ZRP maintains the routing zones through a proactive component called the intra-zone routing protocol (IARP) and is implemented as a modified distance vector scheme. Thus IARP is responsible for maintaining routes within the routing zone.
- Another protocol called the inter-zone routing protocol (IERP) which is responsible for maintaining and discovering the routes to nodes beyond the routing zone.
- This type of process uses a query - response mechanism on-demand basis. IERP is more efficient than standard flooding schemes.
- When a source node send data to a destination which is not in the routing zone, the source initiates a route query packet.
- The latter identified by the tuple <source node ID, request number>. This request is then broadcasted to all the nodes in the source nodes periphery.
- When a node receives this query, it adds its own identification number (ID) to the query. Thus the sequence of recorded nodes presents a route from the current routing zone. Otherwise, if the destination is in the current routing zone of the node, a route reply is sent back to the source along the reverse from the accumulated record.
- A big advantage of this scheme is that a single route request can result in multiple replies of route. The source can determine the quality of these multiple routes based on such parameter as hop count or traffic and choose the best route to be used.

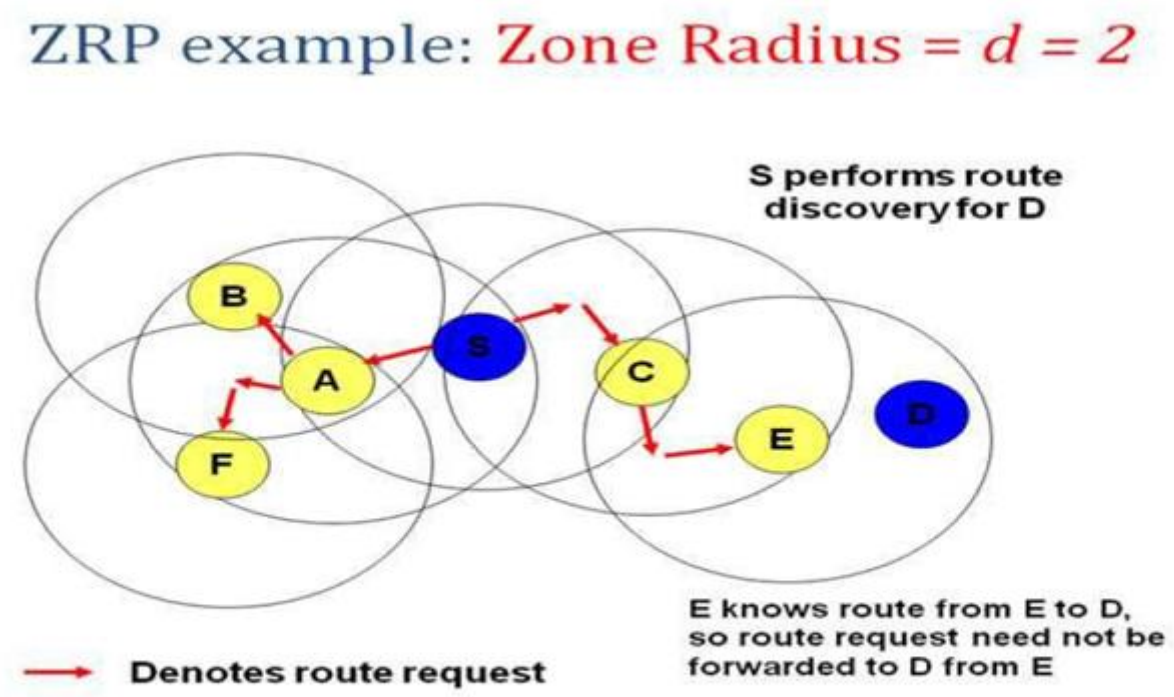


Fig: - Zone Routing Protocols

Introduction to Mobile Ad hoc Networks (MANETs)

Overview of Adhoc Network

Communication between various devices makes it possible to provide unique and innovative services. Although this inter-device communication is a very powerful mechanism, it is also a complex and clumsy mechanism, leading to a lot of complexity in the present-day systems. This not only makes networking difficult but limits its flexibility as well. Many standards exist today for connecting various devices. At the same time, every device has to support more than one standard to make it interoperable between different devices.

Wireless networks can be classified in two types:

- Infrastructure dependent network
- Ad hoc wireless networks (Infrastructure less network)

Ad hoc network is a kind of Wireless Network which can be designed for fulfilling particular purposes that is served by establishment of the whole set up on the fly. Example types of an Ad hoc network are –

- Mobile Ad hoc Network (MANET)
- Vehicular Ad hoc Network (VANET)
- Wireless Sensor Network (WSN)

Mobile Ad hoc Networks (MANETs)

Mobile ad hoc network is a collection of independent mobile nodes that can communicate to each other via radio waves. The mobile nodes can directly communicate to those nodes that are in radio range of each other, whereas others nodes need the help of intermediate nodes to route their packets. These networks are fully distributed, and can work at any place without the aid of any infrastructure. This property makes these networks highly robust.

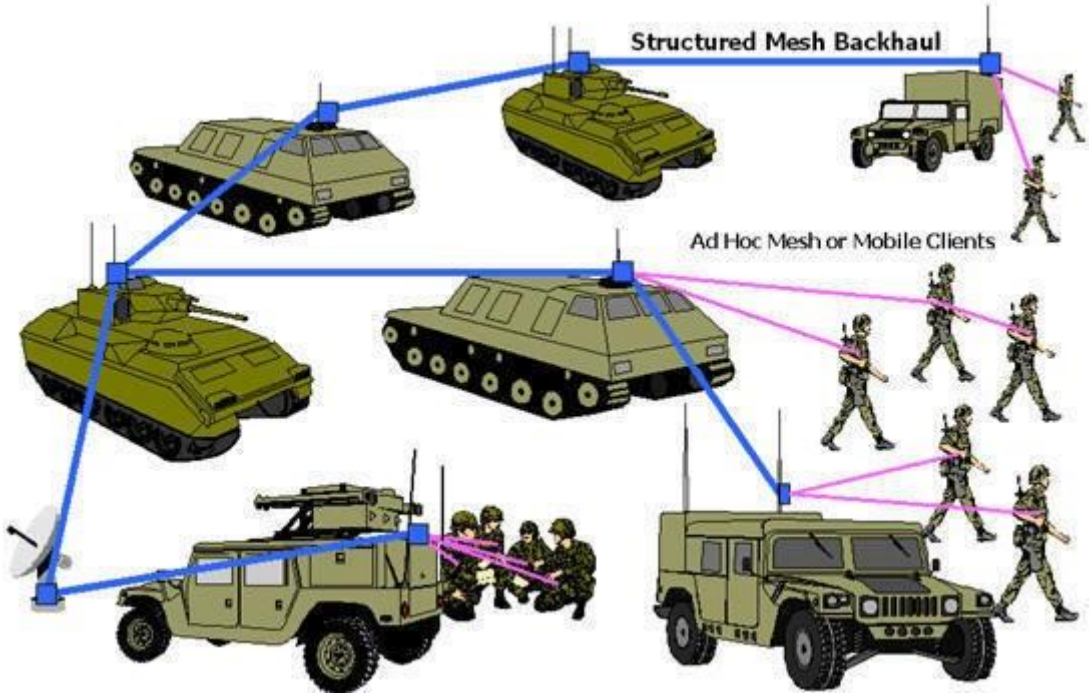
Mobile Ad hoc Networks are autonomous systems which comprise a collection of mobile nodes that use wireless transmission for communication. They are self-organized, self-configured, and self-controlled infrastructure-less networks. This type of network can be set up or deployed anywhere and anytime because it poses very simple infrastructure setup and no or minimal central administration. These networks are mainly used by community users such as military, researchers, business, students, and emergency services. Nodes are using Internet Protocol and IP addresses are assigned to each of the nodes. Individual nodes discover dynamically which other nodes they can communicate with.

Features / Characteristics of MANETs

- Rapidly deployable, self-organized, self-configured, and self-controlled infrastructure- less networks
- Wireless links
- Every computer or device (node) is a router as well as end host
- Nodes are mobile, topology can be very dynamic
- Nodes must be able to relay traffic since communicating nodes might be out of range
- Can be a standalone network or it can be connected to external networks(Internet)
- Radio communication – shared medium

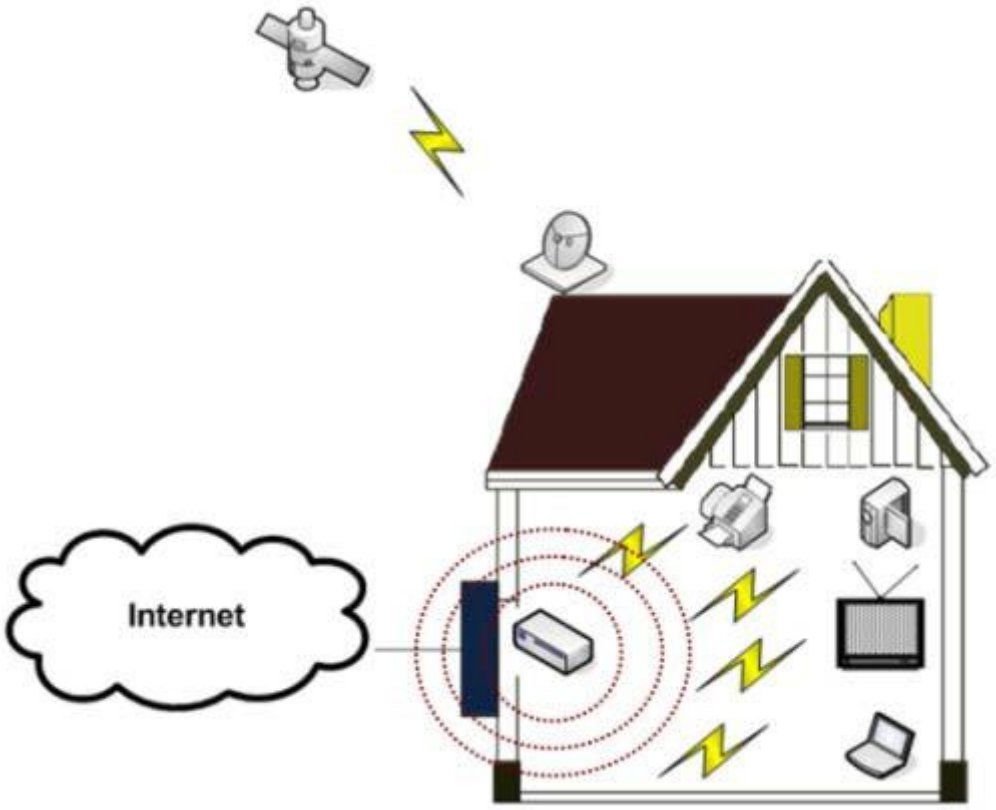
Applications of Wireless Ad hoc Network

Military Purpose & Rescue Mission



Military networks are designed to maintain a low probability of intercept and/or a low probability of detection. Hence, nodes prefer to radiate as little power as necessary and transmit as infrequently as possible, thus decreasing the probability of detection or interception. A lapse in any of these requirements may degrade the performance and dependability of the network.

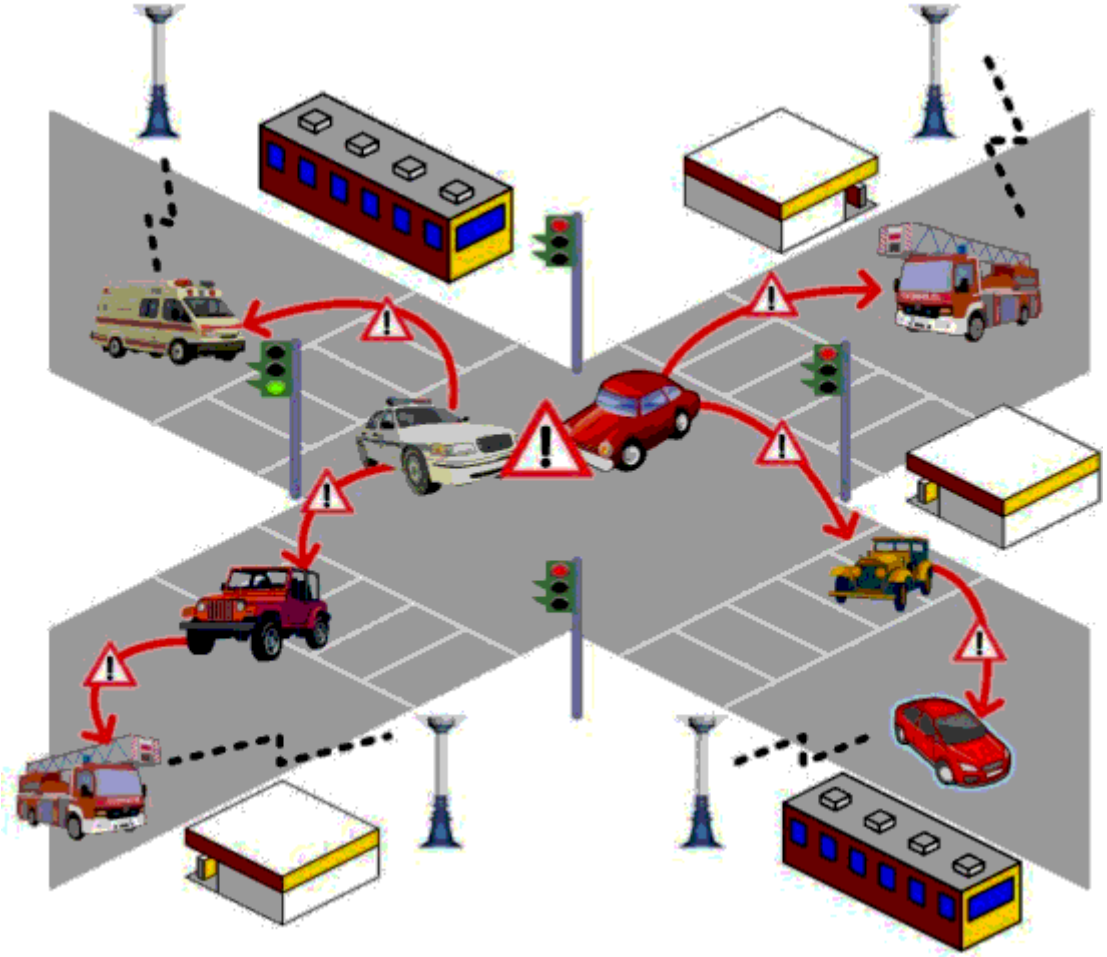
Home Networking



Home networking is the collection of electronic products and systems, enabling remote access and control of those products and systems, and any available contents such as music, video or data. In home networks,

a user wants to accomplish data communication in ways that are affordable, reliable, easy to learn, and easy to use.

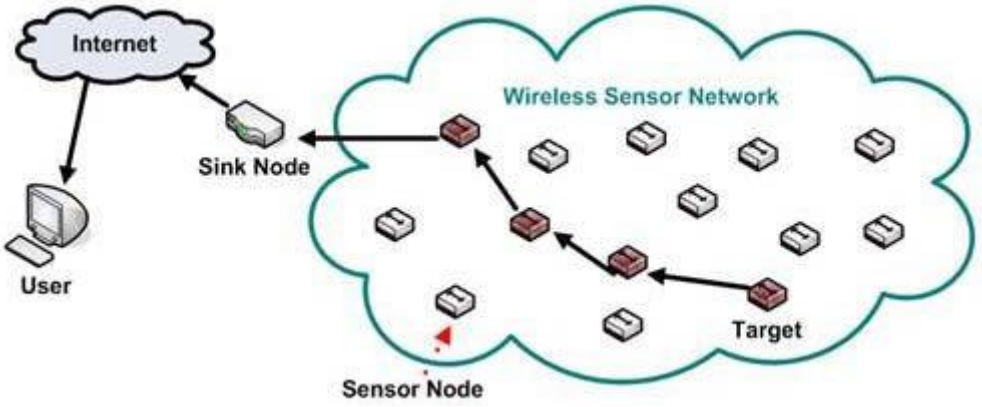
Vehicular Ad hoc network (VANET)



Vehicular Ad hoc Network (VANET) is a subset of mobile ad hoc network, which supports data communications among nearby vehicles and between vehicles and nearby fixed infrastructure, and generally represented as roadside entities. Depending on the range of data communications, nodes in VANET communicate among themselves in type of short-range (vehicle-to-vehicle) or medium-range (vehicle-to-roadside) communications.

In addition, the major application view of VANETs includes real-time and safety applications. Non-safety applications include real-time traffic congestion and routing information, high-speed tolling, mobile infotainment, traffic condition monitoring, and many others. Vehicular safety applications include emergency, collision, car accident, and other safety warnings.

Sensor network



Sensor networks are composed of a large number of small nodes with sensing, computation, and wireless communication capabilities. In sensor networks, sensor nodes are usually scattered and the position of sensor nodes needs not be predetermined. It means that sensor network protocols and algorithms must provide self-organizing capabilities. The features of sensor networks provide a wide range of applications such as health, military, and home. The realization of these and other sensor network applications require wireless ad hoc networking techniques.

Conferencing

Ad hoc network is widely used to create a network in place like Conferences, where for a short time of period the network is formed. One main pc is wired to which the internet facility is given and the rest of the computers or laptops are connectionless. One main computer forms the Ad hoc network and tries to connect with all the rest of the computers to make communication and data transfer possible.

Campus use

Ad hoc Network is also used for Campus use. It's a small area in which network is created or formed whenever it is required. Professor from one building can communicate with the professor of the other building. No need to set up any wired connection.

Limitations & Current Challenges of MANETs

Each node must have full performance. Throughput is affected by system loading. Reliability requires a sufficient number of available nodes. Sparse networks can have problems. Large networks can have excessive latency (time delay), which affects some applications.

The other considerable challenges are pointed out as follows:

- Multihop operation requires a routing mechanism for mobile nodes
- Internet access mechanisms
- Self-configuring networks requires an address allocation mechanism
- Mechanism to detect and act on, merging of existing networks
- Dynamic topology maintenance
- Scaling to large networks
- Limited energy and computing resources
- Security mechanisms

Advantages & Disadvantages of Mobile Ad hoc Networks

It is independent from central network administration. It is having ability of Self-configuring in which nodes are also routers. It is having ability of Self-healing through continuous re configuration. It is having ability of Scalability in which it accommodates the addition of more nodes. It is Flexible which is similar to being able to access the Internet from many different locations. If a Computer is connected with internet and is shut down due to power failure so all the pc which is part of Ad hoc network lost their internet connection. At least one PC should have wired internet connection to provide internet to all other wireless pc.

Routing Protocols for MANETs

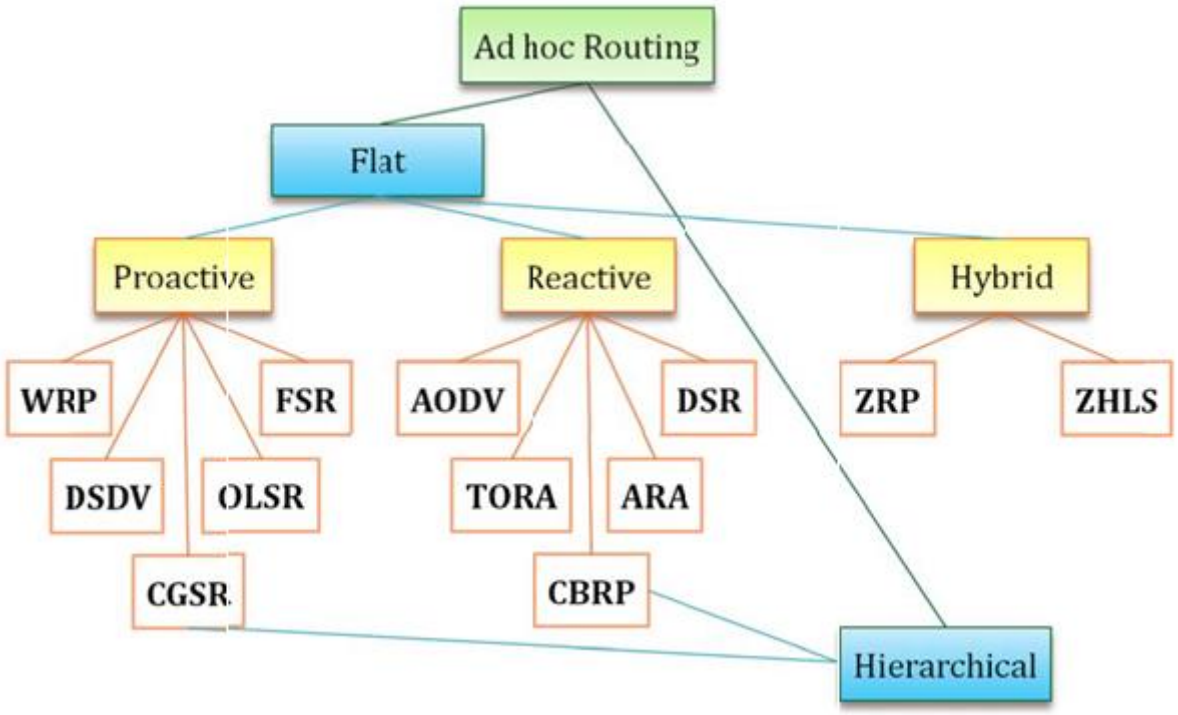
An ad hoc routing protocol is a convention, or standard, that controls how nodes decide which way to route packets between computing devices in a mobile ad hoc network. In ad hoc networks, nodes do not start out familiar with the topology of their networks; instead, they have to discover it. The basic idea is that a new node may announce its presence and should listen for announcements broadcast by its neighbors. Each node learns about nodes nearby and how to reach them, and may announce that it, too, can reach them.

Routing Architecture

The routing architecture of a self-organized network can be either hierarchical or flat. In most self-organized networks, the hosts will be acting as independent routers, which imply that routing architecture should conceptually be flat, that is, each address serves only as an identifier and does not convey any information about one host that is topologically located with respect to any other node. In a flat self-organized network, the mobility management is not necessary because all of the nodes are visible to each other via routing protocols. In flat routing algorithms, the routing tables have entries to all hosts in the self-organized network. However, a flat routing algorithm does not have good scalability. The routing overhead increases rapidly when the network becomes larger. Hence, to control channel reuse spatially (in terms of frequency, time, or spreading code) and reduce routing information overhead, some form of hierarchical scheme should be employed. Clustering is the most common technique employed in hierarchical routing architectures. The idea behind hierarchical routing is to divide the hosts of a self-organized network into a number of overlapping or disjoint clusters. One node is elected as cluster-head for each cluster. cluster. This cluster-head maintains the membership information for the Nodes that are not cluster-heads will, henceforth, be referred to as “ordinary nodes”. When an ordinary node wants to send a packet, the node can send the packet to the cluster-head that routes the packet toward the destination. Cluster-head Gateway Switch Routing (CGSR) and Cluster- Based Routing Protocol (CBRP) belong to this type of routing scheme. Hierarchical routing involves cluster, address, and mobility management.

Classification of Routing Protocols

One of the most popular methods to distinguish mobile ad hoc network routing protocols is based on how routing information is acquired and maintained by mobile nodes. Using this method, mobile ad hoc network routing protocols can be divided, as discussed above, into proactive routing, reactive routing, and hybrid routing.



Proactive or Table-driven Routing Protocol

A proactive routing protocol is also called a "table-driven" routing protocol. Using a proactive routing protocol, nodes in a mobile ad hoc network continuously evaluate routes to all reachable nodes and attempt to maintain consistent, up-to-date routing information. Therefore, a source node can get a routing path immediately if it needs one. In proactive routing protocols, all nodes need to maintain a consistent view of the network topology. When a network topology change occurs, respective updates must be propagated throughout the network to notify the change. Most proactive routing protocols proposed for mobile ad hoc networks have inherited properties from algorithms used in wired networks.

To adapt to the dynamic features of mobile ad hoc networks, necessary modifications have been made on traditional wired network routing protocols. Using proactive routing algorithms, mobile nodes proactively update the network state and maintain a route regardless of whether data traffic exists or not, and the overhead to maintain up-to-date network topology information is high.
Examples are –

- Wireless Routing Protocol (WRP)
- Destination-Sequenced Distance Vector (DSDV)
- Optimized Link State Routing (OLSR) Protocol
- Fisheye State Routing (FSR)
- Topology Broadcast Reverse Forwarding (TBRF)

Reactive or On-demand Routing Protocol

Reactive routing protocols for mobile ad hoc networks are also called "on-demand" routing protocols. In a reactive routing protocol, routing paths are searched only when needed. A route discovery operation invokes a route-determination procedure. The discovery procedure terminates when either a route has been found or no route is available after examination for all route permutations. In a mobile ad hoc network, active routes may be disconnected due to node mobility. Therefore, route maintenance is an important operation of reactive routing protocols.

Compared to the proactive routing protocols for mobile ad hoc networks, less control overhead is a distinct advantage of the reactive routing protocols. Thus, reactive routing protocols have better scalability than proactive routing protocols in mobile ad hoc networks. However, when using reactive routing protocols, source nodes may suffer from long delays for route searching before they can forward data packets.

Examples are –

- Ad hoc On-Demand Distance Vector (AODV)
- Dynamic Source Routing (DSR) Protocol
- Temporally Ordered Routing Algorithm (TORA)
- Cluster-Based Routing Protocol (CBRP)

Various parameter comparison of on demand and table driven routing protocol

Parameters	On Demand	Table Driven
Availability of Routing Information	Available when needed	Always available regardless of need
Routing Philosophy	Flat	Mostly Flat except for CGSR
Periodic route updates	Not Required	Yes
Coping with Mobility	Using Localized route discovery	Inform other nodes to achieve consistent routing tables
Signaling Traffic Generated	Grows with increasing mobility of active nodes	Greater than that of On Demand Routing
QoS Support	Few Can Support QoS	Mainly Shortest Path as QoS Metric

Hybrid Routing Protocol

Hybrid routing protocols are proposed to combine the merits of both proactive and reactive routing protocols and overcome their shortcomings. Normally, hybrid routing protocols for mobile ad hoc networks exploit hierarchical network architectures. The proper proactive routing approach and reactive routing approach are exploited in different hierarchical levels, respectively.

Examples are –

- Zone Routing Protocol (ZRP)
- Zone-Based Hierarchical Link State (ZHLS)
- Distributed Spanning Trees Based Routing Protocol (DST)
- Distributed Dynamic Routing (DDR) Protocol

Routing Protocols in Mobile Ad hoc Networks (MANETs)

Classification of Routing Protocols

An ad hoc routing protocol is a convention, or standard, that controls how nodes decide which way to route packets between computing devices in a mobile ad hoc network. In ad hoc networks, nodes do not start out familiar with the topology of their networks; instead, they have to discover it. The basic idea is that a new node may announce its presence and should listen for announcements broadcast by its neighbors. Each node learns about nodes nearby and how to reach them, and may announce that it, too, can reach them.

One of the most popular methods to distinguish mobile ad hoc network routing protocols is based on how routing information is acquired and maintained by mobile nodes. Using this method, mobile ad hoc network routing protocols can be divided, as discussed above, into proactive routing, reactive routing, and hybrid routing.

Optimized Link State Routing (OLSR) Protocol

OLSR is a table-driven, proactive routing protocol and is based on traditional link state routing algorithm. The typical operation of it is that each node locally monitors the network situation, and periodically floods its local view over the network. By combining all received local views, each node can get a complete picture of the network, and calculate shortest paths to all destinations based on it. The main advantage of link state routing is that it converges faster to correct routing information. However, it is not more robust than distance vector routing (things can go quite wrong when routing updates get lost), and is not necessarily more scalable, because, even though routing update messages are smaller, they need to be flooded over the whole network, and nodes need to locally build a complete picture of the whole network.

Each node selects a set of its neighbor nodes as “multipoint relays” (MPR). In OLSR, only nodes, selected as such MPRs, are responsible for forwarding control traffic, intended for diffusion into the entire network. MPRs provide an efficient mechanism for flooding control traffic by reducing the number of transmissions required. Nodes, selected as MPRs, also have a special responsibility when declaring link state information in the network. Indeed, the only requirement for OLSR to provide shortest path routes to all destinations is that MPR nodes declare link state information for their MPR selectors. Additional available link state information may be utilized, for example for redundancy. Nodes which have been selected as multipoint relays by some neighbor node(s) announce this information periodically in their control messages. Thereby, a node announces to the network that it has reachability to the nodes which have selected it as an MPR.

As a proactive protocol, OLSR is also suitable for scenarios where the communicating pairs change over time: no additional control traffic is generated in this situation because routes are maintained for all known destinations at all times.

Multipoint Relays (MPRs)

The idea of multipoint relays is to minimize the overhead of flooding messages in the network by reducing redundant retransmissions in the same region. Each node in the network selects a set of nodes in its symmetric one-hop neighborhood, which may retransmit its messages. This set of selected neighbor nodes is called the MPR set of that node. The neighbors of node N which are not in its MPR set receive and process broadcast messages but do not retransmit broadcast messages received from node N. Each node

selects its MPR set from among its one-hop symmetric neighbors. This set is selected such that it covers all symmetric strict two-hop nodes. The MPR set of N, denoted as MPR (N), is then an arbitrary subset of the symmetric one-hop neighborhood of N which satisfies the following condition: every node in the symmetric strict two-hop neighborhood of N must have a symmetric link toward MPR (N). Each node maintains information about the set of neighbors that have selected it as an MPR. This set is called the "MPR selector set" of a node. A node obtains this information from periodic Hello messages received from the neighbors. Upon receipt of this MPR selector information, each node calculates and updates its route to each destination.

Protocol Functioning

OLSR is modularized into a "core" of functionality, which is always required, for the protocol to operate, and a set of auxiliary functions. The core specifies, in its own right, a protocol able to provide routing in a stand-alone MANET. Each auxiliary function provides additional functionality, which may be applicable in specific scenarios (e.g., in case a node is providing connectivity between the MANET and another routing domain). All auxiliary functions are compatible, to the extent where any (sub-) set of auxiliary functions may be implemented with the core. Furthermore, the protocol allows heterogeneous nodes—that is, nodes which implement different subsets of the auxiliary functions—to coexist in the network. The purpose of dividing the functioning of OLSR into core functionality and a set of auxiliary functions is to provide a simple and easy-to-comprehend protocol, and to provide a way of only adding complexity where specific additional functionality is required.

Core Functioning

The core functionality of OLSR specifies the behavior of a node, equipped with OLSR interfaces participating in the MANET and running OLSR as a routing protocol. This includes a universal specification of OLSR protocol messages and their transmission through the network, as well as link sensing, topology diffusion, and route calculation. Specifically, the core is made up from the following components.

Packet Format and Forwarding

A universal specification of the packet format and an optimized flooding mechanism serves as the transport mechanism for all OLSR control traffic.

Link Sensing

Link sensing is accomplished through periodic emission of Hello messages over the interfaces through which connectivity is checked. A separate Hello message is generated for each interface. Resulting from link sensing is a local link set describing links between "local interfaces" and "remote interfaces," that is, interfaces on neighbor nodes. If sufficient information is provided by the link layer, this may be utilized to populate the local link set instead of a Hello message exchange.

Neighbor Detection

Given a network with only single interface nodes, a node may deduct the neighbour set directly from the information exchanged as part of link sensing. In a network with multiple interface nodes, additional

information is required to map interface addresses to main addresses (and, thereby, to nodes). This additional information is acquired through Multiple Interface Declaration (MID) messages.

MPR Selection and MPR Signaling

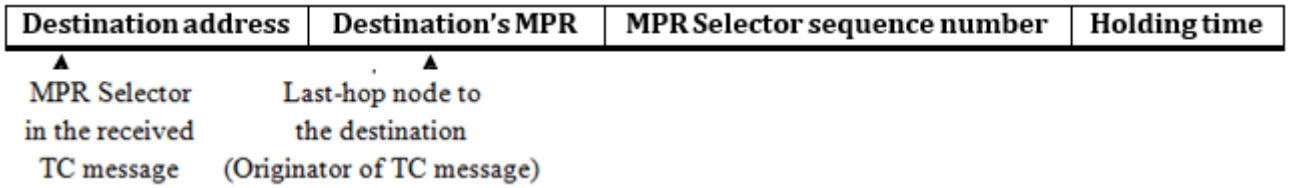
The objective of MPR selection is for a node to select a subset of its neighbours such that a broadcast message, retransmitted by these selected neighbors, will be received by all nodes two hops away. The MPR set of a node is computed such that it, for each interface, satisfies this condition. The information required to perform this calculation is acquired through the periodic exchange of Hello messages.

The MPR selection Algorithm follows below mentioned two steps:

- **Goal:** Select in the 1-neighborhood of u (N1(u)) a set of nodes as small as possible which covers the whole 2-neighborhood of u(N2(u))
- **Step -1:** Select nodes of N1(u) which cover isolated point of N2(u)
- **Step-2:** Select among the nodes of N1(u) not selected at the first step, the node which covers the highest number of points of N2(u) and go on till every points of N2(u) are covered.

Topology Control Message Diffusion

Topology control (TC) messages are diffused with the purpose of providing each node in the network with sufficient link state information to allow route calculation.



TC message contains MPR selector table and the sequence number. Each node maintains a Topology Table based on TC messages which let it to calculate the Routing Table. Upon receipt of a TC message, if there is some entry to the same destination with higher Sequence Number, the TC message is ignored but if there is some entry to the same destination with lower Sequence Number, the topology entry is removed and the new one is recorded. If the entry is the same as in the TC message, then the holding time of this entry is refreshed.

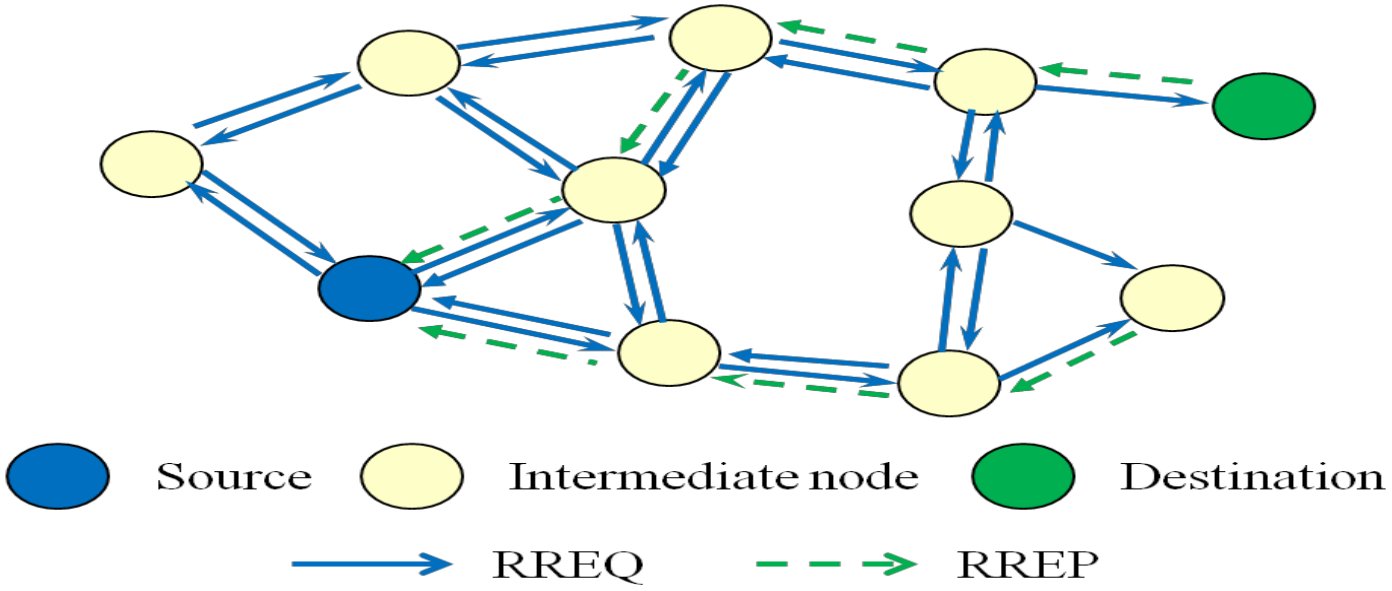
Route Calculation

Given the link state information acquired through periodic message exchange, as well as the interface configuration of the nodes, the routing table for each node can be computed. A routing table contains Destination address, next hop address and Distance. It is recalculated after every change in neighborhood table or in topological table.

Ad hoc On demand Distance Vector (AODV) Protocol

AODV is an on-demand routing protocol, which initiates a route discovery process only when desired by a source node. When a source node wants to send data packets to a destination node but cannot find a route in its routing table, it broadcasts a Route Request (RREQ) message to its neighbors. Its neighbors

then rebroadcast the RREQ message to their neighbors if they do not have a fresh enough route to the destination node. This process continues until the RREQ message reaches the destination node or an intermediate node that has a fresh enough route. Every node has its own sequence number and RREQ ID. Below figure depicts the fundamental Route Discovery Process of AODV.



AODV uses sequence numbers to guarantee that all routes are loop-free and contain the most recent routing information. RREQ ID in conjunction with source IP address uniquely identifies a particular RREQ message. The destination node or an intermediate node only accepts the first copy of a RREQ message, and drops the duplicated copies of the same RREQ message. After accepting a RREQ message, the destination or intermediate node updates its reverse route to the source node using the neighbor from which it receives the RREQ message. The reverse route will be used to send the corresponding Route Reply (RREP) message to the source node. Meanwhile, it updates the sequence number of the source node in its routing table to the maximum of the one in its routing table and the one in the RREQ message.

When the source or an intermediate node receives a RREP message, it updates its forward route to the destination node using the neighbor from which it receives the RREP message. It also updates the sequence number of the destination node in its routing table to the maximum of the one in its routing table and the one in the RREP message. A Route Reply Acknowledgement (RREP-ACK) message is used to acknowledge receipt of a RREP message. Though not required, AODV may utilize the HELLO message to maintain the local connectivity of a node. Route maintenance is done with Route Error (RERR) messages. If a node detects a link break in an active route, it sends out a RERR message to its upstream neighbors that use it as the next hop in the broken route. When a node receives a RERR message from its neighbor, it further forwards the RERR message to its upstream neighbors.

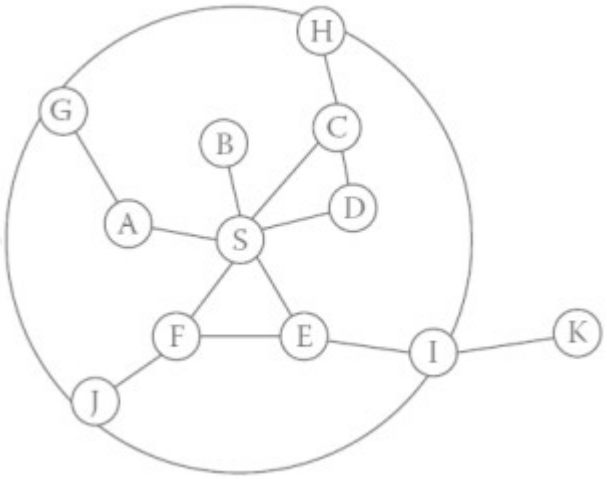
AODV is a stateless protocol; the source node or an intermediate node updates its routing table if it receives a RREP message, regardless of whether it has sent or forwarded a corresponding RREQ message before. If it cannot find the next hop in the reverse routing table, it simply drops the RREP message. Otherwise, it unicasts the RREP message to the next hop in the reverse route. In general, a node may update the sequence numbers in its routing table whenever it receives RREQ, RREP, RERR, or RREP-ACK messages from its neighbors.

Zone Routing Protocol (ZRP)

The ZRP Protocol being a hybrid protocol combines the advantages of the proactive and reactive approaches by maintaining an up-to-date topological map of a zone centered on each node. Within the zone, routes are immediately available. For destinations outside the zone, ZRP employs a route discovery procedure, which can benefit from the local routing information of the zones. Proactive routing uses excess bandwidth to maintain routing information, while reactive routing involves long route request delays. Reactive routing also inefficiently floods the entire network for route determination. ZRP aims to address the problems by combining the best properties of both approaches.

Architecture

The Zone Routing Protocol, as its name implies, is based on the concept of zones. A routing zone is defined for each node separately, and the zones of neighboring nodes overlap. The routing zone has a radius ρ expressed in h hops. The zone thus includes the nodes, whose distance from the node in question is at most h hops. An example of a routing zone is shown below, where the routing zone of S includes the nodes A–I, but not K. In the illustrations, the radius is marked as a circle around the node in question. It should, however, be noted that the zone is defined in hops, not as a physical distance. The nodes of a zone are divided into peripheral nodes and interior nodes. Peripheral nodes are nodes whose minimum distance to the central node is exactly equal to the zone radius ρ. The nodes whose minimum distance is less than ρ are interior nodes.



In above figure of a routing zone having radius 2, the nodes A–F are interior nodes, the nodes G– J are peripheral nodes, and the node K is outside the routing zone. Note that node H can be reached by two paths, one with a length of two hops and one with a length of three hops. The node is, however, within the zone, because the shortest path is less than or equal to the zone radius. The number of nodes in the routing zone can be regulated by adjusting the transmission power of the nodes. Lowering the power reduces the number of nodes within direct reach and vice versa. The number of neighboring nodes should be sufficient to provide adequate reachability and redundancy.

On the other hand, a too large coverage results in many zone members, and the update traffic becomes excessive. Further, large transmission coverage adds to the probability of local contention. ZRP refers to the locally proactive routing component as the Intrazone Routing Protocol (IARP). The globally reactive routing component is named the Interzone Routing Protocol (IERP). IERP and IARP are not specific routing

protocols. Instead, IARP is a family of limited-depth, proactive, link state routing protocols. IARP maintains routing information for nodes that are within the routing zone of the node. Correspondingly, IERP is a family of reactive routing protocols that offer enhanced route discovery and route maintenance services based on local connectivity monitored by IARP.

The fact that the topology of the local zone of each node is known can be used to reduce traffic when global route discovery is needed. Instead of broadcasting packets, ZRP uses a concept called "bordercasting." Bordercasting utilizes the topology information provided by IARP to direct query requests to the border of the zone. The bordercast packet delivery service is provided by the Bordercast Resolution Protocol (BRP). BRP uses a map of an extended routing zone to construct bordercast trees for the query packets. Alternatively, it uses source routing based on the normal routing zone. By employing query control mechanisms, route requests can be directed away from areas of the network that already have been covered. To detect new neighbor nodes and link failures, ZRP relies on a Neighbor Discovery Protocol (NDP) provided by the MAC layer. NDP transmits Hello beacons at regular intervals. Upon receiving a beacon, the neighbor table is updated. Neighbors for which no beacon has been received within a specified time are removed from the table.

Routing

A node that has a packet to send first checks whether the destination is within its local zone using information provided by IARP. Reactive routing is used if the destination is outside the zone. The reactive routing process is divided into two phases: the route request phase and the route reply phase. In the route request phase, the source sends a route request packet to its peripheral nodes using BRP. If the receiver of a route request packet knows the destination, it responds by sending a route reply back to the source. Otherwise, it continues the process by border-casting the packet. The reply is sent by any node that can provide a route to the destination. To be able to send the reply back to the source node, routing information must be accumulated when the request is sent through the network. The information is recorded either in the route request packet or as next-hop addresses in the nodes along the path. In the first case, the nodes forwarding a route request packet append their address and relevant node or link metrics to the packet. When the packet reaches the destination, the sequence of addresses is reversed and copied to the route reply packet. The sequence is used to forward the reply back to the source. In the second case, the forwarding nodes record routing information as next-hop addresses, which are used when the reply is sent to the source.

The zone radius is an important property for the performance of ZRP. If a zone radius of one hop is used, routing is purely reactive and border-casting degenerates into flood searching. If the radius approaches infinity, routing is reactive. The selection of radius is a trade-off between the routing efficiency of proactive routing and the increasing traffic for maintaining the view of the zone.