Shobhit
Institute of Engineering & Technology
Deemed to-be-University
EDUCATION EMPOWERS

Lecture

on

Mobile Computing

By

Dr. Mamta Bansal Rajshree

Professor

Shobhit Institute of Engineering & Technology

(Deemed to be University)

Unit-I

What is Mobile Computing?

Mobile Computing Technology is a **mobile technology** that allows transmission of data of any kind such as written document, voice, picture, video, etc. from one wireless-enabled device to another wireless-enabled device.

These wireless-enabled devices can be a computer, laptop, notepad, mobile phone, pager, sensor, embedded controller, etc.

Mobile computing Technology definition

Mobile Computing Technology Definition 1

"Mobile Computing refers to mobile technology that allows transmission of data, from one wireless-enabled device to another wireless-enabled device without the use of a fixed physical link."

Mobile Computing Technology Definition 2

"Mobile computing can be defined as a human and computer communication while a human is mobile.

It enables the transmission of text, audio, video, and images, etc. The technology includes mobile communication via mobile hardware and mobile software."

Mobile Computing Technology Definition 3

Mobile Computing Technology allows users to use the technology in **mobile** environments. Mobile Computing technology is based on the use of

wireless **computing** and communication devices, such as smart **mobile** phones, tablets, wearable computing devices.

Therefore, the meaning of mobile computing is to communicate any kind of data among users located at any place throughout the world equipped with mobile devices like mobile phones, laptops, notepads, etc.

What are the features of mobile computing Technology

The most important features of mobile computing technology are as follows-

- **Mobility**

Mobility is a fundamental feature of mobile computing technology.

The technology allows transferring data to the user while the user is in a mobile environment means roaming from one place to another place.

It not only supports communication with mobile users but also when devices are mobile doesn't matter the user is the same or different.

- **Wireless Connectivity**

Wireless connectivity is another essential feature of mobile communication.

It means all the mobile devices such as mobile phones, laptops, tablets must have wireless connectivity enabled otherwise no data transfer is possible.

These devices must have all the hardware and software that are essential to the Wi-Fi enabled device.

- **Portability**

Portability means that the devices used to make mobile computing enabled can be ported from one place to another place without any issue.

Mobile phones, laptops, tablets can be ported one location to any other location in any distant area.

- **Interaction**

Interaction refers to communicate and transfer information from one user to another user using wireless-enabled devices.

Users can send textual data, audio, and video files from person to another and one group of people to another group of people.

- **Location Independence**

Location independence means a user can communicate at any distant location from any location.

While sitting in your office you can observe all the cabins of your office and open land areas with the help of mobile computing devices and the Internet of things technology.

While staying in a distant country and sitting in your room you can watch your house located in another country using mobile-enabled devices such as cameras, mobile phones, and laptops, etc.,

What are mobile computing Technology and its importance?

Mobile Computing is an advanced mobile technology that enables users to communicate with one another throughout the world with the help of mobile devices.

Mobile computing has huge importance in today's environment because this is the time of digitization where most of the businesses are digitized and based on mobile devices and mobile communication.

There is no area of our life where we can survive without the use of mobile communication. From personal use to professional use human life is based on mobile communication.

Despite so many security challenges associated with mobile technology, technology is growing day by day.

The areas which are not touched by mobile technology until now are using mobile technology in order to get more growth and to be smarter such as agriculture, small scale industries, music, dance, art, craft, and much more.

What is the use of mobile computing Technology?

Mobile computing Technology applications

Mobile computing is used in huge applications or we can say mobile computing technology is used in each and every application from small scale to large scale applications, from home to business applications and from art to science.

Uses of Mobile Computing Technology

Mobile computing applications- Replacement of Wired Networks

Organizations which were using wired networks for communication before, now replacing their wired connection to wireless connection to make mobile computing technology possible, in order to get growth and improve opportunities in the related domain in which they are dealing.

Mobile computing applications – Education

Education is a big industry where mobile computing technology is growing day by day from child care level to higher levels of education.

Mobile computing technology generates new avenues in the education sector and makes distant and online education possible for both who are involved seriously in the education and research and professionals who want to learn more parallel to their professions.

Mobile computing applications- Entertainment

Entertainment is a big industry that is revolutionized with the use of wireless mobile technology.

Sports, movies, and games all have become richer with the use of mobile computing technology.

New areas also developed in the entertainment industry with the use of mobile technology such as LAN gaming, online games, etc.

Mobile computing applications – Vehicle

Vehicles, cars, buses all are equipped with wireless mobile-enabled devices that allow you to be aware by the traffic status and allow you to plan your journey and route accordingly.

You can get the right times of arrival and departure of buses, trains, plains, etc. Using GPS car navigation systems you can move from one place to another place easily without the use of the exact route of the destination location.

If a journey is very big you can use various ways of entertainment during traveling due to mobile computing technology.

Mobile computing applications in Business

Now, no business can grow and survive without the use of mobile computing technology.

The use of new mobile computing devices and technology can richer your business and you can provide better services to your users whether the business is at home, online or at a market place.

Mobile computing applications in Sports

Every sport whether that is on an open land area or in a closed room area needs advanced mobile technology equipped devices in order to be more advanced and improved and to provide better entertainment facilities to its players, users, and viewers?

Mobile computing applications in Emergencies and Natural Hazards

In situations of emergencies such as accidents, flooding, earthquakes, wars, and natural calamities, etc., wireless mobile technology is the only system that can be survived and can provide services to the victims and facilitators both.

The exact scene can be conveyed as it is to a distant location to the people who can provide supports during the troubles through mobile computing technology. Medical services can be availed in such conditions by the experts.

Issues and Challenges in Mobile Computing Technology

Issues and Challenges in Mobile Computing Technology –

Due to technology growth distances become short, communication becomes easy, data transfer is enriched, but there are so many issues and challenges in the domain that need to be researched and addressed to make mobile computing more secure, robust and reliable. Some important issues are discussed here.

Low bandwidth

Mobile internet access is slower than the fixed desktop connection while using GSM and other advanced technologies such as 3G, 4G, and 5G. Local wireless connection offers Mbit/s of speed and wide wireless connection offers only Kbit/s of speed. There is a requirement of using more bandwidth while using such advanced mobile technologies so that the user can transfer data at a higher speed while the user is mobile.

Lower security

When working with mobile people are completely dependent on the public network which can be easily tracked and hacked by hackers. There is a big problem with the security of data while transferring from one mobile device to another device. Therefore, to protect the data from eavesdropping there is a need for strongly secured algorithms of authentication and security.

Money transaction is a very sensitive area and it is the target on the hackers. Internet frauds related to money are huge. Therefore, more research and development is needed to provide more secure methods to transfer the information.

Transmission interferences

Radio transmission cannot be protected therefore there is higher transmission interference due to electric engines, lightening, high buildings, mountains, weather conditions, etc., all this results in a higher loss of data rate and bit errors.

Shared medium

Radio access is a shared medium because it is just impossible to give dedicated radio access to all the users. However, different techniques are deployed still so many questions are unanswered such as how to provide quality of service to each user sharing radio access.

Ad-hoc networking

Wireless and mobile computing allow ad-hoc networking without a prior set of infrastructure between senders and receivers. This creates several challenges and

issues before the network administration such as the reliable and secure connections between sources to destination.

Dynamic topology again a challenging issue for a reliable connection from one point to another point.

High delays, large delay variation

A serious problem faced by Internet protocols is variations in link characteristics. In wireless transmission delays of various seconds occur that create so many problems in transmission and communication.

Also links are asymmetrical and provide different service qualities depending on the wireless devices.

Regulations and spectrum

Due to technical and political reasons, very limited frequencies are available. This is also a serious issue in mobile computing that need to addressed and researched.

Power consumption

When a power supply is not available mobile devices totally depend on battery power. There is a need to use some resources that can provide power supply at a cheaper cost and with ease.

Potential health hazards:

People use a mobile phone and other mobile devices while driving that again creates problems and prone to accidents. Moreover, mobile devices are found injurious for health if not used carefully. Therefore, there is a need to research and

develop mobile devices which are perfect for health whether you are in mobile or in rest.

Overview of Wireless Telephony - Cellular Concept Mobile Computing Technology

Cellular Concepts refers to the use of a group of cells to provide communication from one place to another place when the user is mobile. A cellular system in mobile computing implements space division multiplexing or SDM. Each transmitter in the cellular system is called a base station.

Base station

The base station covers a specific area that is called a cell.

Cell

Cell radius can vary from tens of meters in building, hundreds of meters in a city, and tens of kilometers in the country.

The shape of a cell depends on the environmental conditions such as type of building, mountains, weather conditions, load, and other conditions. Generally, it is hexagon shape but not an exact hexagon.

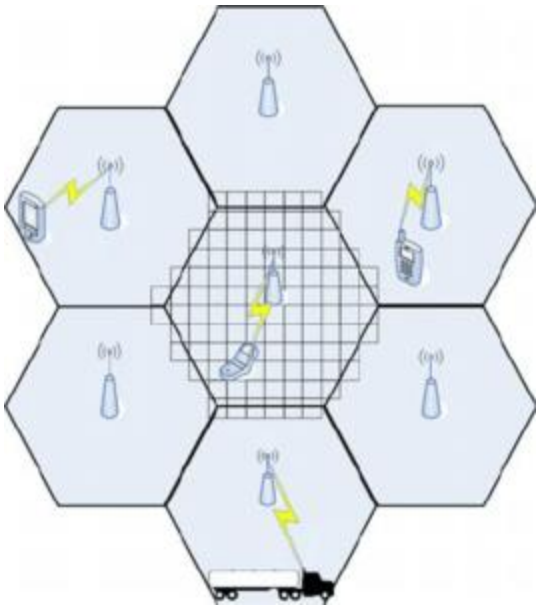


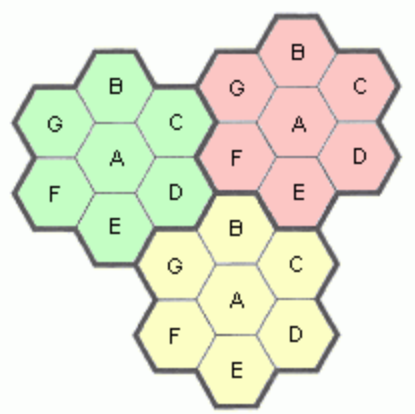
Figure: Cell Structure

Mobile computing used cellular system and it has the following advantages:

- Higher Capacity

Cellular system uses SDM. SDM allows frequency reuse. If one transmitter is far away from another transmitter particularly out of the range of the interference area then another transmitter can use the same frequency as shown in the following figure-

Mobile systems assign one specific frequency to a certain user and that frequency is blocked for the use by other users. But frequency is a scarce resource therefore for their optimum utilization of the same frequency, the same frequency is used for other users using the technique frequency reuse.



- Less Transmission Power

Transmission power is not a big issue for the base station but it is important for the mobile receivers. Keeping cell size small facilitates mobile receivers because as they move far from the base station their receiving capacity reduces and due to small cells after a few distances they can again access receiving power from the cell nearby them and the problem of the mobile receiver is solved.

- Local Interference

When the distance is large between the sender and receiver then interference is also more and difficult to manage. There is only local interference when cells are small and that is easy to handle by the base station and the mobile station.

- Robustness

The cellular system is decentralized therefore more robust as compared to when centralized. If any component fails only that specific area affected and the rest of the part remains unaffected and works efficiently.

Cellular System or having small cells have the following disadvantages

- Infrastructure Requirement

Cellular system establishment needs complex infrastructure, storage registers to locate mobile users in local areas and roaming, antennas, transmitters, receivers, and amplifiers, etc. that are expensive.

- Handover Needed

The mobile system needs handovers when they change cells. This is quite often which further incur a cost.

- Frequency Planning

To avoid the interference between transmitters frequencies are planned carefully. Frequency is a limited resource therefore they are distributed intelligently so that they can be reused without any kind of interference.

Mobile Computing Architecture – GSM Network Architecture

Mobile Computing Architecture – GSM Network Architecture –

The full form of GSM is group special mobile (GSM) and later it was named as a global system for mobile communications (GSM).

It was founded in 1965. GSM is the most successful and popular mobile telecommunication system. GSM is used by over 800 million people and in over 190 countries.

The main aim of GSM was to provide a mobile phone system that allows users to move throughout Europe and allows voice services compatible with the ISDN and PSTN systems.

Mobile Computing Architecture – Versions of GSM

GSM is a particularly second-generation system, replacement of the first-generation analog system. This system was not capable to give high worldwide data rates as promised by the third generation systems.

GSM has initially been deployed in Europe with 890-915 MHz for uplinks and 935-960 MHz downlinks. This GSM system is called GSM 900.

The next version is called GSM 1800 MHz, 1710-1785 MHz uplink and 1805-1880 MHz downlink. This system is called DCS or the digital cellular system-1900.

The next versions of GSM are GSM 400. It is deployed in sparsely populated areas as a replacement of the analog system.

GSM Mobile Services

GSM system has defined three kinds of services:

- Bearer Services
- Tele-Services
- Supplementary services

Bearer Services in GSM

Bearer Services allow transparent and nontransparent, synchronous and nonsynchronous data transmission services.

- **Transparent bearer services in GSM**

Transparent bearer services use the functions of the physical layer to transmit data.

Using transparent and nontransparent services, GSM provides various bearer services for internetworking PSTN, ISDN, and packet-switched public data networks such as X.25. X.25 is available worldwide.

Nontransparent Bearer Services

These services use transparent bearer services using radio link protocol.

This radio link protocol (RLP) includes the high-level data link control (HDLC) mechanism and special selective reject mechanism to trigger retransmission erroneous data.

TeleServices of GSM

GSM provides voice-oriented teleservices. Teleservices include voice transmission, message services, and basic data communication services or PSTN and ISDN services.

- **Emergency Number Service of GSM**

This service is free of cost and essential for all the service providers. Emergency number service is of the highest priority service and pre-empting other connections.

The service is automatically set up with the closest emergency center.

- **Short Message Service**

SMS is a simple short message service. It allows 160 characters of messages.

SMS services do not use the standard data channel of GSM. It uses unused capacity in the channels.

- **Supplementary services**

In addition to teleservices and bearer services, GSM providers provide various supplementary services.

Examples o supplementary services are as follows-

1. Identification
2. Call redirection
3. Forward incoming calls
4. Close user group service – Through a company-specific GSM sub-network facility only members of a group can communicate.
5. Multi-party communication service

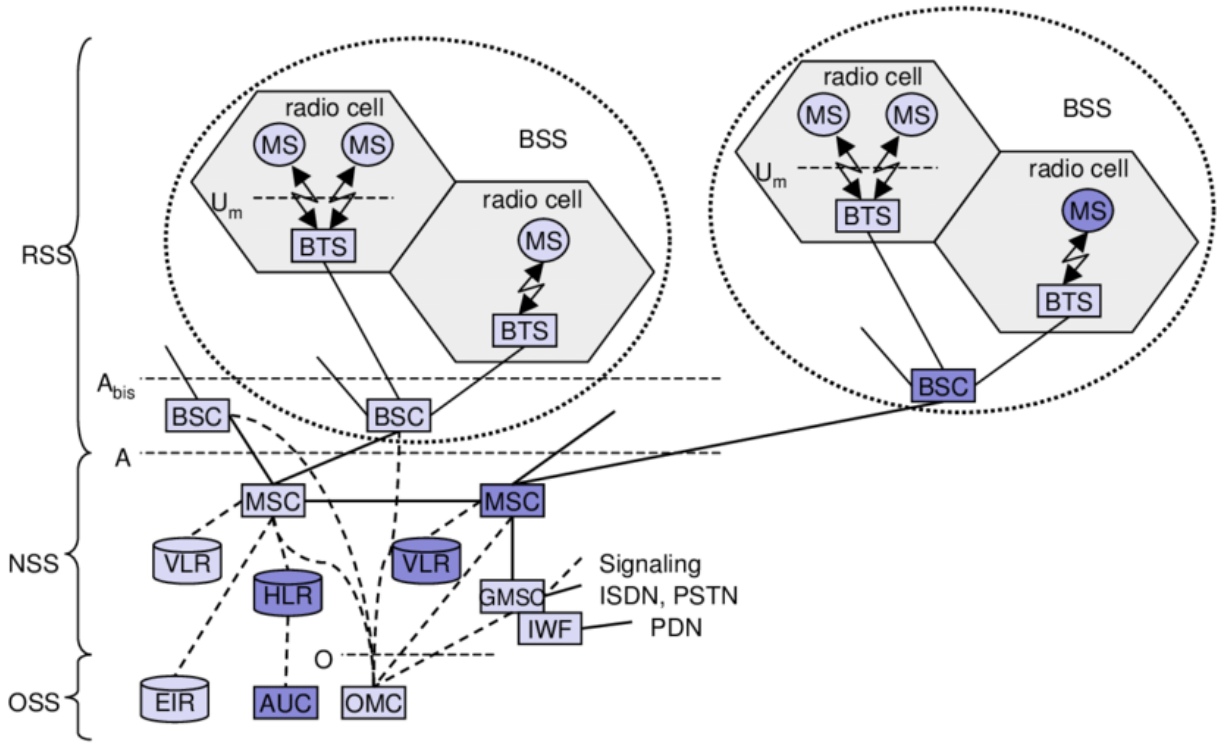
GSM System Architecture

GSM has a complex hierarchical system architecture. It includes various entities, interfaces, and acronyms.

GSM consists of three subsystems as given below –

1. Radio Subsystem
2. Network Switching subsystem

3. Operation subsystem



GSM – Radio Sub System

Radio subsystem comprises the entire radio specific entities such as

- A mobile station (ms) and
- Base station subsystem (BSS)

RSS and NSS connected with “a” interface and further connected with OSS via the “o” interface.

All the subsystems of GSM are as follows-

1. Base station subsystem – BSS
2. Base transceiver subsystem – BTS
3. Base station controller -BSC
4. Mobile station – ms

Base station subsystem – BSS

- A BSS comprises many base station subsystems.
- Each BSS is controlled by BSC or base station controller.
- A BSS does all functions that are necessary to manage and maintain radio connections to a mobile station, for instance, coding and decoding of voice, and rate adaptation to and from the wireless network.
- A BSS contains several BSC and BTS.

Base transceiver subsystem – BTS

The base transceiver comprises all radio entities and equipment such as

- Antennas, Signal processing, Amplifiers, etc. necessary for radio transmission
- A BTS form a radio cell or using sectorized antennas several cells and that are connected to the mobile station via the Um interface (ISDN U Interface for mobile) and to the BSC via the Abis interface.
- The Um-interface contains all the entities required for wireless transmission such as TDMA, FDMA, etc.

Base station controller –BSC

- The base station controller manages all the BTS.
- It reserves radio frequencies, handles handovers, from one BTS to another BTS within the BSS.

Mobile station – ms

- The mobile station comprises user equipment and entities as well as the software required for communication with the GSM network system.
- The mobile station consists of user-independent hardware and software and subscriber entity module or 'Sim' that store all the user-related data that is relevant for the GSM system.

Network and Switching Subsystem

Network and Switching subsystem connects the wireless with the standard public network.

NSS performs handovers between different BSS.

NSS performs following functions-

- Worldwide localization of mobile users
- Support charging, accounting, and roaming of users between various providers in different countries

NSS consists of following switches and databases

- Mobile services switching center –MSC

- MSCs are high-performance digital ISDN switches.
- MSC manages several BSCs in geographical regions
- MSC setups connections to other MSC via 'A' interface
- MSC has some more connections to other fixed networks such as connection with PSTN and ISDN.
- Using additional interworking functions MSC is connected to Public Data Network (PDN) also, such x.25.
- MSC handles all the signaling required for connection setup, connection release and handovers of connections to other MSCs.

Home location register – HLR

The home location register is the most important database in the GSM system. It stores all the user-specific data such as-

- Mobile Subscriber ISDN number – MSISDN
- Information on subscriber services such as forwarding, roaming restrictions, and the international mobile subscriber identity – IMSI
- Dynamic information is also needed such as the current location area (LA) of the mobile station.
- As soon as the mobile station leaves its current location area, the information of the HLR is immediately updated. This information is needed to localize the mobile station in the worldwide GSM network.
- HLR also stores present MSC and VLR.
- All these user-specific information elements only exist once for each user in a single HLR, that also facilitates charging and accounting.

Visitor Location Register – VLR

- The visitor location register is associated with each MSC.
- VLR is dynamic database stores all important information of the mobile user such as its current location area, and current MSC.
- Whenever a new user changes its location area and MSC, VLR copies its data from its HLR.
- VlRs capable to store the data of millions of customers.

Operation Subsystem – OSS

OSS is the third subsystem of GSM. It contains all the necessary functions for network operations and maintenance.

It contains the following entities-

- Operation and Maintenance Center
- Authentication Center
- Equipment Entity Register

Operation and Maintenance Center

Operation and Maintenance (OMC) monitors and controls network entities via O interface.

OMC management functions are as below-

- Traffic monitoring
- Status reporting of network entities
- Subscriber management

- Security management
- Accounting and billing

Authentication Center (AUC)

AUC is situated in the protected part of the home location register.

AUC protect user identity and data transmission.

It contains algorithms for user authentication.

Equipment Entity Register

EIR is a data repository that keeps records of all the equipment or devices related to GSM.

It also keeps records of valid devices including mobile phones and also invalid devices or stolen devices.

GSM Radio Interface

In GSM 900, 124 channels are used. Each channel is 200 kHz wide.

GSM 1800 uses 374 channels.

GSM 900

In GSM data is transmitted in small portions called bursts that are used for data transmission inside a time slot for both user data as well as signaling data.

The burst is only 546.5 us long and contains 148 bits. The 30.5 us are used as guard space to avoid overlapping with other bursts.

The details of 148 bits are as below-

- The first and last 3 bits of a normal burst is set to 0 and can be used to improve the receiver performance
- The training sequence in the middle of a slot is used to adapt the parameters of the receiver to the current path propagation characteristics and to select the strongest signal in the case of multipath propagation.
- A flag S indicates whether the data field is user data or network data.
- 57 bits on both the side of the training sequence are used for user data.

How GSM works

GSM Logical Channels and Frame Hierarchy

GSM uses two kinds of channels

- **Logical Channels**
- **Traffic Channels**

GSM Traffic Channels – GSM TCH

GSM uses traffic channels to transmit voice and fax data.

Two types of TCHs have been defined in the GSM.

Full-rate TCH – TCH/F

Half-rate TCH – TCH/h

- Full-rate TCH – TCH/F

A TCH/F has a data rate of 22.8 kbits/s.

- Half-rate TCH – TCH/h

TCH/h has data rate of 11.4 kbits/s.

Voice codecs used in the beginning of GSM standardization consumed 13 kbit/s. Remaining capacity of the TCH/F (22.8 kbit) was used for error correction.

Improved codecs give better voice quality in terms of speed and used with TCH/h, however voices quality decreased after the use of improved codecs.

The standard codecs system for voice are called full rate or FR,13 kbits/s and half rate or HR, 5.6 kbit/s.

A new codec system enhanced full rate or EFR, provides better voice quality than full rate codec if transmission error is low.

GSM Control channels – CCH

Different control channels are used in the GSM network architecture to control the following –

- Medium access
- Allocation of traffic channels
- Mobility management

Three groups of Control channels – CCH are defined and these groups of channels have further sub channels

Broadcast control channel – BCCH

Common control channel – CCCH

Dedicated control channel – DCCH

Broadcast control channel – BCCH

A base transceiver system or BTS uses a broadcast control channel to provide information to all mobile stations within a cell.

Information transmitted in this channel is as follows-

- Cell identifier
- Options present within the cell about frequency hopping
- Frequency available inside the cell and in the neighboring cells
- BTS gives frequency correction information via frequency correction channel (FCCH) and also information about time synchronization via synchronization channel (SCH).

FCCH and SCH are sub-channels of BCCH.

Common control channel – CCCH

The common control channel exchanged all the information about the connection setup between a mobile station and BTS.

For calls toward the direction of the Mobile station, BTS uses a paging channel (PCH) for paging the correct mobile station.

If a mobile station wants to set up the connection it uses a random access channel or RACH, to send data to the BTS.

RACH implements multiple accesses using slotted Aloha. All the mobile stations within a cell may access RACH channel.

If any collision occurs with other mobile stations in a GSM system, the BTS uses the access grant channel or AGCH to signal a mobile station that it can use a TCH or SDCCH for further connection setup.

Dedicated control channel – DCCH

Dedicated control channels are bidirectional channels while all other channels mentioned above are unidirectional channels.

As long as a mobile station has not set up a TCH with the BTS, the mobile station uses the Standalone dedicated control channel (SDCCH) at a low data rate of 782bit/s for signaling. This includes signaling for authentication, registration or other data required for setup TCH.

TCH and SDCCH have a slow associate dedicated control channel (SACCH) also which is used for exchanging system information like system quality and signal power level.

If more signaling data required to be shared a GSM uses fast associated dedicated control channel (FACCH).

FACCH uses the time slots that are otherwise used by TCH.

This is required in the case of handovers where BTS and Mobile station needs to share a large amounts of data in minimum time.

UMTS In Mobile Computing –

Universal Mobile Telecommunications System (UMTS) is a third-generation mobile **cellular system** developed by ETSI in 1998 based on the **technology of GSM**.

UMTS employs Wideband Code Division Multiple Access (W-CDMA) radio access technology.

To avoid high costs, UMTS reuses as much infrastructure as possible while introducing new services and higher data rates based on CDMA technology.

The initial installations of UMTS used GSM/GPRS infrastructure and offer only moderate data rates.

What are UMTS SERVICES?

- UMTS offers greater spectral efficiency and bandwidth for mobile network operations.
- The technology used in UMTS is also referred to as Freedom of Mobile Multimedia Access (FOMA) or 3GSM.
- UMTS is suitable for a bigger framework developed in mid-nineties by ETSI, called Global Multimedia Mobility (GMM).

- UMTS provides several **bearer services,** real-time and non-real-time services, circuit and packet-switched transmission and many different data rates.

UNIT-II

CDMA

What Is Cdma In Mobile Computing – CDMA refers to Code Division Multiple Access. CDMA is cellular technology.

In CDMA there are two main systems

- **Base Station (BS)**
- **mobile subscribers or users.**

In CDMA Unique codes are assigned to the users for communication with BS or base station.

CDMA allows multiple transmitters to send data over a single channel simultaneously.

This makes it possible to share the same bandwidth by multiple users at the same time.

How CDMA works

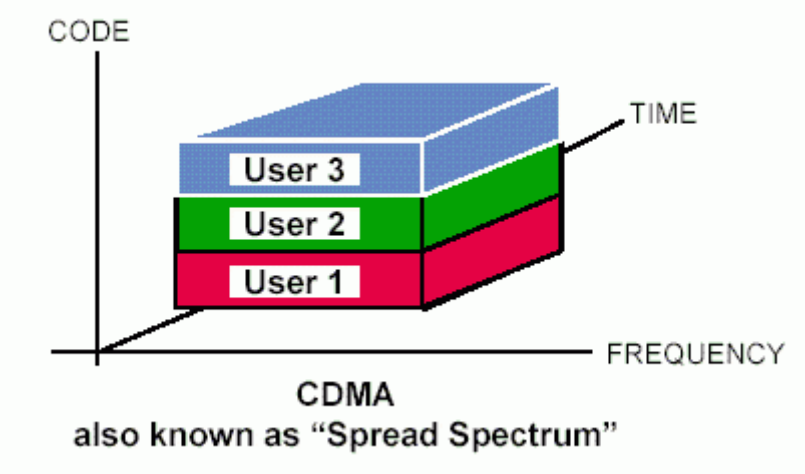
Codes with specific characteristics can be applied to the transmission to enable the use of **code division multiplexing (CDM)**.

Code division multiple access (CDMA) systems use exactly these codes to separate different users in code space and to enable access to a shared medium without interference.

In CDMA entire bandwidth is being used by multiple users and each user has their unique codes to recover the data. The system works based on the spread spectrum concept.

The main problem is how to find “good” codes and how to separate the signal from noise generated by other signals and the environment.

CDMA can be used in combination with FDMA/TDMA access schemes to increase the capacity of a cell.



GPRS - General Packet Radio Service – Definition & Architecture

GPRS refers to General Packet Radio Service.

GPRS provides packet mode data transfer service over cellular network system.

GPRS uses the existing network resources more efficiently for packet mode application and also provide quality of service.

GPRS provides unicast, multicast and broadcast services to its users.

Goal of GPRS

The main aim of GPRS is to allow more efficient and cheaper packet data transfer service for Internet application that particularly requires packet data transfer.

GPRS network providers specially support the system by charging on volume and not on connection time as is used in traditional GSM data services.

The main goals of GPRS are as below-

- Continuous and consistent Internet Packet service
- This is a cheaper service as compare to circuit switched network service
- It has a open architecture
- GPRS service innovations are independent from the infrastructure

GPRS Services

The main advantage of GPRS is the 'always on' feature of GPRS.

There is no need of any connection set up prior to data transfer.

But GPRS needs some additional network components including software and hardware components to transfer the data from source to destination.

GPRS provides following services to its users

- Instant messaging and presence
- Multimedia messaging service

- Point-to-Point and Point-to-Multipoint services
- SMS messaging and broadcasting

Advantages of GPRS

- **Mobility**

  GPRS has the capacity to manage a consistent voice transfer and information interchanges while the user is moving.

- **Cost Efficient**

  Communication via GPRS technology is cheaper than the GSM network system.

- **Availability**

  GPRS allows users to get connectivity whenever they need regardless of the distance, and location.

- **Localization**

  GPRS enables users to receive data at their present location.

- **Easy Billing**

  GPRS packet transmission service provides an easier billing system than that provided by the GSM circuit switched system

Bluetooth

Bluetooth wireless technology is a short range communications technology intended to replace the cables connecting portable unit and maintaining high levels of security.

Bluetooth technology is based on **Ad-hoc technology** also known as **Ad-hoc Pico nets**, which is a local area network with a very limited coverage.

The usage of Bluetooth has widely increased for its special features.

- Bluetooth offers a uniform structure for a wide range of devices to connect and communicate with each other.
- Bluetooth technology has achieved global acceptance such that any Bluetooth enabled device, almost everywhere in the world, can be connected with Bluetooth enabled devices.
- Low power consumption of Bluetooth technology and an offered range of up to ten meters have covered the way for several usage models.
- Bluetooth offers interactive conference by establishing an adhoc network of laptops.
- Bluetooth usage model includes cordless computer, intercom, cordless phone and mobile phones.

Piconets and Scatternets

Bluetooth enabled electronic devices connect and communicate wirelessly through short range devices known as **Piconets**.

Bluetooth devices exist in small ad-hoc configurations with the ability to act either as master or slave the specification allows a mechanism for **master** and **slave** to switch their roles.

Point to point configuration with one master and one slave is the simplest configuration.

When more than two Bluetooth devices communicate with one another, this is called a **PICONET**.

A Piconet can contain up to seven slaves clustered around a single master.

The device that initializes establishment of the Piconet becomes the **master**.

The master is responsible for transmission control by dividing the network into a series of time slots amongst the network members, as a part of **time division multiplexing** scheme

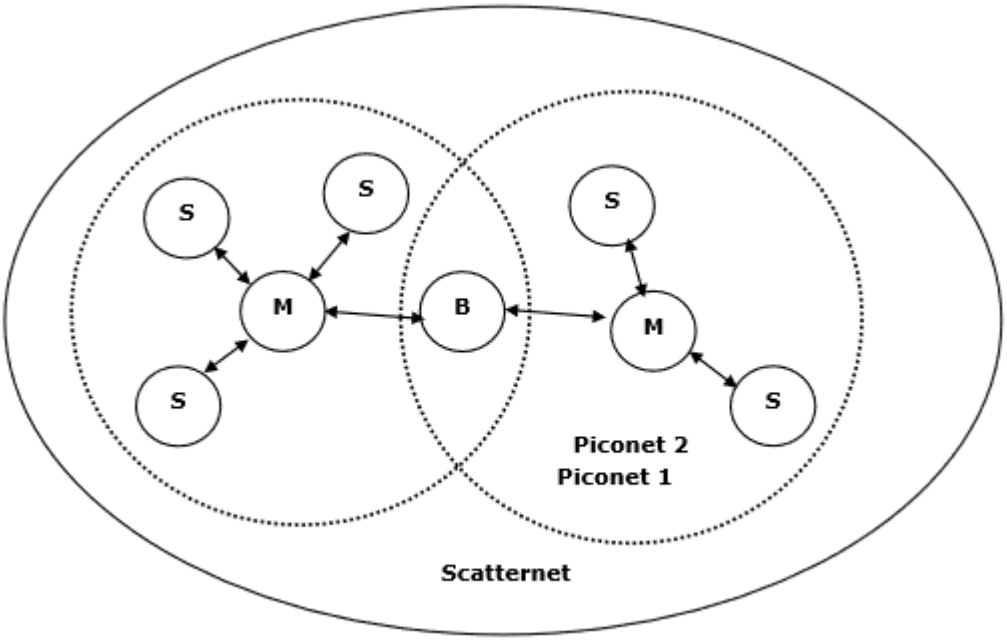


Figure: Piconets and Scatternets

The features of Piconets are as follows −

- Within a Piconet, the timing of various devices and the frequency hopping sequence of individual devices is determined by the clock and unique **48-bit address** of master.
- Each device can communicate simultaneously with up to seven other devices within a single Piconet.
- Each device can communicate with several piconets simultaneously.
- Piconets are established dynamically and automatically as Bluetooth enabled devices enter and leave piconets.
- There is no direct connection between the slaves and all the connections are essentially master-to-slave or slave-to-master.
- A device can be a member of two or more piconets, jumping from one piconet to another by adjusting the transmission regime-timing and frequency hopping sequence dictated by the master device of the second piconet.
- It can be a slave in one piconet and master in another. It however cannot be a master in more than onc piconet.
- Devices resident in adjacent piconets provide a bridge to support inner-piconet connections, allowing assemblies of linked piconets to form a physically extensible communication infrastructure known as **Scatternet**.

WAP - Wireless Application Protocol

WAP stands for Wireless Application Protocol.

WAP represents a group of protocols rather than a single protocol.

WAP aims at integrating a simple lightweight browser also known as a micro-browser into handheld devices, thus requiring minimal amounts of resources such as **memory** and **CPU**.

The primary objectives of the WAP protocols are the following.

- Independence from the wireless network standards
- Interoperability among service providers
- Overcoming the shortfalls of the wireless medium
- Overcoming the drawbacks of handheld devices
- Increasing efficiency and reliability
- Providing security, scalability, and extensibility

The WAP Model

WAP adopts a client-server approach.

It specifies a proxy server that acts as an interface between the wireless domain and core wired network.

This proxy server, also known as a **WAP gateway**, is responsible for a wide variety of functions such as protocol translation and optimizing data transfer over the wireless medium.

Wireless network parts consist of −

- Content provider (Application or origin server)
- Mobile device (WAP client)
- WAP gateway
- WAP proxy

WAP Client

The three sections to be mentioned regarding WAP client are WAE user agent, WTA user agent and WAP stack.

- **WAE user agent** – Wireless application environment user agent is the browser that renders the content for display.
- **WTA user agent** – Wireless telephony application agent receives compiled WTA files from WTA server and executes them.
- **WAP stack** – WAP stack allows the phone to connect to the WAP gateway using the WAP Protocols.

Application Server

The elements in the network where the information (web, WAP) applications reside are WAP proxy, WAP gateway or WAP server –

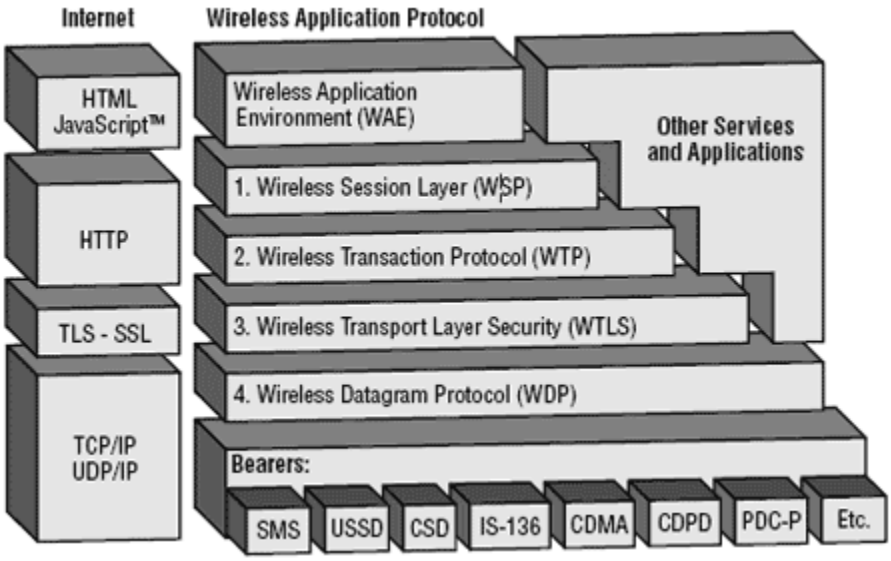
- **Proxy** – This is an intermediary element acting both as a client and as a server in the network. It is located between client and server.

  The client sends requests to it and it retrieves and caches the information needed by contacting the origin Server.

- **Gateway** – This is an intermediary element usually used to connect two different types of networks.

The WAP Protocol Stack

WAP protocol stack is shown in the following figure –



Application Layer

The application layer provides an application environment intended for the development and execution of portable application and services WAE consists of two different user agents located on client side.

The WAE user agent consists of browser and the text message editor along with the WTA user agent.

Session Layer

The session layer supplies methods for the organized exchange of content between Client/Service applications.

WAP contains the following components −

- **Connection Oriented Session Services** − These operate over WTP.
- **Connectionless Session Services** − These operate directly over WDP.
- **Session services** − These functionalities help to set up a connection between a client and server using primitive messages.

Transaction Layer

Provides different methods for performing transactions to varying degree of reliability

Security Layer

It provides services that ensure privacy, server authentication, client authentication and data integrity.

Transport Layer

This is the bottom layer, connected with the bearer service offered by the operator. Bearer services are the communication between the mobile phone and the base stations. They include **SMS, GSM, GPRS, CDMA, FDMA, TDMA**, etc

The physical layer prepares the data to be sent from the mobile device over the air services and sends the data using bearer service implemented in the network that the device is operating in.

Spread spectrum

Spread spectrum techniques involve spreading the bandwidth needed to transmit data.

Spreading the bandwidth has several advantages.

- The sender spreads the signal means converts the narrowband signal into a broadband signal.
- The energy needed to transmit the signal is the same, but it is now spread over a comparatively larger frequency range.
- The power level of the spread signal can be lower than that of the original narrowband signal without any lose of data.
- Depending on the generation and reception of the spread signal, the power of the user signal can be as low as the background noise there.
- It makes now difficult to differentiate between the user signal and the background noise and therefore difficult to detect the original signal.
- During the transmission process, narrowband and broadband interference added to the signal level.
- The sum of interference and original signal is obtained and received. The receiver knows how to dispreads the signal.
- The receiver converts the spread original signal into a narrowband signal again, while spreading the narrow and noise interference and leaving the broadband interference.
- The receiver node applies a bandpass filter to cut off frequencies of left and right narrowband signals.

At the end, the receiver restructure the original data because the power level of the original signal is high, means the signal is much stronger than the remaining noise interferences.

Spreading the spectrum is done via two ways as shown in the following two sections.

- Direct sequence spread spectrum (DSSS)

Direct sequence spread spectrum (DSSS) system takes a user bit stream and performs an (XOR) function.

The result is either the sequence of - 0110101 (if user bit equals to 0) or the its complement 1001010 (if the user bit is equals to 1).

The DSSS receiver process is more complex than the transmitter process. The receiver only performs the inverse functions of the two transmitter modulations steps.

However, noise and multi-path propagation processes need some additional processes to rebuild the original data signals.

- Frequency hopping spread spectrum

This is frequency hopping technique, where the users are made to change the frequencies of usage, from one to another in a specified time interval, hence called as **frequency hopping**.

For example, a frequency was allotted to sender 1 for a particular period of time.

Now, after a while, sender 1 hops to the other frequency and sender 2 uses the first frequency, which was previously used by sender 1. This is called as **frequency reuse**.

The frequencies of the data are hopped from one to another in order to provide a secure transmission.

The amount of time spent on each frequency hop is called as **Dwell time**.

Introduction to 802.11a OFDM

Orthogonal frequency division multiplexing (OFDM) is a technique of digital signal modulation.

In this technique a single data stream is split across several separate narrowband channels at different frequencies in order to reduce interferences and crosstalk.

The original data stream bits sent serially (one after the other) and they are transmitted in parallel means several at once on separate channels, but at a lower speed, each sub-streams relative to the original signal.

OFDM technology

OFDM technology was first implemented in the 1960s and 1970s during the research in order to minimizing the interference or noise among the channels that are near to each other in frequency.

It is applied to achieve accurate data transmission in situations which are prone to interferences and signal corruption when more conventional modulation schemes are used.

OFDM is used in the following-Wi-Fi,

- 4G wireless communications,
- Digital television and radio broadcast services

Mobile IP

Mobile IP (or MIP) is an Internet Engineering Task Force (IETF) standard communications protocol that is designed to allow **mobile** device users to move from one network to another while maintaining a permanent **IP** address.

It enables the transfer of information to and from mobile computers, such as laptops and wireless communications.

The mobile computer can change its location to a foreign network and still access and communicate with and through the mobile computer's home network.

Mobile IP for better Mobility

Mobile IP – A technology which supports mobile data and applications that are dealing with wireless connectivity. A user may now disconnect his computer in the office and reconnect from another site within the same office or elsewhere.

Components of a Mobile IP Network

Mobile IP has three major components as mentioned below –

- **Mobile Node** – A device such as a cell phone, personal digital assistant, or laptop whose software enables network roaming capabilities.
- **The Home Agent** – A router on the home network serving as the anchor point for communication with the mobile node; its tunnel packets from a device on the Internet, called a correspondent node, to the roaming mobile node.
- **The Foreign Agent** – 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.

The Mobile IP process has three main phases –

Phase I: Agent Discovery

This is the phase where mobile node discovers its foreign and home agents.

A mobile node first determines its connected location by using ICMP router discovery messages.

If it's connected location is with the local network, then the normal IP routing is used for the communication.

When a mobile node determines that it has moved to a foreign network it obtains a care-of address from the foreign agent reflecting its current location.

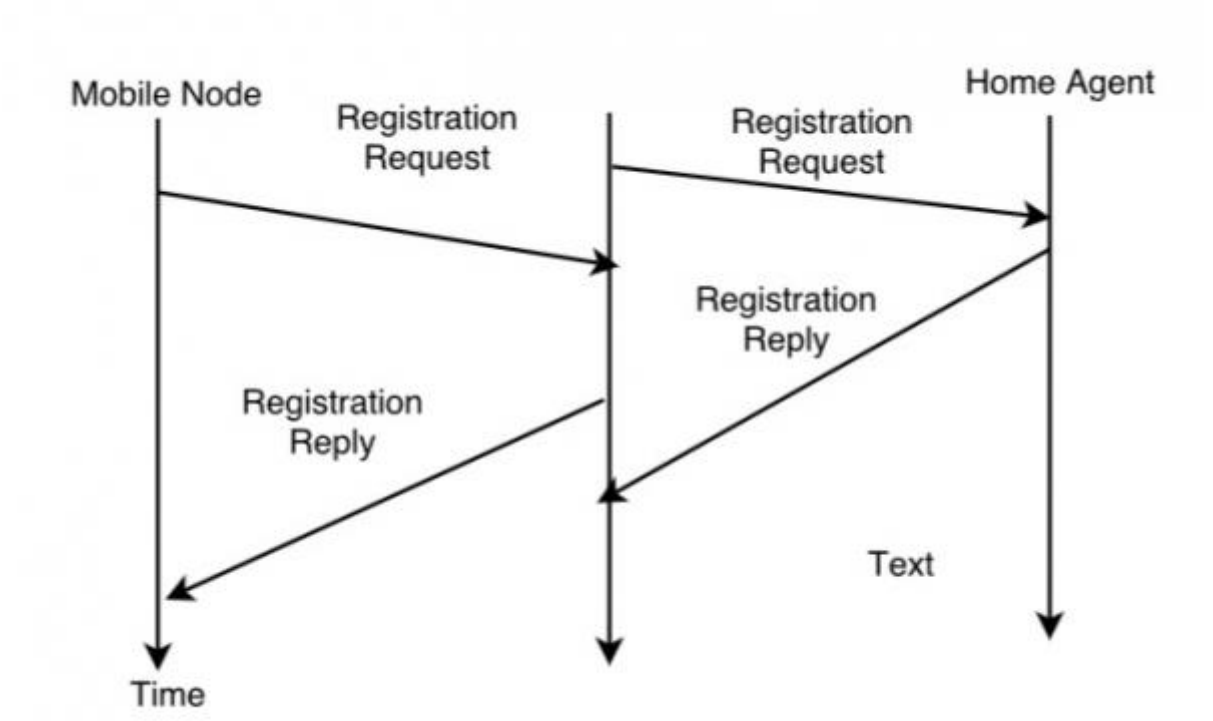
Phase II: Registration

This the phase, where a mobile node registers its current location with the foreign agent and the home agent.

If the connected location is identified as foreign location, then the mobile node looks for a foreign agent and registers itself with the foreign location and the foreign agent, in turn, notifies the home agent and creates a tunnel (called tunneling) between itself and the home agent.

During this phase, the Mobile node sends a registration request message to the foreign agent which forwards the message to the home agent. The home agent sends back a reply after updating its registration table with the home address and "care-of" address mapping.

The flow of these messages is described in the figure below.



Thus, a successful Mobile IP registration sets up the routing mechanism for transporting packets to and from the mobile node as it roams.

Phase III: Tunneling

This is the phase where a reciprocal tunnel is set up by the home agent to the care-of address to route packets to the mobile node as it roams.

The method by which mobile IP receives information from a network is called tunneling.

It has two primary functions –

- Encapsulation of the data packet to reach the tunnel endpoint.
- Decapsulation, when the packet is delivered at that endpoint.

After the registration phase, the home agent now encapsulates all the packets intended for the mobile node and forwards those packets through the tunnel to the foreign agent.

The foreign agent de-encapsulates the packet and forwards them to the mobile node.

Wireless LAN

A wireless LAN is a wireless computer network that links two or more devices using wireless communication to form a local area network within a limited area such as a home, school, computer laboratory, campus, or office building.

Wireless LANs are those Local Area Networks that use high frequency radio waves instead of cables for connecting the devices in LAN.

Users connected by WLANs can move around within the area of network coverage.

Most WLANs are based upon the standard IEEE 802.11 or WiFi.

IEEE 802.11 Architecture

The components of an IEEE 802.11 architecture are as follows

1) **Stations (STA)** − Stations comprise all devices and equipments that are connected to the wireless LAN.

A station can be of two types:

- **Wireless Access Point (WAP)** − WAPs or simply access points (AP) are generally wireless routers that form the base stations or access.
- **Client.** − Clients are workstations, computers, laptops, printers, smartphones, etc.

Each station has a wireless network interface controller.

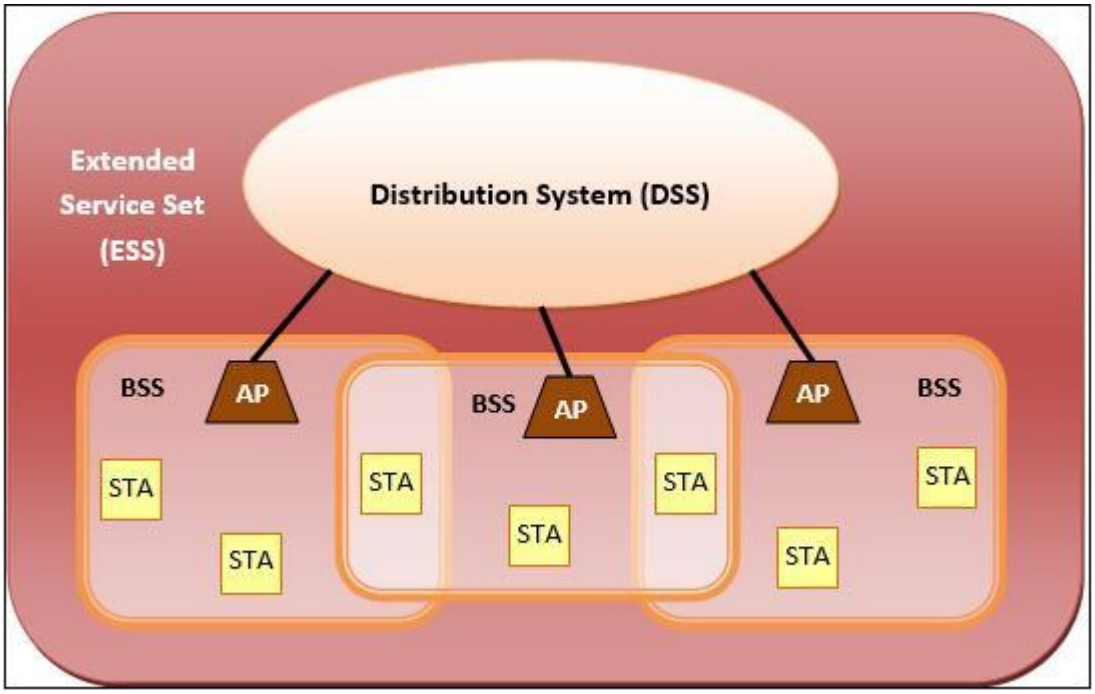
2) **Basic Service Set (BSS)** −A basic service set is a group of stations communicating at physical layer level.

BSS can be of two categories depending upon mode of operation:

- **Infrastructure BSS** − Here, the devices communicate with other devices through access points.
- **Independent BSS** − Here, the devices communicate in peer-to-peer basis in an ad hoc manner.

3) Extended Service Set (ESS) − It is a set of all connected BSS.

4) Distribution System (DS) − It connects access points in ESS.



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.
- They provide clutter free homes, offices and other networked places.
- The LANs are scalable in nature, i.e. devices may be added or removed from the network at a greater ease than wired LANs.
- The system is portable within the network coverage and access to the network is not bounded by the length of the cables.
- Installation and setup is much easier than wired counterparts.
- The equipment and setup costs are reduced.

Disadvantages of WLANs

- Since radio waves are used for communications, the signals are noisier with more interference from nearby systems.

- Greater care is needed for encrypting information. Also, they are more prone to errors. So, they require greater bandwidth than the wired LANs.
- WLANs are slower than wired LANs.

UNIT-III

Mobile Database - Data Management Issues

Data management technology that can support easy data access from and to mobile devices is among the main concerns in mobile information systems.

Mobile computing may be considered a variation of distributed computing.

The two scenarios in which mobile databases are distributed are:

- Among the wired components, the entire database is distributed, possibly with full or partial replication.
- A base station or fixed host manages its own database with a DBMS like functionality, with additional functionality for locating mobile units and additional query and transaction management features to meet the requirements of mobile environments

Among the wired and wireless components, the database is distributed.

Among the base stations or fixed hosts and mobile units, the data management responsibility is shared.

Here are some of the issues which arise in **data management** of the mobile databases:

1. Mobile database design

Because of the frequent shutdown and for handling the queries, the global name resolution problem is important.

2. Security

The data which is left at the fixed location is more secure as compared to mobile data. That is mobile data is less secure.

Data are also becoming more volatile and techniques must be able to compensate for its loss.

The most important thing needed in this environment is the authorizing access to critical data and proper techniques.

3. Data distribution and replication

Uneven distribution of data among the mobile units and the base stations take place here. Higher data availability and low cost of remote access is there in data distribution and replication. The problem of Cache management is compounded by the consistency constraints. The most updated data and frequently accessed data is provided by the Caches to the mobile units. It process their own transactions. There is most efficient access of data and higher security is available.

Replication issues

There is increase of costs for updates and signaling due to increase in number of replicas. Mobile hosts can move anywhere and anytime.

5. Division of labor

There is a certain change in the division of labour in query processing because of certain characteristics of the mobile environment. There are some of the cases in which the client must function independently of the server.

6. Transaction models

In mobile environment, the issues of correctness of transactions and fault tolerance

are aggravated. All transactions must satisfy the ACID properties, these are atomic, consistency, isolation, and durability.

Depending upon the movement of the mobile unit, possibly on multiple data sets and through several base station, a mobile transaction is executed sequentially. When the mobile computers are disconnected, ACID properties gets hard to enforce. Because of the disconnection in mobile units, there is expectation that a mobile transaction will be lived long.

7. Recovery and fault tolerance

Fault tolerance is the ability of a system to perform its function correctly even in the presence of internal faults. Faults can be classified in two types:

transient and permanent. Without any apparent intervention, a transient fault will be eventually disappeared but a permanent fault will remain unless it is removed by some external agency.

The mobile database environment must deal with site, transaction, media, and communication failures. Due to limited battery power there is a site failure at MU. If a voluntary shutdown occurs in MU, then it should not be treated as a failure. Whenever Mu crosses the cells, most frequently there will be transaction failures during handoff. Due to failure of MU, there is a big cause of network partitioning and affection of the routing algorithms. The characterization of mobile computing is done by:

- Limiting resource availability
- Frequent disconnection
- High mobility

- Low bandwidth

8. Location based service

One of the most challenging tasks which must be undertaken is determining the location of mobile users, which must be undertaken in order to enable a location based service. A cache information becomes sale when clients move location dependent. Eviction techniques are important in this case. Issues that arises in location and services are:

- User Privacy
- Diverse mobile mapping standards
- Market capability
- Interoperability

9. Query processing

Because of the mobility and rapid resource changes of mobile units, Query optimization becomes the most complicated.

Communication cost is the most important in distributed environments. It is possible to formulate location dependent queries. There is difficulty in estimating the communication costs in distributed environments because the mobile host may be situated in different locations. There is a requirement of dynamic optimization strategies in the mobile distributed context.

Data Replication

Data Replication in Mobile Computing

Data replication is the process of making copies of data stored on various sites in order to improve -

- reliability,
- efficiency,
- robustness,
- simpler transaction,
- fault tolerance and
- reduce network load

Goals of data replication

Data replication is performed with the purpose to

- Increase the availability of data.
- Speed up the query evaluation.

Types of data replication

There are two types of data replication

1. Synchronous Replication:

In synchronous replication, the replica of the database is modified immediately after changes are made in the relation table.
So there is no difference between original data and replicated data table.

2. Asynchronous replication:

In asynchronous replication, the replica will be modified after commit action is fired on to the database.

Replication Schemes

The three replication schemes are as follows:

1. Full Replication scheme

In full replication scheme, the database is available at all the locations to ease the user in communication network

Advantages of full replication

- It gives high availability of data. In this scheme database is available to at each location.
- It support faster execution of queries.

Disadvantages of full replication

- In full replication scheme concurrency control is difficult to achieve in full replication.
- During updating each and every site need to be updated therefore update operation is slower.

2. No Replication

No replication means, each fragment is stored exactly at one location only.

Advantages of no replication

- Concurrency can be easily minimized.
- Easy recovery of data becomes easy.

Disadvantages of no replication

- Poor availability of data.
- Slows down the query execution process, because multiple clients are accessing the same data at the same server.

3. Partial replication

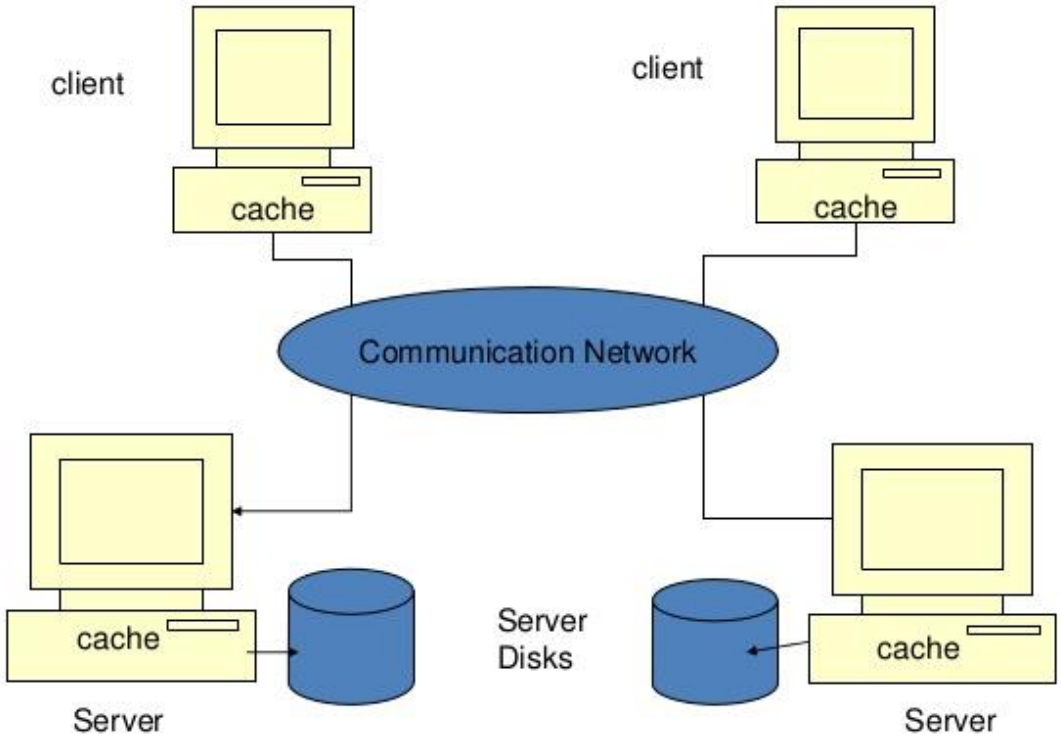
Partial replication scheme means only part of the or data fragments are replicated.

File System

The general goal of a file system is to support efficient, transparent, and consistent access to files, no matter where the client requesting files or the server(s) offering files are located.

In Distributed File System clients or users access files and folders that are provided by the file servers.

The client gives the request and the server provides a service to the client through a file service interface. Service fulfills the request of the client by providing the requested file view, part of the file view or folders.



Architecture of a distributed file system: client-server model

The important challenges of distributed systems apply to DFS are as below:

Transparency of Dfs

Dfs must be completely transparent to ensure the following to its users

- **Location**

  A Dfs user or client cannot informed be about the exact location of file the client is accessing

- **Migration**

  In DFS a file can be transparently moved from one server to another

- **Replication:**

  In Dfs multiple copies of the same file may present

- **Concurrency:**

  At the same time, multiple clients can access the same file in DFS.

- **Consistency:**

  Due to replicated copies of a file and permitted concurrent access to files inconsistency may occur in the data

- **Security:**

  In DFS each and every client must authenticate themselves before using a file and servers must decide whether the clients are authorized to do the requested operation.

  Also, communication between clients and the file server must be done with complete security.

- **Fault tolerance:**

  In DFS all the clients should be able to work continuously if a file server crashes. Similarly, data must not be lost. Also, a restarted file server must be recovered to a valid state.

What is a distributed file system with example

Examples of Dfs

- **Andrew File System (AFS)**
- **Coda**
- **Ficus**
- **Rover**
- **MI-NFS**

Coda

Coda is a distributed file system developed as a research project at Carnegie Mellon University since 1987 under the direction of Mahadev Satyanarayanan.

It descended directly from an older version of Andrew File System and offers many similar features.

The predecessor of many distributed file systems that can be used for mobile operation is the Andrew file system (AFS, (Howard, 1988)).
Coda is the successor of AFS and offers two different types of replication: server replication and caching on clients.
Disconnected clients work only on the cache, i.e., applications use only cached replicated files.

Coda is a transparent extension of the client's cache manager. Its general architecture is valid for most of today's mobile systems that utilise a cache.

It has many features that are very desirable for network file systems.
Currently, Coda has several features not found elsewhere.

1. **disconnected** operation for mobile computing
2. is **freely available** under a liberal license
3. **high performance** through client side persistent caching
4. **server replication**
5. **security** model for authentication, encryption and access control
6. **continued operation during partial network failures** in server network
7. network **bandwith adaptation**
8. good **scalability**
9. **well defined semantics** of sharing, even in the presence of network failures

Ficus DFS

Ficus is a distributed file system, which is not based on a client/server approach (Popek, 1990), (Heidemann, 1992).
Ficus allows the optimistic use of replicates, detects write conflicts, and solves conflicts on directories.
Ficus uses **gossip protocol**s, an idea many other systems took over later.
A mobile computer does not necessarily need to have a direct connection to a server.

With the help of other mobile computers, it can propagate updates through the network until it reaches a fixed network and the server.

Thus, changes on files propagate through the network step-by-step.

Ficus tries to minimize the exchange of files that are valid only for a short time, e.g. temporary files.

A critical issue for gossip protocols is how fast they propagate to the client that needs this information and how much unnecessary traffic it causes to propagate information to clients that are not interested.

Mio-NFS

The system mobile integration of NFS (MIo–NFS) is an extension of the Network File System (NFS, (Guedes, 1995)).

In contrast to many other systems, MIo-NFS uses a pessimistic approach with tokens controling access to files.

Only the token-holder for a specific file may change this file, so Mio-NFS avoids write conflicts.

Mio-NFS supports three different modes:

- **Connected:** The server handles all access to files as usual.
- **Loosely connected:** Clients use local replicates, exchange tokens over the network, and update files via the network.
- **Disconnected:** The client uses only local replicates. Writing is only allowed if the client is token-holder.

Rover DFS

Compared to Coda, the Rover platform uses another approach to support mobility (Joseph, 1997a and 1997b).
Instead of adapting existing applications for mobile devices, Rover provides a platform for developing new, mobility aware applications.

Two new components have been introduced in Rover-

Relocatable dynamic objects – these are objects that can be dynamically loaded into a client computer from a server (or vice-versa) to reduce client-server communication.
A trade-off between transferring objects and transferring only data for objects has to be found.
If a client needs an object quite often, it makes sense to migrate the object.

Object migration for a single access, on the other hand, creates too much overhead.

Queued remote procedure calls allow for non-blocking RPCs even when a host is disconnected.
Requests and responses are exchanged as soon as a connection is available again.
Conflict resolution is done in the server and is application specific.

UNIT-IV

Tcp Ip Protocol - TCP IP Protocol Over Wireless

The Transmission Control Protocol (TCP) is one of the most important and core protocols of the Internet protocol suite. It is referred to as TCP/IP.

Main features of TCP Protocol

- TCP is reliable protocol
- TCP protocol guarantees in-order delivery of data.
- The protocol covers congestion control and flow control mechanisms also.
- TCP supports a variety of important internet application protocols and corresponding applications such as World Wide Web, e-mail, File Transfer Protocol and Secure Shell.

In the Internet protocol suite, TCP is the intermediate layer. It is between the Internet layer and application layer.

The major responsibilities of TCP in an active session are as follows-

• TCP allows reliable in-order transport of data without loss of any data.

• TCP manages congestions in the networks system without any kind of degradation of the network performance

• The protocol controls a packet flow between the source who transmit data and the receiver. It also keep track that the data should not exceed the receiver's capacity.

TCP applies a number of mechanisms to attain high performance and avoid congestion problem.

TCP mechanisms control the rate of data entering into the network, and keeping the data flow below the prescribed rate in order to avoid congestion collapse.

There are a number of mechanisms in TCP that control the competence of TCP in a mobile environment.

Acknowledgments for information sent, or lack of acknowledgments, are used by source or the senders to perfectly interpret network condition between the TCP sender and TCP receiver.

Congestion Control

A transport layer protocol like TCP has been designed for fixed wired networks with fixed end system nodes.

Congestion may appear in any kind of network system and even in carefully designed network also.

The packet-buffers of a router are jam-packed and the router cannot forward the packets fast enough because the sum of the incoming rates of the packets destined for one output link is higher than the power of the output link.

In this condition the router can drop the packets. The dropped packet is lost for the communication, and the receiver notices a break in the packet stream. Now the receiver node does not directly inform the sender node about which packet is missing, but continues to acknowledge all in sequence packets including the missing packet.

The sender node checks the missing acknowledgement for the missed packet and thinks a packet loss due to congestion problem.

Now, retransmitting the lost packet and continuing at the previous sending rate would now be risky, because it might increase the congestion problem.

To reduce congestion, TCP slows down the transmission rate considerably. All other TCP connections considering the same congestion and do exactly the same and hence, the congestion is resolved speedily.

Slow start

TCP's reaction to a missing acknowledgement is quite drastic, but it is necessary to get rid of congestion quickly. The behavior TCP shows after the detection of congestion is called slow start.

The sender always calculates a congestion window for a receiver. The start size of the congestion window is one segment (TCP packet). The sender sends one packet and waits for acknowledgement. If this acknowledgement arrives, the sender increases the congestion window by one, now sending two packets (congestion window = 2). This scheme doubles the congestion window every time the acknowledgements come back, which takes one round trip time (RTT). This is called the exponential growth of the congestion window in the slow start mechanism.

But doubling the congestion window is too dangerous. The exponential growth stops at the congestion threshold.

As soon as the congestion window reaches the congestion threshold, further increase of the transmission rate is only linear by adding 1 to the congestion window each time the acknowledgements come back.

Linear increase continues until a time-out at the sender occurs due to a missing acknowledgement, or until the sender detects a gap in transmitted data because of continuous acknowledgements for the same packet.

In either case the sender sets the congestion threshold to half of the current congestion window. The congestion window itself is set to one segment and the sender starts sending a single segment. The exponential growth starts once more up to the new congestion threshold, then the window grows in linear fashion.

Fast retransmit/fast recovery

The congestion threshold can be reduced because of two reasons.

1. First one is if the sender receives continuous acknowledgements for the same packet. It informs the sender that the receiver has got all the packets upto the acknowledged packet in the sequence and also the receiver is receiving something continuously from the sender.

The gap in the packet stream is not due to congestion, but a simple packet loss due to a transmission error.

The sender can now retransmit the missing packet(s) before the timer expires.

This behavior is called fast retransmit.

It is an early enhancement for preventing slow-start to trigger on losses not caused by congestion.

The receipt of acknowledgements shows that there is no congestion to justify a slow start.

The sender can continue with the current congestion window. The sender performs a fast recovery from the packet loss. This mechanism can improve the efficiency of TCP dramatically.

2. The other reason for activating slow start is a time-out due to a missing acknowledgement. TCP using fast retransmit/fast recovery interprets this congestion in the network and activates the slow start mechanism.

Problems with Traditional TCP in wireless environments

Slow Start mechanism in fixed networks decreases the efficiency of TCP if used with mobile receivers or senders.

Error rates on wireless links are orders of magnitude higher compared to fixed fiber or copper links. This makes compensation for packet loss by TCP quite difficult. Mobility itself can cause packet loss.

There are many situations where a soft handover from one access point to another is not possible for a mobile end-system.

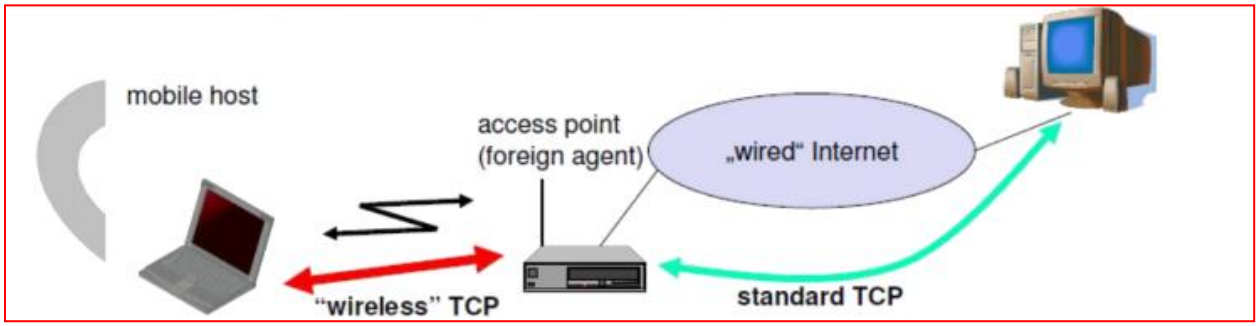
Standard TCP reacts with slow start if acknowledgements are missing, which does not help in the case of transmission errors over wireless links and which does not really help during handover.

This behavior results in a severe performance degradation of an unchanged TCP if used together with wireless links or mobile nodes

Classical TCP Improvements

Indirect TCP (I-TCP)

Indirect TCP segments a TCP connection into a fixed part and a wireless part. The following figure shows an example with a mobile host connected via a wireless link and an access point to the 'wired' internet where the correspondent host resides.

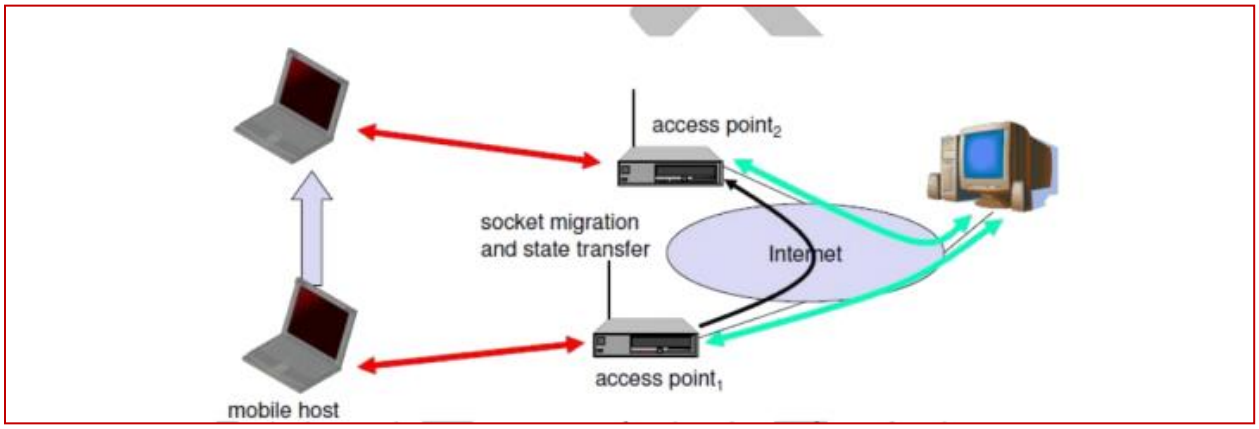


Standard TCP is used between the fixed computer and the access point. No computer in the internet recognizes any changes to TCP. Instead of the mobile host, the access point now terminates the standard TCP connection, acting as a proxy. This means that the access point is now seen as the mobile host for the fixed host and as the fixed host for the mobile host.

Between the access point and the mobile host, a special TCP, adapted to wireless links, is used.

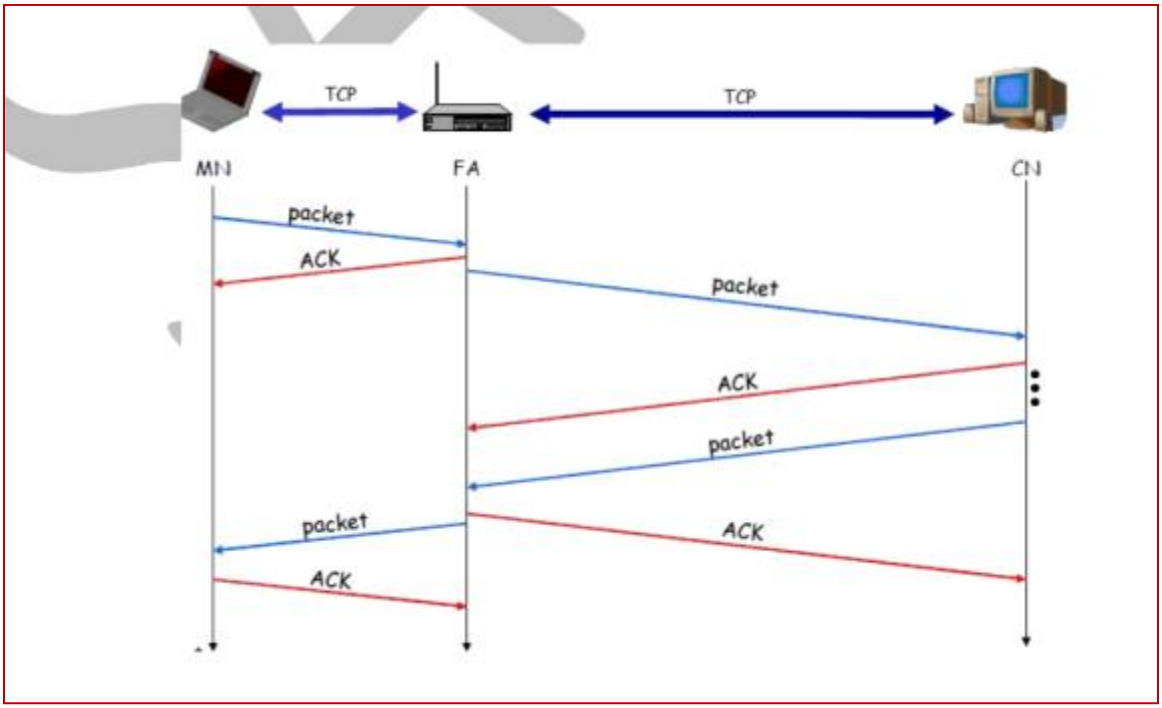
The foreign agent acts as a proxy and relays all data in both directions. If CH (correspondent host) sends a packet to the MH, the FA acknowledges it and forwards it to the MH. MH acknowledges on successful reception, but this is only used by the FA. If a packet is lost on the wireless link, CH doesn't observe it and

FA tries to retransmit it locally to maintain reliable data transport. If the MH sends a packet, the FA acknowledges it and forwards it to CH. If the packet is lost on the wireless link, the mobile hosts notice this much faster due to the lower round trip time and can directly retransmit the packet. Packet loss in the wired network is now handled by the foreign agent.



Socket and state migration after handover of a mobile host

During handover, the buffered packets, as well as the system state (packet sequence number, acknowledgements, ports, etc), must migrate to the new agent. No new connection may be established for the mobile host, and the correspondent host must not see any changes in connection state. Packet delivery in I-TCP is shown below:



Advantages of I-TCP

No changes in the fixed network necessary, no changes for the hosts (TCP protocol) necessary, all current optimizations to TCP still work

Simple to control; mobile TCP is used only for one hop between, e.g., a foreign agent and mobile host.

1. transmission errors on the wireless link do not propagate into the fixed network

2. therefore, a very fast retransmission of packets is possible, the short delay on the mobile hop s known

Disadvantages of I-TCP

- Loss of end-to-end semantics:- an acknowledgement to a sender no longer means that a receiver really has received a packet, foreign agents might crash.

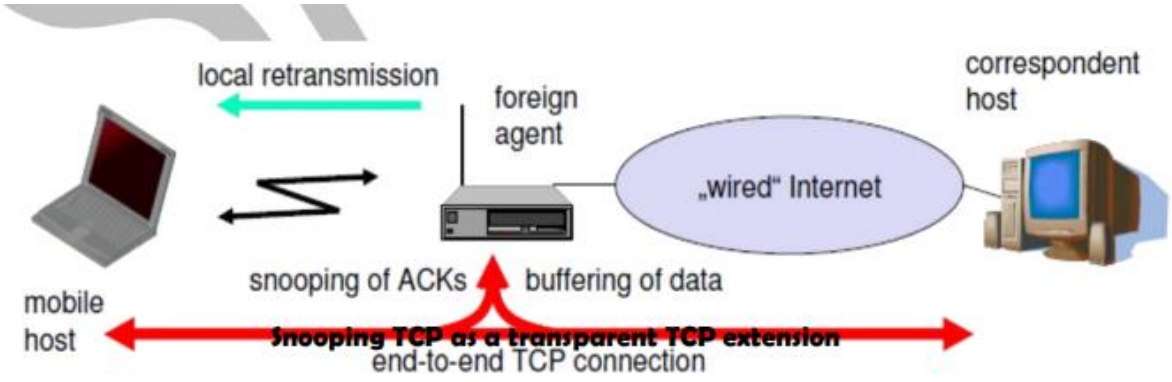
- Higher latency possible:- due to buffering of data within the foreign agent and forwarding to a new foreign agent
- Security issue:- The foreign agent must be a trusted entity

Snooping TCP

The main drawback of I-TCP is the segmentation of the single TCP connection into two TCP connections, which loses the original end-to-end TCP semantic.

A new enhancement, which leaves the TCP connection intact and is completely transparent, is Snooping TCP.

The main function is to buffer data close to the mobile host to perform fast local retransmission in case of packet loss.



Here, the foreign agent buffers all packets with destination mobile host and additionally 'snoops' the packet flow in both directions to recognize acknowledgements.

The foreign agent buffers every packet until it receives an acknowledgement from the mobile host.

If the FA does not receive an acknowledgement from the mobile host within a certain amount of time, either the packet or the acknowledgement has been lost.

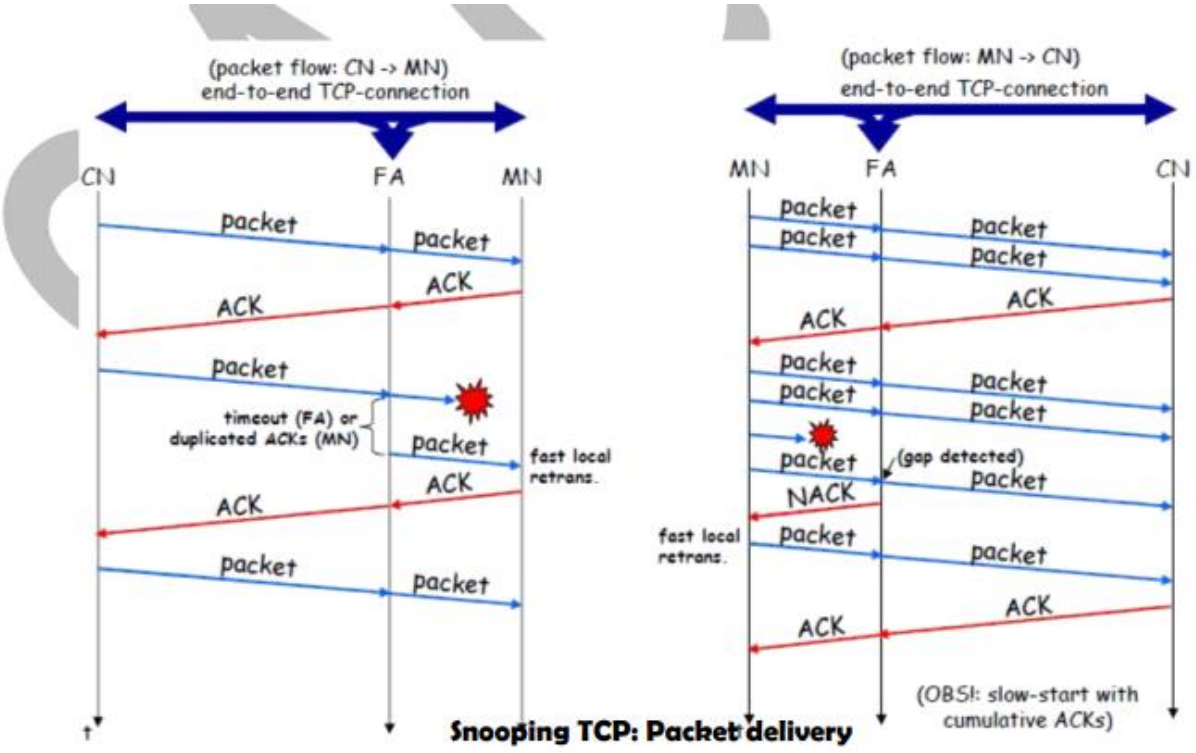
Alternatively, the foreign agent could receive a duplicate ACK which also shows the loss of a packet.

Now, the FA retransmits the packet directly from the buffer thus performing a faster retransmission compared to the CH. For transparency, the FA does not acknowledge data to the CH, which would violate end-to-end semantic in case of a FA failure.

The foreign agent can filter the duplicate acknowledgements to avoid unnecessary retransmissions of data from the correspondent host.

If the foreign agent now crashes, the time-out of the correspondent host still works and triggers a retransmission. The foreign agent may discard duplicates of packets already retransmitted locally and acknowledged by the mobile host. This avoids unnecessary traffic on the wireless link.

For data transfer from the mobile host with destination correspondent host, the FA snoops into the packet stream to detect gaps in the sequence numbers of TCP. As soon as the foreign agent detects a missing packet, it returns a negative acknowledgement (NACK) to the mobile host. The mobile host can now retransmit the missing packet immediately. Reordering of packets is done automatically at the correspondent host by TCP.



Advantages of snooping TCP:

- The end-to-end TCP semantic is preserved.
- Most of the enhancements are done in the foreign agent itself which keeps correspondent host unchanged.
- Handover of state is not required as soon as the mobile host moves to another foreign agent. Even though packets are present in the buffer, time out at the CH occurs and the packets are transmitted to the new COA.
- No problem arises if the new foreign agent uses the enhancement or not. If not, the approach automatically falls back to the standard solution.

Disadvantages of snooping TCP

- Snooping TCP does not isolate the behavior of the wireless link as well as I - TCP. Transmission errors may propagate till CH.
- Using negative acknowledgements between the foreign agent and the mobile host assumes additional mechanisms on the mobile host.

Mobile TCP

Both I-TCP and Snooping TCP does not help much, if a mobile host gets disconnected.

The M-TCP (mobile TCP) approach has the same goals as I-TCP and snooping TCP: to prevent the sender window from shrinking if bit errors or disconnection but not congestion cause current problems.

M-TCP wants to improve overall throughput, to lower the delay, to maintain end-to-end semantics of TCP, and to provide a more efficient handover.

Additionally, M-TCP is especially adapted to the problems arising from lengthy or frequent disconnections.

M-TCP splits the TCP connection into two parts as I-TCP does. An unmodified TCP is used on the standard host-supervisory host (SH) connection, while an optimized TCP is used on the SH-MH connection.

The SH monitors all packets sent to the MH and ACKs returned from the MH. If the SH does not receive an ACK for some time, it assumes that the MH is disconnected. It then chokes the sender by setting the sender's window size to 0. Setting the window size to 0 forces the sender to go into persistent mode, i.e., the

state of the sender will not change no matter how long the receiver is disconnected. This means that the sender will not try to retransmit data. As soon as the SH (either the old SH or a new SH) detects connectivity again, it reopens the window of the sender to the old value. The sender can continue sending at full speed. This mechanism does not require changes to the sender's TCP.

Advantages of M-TCP:

- It maintains the TCP end-to-end semantics. The SH does not send any ACK itself but forwards the ACKs from the MH.
- If the MH is disconnected, it avoids useless retransmissions, slow starts or breaking connections by simply shrinking the sender's window to 0.
- As no buffering is done as in I-TCP, there is no need to forward buffers to a new SH. Lost packets will be automatically retransmitted to the SH.

Disadvantages of M-TCP:

As the SH does not act as proxy as in I-TCP, packet loss on the wireless link due to bit errors is propagated to the sender. M-TCP assumes low bit error rates, which is not always a valid assumption.

Transmission/time-out freezing

Often, MAC layer notices connection problems even before the connection is actually interrupted from a TCP point of view and also knows the real reason for the interruption.

The MAC layer can inform the TCP layer of an upcoming loss of connection or that the current interruption is not caused by congestion.

TCP can now stop sending and ‘freezes’ the current state of its congestion window and further timers.

If the MAC layer notices the upcoming interruption early enough, both the mobile and correspondent host can be informed. With a fast interruption of the wireless link, additional mechanisms in the access point are needed to inform the correspondent host of the reason for interruption. Otherwise, the correspondent host goes into slow start assuming congestion and finally breaks the connection.

As soon as the MAC layer detects connectivity again, it signals TCP that it can resume operation at exactly the same point where it had been forced to stop. For TCP time simply does not advance, so no timers expire.

Advantages:

It offers a way to resume TCP connections even after long interruptions of the connection.

It can be used together with encrypted data as it is independent of other TCP mechanisms such as sequence no or acknowledgements

Disadvantages:

Lots of changes have to be made in software of MH, CH and FA.

Selective retransmission

A very useful extension of TCP is the use of selective retransmission.

TCP acknowledgements are cumulative, i.e., they acknowledge in-order receipt of packets up to a certain packet. A single acknowledgement confirms reception of all packets upto a certain packet. If a single packet is lost, the sender has to retransmit everything starting from the lost packet (go-back-n retransmission). This obviously wastes bandwidth, not just in the case of a mobile network, but for any network.

Using selective retransmission, TCP can indirectly request a selective retransmission of packets. The receiver can acknowledge single packets, not only trains of in-sequence packets. The sender can now determine precisely which packet is needed and can retransmit it.

The advantage of this approach is obvious: a sender retransmits only the lost packets. This lowers bandwidth requirements and is extremely helpful in slow wireless links.

The disadvantage is that more complex software on the receiver side is needed. Also more buffer space is needed to re-sequence data and to wait for gaps to be filled.

Transaction-oriented TCP

Assume an application running on the mobile host that sends a short request to a server from time to time, which responds with a short message and it requires reliable TCP transport of the packets.

For it to use normal TCP, it is inefficient because of the overhead involved.

Standard TCP is made up of three phases:

setup, data transfer and release. First, TCP uses a three-way handshake to establish the connection. At least one additional packet is usually needed for transmission of

the request, and requires three more packets to close the connection via a three-way handshake. So, for sending one data packet, TCP may need seven packets altogether. This kind of overhead is acceptable for long sessions in fixed networks, but is quite inefficient for short messages or sessions in wireless networks.

This led to the development of transaction-oriented TCP (T/TCP).

Transaction-oriented TCP (T/TCP) - T/TCP can combine packets for connection establishment and connection release with user data packets. This can reduce the number of packets down to two instead of seven.

The obvious advantage for certain applications is the reduction in the overhead which standard TCP has for connection setup and connection release.

Disadvantage is that it requires changes in the software in mobile host and all correspondent hosts.

This solution does not hide mobility anymore. Also, T/TCP exhibits several security problems.

UNIT-V

What is a Router - Types of Routing Algorithms

In order to transfer the packets from source to the destination, the network layer must determine the best route through which packets can be transmitted.

- Whether the network layer provides datagram service or virtual circuit service, the main job of the network layer is to provide the best route. The routing protocol provides this job.
- The routing protocol is a routing algorithm that provides the best path from the source to the destination.
- The best path is the path that has the "least-cost path" from source to the destination.
- Routing is the process of forwarding the packets from source to the destination but the best route to send the packets is determined by the routing algorithm.

Classification of a Routing algorithm

The Routing algorithm is divided into two categories:

- Adaptive Routing algorithm
- Non-adaptive Routing algorithm

Adaptive Routing algorithm

- An adaptive routing algorithm is also known as dynamic routing algorithm.
- This algorithm makes the routing decisions based on the topology and network traffic.

- The main parameters related to this algorithm are hop count, distance and estimated transit time.

An adaptive routing algorithm can be classified into three parts:

- **Centralized algorithm:** It is also known as global routing algorithm as it computes the least-cost path between source and destination by using complete and global knowledge about the network. This algorithm takes the connectivity between the nodes and link cost as input, and this information is obtained before actually performing any calculation. **Link state algorithm** is referred to as a centralized algorithm since it is aware of the cost of each link in the network.
- **Isolation algorithm:** It is an algorithm that obtains the routing information by using local information rather than gathering information from other nodes.
- **Distributed algorithm:** It is also known as decentralized algorithm as it computes the least-cost path between source and destination in an iterative and distributed manner. In the decentralized algorithm, no node has the knowledge about the cost of all the network links. In the beginning, a node contains the information only about its own directly attached links and through an iterative process of calculation computes the least-cost path to the destination. A Distance vector algorithm is a decentralized algorithm as it never knows the complete path from source to the destination, instead it knows the direction through which the packet is to be forwarded along with the least cost path.

Non-Adaptive Routing algorithm

- Non Adaptive routing algorithm is also known as a static routing algorithm.
- When booting up the network, the routing information stores to the routers.
- Non Adaptive routing algorithms do not take the routing decision based on the network topology or network traffic.

The Non-Adaptive Routing algorithm is of two types:

Flooding: In case of flooding, every incoming packet is sent to all the outgoing links except the one from it has been reached. The disadvantage of flooding is that node may contain several copies of a particular packet.

Random walks: In case of random walks, a packet sent by the node to one of its neighbors randomly. An advantage of using random walks is that it uses the alternative routes very efficiently.

Basis Of Comparison	**Adaptive Routing algorithm**	**Non-Adaptive Routing algorithm**
Define	Adaptive Routing algorithm is an algorithm that constructs the routing table based on the network conditions.	The Non-Adaptive Routing algorithm is an algorithm that constructs the static table to determine which node to send the packet.
Usage	Adaptive routing algorithm is used by dynamic routing.	The Non-Adaptive Routing algorithm is used by static routing.
Routing	Routing decisions are made	Routing decisions are the static

decision	based on topology and network traffic.	tables.
Categorization	The types of adaptive routing algorithm, are Centralized, isolation and distributed algorithm.	The types of Non Adaptive routing algorithm are flooding and random walks.
Complexity	Adaptive Routing algorithms are more complex.	Non-Adaptive Routing algorithms are simple.

Distance Vector Routing Algorithm

- **The Distance vector algorithm is iterative, asynchronous and distributed.**
    - **Distributed:** It is distributed in that each node receives information from one or more of its directly attached neighbors, performs calculation and then distributes the result back to its neighbors.
    - **Iterative:** It is iterative in that its process continues until no more information is available to be exchanged between neighbors.
    - **Asynchronous:** It does not require that all of its nodes operate in the lock step with each other.
- The Distance vector algorithm is a dynamic algorithm.
- It is mainly used in ARPANET, and RIP.
- Each router maintains a distance table known as **Vector**.

Three Keys to understand the working of Distance Vector Routing Algorithm:

- **Knowledge about the whole network:** Each router shares its knowledge through the entire network. The Router sends its collected knowledge about the network to its neighbors.
- **Routing only to neighbors:** The router sends its knowledge about the network to only those routers which have direct links. The router sends whatever it has about the network through the ports. The information is received by the router and uses the information to update its own routing table.
- **Information sharing at regular intervals:** Within 30 seconds, the router sends the information to the neighboring routers.

Distance vector routing is an asynchronous algorithm in which node x sends the copy of its distance vector to all its neighbors. When node x receives the new distance vector from one of its neighboring vector, v, it saves the distance vector of v and uses the Bellman-Ford equation to update its own distance vector.

The node x has updated its own distance vector table by using the above equation and sends its updated table to all its neighbors so that they can update their own distance vectors.

Link State Routing

Link state routing is a technique in which each router shares the knowledge of its neighborhood with every other router in the internetwork.

The three keys to understand the Link State Routing algorithm:

- **Knowledge about the neighborhood:** Instead of sending its routing table, a router sends the information about its neighborhood only. A router broadcast its identities and cost of the directly attached links to other routers.
- **Flooding:** Each router sends the information to every other router on the internetwork except its neighbors. This process is known as Flooding. Every router that receives the packet sends the copies to all its neighbors. Finally, each and every router receives a copy of the same information.
- **Information sharing:** A router sends the information to every other router only when the change occurs in the information.

Link State Routing has two phases:

Reliable Flooding

- **Initial state:** Each node knows the cost of its neighbors.
- **Final state:** Each node knows the entire graph.
- The Link state routing algorithm is also known as Dijkstra's algorithm which is used to find the shortest path from one node to every other node in the network.
- The Dijkstra's algorithm is an iterative, and it has the property that after k^{th} iteration of the algorithm, the least cost paths are well known for k destination nodes.

Types of Network Routing Algorithms

Recently vast research is done on ad-hoc network routing protocols. All the routing algorithms or protocols are categorized into three categories-

- • Flat routing,

- • Hierarchical routing,
- • Geographic-position-assisted routing

Flat routing

Flat routing protocols includes all those routing protocols that do not setup hierarchical approach in which one node works as a head of the region. In flat routing protocols all the nodes works at the same level.

Flat routing is further divided into two subcategories:

- Proactive Network Routing
- Reactive Network Routing Protocols

Proactive Network Routing Protocols

It set up tables required for routing not considering of any traffic that would require routing functions. DSDV is proactive protocol. Many other routing protocols belonging to this category are based on a link-state algorithm that is used for fixed networks.

Link-state algorithms flood their information about neighbors from time to time or when needed. In mobile ad-hoc environments this method exhibits severe drawbacks related extra overhead on the nodes due to flooding.

Reactive Network Routing Protocols

Reactive protocols try to avoid this problem by setting up a path between the sender and the receiver only if a communication is waiting.

The two most important protocols are **dynamic source routing** and **ad-hoc on-demand distance vector protocol**.

AODV acquires and maintains routes only on demand like DSR does.

Hierarchical Network Routing

Routing algorithms like DSDV, AODV, and DSR suitable for a smaller number of

nodes and they depend on the mobility of the nodes.

For larger networks, clustering of nodes and using different routing algorithms between and within clusters can be a scalable and efficient solution of routing problems in the ad-hoc network.

The inspiration behind this technique is the locality property. Locality property means that if a cluster can be established, nodes remain within the cluster, only some nodes change the clusters.

If the topology within a cluster changes, only nodes of the cluster have to be informed about the change and nodes of other clusters only required to know how to route the cluster.

Geographic position assisted routing approach

Geographic-position-assisted routing scheme based on the geographic position of the mobile nodes. If mobile nodes know their geographical position this can be used for routing algorithms also. It improves the overall performance of routing algorithms.

One way to obtain geographic position information is via the global positioning system (GPS).

Therefore various schemes are there to create and maintain **network routing**. Still more optimization is needed to save time and space particularly in the ad-hoc **network routing**.