TABLE OF CONTENTS

S.NO	CHAPTER No.	TOPICS	PAGE NO.
		UNIT-I MOBILE NETWORKS	
1	1.1	Cellular Wireless Networks	1
2	1.2	GSM	16
3	1.3	Architecture	20
4	1.4	Protocols	24
5	1.5	Connection Establishment	26
6	1.6	Frequency Allocation	31
7	1.7	Routing	33
8	1.8	Mobility Management	34
9	1.9	Security	38
10	1.10	GPRS	41
		UNIT-II WIRELESS NETWORKS	
11	2.1	Wireless LANs and PANs	50
12	2.2	IEEE 802.11 Standard	52
13	2.3	Architecture	53
14	2.4	Services	55
15	2.5	Network	56
16	2.6	HiperLAN	56
17	2.7	Blue Tooth	68
18	2.8	Wi-Fi	75
19	2.9	WiMAX	79
		UNIT-III ROUTING	
20	3.1	Mobile IP	85
21	3.2	DHCP	87
22	3.3	AdHoc	89
23	3.4	Proactive and Reactive Routing Protocols	95
24	3.5	Multicast Routing.	102
		UNIT-IV TRANSPORT AND APPLICATION LAYERS	
25	4.1	Mobile TCP	113
26	4.2	WAP	115
27	4.3	Architecture	117
28	4.4	WWW Programming Model	118
29	4.5	WDP	120
30	4.6	WTLS	124
31	4.7	WTP	125
32	4.8	WSP	125
33	4.9	WAE	127
34	4.10	WTA Architecture	129
35	4.11	WML	130
36	4.12	WML Scripts	132

UNIT-V PERVASIVE COMPUTING			
37	**5.1**	**Pervasive computing infrastructure**	**133**
38	**5.2**	**applications**	**137**
39	**5.3**	**Device Technology**	**141**
40	**5.4**	**Hardware, Human-machine Interfaces**	**143**
41	**5.5**	**Biometrics, and Operating systems**	**148**
42	**5.6**	**Device Connectivity**	**150**
43	**5.7**	**Protocols, Security, and Device Management**	**155**
44	**5.8**	**Pervasive Web Application architecture**	**160**
45	**5.9**	**Access from PCs and PDAs**	**163**
46	**5.10**	**Access via WAP**	**166**

CS2402 MOBILE AND PERVASIVE COMPUTING L T P C
3 0 0 3

UNIT I MOBILE NETWORKS 9
Cellular Wireless Networks – GSM – Architecture – Protocols – Connection Establishment – Frequency Allocation – Routing – Mobility Management – Security – GPRS.

UNIT II WIRELESS NETWORKS 9
Wireless LANs and PANs – IEEE 802.11 Standard – Architecture – Services –Network – HiperLAN – Blue Tooth- Wi-Fi – WiMAX

UNIT III ROUTING 9
Mobile IP – DHCP – AdHoc– Proactive and Reactive Routing Protocols – Multicast Routing.

UNIT IV TRANSPORT AND APPLICATION LAYERS 9
Mobile TCP– WAP – Architecture – WWW Programming Model– WDP – WTLS – WTP – WSP – WAE – WTA Architecture – WML – WMLScripts.

UNIT V PERVASIVE COMPUTING 9
Pervasive computing infrastructure-applications- Device Technology - Hardware, Human-machine Interfaces, Biometrics, and Operating systems– Device Connectivity – Protocols, Security, and Device Management- Pervasive Web Application architecture- Access from PCs and PDAs - Access via WAP

TOTAL: 45 PERIODS

TEXT BOOKS:
1. Jochen Schiller, "Mobile Communications", PHI, Second Edition, 2003.
2. Jochen Burkhardt, Pervasive Computing: Technology and Architecture of Mobile Internet Applications, Addison-Wesley Professional; 3rd edition, 2007

REFERENCES:
1. Frank Adelstein, Sandeep KS Gupta, Golden Richard, Fundamentals of Mobile and Pervasive Computing, McGraw-Hill 2005
2. Debashis Saha, Networking Infrastructure for Pervasive Computing: Enabling Technologies, Kluwer Academic Publisher, Springer; First edition, 2002
3. Introduction to Wireless and Mobile Systems by Agrawal and Zeng, Brooks/ Cole (Thomson Learning), First edition, 2002
4. Uwe Hansmann, Lothar Merk, Martin S. Nicklons and Thomas Stober, Principles of Mobile Computing, Springer, New York, 2003.

UNIT I MOBILE NETWORKS

Cellular Wireless Networks – GSM – Architecture – Protocols – Connection Establishment – Frequency Allocation – Routing – Mobility Management – Security – GPRS.

PREREQUISTIES DISCUSSION:

In this unit introduces the field of mobile and wireless communication, presents a short history and challenges for research, and concludes with a market vision, which shows the potential of mobile technology.

1.1 CELLULAR WIRELESS NETWORK

CONCEPTS:

INTRODUCTION TO CELLULAR NETWORK

Cellular network or mobile network is a radio network scattered over land areas called **cells**. A cell is coverage of base station connected to other stations via wire or fiber or wirelessly through switching centre. Cellular network employ Space division multiplexing (SDM). Cellular network consist of

- Cellular Base station.
- Mobile telephone switching offices (MTSO).
- Mobile communication devices.

Base station: Each cell is served by at least one fixed-location transceiver, known as a cell site or base station. It contains a radio transceiver and controller and provides radio communication to mobile units located in cell.

Mobile telephone switching offices (MTSO): The MTSO links calls together using traditional copper, fiber optic, or microwave technology. It also allows mobile communication devices in the cell to dial out and alerts devices in the cell of incoming calls. The MTSO monitors the quality of the communications signal and transfers the call to another base station which is better suited to provide communication to the mobile device.

Mobile communication devices: The mobile communication devices consist of hand held phones, car phones, notebook computers, palm-top computers, and portable data collection devices. When these mobile units communicate to the network, they must register with the system by subscribing to a carrier service.

If the cell size is preferred as circle, then a overlap and gap occurs. The cell which gives the actual radio coverage is called **footprint of a cell**. It might so happen that either there may be an overlap between any two such side by side circles or there might be a gap between the coverage areas of two adjacent circles. In a cellular radio system, a land region to be supplied with radio service is

divided into regular shaped cells, even though hexagonal cells are conventional. A regular shape for cellular design over a territory which can be served by 3 regular Polygons, namely, equilateral triangle, square and regular hexagon, which can cover the entire area without any overlap and gaps.

Along with its regularity, a cell must be designed such that it is most reliable too, i.e., it supports even the weakest mobile with occurs at the edges of the cell. For any distance between the center and the farthest point in the cell from it, a regular hexagon covers the maximum area. In a cellular network, each cell uses a dissimilar set of frequencies from neighboring cells, to avoid interference and provide assured bandwidth within each cell.

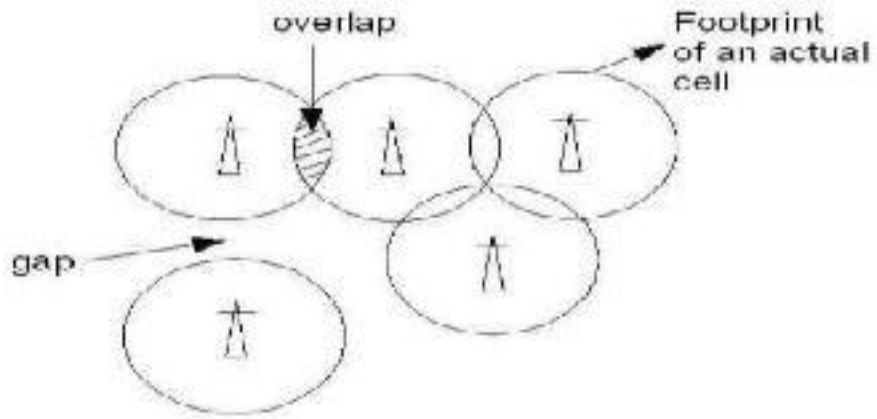


Fig:1.1 Foot Print of cell showing overlap and gaps

Cell radius differs from tens of meters in buildings, and hundreds of meters in cities, up to tens of kilometers in the landscape. The set of frequencies can be reused in other cells, provided that the same frequencies are not reused in adjacent nearby cells as that would cause co-channel interference.

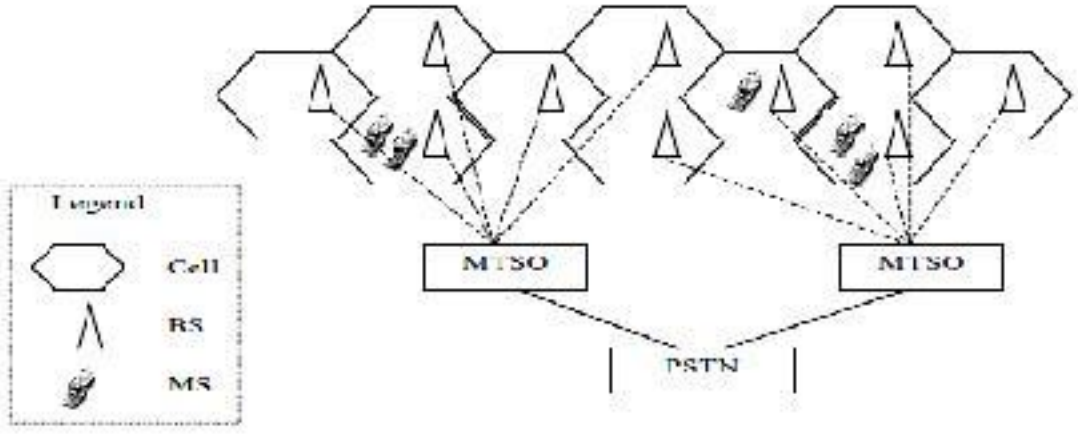


Fig:1.2 Typical cellular network

FUNDAMENTAL CONCEPT IN CELLULAR TECHNOLOGY

The radio spectrum contains many bands that are allocated and used for commercial, personal, and military applications. Fifty (50) MHz of spectrum allocated to cellular networks exists in the 824-849 MHz and the 869-894 MHz bands (Pagett, 1995). These bands are then further subdivided into 832 channels allowing many users in the same area to simultaneously access the network. Types of cellular network access are:

- Advanced mobile phone system (AMPS).
- Time division multiple access (TDMA).
- Code division multiple access (CDMA).

Cells are combined together to form clusters. There are 2 types of formation of clusters

- Three cells forming a cluster.
- Seven cells forming a cluster.

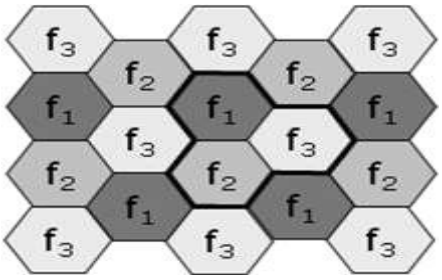


Fig 1.3 Three cell forming Cluster

The set of frequencies can be reused in other cells, provided that the same frequencies are not reused in adjacent nearby cells as that would cause co-channel interference. So never use same frequencies at same time with in the interference range.

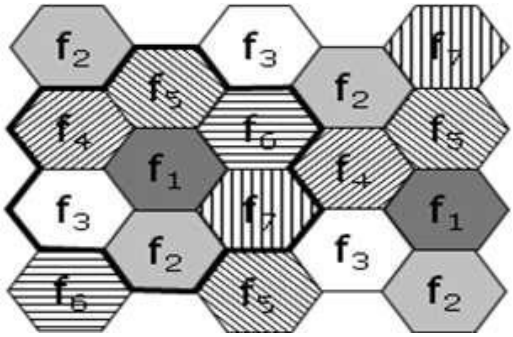


Fig 1.4 Seven cell forming cluster

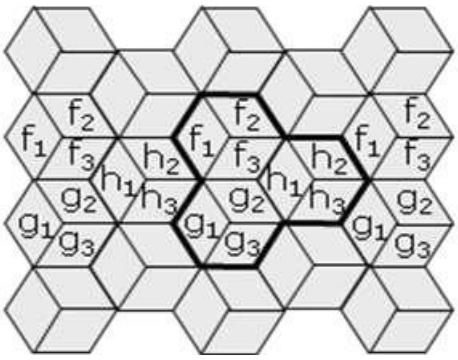


Fig 1.5 Cells with sectorized antenna

Sectorized antenna is an another method to reduce the interference.

Channel Assignment Strategies:

- Fixed frequency assignment.
- Dynamic frequency assignment.

Fixed frequency assignment:

If certain frequency are assigned to certain cells then it is called **fixed channel allocation (FCA)**. In fixed channel assignment strategy each cell is allocated a fixed number of voice channels. The problem related is dissimilar traffic load occurs in dissimilar cell.

Dynamic frequency assignment:

If the frequency is borrowed and assigned to cells then it is called **dynamic channel allocation(DCA)**. In dynamic channel assignment strategy channels are temporarily assigned for use in cells for the duration of the call. As the frequencies are recurring, the transmission power is restricted to stay away from interference with subsequent cell using the same frequency.

If the cell has heavy traffic and its neighboring cell has less load then the frequencies can be borrowed and assigned to the cell having heavy load. This is called **borrowing channel allocation(BCA).**The idea of breathe is that the cell can cover bulky area under light load and its size shrink under heavy load.CDM system faces a problem of cell size depending upon load.

Handoff process:

Handoff is an important task in cellular network. when a MS moves into another cell, while the conversation is still in progress, the MSC automatically transfers the call to a new FDD channel without disturbing the conversation. This process is called as **handoff.**

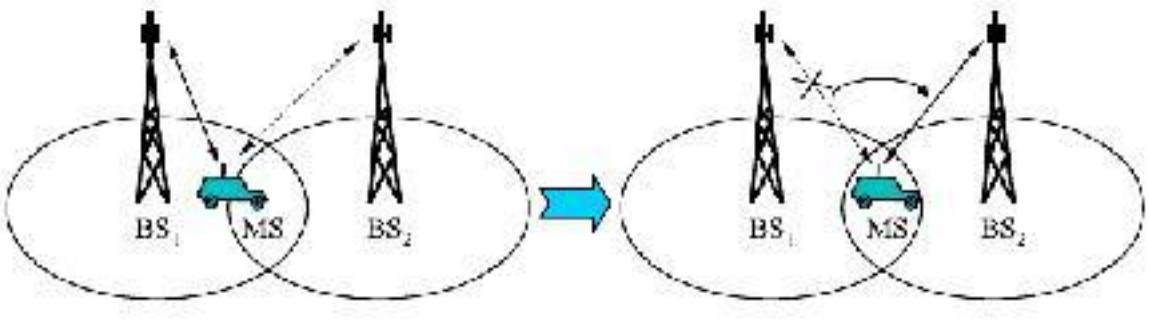


Fig 1.6a. Before handoff **b. After handoff**

Handoff performance metrics:

- The probability of a new call being blocked is referred as ***call blocking Probability.***
- The probability that the call is ended due to handoff is the ***call dropping probability.***
- The probability that a admitted call is not dropped before ended is the *call* ***completion probability.***
- The probability that a handoff is executed while the response conditions are inadequate is referred as ***Probability of unsuccessful handoff.***
- The probability that a handoff cannot be effectively completed is the ***handoff blocking probability.***
- The probability that a handoff occur earlier than call termination is the ***handoff probability.***
- The number of handoffs per unit time is referred as ***rate of handoff.***
- The duration of moment during a handoff in which a mobile is not connected to either base station is the ***Interruption duration.***
- Distance the mobile moves from the point at which the handoff should occur to the point at which it does occur is the ***handoff delay.***

Handoff strategy used to determine instant of handoff:

- Relative signal strength
- Prediction techniques
- Relative signal strength with hysteresis and threshold

- Relative signal strength with hysteresis
- Relative signal strength with threshold

Frequency reuse:

Frequency reuse, or, frequency planning, is a technique of reusing frequencies and channels within a communication system to improve capacity and spectral efficiency.

The characteristic of a cellular network is the ability to re-use frequencies to increase both coverage and capacity. Adjacent cells must use different frequencies to avoid interference, however there is no difficult with two cells sufficiently far apart in use on the same frequency. 10 to 50 frequencies are assigned to each cell. The elements that determine frequency reuse are the reuse distance and the reuse factor. The reuse distance, *D* is calculated as

$$D = R\sqrt{3N},$$

R – Cell radius and *N*- number of cells per cluster.

The rate at which the same frequency can be used in the network is called frequency reuse factor. It is *1/K*

K – Number of cells which cannot use the same frequencies for transmission.

Common values for frequency reuse factor are 1/3, 1/4, 1/7, 1/9 and 1/12.Factors limiting Frequency reuse are co-channel interference and adjacent channel interference.

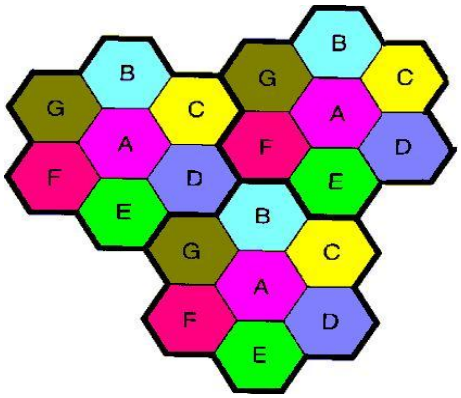


Fig 1.7 Example of frequency reuse

Each Colour/letter uses the same frequency band.

ADVANTAGES OF CELLULAR SYSTEMS WITH MINIATURE CELLS

1. **Higher capacity:** If SDM is employed it allows frequency reuse. If two transmitter is far away then frequency reuse is possible. If the cell is small, then more number of
   users are allowed.
2. **Less transmission power:** A mobile or hand held devices needs more power. The devices which are closer to Base station needs less power for transmission.
3. **Local interference:** If there is a large distance between sender and receiver then there will be more interference. If the cell size is small then mobile station and base station need to deal with local interference.
4. **Robustness:** The cellular systems are decentralized. If one antenna fails it will affect small region.

LIMITATIONS:

1. **Infrastructure needed:** Complex infra structure is needed to connect all base station which include antenna, switches, location register which make the system expensive.
2. **Handover needed:** Handling over the mobile from one cell to another if the signal strength decreases when the mobile station moves far away from BTS. If the cell size is small, then handover take place.
3. **Frequency planning:** Frequencies have to be distributed carefully to avoid interference between transmitter using same frequencies. Limited number of frequencies are available hence interference should be avoided.

CELLULAR OPERATION:

Cellular network organization uses low power transmitter(100W or less).The areas are divided into cells. Each cell is served by its own antenna and a base station consisting of transmitter, receiver, and control unit.

There are three basic devices they are:

- A mobile station(MS)
- A base transceiver Station(BS)
- A Mobile Telecommunications Switching Office (MTSO)

Base station include an antenna, a controller, and a number of receivers.

Base station is at center of each cell. Base station is connected to MTSO. One MTSO serve as multiple Base station. The link between MTSO to BS is by wire or wireless.MTSO connects calls between mobile units and from mobile to fixed telecommunications network .It assigns voice channel and performs handoffs and monitors calls (billing).

Two channels are available between mobile unit and BS, they are:

1. **Control channel:** They are used to exchange information and perform setup and maintaining calls. It establishes a relationship between Mobile unit and nearest BS.
2. **Traffic channel:** It carries voice or data connection between users.

Public Land Mobile Network (PLMN) refer to a cellular network that has land and radio based sections.

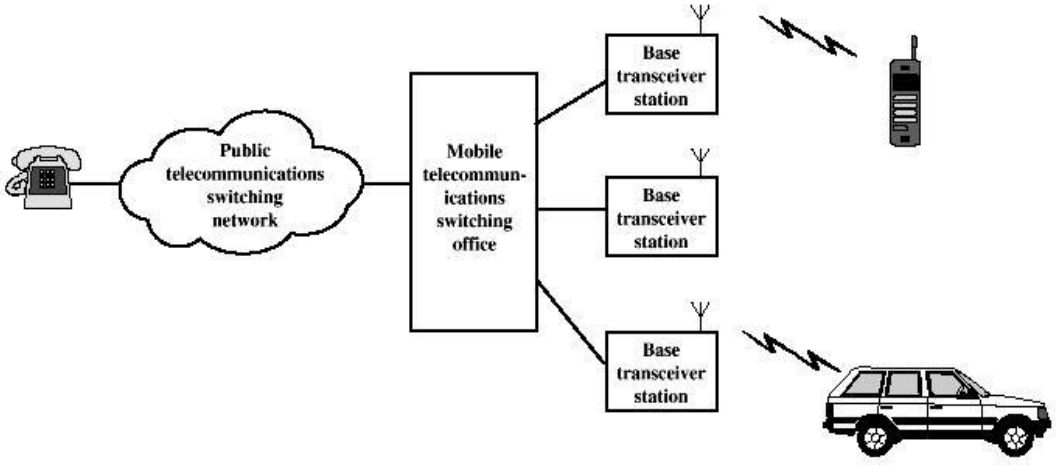


Fig 1.8 Overview of cellular network

This network consist of:

- **Mobile station (MS)** is a device used for communication over the network.
- **Base station transceiver (BST)** is a transmitter/receiver that are used to transmit/receive signals over the network.
- **Mobile switching center (MSC)** is used to Sets up and maintain calls made over the network.
- **Base station controller (BSC)** which provides a Communication between a group of BSTs and a single MSC is controlled by the BSC
- **Public switched telephone network (PSTN)** Consist of Section of the network that is land base.

Steps in MTSO controlled call connecting mobile units:

1. ***Mobile unit initialization*** scans and choose strongest set up control channel and automatically pick up a BS antenna of cell. Handshake is used to spot user and register location. Scan is recurring to allow for movement of change of cell.
2. ***Mobile originated call*** check if the set up channel is free and Send number on pre-selected channel.
3. In ***Paging*** MTSO attempts to connect to mobile unit. Depending on called mobile number the paging message will be sent to BSs. By using the setup channel Paging signal is transmitted.
4. In ***call accepted***, the Mobile unit recognizes the number on the set up channel and responds to BS which in turn send response to MTSO. Then the MTSO sets up a circuit between calling and called BSs and select a available traffic channel within cells and notifies BSs. The BSs notify mobile unit of channel.
5. In ***Ongoing call*** the Voice/data is exchanged through respective BSs and MTSO.
6. If the signal strength decreases as the mobile moves out of range from BTS it is called ***handoff***. And the traffic channel changes to the one assigned to new BS.

Other Functions:

1. ***Call blocking:*** On mobile-initiated call stage, if all the traffic channels are busy, the mobile tries again and again. After numeral retries, a busy tone will be returned.
2. ***Call termination:*** The User will hang up, MTSO is informed and the traffic channels at two BSs are released.
3. ***Call drop:*** If the BS cannot maintain a required signal strength then call drop will occur and the traffic channel is dropped and MTSO informed.
4. ***Calls to/from fixed and remote mobile subscriber:*** Here the MTSO connects to PSTN and can connect to mobile user and fixed subscriber via PSTN. MTSO can also connect to remote MTSO via PSTN or via dedicated line.

Mobile Radio Propagation Effects: Signal strength between BS and mobile unit is strong enough to maintain signal quality at the receiver. Signal propagation effects may interrupt the signal and causes error. This is called **fading**.

Power control:

Design issues making it advantageous to include dynamic power control in cellular systems.For effective communication, the power received must be sufficiently above the background noise. It is advantageous to minimize the power in the transmitted signal from the mobile. Thus it reduce co-channel interference, save battery power and alleviate health concern.Types of power control:

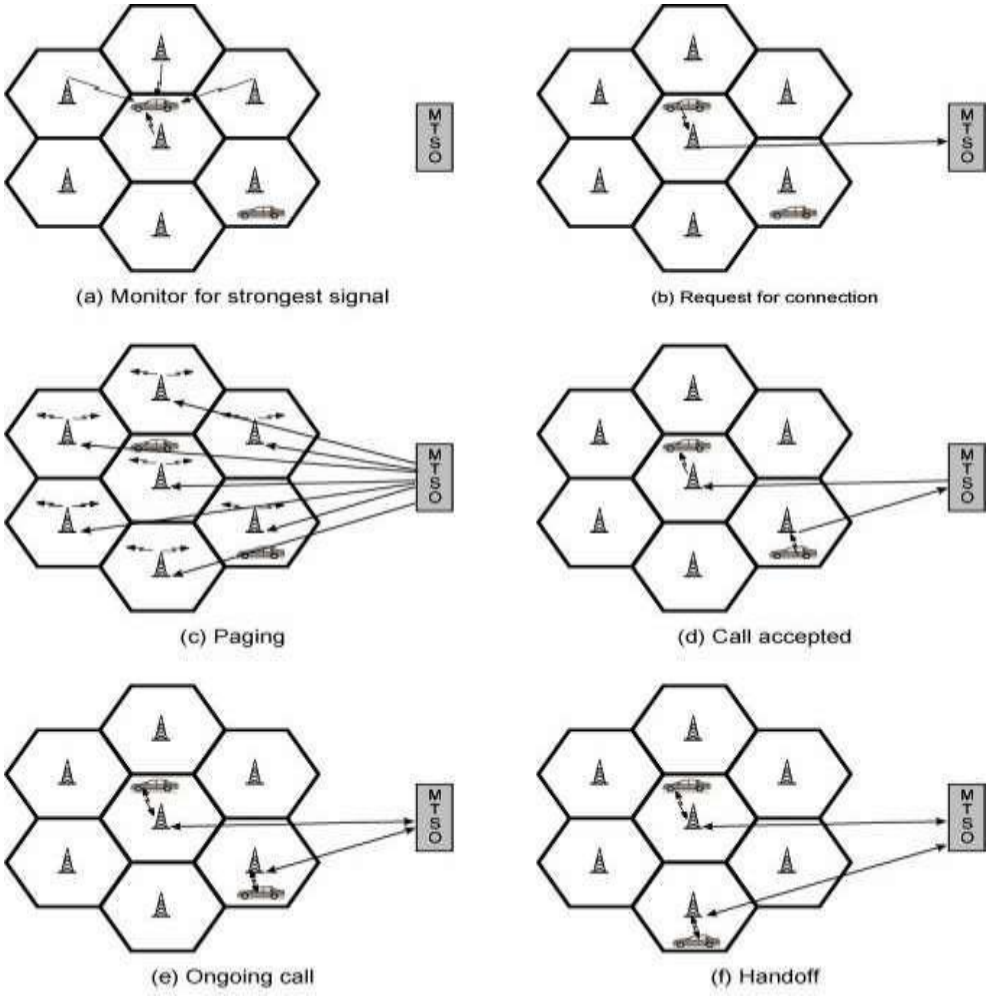


Fig 1.9 Call stages

Open-loop power control:

It depends solely on mobile unit. There is no feedback from BS. Open loop is not as accurate as closed loop, but it can react quicker to fluctuate in signal strength.

Closed-loop power control:

Based on performance metric the signal strength is adjusted in

reverse channel.BS makes power tuning decision and communication to mobile on control channel.

Traffic Engineering:

Traffic engineering is a method of optimizing the performance of a telecommunication network by vigorously analyzing, predicting and regulating the behavior of data transmitted over that network. Traffic engineering is also known as tele traffic engineering and traffic management. The method of traffic engineering can be applied to networks of all kinds, together with the PSTN (public switched telephone network), LANs (local area networks), WANs (wide area networks), cellular telephone networks, proprietary business and the Internet .For N simultaneous user capacity and L subscribers

$L < N$ – non-blocking system, $L > N$ – blocking system.

Traffic Intensity:

Load accessible to a system:

$$A=\lambda h$$

Where

λ -mean rate of calls attempted per unit time

h -mean holding time per successful call

A -average number of calls arriving during average holding period, for normalized

1.2 GSM

INTRODUCTION

GSM was formally known as Groupe Speciale Mobile (found in1982) and now it is abbreviated as Global System for mobile communications. It is a standard set developed by the European Telecommunications Standards Institute (ETSI) to describe protocols for second generation (2G) digital cellular networks used by mobile phones. It became the de facto global standard for mobile communications with over 80% market share. The GSM standard was developed as a replacement for first generation (1G) analog cellular networks, and originally described a digital, circuit switched network optimized for full duplex

voice telephony. Further improvements were made when the 3GPP developed third generation (3G) UMTS standards followed by fourth generation (4G) LTE Advanced standards.

The primary goal of GSM was to provide a mobile phone system that allows user to roam throughout Europe and provides voice services compatible to ISDN and other PSTN systems. GSM has initially been deployed in Europe using 890-915MHz for uplinks and 935-960 for downlinks.

- GSM 1800: Otherwise called as Digital cellular Systems DCS 1800 Uplink → 1710 to 1785 MHz
  Downlink → 1805 to 1880 MHz
- GSM 1900 : Otherwise called as Personal Communication Service PLS 1900
  Uplink → 1850 to 1910 MHz
  Downlink → 1930 to 1990 MHz
- GSM 400: Uplink → 450.4 to 478 MHz
  Downlink → 460 to 496 MHz
- GSM Rail is used in European Countries and railroad systems.

FEATURES OF GSM RAIL

It offers 19 exclusive channels for voice and data traffic. The special features like emergency calls, voice group call service etc. are available. Calls are prioritized. Mostly used to control the trains, switches, signals, gates.

GSM SERVICES

- GSM allows the integration of voice and data services and also the internetworking with the existing network.
- There are three types of services offered by GSM

- Bearer Service
- Tele Service
- SupplementaryService

REFERENCE MODEL FOR GSM SERVICES

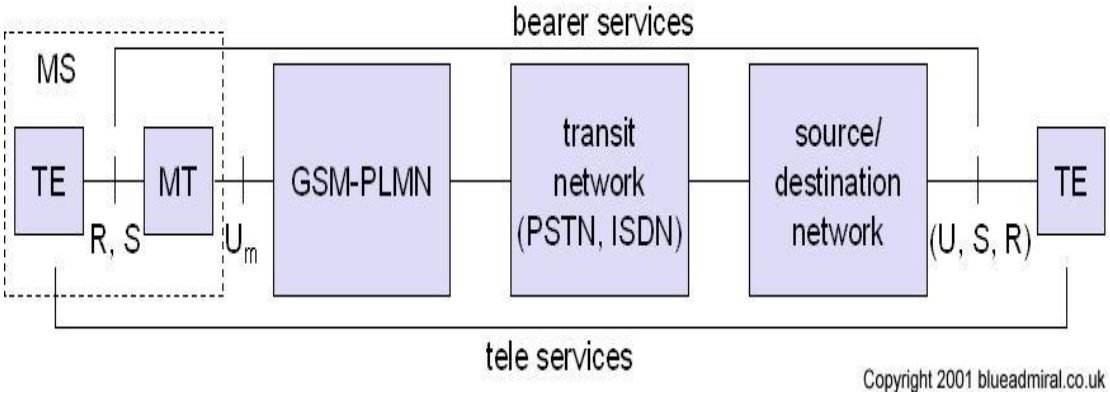


Fig 2.1 GSM services

EXPLANATION:

A mobile station MS is connected to the GSM public land mobile network PLMN via Um interface. PLMN is the infrastructure needed for the GSM networks. This network is connected to transit networks(eg.)PSTN,ISTN,etc. There will be additional network the source/ destinations network before another terminal TE is connected.

BEARER SERVICES:

Bearer Services comprises of all the services that enables the transparent transmission of data between the interface to the network. It permits transparent/non transparent , synchronous and asynchronous data transmission.

✓ **TRANSPARENT BEARER SERVICES:**

This services uses the functions of physical layer to transmit data. Data transmission has a constant delays and throughout if no error occurs but not in real time. FEC is used to increase the transmission quality. It does not try to recover the lost data in case of handover.

✓ **NON TRANSPARENT BEARER SERVICES**:

It uses the protocols of layers data link and network to transmit data. These services uses transparent bearer service radio link protocol (RLP). RLP has mechanisms of high level data link control HDLC. It allows retransmission of erroneous data by using elective reject mechanisms.

TELE SERVCIES:

Tele services are application specific and need all the 7 layers of ISO/OSI reference model. Services are specified end to end. There tele services are voice oriented tele service. They are encrypted voice transmission, message services and data communication with terminals from PSTN/ISDN.

There are some important services:

- Telephony services: It has high quality digital voice transmission.
- Emergency Number: Mandatory service for all service providers. Its of free of charge. This connection has the highest priority with pre-emption.
- Short Message Service: It is used fro simple message transfer with the maximum of 160 characters. SMS does not use the standard data channels of GSM uses the signaling channels. Sending andg receiving SMS is possible during the data/ voice transmission.
- Enhanced Message Service : It is used for large message size with 760 characters, animated pictures, small images can be transmitted.
- Multimedia Message Service : It is used to transmit large pictures of GIF/ JPEG , video clips.
- Group 3 fax : Fax data is transmitted as digital data over analog telephone network using modem.

SUPPLEMENTARY SERVICES:

✓ User Identification

- Call Redirecting / Forwarding
- Closed User Group
- Multiparty communication

GSM ARCHITECTURE

The architecture of GSM comes in hierarchy, consisting of many entities, interfaces and subsystems.

The GSM system consist of three subsystems namely,

- ✓ The Radio Subsystems(RSS)
- ✓ Network and Switching Subsystems(NSS)
- ✓ Operation Subsystem(OSS)

The customer is able to notice few components of the network viz. Mobile Station and Antenna of the Base Transceiver Station(BTS). Remaining entities are not visible.

1.RADIO SUBSYSTEM:

As the name implies, the radio subsystem (RSS) comprises all radio specific entities. i.e. the mobile stations(MS) and the base station subsystem(BSS).

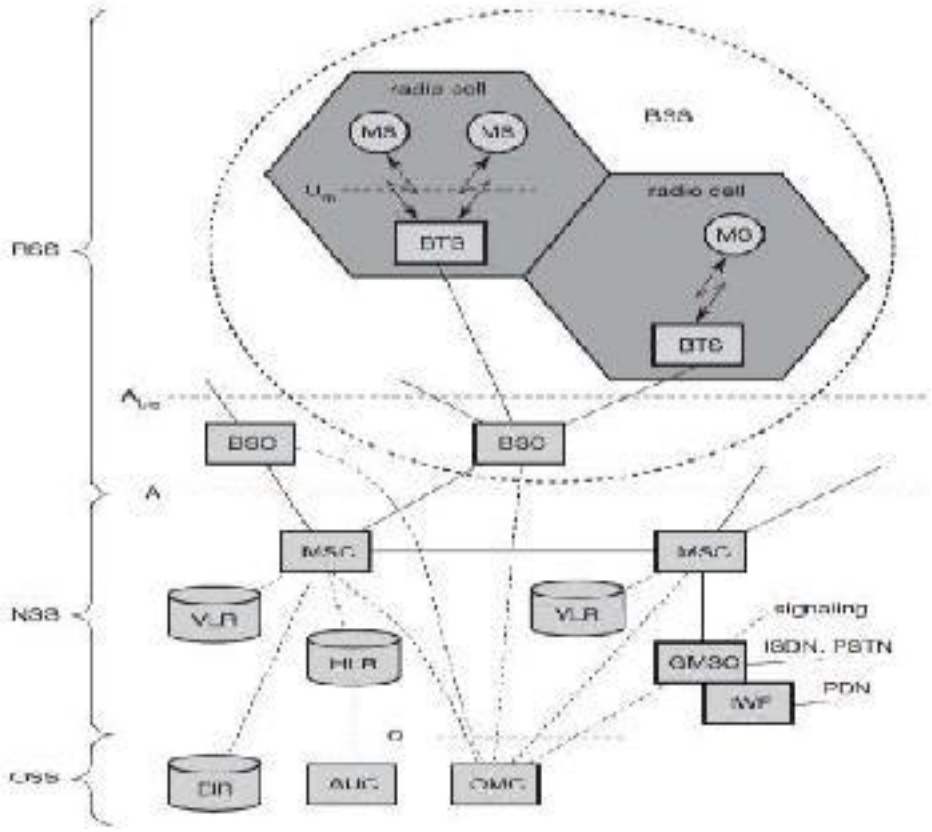


Fig 2.2 GSM Architecture

As they are in same frequency they form a cell . The components of RSS are

- ✓ Mobile station
- ✓ Base Transceiver Station
- ✓ Base Station Subsystem
- ✓ Base Station Controller

MOBILE STATION: (MS)

MS has all user equipment and software needed for mobile communications. It has user independent hardware and software. Subscriber Identity Module (SIM) stores all user specific data. Mobile Station can be identified as International Mobile equipment identity (IMEI). The sim card

Authentication key k, International Mobile subscriber identity (IMSI). It also has Identifiers and tables. The current location of MS is found using Temporary Mobile Subscriber Identity(TMSI) . With TMSI and Location Area Identification (LAI) the current location can be identified.

BASE TRNSCEIVER STATION: (BTS)

BTS contains the equipment for transmitting and receiving of radio signals,antennas and equipment for encrypting and decrypting communications with the Base station controller(BSC). A BTS is controlled by ap parent BSC via the base station control function(BSCF). The BCFis implemented as a discrete unit or even incorporated in a TRX in compact base stations. The BCF provides an operations and maintenance (O&M) connection to the Network Management System(NMS)and manage operational state of each TRX as well as soft handling and alarm collection.

The function of BTS vary depending on the cellular technology used and cellula telephone provider. There are vendors in which the BTS is a plain transceiver which receives information from MS through Um(Air Interface) and then it converts into TDM based interface, the Abis and it sends it towards the BSC. A GSM cell can measure between some 100m and 35km depending on the environment.

BASE STATION SUBSYSTEM: (BSS)

The base station subsystem is the section of traditional cellular telephone network which is responsible for handling traffic and signaling between a mobile phone and the network switching subsystem. The BSS carries out transcoding of speech channels, allocation of radio channels to mobile phones, paging, quality management of transmission and reception over the Air interface and many other tasks related to the radio network.

BASE STATION CONTROLLER : (BSC)

The BSC basically manages the BTSs. It reverses radio frequencies and handles the handover from one BTS to another within BSS, and performs paging of the MS. The BSC also multiplexes the radio channels onto the fixed network connections at A interface.

b) NETWORK AND SWITCHING SUBSYSTEM:

This network and switching subsystem is the heart of GSM. Their function are to connect wireless network with standard public network, performs

handover between different BSS, localization (to locate the mobile station),Charging, Accounting and roaming of users.The NSS contains the following switches and databases.

MOBILE SERVIES SWITCHING CENTER(MSC):

MSC are digital ISDN switches. It establishes connection qith other MSC and BSC via A interfacet Gateway MSC connects to fixed networks(eg.) PSTN, ISDN. With the help of Internet Working Functions, MSC can connect to public data Network PDN. It handles all signaling needed for connection setup, connection release and handover.

HOME LOCATION REGISTER : (HLR)

It is an important database. It stores user relevant information and also has static information and dynamic information.

Static Information:

Mobile subscriber ISDN number is available. It has subscribed services for a particular number. It is also an international mobile subscriber identity.

Dynamic Information :

It is a current location area(LA) of MS. It consist of Mobile subscriber roaming number (MSRN), VLR and MSC in it. When MS leaves the current LA, then the information is updated in HLR. The Usage of the information is to locate the user.

VISITOR LOCATION REGISTER: (VLR)

The VLR is associated to each MSC. It is a dynamic database. It stores all the information needed for the MS currently in LA. If new MS comes to LA then the VLR is responsible for copying the information needed from HLR**.**

c) OPERATION SUBSYSTEM :

The third part of a GSM system is operation subsystem (OSS) which contains the necessary functions for network operation and maintenance. The OSS possesses network entities of its own and accesses other entities via SS7 signaling. The following entities have been defined:

OPERATION AND MAINTENANCE CENTER (OMC):

The OMC monitors and controls all other network entities via the O interface (SS7 with X.25). Typical OMC management functions are traffic monitoring, status reports of network entities, subscriber and security management, or accounting and billing. OMCs use the concept of telecommunication management network (TMN) as standardized by the ITU-T. Authentication centre (AuC): As the radio interface and mobile stations are particularly vulnerable, a separate AuC has been defined to protect user identity and data transmission. The AuC contains the algorithms for authentication as well as the keys for encryption and generates the values needed for user authentication in the HLR.

EQUIPMENT IDENTITY REGISTER (EIR):

The EIR is a database for all IMEIs, i.e., it stores all device identifications registered for this network. As MSs are mobile, they can be easily stolen. With a valid SIM, anyone could use the stolen MS. The EIR has a blacklist of stolen (or locked) devices. In theory an MS is useless as soon as the owner has reported a theft. Unfortunately, the blacklists of different providers are not usually synchronized and the illegal use of a device in another operator‟s network is possible (the reader may speculate as to why this is the case). The EIR also contains a list of valid IMEIs (white list), and a list of malfunctioning devices (gray list).

1.3 GSM PROTOCOL SUITES :

The layers are

PHYSICAL LAYER:

The physical layer handles all radio specific functions.

Functions :

1. Creation of burst in any one of 5 format.

2. Multiplexing burst into a TDMA frame.
3. Synchronization with BTS.
4. Detection of idle channel.
5. Channel Quality measurement.
6. Channel coding and error detection and correction.

The Um interfaces use GMSK for modulation and perform encryption and decryption.

The protocol architecture of GSM is shown below:

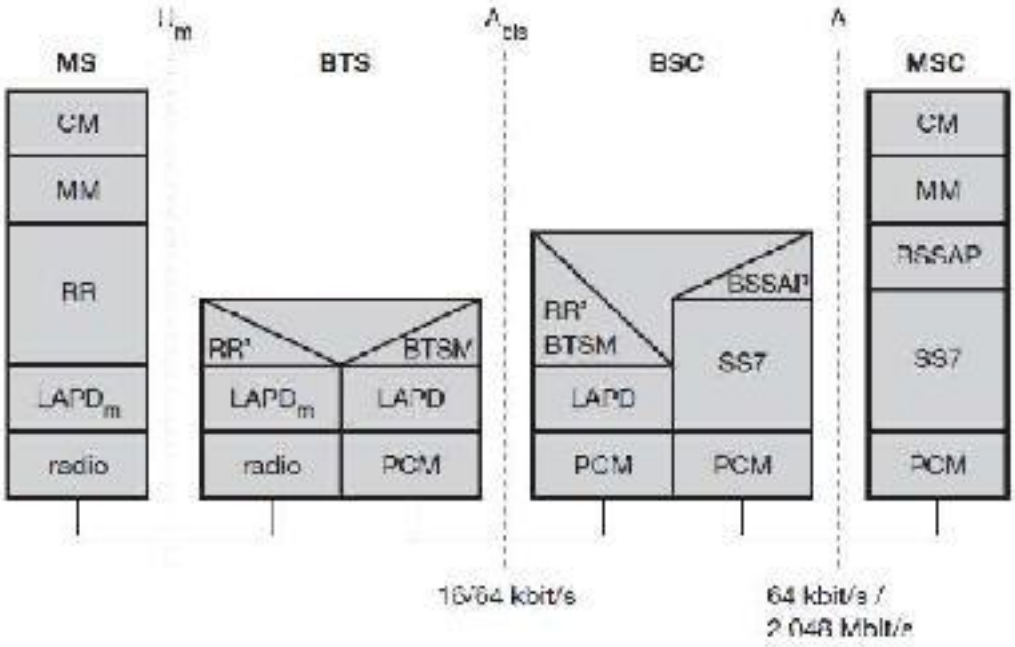


Fig :2.3 Protocol Architecture

LAYER 2 :

For signaling between entitites in GSM network this layer is used. The protocol used is LAPDM. LAPD stands for link access procedure for D channel. LAPDM has no buffers has to follow Um interface patterns. The functions of the layer are namely:

1. Reliable data transfer
2. Reseqeuncing of data

3. Flow control

LAYER 3 : NETWORK LAYER

The network layer has sublayers. They are,

1. **RADIO RESOURCE MANAGEMENT:**

This is the lowest sub layer and it's a part of RR and RR' is implemented by BSC. The function of RR are Setup, Maintenance, Release of radio channels. RR directly access the physical layer. It supports BTS management. The function of RR' are supported by BSC via BSTM.

2. **MOBILITY MANAGEMENT:**

The main function of Mobility management are Registration, Authentication, Identification, Location Updating, Providing TMSI, IMSI.

LAYER 4 : CALL MANAGEMENT:

This layer contains three entities. They are Call control, SMS, Supplementary services. Call control provides point to point connection between two terminals and also used for call clearance, change of call parameters. SMS allows messages transfer using control channels. The supplementary services discussed already is to be reproduced here.

1.4 CONNECTION ESTABLISHMENT:

The number dialed to reach a mobile subscriber (MSISDN) contains no information at all about the current location of the subscriber. In order to establish a complete connection to a mobile subscriber, however, one must determine the current location and the locally responsible switch (MSC). In order to be able to route the call to this switch, the routing address to this subscriber (MSRN) has to be obtained. This routing address is assigned temporarily to a subscriber by its currently associated VLR. At the arrival of a call at the GMSC,

the HLR is the only entity in the GSM network which can supply this information, and therefore it must be interrogated for each connection setup to a mobile subscriber. An ISDN switch recognizes from the MSISDN that the called subscriber is a mobile subscriber, and therefore can forward the call to the GMSC of the subscriber's home PLMN based on the CC and NDC in the MSISDN. This GMSC can now request the current routing address (MSRN) for the mobile subscriber from the HLR using the MAP . By way of the MSRN the call is forwarded to the local MSC, which determines the TMSI of the subscriber and initiates the paging procedure in the relevant location area . After the MS has responded to the paging call, the connection can be switched through.Several variants for determining the route and interrogating the HLR exist, depending on how the MSRN was assigned and stored, whether the call is national or international and depending on the capabilities of the associated switching centers.

To locate a MS and to address the MS, several numbers are needed:

MOBILE STATION INTERNATIONAL ISDN NUMBER (MSISDN):

The only important number from the user of GSM is phone number. Remember that the phone number is not associated with certain device but with the SIM, which is personalized for user. The number consist of country code(CC) as eg.+49 179 1234567 with 49 for germany. National Destination Code (NDC) is used to locate the network provider and Subscriber number.

INTERNATIONAL MOBILE SUBSCRIBER IDENTITY(IMSI):

GSM uses the IMSI for internal unique identification of a subscriber. IMSI consist of a mobile country code (MCC) and mobile network code(MNC) and finally the mobile subscriber identification number(MSIN).

TEMPORARY MOBILE SUBSCRIBE IDENTITY(TMSI):

To hide the IMSI, which would give away the exact identity of user signaling over the air interface, GSM uses the 4 byte.TMSI is selected by current VLR and is only valid temporarily and within location area of VLR.

MOBILE STATION ROAMING NUMBER (MSRM):

Another temporary address which hides the identity and location of a subscriber is MSRN. The VLR generates this address on request from MSC, and the address is stored in the HLR . MSRN contains current visitory and visitor national destination code(VNDC).

MESSAGE TERMINATED CALL (MTC):

This figure shows the basic steps needed to connect the calling station with the mobile user.

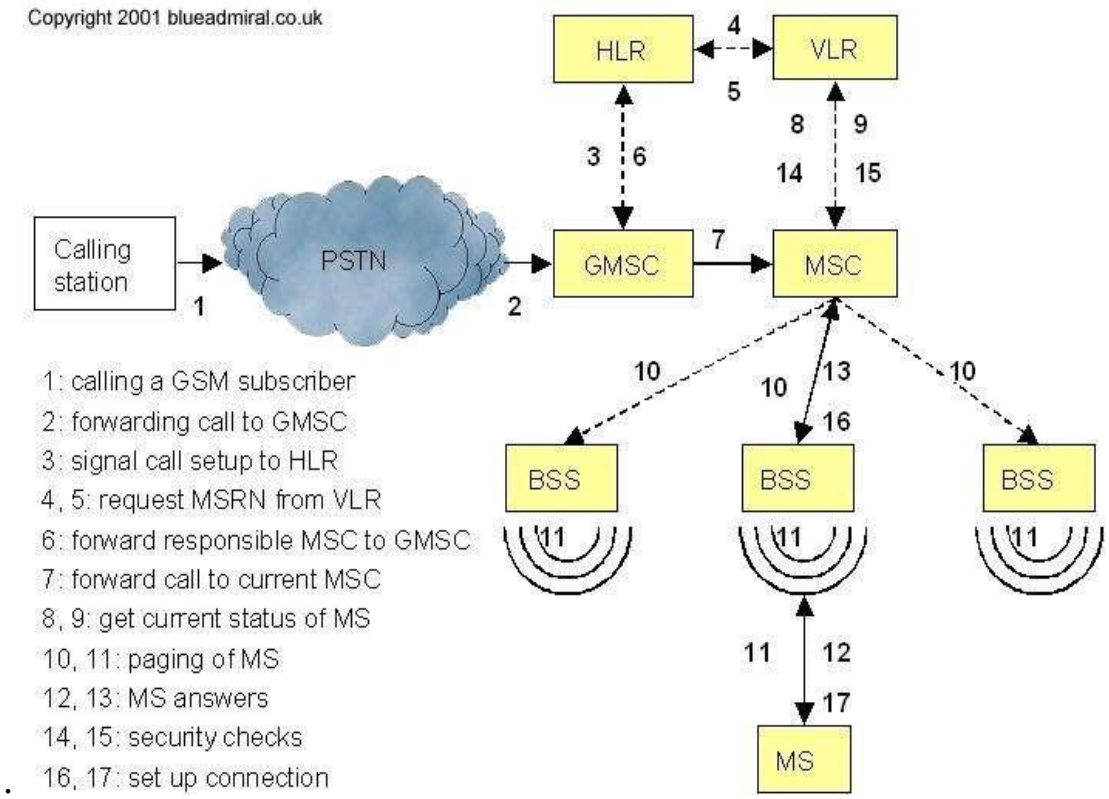


Fig :2.4 Message Terminated Call (MTC)

Step 1: A user dials the phone number of GSM subscriber. The fixed network PSTN notices that the number belongs to the user in the GSM network and forwards the call setup to the Gateway MSC .

Step 2: The GMSC identifies the HLR for the subscriber and signals the call setup to the HLR.

Step 3: The HLR now checks whether the number exists and whether the user has subscribed to the requested services, and the requests an MSRN from the current VLR.

Step 4 : After receiving the request from MSRN

Step 5 : HLR can determine the MSC responsible for the MS and forwards this information to GMSC

Step 6 : GMSC can now forward the call setup request to MSC indicated.

Step 7: From this, MSC is responsible for all further steps. First it requests the current status of MS from VLR.

Step 8 : If MS is available, then MSC initiates paging in all cells it is responsible for as searching for the right cell would be too time consuming.

Step 9: This approach puts some load on signalling channels so optimization exist.

Step 10 : Location area (LA) can be determined.

Step 11 : The BTSs of all BSSs transmit this paging signal to MS.

Step 12 & 13 : If MS answers (12 and 13) the VLR has to perform security checks set up be encryption techniques.

Steps 14 to 17 : The VLR signals to MSC to setup a connection to MS.

MESSAGE ORIGINATED CALL (MOC):

It is simpler to perform message originated call(MOC) compared to MTC.

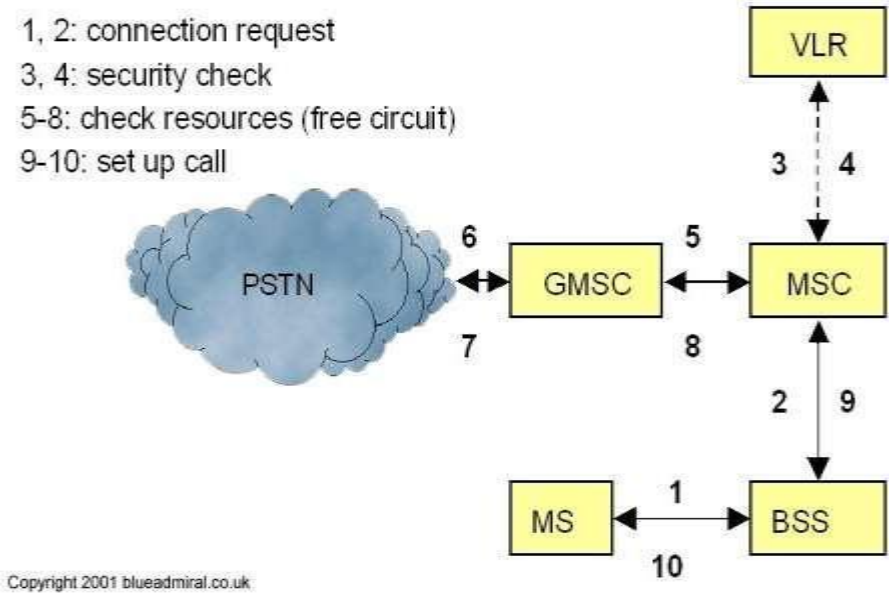


Fig 2.5 Message Originated Call (MOC)

The basic steps for MOC are,

Step 1: MS transmits a request for a new connection.

Step 2: BSS forwards the request to MSC.

Steps 3 & 4 : MSC then checks if the user is allowed to set up a call with the requested service(3 and 4) and checks the availability of resources through GSM network into PSTN.

Steps 5to 8 : If all resources are available, MSC sets up a connection between MS and fixed network.

Steps 9 & 10 : Its set up a call with the help of BSS and MS.

MESSAGE FLOW FOR MTC AND MOC :

In addition to the above steps mentioned above , the other messages are exxchanged between an MS and BTS during connection setup. These messages can be quite often heard in radios or badly sheileded speakers as crackling noise before the phone rings. Figure shows the message for an MTC and MOC. Paging is only necessary for an MTC, then similar message exchanges follow. The next step which are needed for a communication security comprises the authentication of MS and switching to encrytpted communication

The following steps which are mentioned in the figure denotes th euse of MSC and MOC. If someone is calling the MS, it answers now with ‘alerting’ that MS is ringing and with ‘connect’ that the user has pressed the connect button. The same actions happen the other way round if MS has initiated the call. Af.ter connection acknowledgement both parties are exchanged.

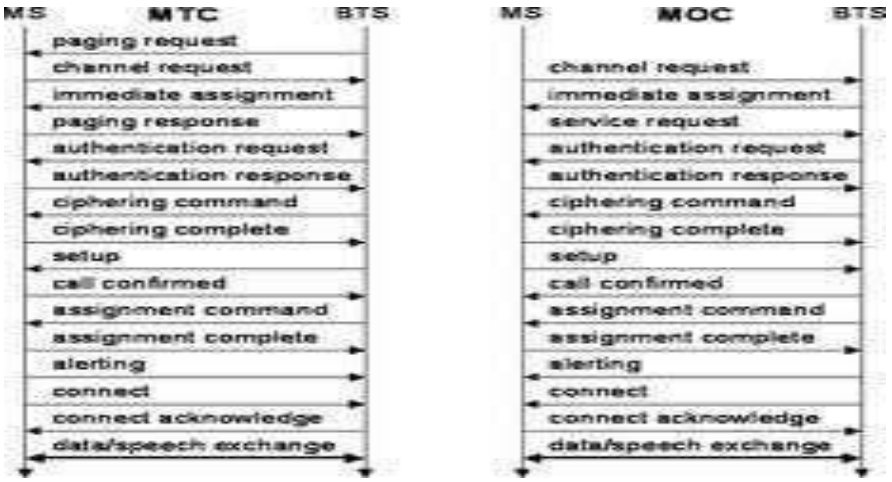


Fig 2.6 Message flow for MTC and MOC

1.5 FREQUENCY ALLOCATION:

Radio transmission can take place using many different frequency bands. Each frequency band exhibits certain advantages and disadvantages. Figure gives a rough overview of the frequency spectrum that can be used for data transmission. The figure shows frequencies starting at 300 Hz and going up to over 300 THz. Directly coupled to the frequency is the wavelength λ via the equation: $\lambda = c/f$, where $c \cong 3 \cdot 108$ m/s (the speed of light in vacuum) and f the frequency. For traditional wired networks, frequencies of up to several hundred kHz are used for distances up to some km with twisted pair copper wires, while frequencies of several hundred MHz are used with coaxial cable (new coding schemes work with several hundred MHz even with twisted pair copper wires over distances of some 100 m).

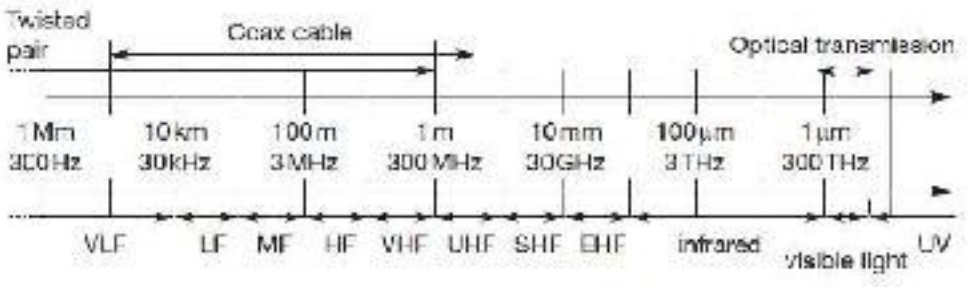


Fig 2.7 Frequency allocation

Fiber optics are used for frequency ranges of several hundred THz, but here one typically refers to the wavelength which is, e.g., 1500 nm, 1350 nm etc. (infra red). Radio transmission starts at several kHz, the **very low frequency (VLF)** range. These are very long waves. Waves in the **low frequency (LF)** range are used by submarines, because they can penetrate water and can follow the earth‟s surface. Some radio stations still use these frequencies, e.g., between 148.5 kHz and 283.5 kHz in Germany. The **medium frequency (MF)** and **high frequency (HF)** ranges are typical for transmission of hundreds of radio stations either as amplitude modulation **(AM)** between 520 kHz and 1605.5 kHz, as short wave **(SW)** between 5.9 MHz and 26.1 MHz, or as frequency modulation **(FM)** between 87.5 MHz and 108 MHz. The frequencies limiting these ranges are typically fixed by national regulation and, vary from country to country. Short waves are typically used for (amateur) radio transmission around the world, enabled by reflection at the ionosphere

. Transmit power is up to 500 kW – which is quite high compared to the 1 W of a mobile phone. As we move to higher frequencies, the TV stations follow. Conventional analog TV is transmitted in ranges of 174–230 MHz and 470–790 MHz using the very high frequency **(VHF)** and ultra high frequency **(UHF)** bands. In this range, digital audio broadcasting (DAB) takes place as well (223–230 MHz and 1452–1472 MHz) and digital TV is planned or currently being installed (470– 862 MHz), reusing some of the old frequencies for analog TV. UHF is also used for mobile phones with analog technology (450–465 MHz), the digital GSM (890–960 MHz, 1710–1880 MHz), digital cordless telephones following the DECT standard (1880–1900 MHz), 3G cellular systems following the UMTS standard (1900–1980 MHz, 2020–2025 MHz, 2110–2190 MHz) and many more. VHF and especially

UHF allow for small antennas and relatively reliable connections for mobile telephony. **Super high frequencies (SHF)** are typically used for directed

microwave links (approx. 2–40 GHz) and fixed satellite services in the C-band (4 and 6 GHz), Ku-band (11 and 14 GHz), or Ka-band (19 and 29 GHz). Some systems are planned in the **extremely high frequency (EHF)** range which comes close to infra red. All radio frequencies are regulated to avoid interference, e.g., the German regulation covers 9 kHz–275 GHz. The next step into higher frequencies involves optical transmission, which is not only used for fiber optical links but also for wireless communications. **Infra red (IR)** transmission is used for directed links, e.g., to connect different buildings via laser links. The most widespread IR technology, infra red data association (IrDA), uses wavelengths of approximately 850–900 nm to connect laptops, PDAs etc. Finally, visible light has been used for wireless transmission for thousands of years. While light is not very reliable due to interference, but it is nevertheless useful due to built-in human receivers. Powered by

1.6 ROUTING :

A satellite system together with gateways and fixed terrestrial networks as shown in Figure 5.1 has to route data transmissions from one user to another as any other network does. Routing in the fixed segment (on earth) is achieved as usual, while two different solutions exist for the satellite network in space. If satellites offer ISLs, traffic can be routed between the satellites. If not, all traffic is relayed to earth, routed there, and relayed back to a satellite. Assume two users of a satellite network exchange data. If the satellite system supports ISLs, one user sends data up to a satellite and the satellite forwards it to the one responsible for the receiver via other satellites. This last satellite now sends the data down to the earth. This means that only one uplink and one downlink per direction is needed. The ability of routing within the satellite network reduces the number of gateways needed on earth. If a satellite system does not offer ISLs, the user also sends data up to a satellite, but now this satellite forwards the data to a gateway on earth. Routing takes place in fixed networks as usual until another gateway is reached which is responsible for the satellite above the receiver. Again data is sent up to the satellite which forwards it down to the receiver. This solution requires two uplinks and two downlinks. Depending on the orbit and the speed of routing in the satellite network compared to the terrestrial network, the solution with ISLs might offer lower latency. The drawbacks of ISLs are higher system complexity due to additional antennas and routing hard- and software for the satellites.

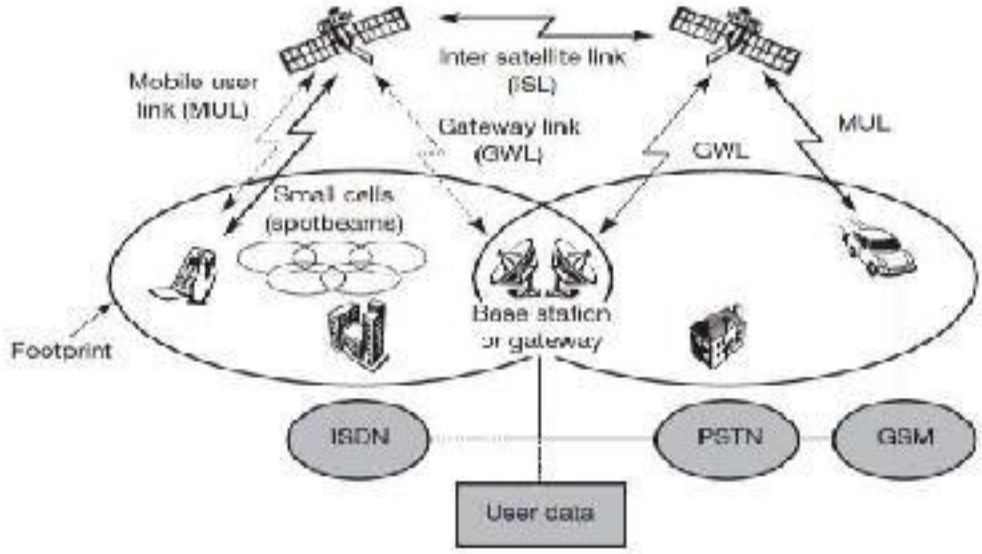


Fig 2.8 Routing

1.7 MOBILITY MANAGEMENT:

Mobility management is one of the major functions of a GSM or a UMTS network that allows mobile phones to work. The aim of mobility management is to track where the subscribers are, allowing calls, SMS and other mobile phone services to be delivered to them. Location update procedure.

A GSM or UMTS network, like all cellular networks, is a radio network of individual cells, known as base stations. Each base station covers a small geographical area which is part of a uniquely identified location area. By integrating the coverage of each of these base stations, a cellular network provides a radio coverage over a much wider area. A group of base stations is named a location area, or a routing area.

The location update procedure allows a mobile device to inform the cellular network, whenever it moves from one location area to the next. Mobiles are responsible for detecting location area codes. When a mobile finds that the location area code is different from its last update, it performs another update by sending to the network, a location update request, together with its previous location, and its Temporary Mobile Subscriber Identity (**TMSI**).

There are several reasons why a mobile may provide updated location information to the network. Whenever a mobile is switched on or off, the network may require it to perform an IMSI attach or IMSI detach location update procedure. Also, each mobile is required to regularly report its location at a set time interval using a **periodic location update** procedure. Whenever a mobile

moves from one location area to the next while not on a call, a **random location update** is required. This is also required of a stationary mobile that reselects coverage from a cell in a different location area, because of signal fade. Thus a subscriber has reliable access to the network and may be reached with a call, while enjoying the freedom of mobility within the whole coverage area.

When a subscriber is paged in an attempt to deliver a call or SMS and the subscriber does not reply to that page then the subscriber is marked as absent in both the Mobile Switching Center / Visitor Location Register (MSC/VLR) and the Home Location Register (HLR) (Mobile not reachable flag MNRF is set). The next time the mobile performs a location update the HLR is updated and the mobile not reachable flag is cleared.

TMSI

The Temporary Mobile Subscriber Identity (TMSI) is the identity that is most commonly sent between the mobile and the network. TMSI is randomly assigned by the VLR to every mobile in the area, the moment it is switched on. The number is local to a location area, and so it has to be updated each time the mobile moves to a new geographical area.

The network can also change the TMSI of the mobile at any time. And it normally does so, in order to avoid the subscriber from being identified, and tracked by eavesdroppers on the radio interface. This makes it difficult to trace which mobile is which, except briefly, when the mobile is just switched on, or when the data in the mobile becomes invalid for one reason or another. At that point, the global "international mobile subscriber identity" (IMSI) must be sent to the network. The IMSI is sent as rarely as possible, to avoid it being identified and tracked.

A key use of the TMSI is in paging a mobile. "Paging" is the one-to-one communication between the mobile and the base station. The most important use of broadcast information is to set up channels for "paging". Every cellular system has a broadcast mechanism to distribute such information to a plurality of mobiles. Size of TMSI is 4 octet with full hex digits and can't be all 1 because the SIM uses 4 octets with all bits equal to 1 to indicate that no valid TMSI is available.

ROAMING

Roaming is one of the fundamental mobility management procedures of all cellular networks. Roaming is defined as the ability for a cellular customer to automatically make and receive voice calls, send and receive data, or access other services, including home data services, when travelling outside the geographical coverage area of the home network, by means of using a visited

network. This can be done by using a communication terminal or else just by using the subscriber identity in the visited network. Roaming is technically supported by mobility management, authentication, authorization and billing procedures.

LOCATION AREA

A "location area" is a set of base stations that are grouped together to optimise signalling. Typically, tens or even hundreds of base stations share a single Base Station Controller (BSC) in GSM, or a Radio Network Controller (RNC) in UMTS, the intelligence behind the base stations. The BSC handles allocation of radio channels, receives measurements from the mobile phones, controls handovers from base station to base station.

To each location area, a unique number called a "location area code" is assigned. The location area code is broadcast by each base station, known as a "base transceiver station" BTS in GSM, or a Node B in UMTS, at regular intervals. In GSM, the mobiles cannot communicate directly with each other but, have to be channeled through the BTSs. In UMTS networks, if no Node B is accessible to a mobile, it will not be able to make any connections at all.

If the location areas are very large, there will be many mobiles operating simultaneously, resulting in very high paging traffic, as every paging request has to be broadcast to every base station in the location area. This wastes bandwidth and power on the mobile, by requiring it to listen for broadcast messages too much of the time. If on the other hand, there are too many small location areas, the mobile must contact the network very often for changes of location, which will also drain the mobile's battery. A balance has therefore to be struck.

ROUTING AREA

The routing area is the PS domain equivalent of the location area. A "routing area" is normally a subdivision of a "location area". Routing areas are used by mobiles which are GPRS-attached. GPRS is optimized for "bursty" data communication services, such as wireless internet/intranet, and multimedia services. It is also known as GSM-IP ("Internet Protocol") because it will connect users directly to Internet Service Providers (ISP).

The bursty nature of packet traffic means that more paging messages are expected per mobile, and so it is worth knowing the location of the mobile more accurately than it would be with traditional circuit-switched traffic. A change from routing area to routing area (called a "Routing Area Update") is done in an almost identical way to a change from location area to location area. The main

differences are that the "Serving GPRS Support Node" (SGSN) is the element involved.

TRACKING AREA

The tracking area is the LTE counterpart of the location area and routing area. A tracking area is a set of cells. Tracking areas can be grouped into lists of tracking areas (TA lists), which can be configured on the User equipment. Tracking area updates are performed periodically or when the UE moves to a tracking area that is not included in its TA list.

Operators can allocate different TA lists to different UEs. This can avoid signaling peaks in some conditions: for instance, the UEs of passengers of a train may not perform tracking area updates simultaneously. On the network side, the involved element is the Mobility Management Entity.

HANDOVER

Handover means handing over the mobile from one cell to another cell. There are two reasons for handover.

They are,

(1)

- ✓ When a mobile station moves out of the range of BTS the signal level decreases continuously and falls below the minimal requirements for communication.
- The error rate increases due to interference.
- The quality of radio link decrease.

(2)

- The traffic in one cell is too high, shifting of some MS to other cells with lower load. This is called load balancing.
- The number of handover will be more when the cell size is small.
- Due to handover the calls should not get to cutoff which is called as call drop.

TYPES OF HANDOVER :

(1) INTRA CELL HANDOVER:

✓ With in a cell, narrow band interference can cause transmission at a certain frequency impossible.

✓ The BSC decides to change the carrier frequency.

(2) INTER CELL, INTRA BSC HANDOVER:

✓ The mobile station moves from one cell to another but remains with in the same BSC.

✓ The BSC performs a handover, assigns a new radio channel in the new cell **and** releases the old one.

(3) **INTER BSC, INTRA MSC HANDOVER:**

- The BSC controls only limited cells.
- Handover needs to be done between different BSC.
- This is controlled by MSC.

☎🗐☽ **INTER MSC HANDOVER:**

✓ A handover is needed between 2 cells which belong to difference MSC.

- Both MSC performs the handover together.

1.8 SECURITY :

GSM offers several security services using confidential information stored in the AuC and in the individual SIM (which is plugged into an arbitrary MS). The SIM stores personal, secret data and is protected with a PIN against unauthorized use. (For example, the secret key Ki used for authentication and encryption procedures is stored in the SIM.) The security services offered by GSM are explained below:

● **ACCESS CONTROL AND AUTHENTICATION:** The first step includes the authentication of a valid user for the SIM. The user needs a secret PIN to access the SIM. The next step is the subscriber authentication .

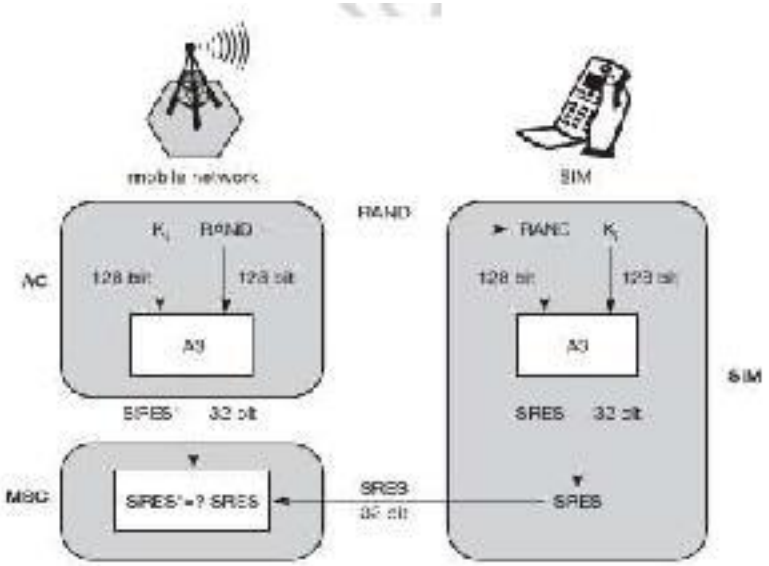
● **CONFIDENTIALITY:** All user-related data is encrypted. After authentication, BTS and MS apply encryption to voice, data, and signaling. This confidentiality exists only between MS and BTS, but it does not exist end-to-end or within the whole fixed GSM/telephone network.

● **ANONYMITY:** To provide user anonymity, all data is encrypted before transmission, and user identifiers (which would reveal an identity) are not used over the air. Instead, GSM transmits a temporary identifier (TMSI), which is newly assigned by the VLR after each location update. Additionally, the VLR can change the TMSI at any time. Three algorithms have been specified to provide security services in GSM. Algorithm A3 is used for authentication, A5 for encryption, and A8 for the generation of a cipher key. In the GSM standard only algorithm A5 was publicly available, whereas A3 and A8 were secret, but standardized with open interfaces. Both A3 and A8 are no longer secret, but were published on the internet in 1998. This demonstrates that security by obscurity does not really work. As it turned out, the algorithms are not very strong. However, network providers can use stronger algorithms for authentication – or users can apply stronger end-to-end encryption. Algorithms A3 and A8 (or their replacements) are located on the SIM and in the AuC and can be proprietary. Only A5 which is implemented in the devices has to be identical for all providers.

AUTHENTICATION

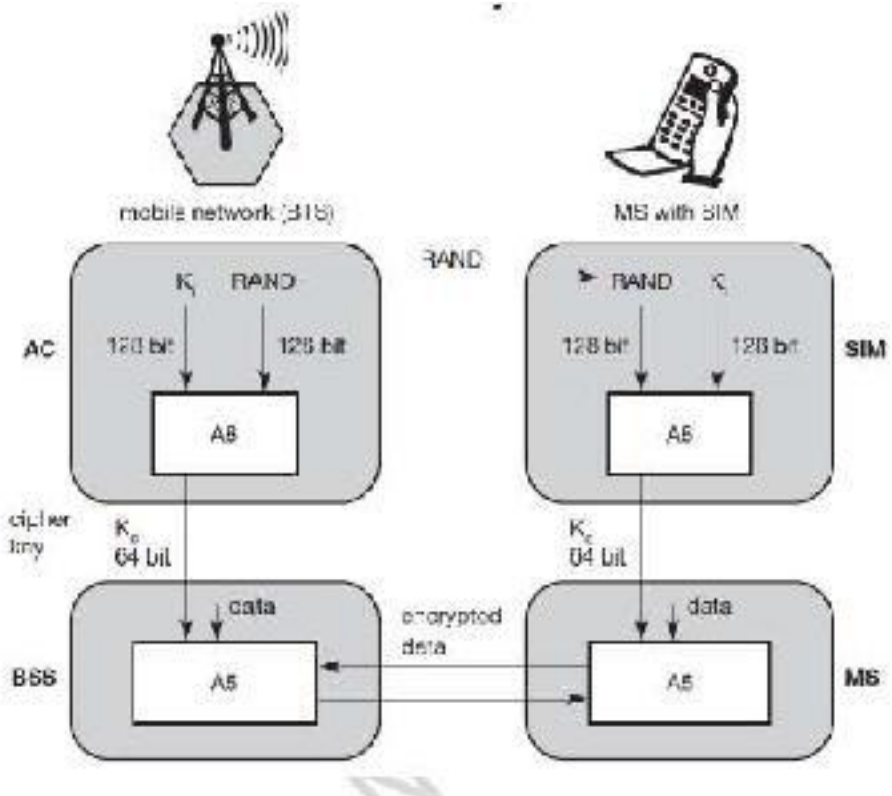
Before a subscriber can use any service from the GSM network, he or she must be authenticated. Authentication is based on the SIM, which stores the individual authentication key Ki, the user identification IMSI, and the algorithm used for authentication A3. Authentication uses a challenge-response method: the access control AC generates a random number RAND as challenge, and the SIM within the MS answers with SRES (signed response) as response. The AuC performs the basic generation of random values RAND, signed responses SRES, and cipher keys Kc for each IMSI, and then forwards this information to the HLR. The current VLR requests the appropriate values for RAND, SRES, and Kc from the HLR.

For authentication, For authentication, the VLR sends the random value RAND to the SIM. Both sides, network and subscriber module, perform the same operation with RAND and the key Ki, called A3. The MS sends back the SRES generated by the SIM; the VLR can now compare both values. If they are the same, the VLR accepts the subscriber, otherwise the subscriber is rejected.



ENCRYPTION :

To ensure privacy, all messages containing user-related information are encrypted in GSM over the air interface. After authentication, MS and BSS can start using encryption by applying the cipher key Kc (the precise location of security functions for encryption, BTS and/or BSC are vendor dependent). Kc is generated using the individual key Ki and a random value by applying the algorithm A8. Note that the SIM in the MS and the network both calculate the same Kc based on the random value RAND. The key Kc itself is not transmitted over the air interface.



MS and BTS can now encrypt and decrypt data using the algorithm A5 and the cipher key Kc. As Figure shows, Kc should be a 64 bit key which is not very strong, but is at least a good protection against simple eavesdropping. However, the publication of A3 and A8 on the internet showed that in certain implementations 10 of the 64 bits are always set to 0, so that the real length of the key is thus only 54 consequently, the encryption is much weaker.

1.9 GPRS

INTRODUCTION:

General packet radio service (GPRS) is a packet oriented mobile data service available to users of the 2G cellular communication systems, 3G systems and GSM. GPRS re-use the existing GSM infrastructure. It interworked with existing circuit-switched services. It is based on standardized open interfaces.

GPRS usage is typically charged based on volume of data transferred, contrasting with circuit switched data, which is usually billed per minute of connection time. 5 GB per month for a fixed fee or on a pay-as-you-use basis. Usage above the bundle cap is either charged per megabyte or disallowed.

GPRS is a best effort service, implying variable throughput and latency that depend on the number of other users sharing the service concurrently, as opposed to circuit switching, where a certain quality of service (QoS) is guaranteed during the connection. In 2G systems, GPRS provides data rates of 56–114 kbit/second. 2G cellular technology combined with GPRS is sometimes described as *2.5G*, that is, a technology between the second (2G) and third (3G) generations of mobile telephony. It provides moderate-speed data transfer, by using unused time division multiple access (TDMA) channels in, for example, the GSM system. GPRS is integrated into GSM Release 97 and newer releases.

GPRS provides two services:

1. Point-to-point **(PTP)**

2. Point-to-multipoint **(PTM)**

In point-to-point packet delivery service, in which packet is transfer between two users and in point-to-multipoint **(PTM)** service, in which packet is delivering to multiple destinations within one service request.

In PTP versions are PTP Connection oriented Network service **(PTP-CONS)**, which establish a logical relation in between users. Multiple packets are sent between single source and a single destination. Other version is the PTP Connectionless Network Service **(PTP-CLNS)**, which does not require a logical link between users. Packets are sent between a single source and a single destination. Each packet is independent of its predecessor and successor.

QoS-profile can be specified by the users of the GPRS. It is maintained in the PDP context. **PDP Context** is nothing but which is created in each communication session. QoS-profile is used to indicate the network and radio resources required for data transmission. It has the attributes such as service precedence (high,normal,low),reliability class, delay class, peak throughput class, mean throughput class. GPRS must allocate radio resources to fulfill these user specifications. GPRS network is suffered by the following delays such as channel access delay, coding for error correction and transfer delay in the fixed part and wireless part of the network. GPRS also includes several security services namely authentication, user identity confidentiality, access control and user information confidentiality.

Main benefits

Resources are reserved only when needed and charged accordingly. Connection setup times are reduced. It will enable new service opportunities. It has High Speed (Data Rate 14.4 – 115 kbps). It uses the efficient radio bandwidth (Statistical Multiplexing).Circuit switching & Packet Switching can be used in parallel. It has Constant connectivity.

Characteristics of GPRS:

1. GPRS uses packet switched resource allocation.
2. Flexible channel allocation.
3. Support for leading internet communication protocols.

GPRS Terminal Classes:

1. **Class A**

It can be connected to GPRS service and GSM service (voice, SMS), using both at the same time. Such devices are known to be available today.

2. **Class B**

It can be connected to GPRS service and GSM service (voice, SMS), but using only one or the other at a given time. During GSM service (voice call or SMS), GPRS service is suspended, and then resumed automatically after the GSM

service (voice call or SMS) has concluded. Most GPRS mobile devices are Class B.

3. **Class C**

They are connected to either GPRS service or GSM service (voice, SMS). Must be switched manually between one or the other service.

GPRS ARCHITECTURE

In order to understand the GPRS network architecture, some fundamental GSM terminology is necessary. This section describes some of the main components of the GSM network.

GPRS Networks

GPRS architecture has two network elements, which are called as **GPRS support nodes (GSN)**.They are,

1. Gateway GPRS Support Node(GGSN)

2. Serving GPRS Support Node (SGSN)

All GSNs are integrated into the standard GSM architecture and many interfaces (see figure 1). The network elements are **gateway GPRS support node (GGSN)** is provisioned by router, which supports traditional gateway functionality. It is the interworking unit between the GPRS network and external **packet data networks (PDN).** This node contains routing information for GPRS users. It performs address conversion and tunnels data to a user via encapsulation.

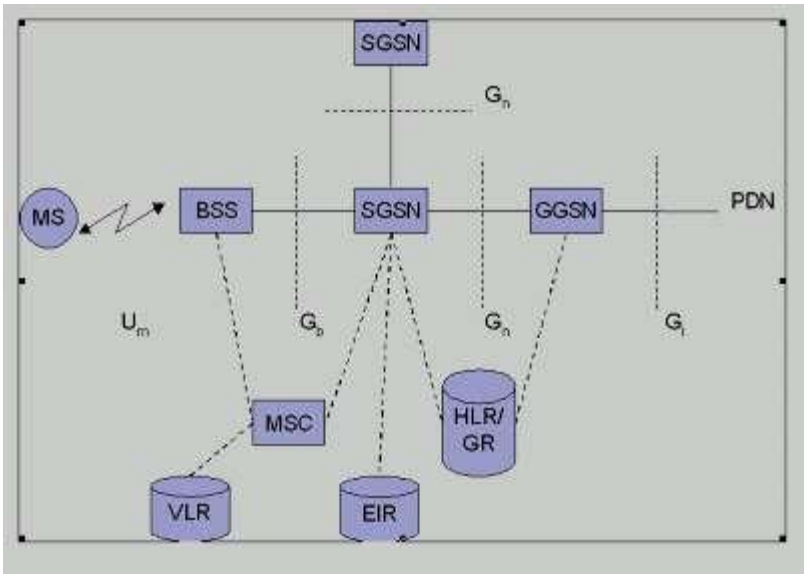


Fig 3.1 GPRS Architecture Reference Model

The other element is the **serving GPRS support node (SGSN)** which connects BSS and GGSN. It supports the MS via the G_b interface. It requests the user address from the **GPRS register (GR).** It keeps track of the individual MSs' location, is in charge for collecting billing information. It performs many security functions.

Packet Control Unit (PCU)

The PCU separates the circuit switched and packet switched traffic from the user and sends them to the GSM and GPRS networks respectively. It also performs most of the radio resource management functions of the GPRS network. The PCU can be either located in the BTS, BSC, or some other point between the MS and the MSC. There will be at least one PCU that serves a cell in which GPRS services will be available. Frame Relay technology is being used at present to interconnect the PCU to the GPRS core.

GPRS interfaces

Um between an MS and the GPRS fixed network part. The Um is the access interface the MS uses to access the GPRS network. The radio interface to the BTS is the same interface used by the existing GSM network with some GPRS specific changes.

Gb between a SGSN and a BSS. The Gb interface carries the GPRS traffic and signaling between the GSM radio network (BSS) and the GPRS network. Frame Relay based network services is used for this interface.

Gn between two GSNs within the same PLMN. The Gn provides a data and signalling interface in the Intra-PLMN backbone. The GPRS Tunnelling Protocol (GTP) is used in the Gn (and in the Gp) interface over the IP based backbone network.

Mobile Station

A GSM subscriber needs a terminal called **Mobile Station (MS).** It is used to connect to the network using the radio interface $U_{m.}$ In idle mode an MS is not reachable and all contexts will be deleted. In the standby state there is only movement across routing areas which is updated to the SGSN. Before sending any data over the GPRS network, an MS must attach to it, following the procedures of the **mobility management**. The attachment procedure includes assigning a temporal identifier, called a **temporary logical link identity (TLLI)**, and a **ciphering key sequence number (CKSN)** for data encryption.

GPRS BSS

Base Station Subsystem (BSS) which performs radio-related functions. BSS contains Base Transceiver Stations (BTS) and Base Station Controllers (BSC).

BTS which provides new GPRS channel coding schemes through Channel Codec Unit (CCU). The BTS handles the radio interface to the MS. It consists of radio equipment (transceivers and antennas) required to service each cell in the

network.

BSC forwards the Circuit-switched calls to MSC and the Packet-switched data to SGSN. The BSC provides the control functions and physical links between the MSC and the BTS. A number of BSCs are served by one MSC while several BTSs can be controlled by one BSC.

The Network Switching Subsystem

The NSS is responsible for call control, service control and subscriber mobility manage

a) Mobile Switching center (MSC)

MSC is in charge for telephony switching functions of the network. It also performs authentication to verify the user's identity. It ensures the confidentiality of the calls. The Authentication Center (AuC) provides the necessary parameters to the MSC to perform the authentication procedure. The AuC is shown as a separate logical entity but is generally integrated with the HLR. The Equipment Identity Register (EIR) is on the other hand a database that contains information about the identity of the mobile equipment. It prevents calls from unauthorized or stolen MSs.

b)Home Location register (HLR)

HLR is a database used to store and manage permanent data of subscribers. HLR is used to map an MS to one or more GGSNs. It is used to update the SGSN of the MS. It is also used to store the fixed IP address and QoS profile for a transmission path.

c) Visitor location register (VLR)

VLR is a database used to store temporary information about the subscribers. It is needed by the MSC in order to service visiting subscribers. The MSC and VLR are commonly integrated into one single physical node and the term MSC/VLR is used instead. When a subscriber enters a new MSC area, a copy of all the ne) cessary information is downloaded from the HLR into the VLR. The VLR keeps this information so that calls of the subscriber can be processed without having to interrogate the HLR (which can be in another PLMN) each time. The temporary information is cleared when the mobile station roams out of the service area.

d)Equipment identity register (EIR)

EIR is also a database that encloses information about the identity of the mobile equipment. It prevents calls from unauthorized or stolen MSs.

GPRS Mobility Management States

a)Idle State

A MS in the idle state is not traceable and can only receive PTM-M transmissions such as general broadcast events destined to a specific geographical area. The MS needs to perform the attach procedure in order to connect to the GPRS network and become reachable.

b)Ready State

Data is sent or received in this state. The MS informs the SGSN when it changes cells. The MS may explicitly request (or can be forced by the network) to detach in which case it moves to Idle. A timer monitors the Ready state and upon its expiry, the MS is put on Standby. The timer insures that resources are not wasted by an inactive MS.

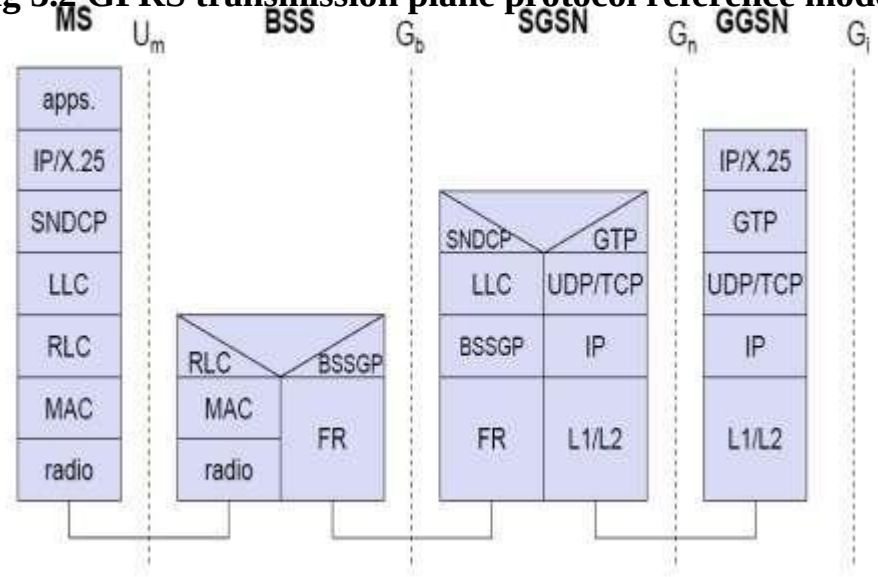
c)Standby State

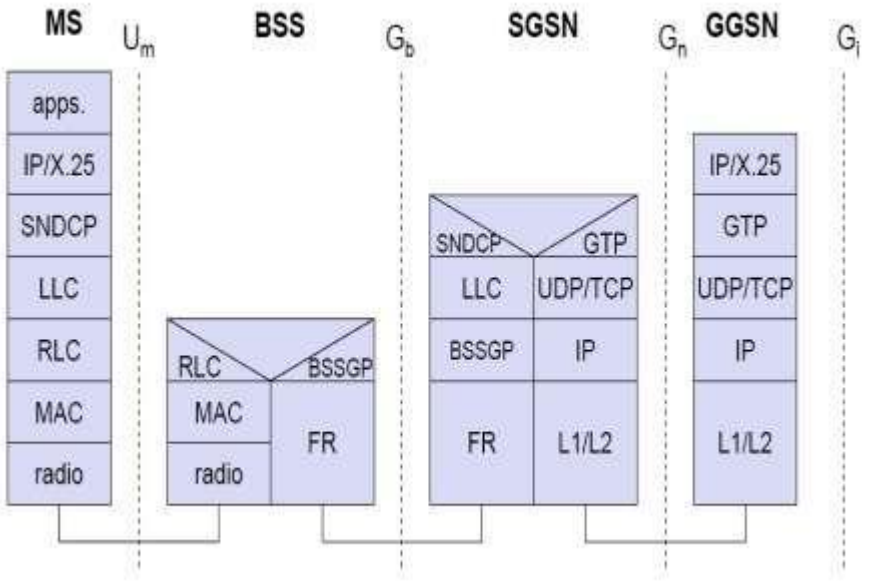
A connected MS which is inactive is put in the Standby state. Moving back to Ready can be triggered by sending data or signalling information from the MS to the SGSN. Upon arrival of data destined to the MS, the SGSN pages the latter and a response to the page moves the MS back to the Ready state. The MS may wish (or can be forced by the network) to terminate the connection by requesting to detach in which case it returns to Idle. A timer is used by the SGSN to monitor the tracking of the MS, and when it expires, the MS is detached and is considered unreachable

GPRS PROTOCOL ARCHITECTURE

A GPRS network introduces many new protocols designed to convey user data in a reliable and secure way. The protocol architecture is implemented for the transmission and signaling planes in GPRS. **Transmission plane protocols** are used for the transmission of user data and control functions. **Signaling plane protocols** are used to convey signaling information that controls and supports the transmission plane functions. (See figure 2)**.**

Fig 3.2 GPRS transmission plane protocol reference model





Transmission protocols in the Um interface

a) Physical layer

The physical layer can be divided into the Radio Frequency (RF) layer and the Physical Link layer.

The **Radio Frequency (RF)** is the normal GSM physical radio layer. Among other things the RF layer specifies the carrier frequency characteristics and GSM radio channel structures. It uses the radio modulation scheme for the data. The GSM RF physical layer is used for GPRS with the possibility for future modifications.

The **Physical Link layer** supports multiple MSs sharing a single physical channel and provides communication between the MSs and the network. Network controlled handovers are not used in the GPRS service. Instead, routing area updates and cell updates are used.

b)Medium Access Control (MAC)

MAC protocol handles the channel allocation and the multiplexing. The RLC and the MAC together form the OSI Layer 2 protocol for the U_m interface. The radio interface at U_m need GPRS which does not require changes compared to GSM.

c) Radio Link Control (RLC)

RLC protocol establishes a reliable radio link to the upper layers. It also works either in acknowledged or unacknowledged modes.

Logical Link Control (LLC)

LLC layer establishes a secure and reliable logical link between the MS and the SGSN for upper layer protocols. It works either in acknowledged or unacknowledged modes. The data confidentiality is ensured by using ciphering functions.

Subnetwork dependent convergence protocol (SNDCP)

SNDCP is used to transfer data packets between SGSN and MS. It is used to provide multiplexing of several connections of network layer onto one logical connection of underlying LLC layer. It provides functions that help to improve channel efficiency. This is achieved by means of compression techniques. Data Link layer is divided into LLC layer and RLC/MAC Layer.

Transmission protocols in the Gb interface
a)Physical Layer Protocol

Several physical layer configurations and protocols are possible at the Gb interface and the physical resources are allocated by Operation & Maintenance (O&M) procedures. Normally a G703/704 2Mbit/s connection is provided.

b) Network Services layer

The Gb interface Network Services layer is based on Frame Relay. Frame Relay virtual circuits are established between the SGSN and BSS. LLC PDUs from many users are statistically multiplexed onto these virtual circuits. These virtual circuits may traverse a network of Frame Relay switching nodes, or may just be provided on a point to point link between the BSC and the SGSN.

Base station subsystem GPRS protocol (BSSGP)

BSSGP is used to deliver routing and QoS-related information between the BSS and SGSN. It is to enable two physically distinct nodes, the SGSN and BSS. It is to operate node management control functions. There is a one-to-one relationship between the BSSGP protocol in the SGSN and in the BSS. If one SGSN handles multiple BSSs, the SGSN has to have one BSSGP protocol device for each BSS. BSSGP does not perform error correction and works on top of a frame relay (FR) network.

Transmission protocols in the Gn interface

a)Layer 1 and Layer 2

The **L1** and the **L2** protocols are vendor dependent OSI layer 1 and 2 protocols. It carries the IP datagrams for the GPRS backbone network between the SGSN and the GGSN.

b) Internet Protocol (IP)

The Internet Protocol (IP) datagram in the Gn interface is only used in the GPRS backbone network. The GPRS backbone (core) network and the GPRS subscribers use different IP addresses. This makes the GPRS backbone IP network invisible to the subscribers and vice versa. The GPRS backbone network carries the subscriber IP or X.25 traffic in a secure GPRS tunnel.

c) TCP or UDP

TCP or UDP are used to carry the GPRS Tunnelling Protocol (GTP) PDUs across the GPRS backbone network. TCP is used for user X.25 data and UDP is used for user IP data and signalling in the Gn interface.

d) GPRS tunneling protocol (GTP)

GTP is the basis for tunnel signaling. It uses two transport protocols such as reliable TCP and non-reliable UDP**.** The GPRS Tunnelling Protocol (GTP) allows multi-protocol packets to be tunnelled through the GPRS backbone between GPRS Support Nodes (GSNs).

SIGNIFICANCE:

This unit is important for cellular telephone technique and GSM networks. and also include security related techniques and recent technology in mobile computing

UNIT II WIRELESS NETWORKS

Wireless LANs and PANs – IEEE 802.11 Standard – Architecture – Services –Network – HiperLAN – Blue Tooth- Wi-Fi – WiMAX

PREQUISTIES DISCUSSION:

In this unit follows the classical layers of communication systems and explains the basics of wireless technology from a computer science point of view. Topics in this chapter are signal propagation, multiplexing, and modulation. Profound electrical engineering knowledge is not required; however, it is necessary to comprehend the basic principles of wireless transmission to understand the design decisions of higher layer communication protocols and applications.

2.1 WIRELESS NETWORKS

WIRELESS LANs

The increased demands for mobility and flexibility in our daily life are demands that lead the development from wired LANs to wireless LANs (WLANs).

WLANS use electromagnetic radio waves to transport data between computers in a Local Area Network (LAN), without the limitations set by "hard wired network cable or phone wire connection". Whilst simple optical links are commercially available, radio is presently more useful since it is not strictly restricted to line-of-sight paths.

Radio waves are often called radio carriers when they are used to carry information. The data to be transported is superimposed on the radio carrier by various modulation techniques which allow the data to be faithfully reconstructed at the receiving end. Once data is superimposed (modulated) onto the radio carrier, this combined "radio channel" now occupies more than a single frequency since the frequency components or spectra of the modulating data add frequency bandwidth to the basic carrier (in direct proportion to its information content or bit rate). The frequency range which is needed to accommodate a radio signal with any given modulation bandwidth is called a channel. Radio receiver techniques can select one radio channel while efficiency rejecting signals on other frequencies. Many radio signals to and from many users can thereby co-exist in the same place and time without interfering with each other if the radio waves are transmitted at minimum necessary power within different radio channels.

ADVANTAGES OF WLAN OVER WIRED LAN

- **Flexibility :** With in radio coverage nodes can communicate without further restriction**.**
- **Planning :** Wireless ad hoc network allow communication without planning whereas wired network needs wiring plans.
- **Design :** Wireless Network allows for the design of small independent devices.
- **Robustness :** Wireless network can survive disaster. If the wireless devices survive people can communicate.

- **Cost :** Adding additional users to a wireless network will be increase the cost. But where as with fixed network addition of an user will lead into unplugging and plugging. Wireless communications do not wear out.

DISADVANTAGES

- **QOS:**

  0 Wireless offers lower quality than that of wired. The reasons are

  1 The lower bandwidth due to limitations in radio transmission.

  2 High error rate due to interference.

  3 Higher delay due to error correction and detection mechanisms.

- **PROPRIETARY SOLUTION :**

  0 Many companies have comeup with proprietary solutions offering standardized functionality.

  1 This is due to slow standardization procedures.

- **RESTRICTION :**

  0 The wireless products need to comply with national regulations.

  1 WLAN are limited to low power senders and certain license free frequency hand which are not same world wide.

- **SAFETY AND SECURITY :**

  0 The radio waves are used for data transmission. They will interfere with other equipment. Precaution have to be taken to prevent safety hazards.i

- As it is via radio transmissions eaves dropping is possible.

DESIGN GOALS

- **Global operation :** While the product is being sold in all the countries, national and international frequency regulations should be considered.

- **Low Power :** Devices communicating via WLAN are also wireless devices. These devices run on battery power – while designing a WLAN these aspects should also be considered.

- **License free Operation :** The equipment must operate in a license free band such as 2.4 GHz ISM Band

- **Robust Transmission Technology :** WLAN operate under difficult conditions. As they are radio transmission many other electrical devices can interfere with them.

- **Simplified Spontaneous Co-operation :** WLAN should not require complicated startup routines , but should run spontaneously after power up.

- **Easy to use :** WLAN's are mad for simple use. They should be like plug and play.

- **Protection of Investment :** For transmission from wired to wireless, simple bridging should be enough to interoperate.

- **Safety and Security :** WLAN should be safe to operate. When low radiation are used. The network should consider user privacy and security.

- **Transparency :** Existing applications should continue to run over WLAN with trade off to higher delay and lower bandwidth.

2.2 IEEE 802.11 STANDARD

The IEEE standard 802.11 (IEEE, 1999) specifies the most famous family of WLANs in which many products are available. As the standard's number indicates, this standard belongs to the group of 802.x LAN standards, e.g.,

802.3 Ethernet or 802.5 Token Ring. This means that the standard specifies the physical and medium access layer adapted to the special requirements of wireless LANs, but offers the same interface as the others to higher layers to maintain interoperability. The primary goal of the standard was the specification of a simple and robust WLAN which offers time-bounded and asynchronous services.

2.3 SYSTEM ARCHITECTURE

Wireless networks can exhibit two different basic system architectures as infrastructure-based or ad-hoc. Figure shows the components of an infrastructure and a wireless part as specified for IEEE 802.11. Several nodes, called stations (STAi), are connected to access points (AP). Stations are terminals with access mechanisms to the wireless medium and radio contact to the AP. The stations and the AP which are within the same radio coverage form a basic service set (BSSi). The example shows two BSSs – BSS1 and BSS2 – which are connected via a distribution system. A distribution system connects several BSSs via the AP to form a single network and thereby extends the wireless coverage area. This network is now called an extended service set (ESS) and has its own identifier, the ESSID. The ESSID is the 'name' of a network and is used to separate different networks. Without knowing the ESSID (and assuming no hacking) it should not be possible to participate in the WLAN. The distribution system connects the wireless networks via the APs with a portal, which forms the interworking unit to other LANs.

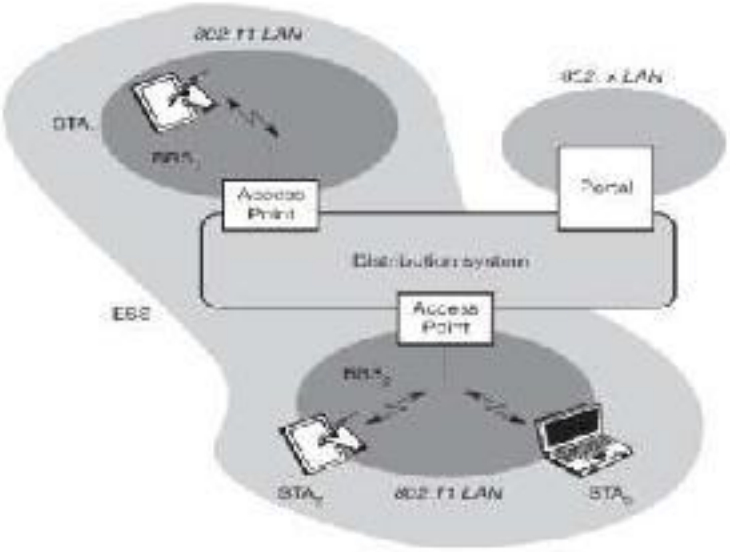


Fig 2.1 Architecture of infrastructure based IEEE 802.11

The architecture of the distribution system is not specified further in IEEE 802.11. It could consist of bridged IEEE LANs, wireless links, or any other networks. However, distribution system services are defined in the standard Stations can select an AP and associate with it. The APs support roaming (i.e., changing access points), the distribution system handles data transfer between the different APs. APs provide synchronization within a BSS, support power management, and can control medium access to support time-bounded service. In addition to infrastructure-based networks, IEEE 802.11 allows the building of ad-hoc networks between stations, thus forming one or more independent BSSs (IBSS) as shown in Figure 7.4. In this case, an IBSS comprises a group of stations using the same radio frequency. Stations STA1, STA2, and STA3 are in IBSS1, STA4 and STA5 in IBSS2. This means for example that STA3 can communicate directly with STA2 but not with STA5. Several IBSSs can either be formed via the distance between the IBSSs or by using different carrier frequencies (then the IBSSs could overlap physically). IEEE 802.11 does not specify any special nodes that support routing, forwarding of data or exchange of topology information as, e.g., HIPERLAN 1 or Bluetooth.

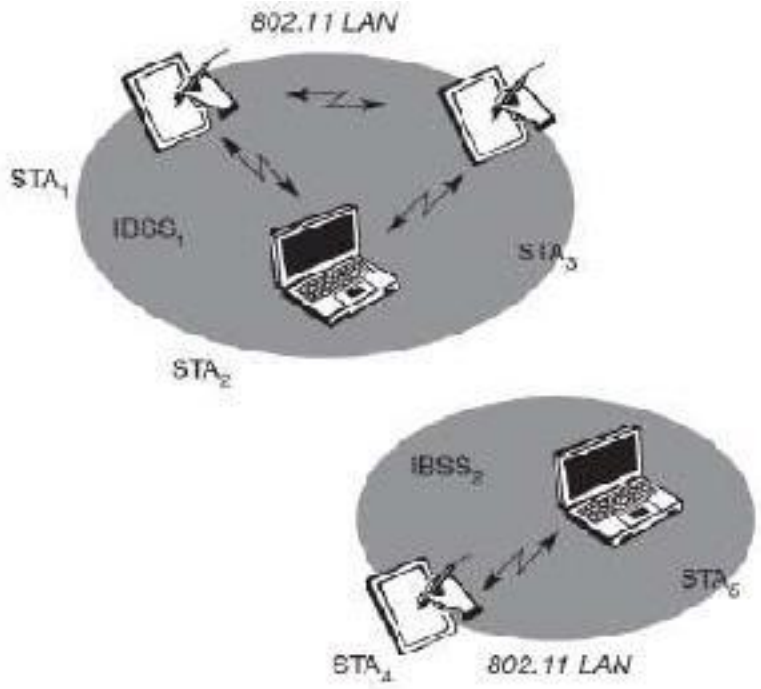


Figure 2.2 Architecture of Ad-hoc Wireless LANs

2.4 SERVICES

The MAC layer has to fulfill several tasks. First of all, it has to control medium access, but it can also offer support for roaming, authentication, and power conservation. The basic services provided by the MAC layer are the mandatory **asynchronous data service** and an optional **time-bounded service**. While 802.11 only offers the asynchronous service in ad-hoc network mode, both service types can be offered using an infrastructure-based network together with the access point coordinating medium access. The asynchronous service supports broadcast and multi-cast packets, and packet exchange is based on a 'best effort' model, i.e., no delay bounds can be given for transmission. The following three basic access mechanisms have been defined for IEEE 802.11: the mandatory basic method based on a version of CSMA/CA, an optional method avoiding the hidden terminal problem, and finally a contention- free polling method for time-bounded service. The first two methods are also summarized as **distributed coordination function (DCF)**, the third method is called **point coordination function (PCF**). DCF only offers asynchronous service, while PCF offers both asynchronous and time-bounded service but needs an access point to control medium access and to avoid contention. The MAC mechanisms are also called **distributed foundation wireless medium access control (DFWMAC)**. For all access methods, several parameters for controlling the waiting time before medium access are important. Figure 7.9 shows the three different parameters that define the priorities of medium access. The values of the parameters depend on the PHY and are defined in relation to a **slot** time. Slot time is derived from the medium propagation delay, transmitter delay, and other PHY dependent parameters. Slot time is 50 μs for FHSS and 20 μs for DSSS. The medium, as shown, can be busy or idle (which is detected by the CCA). If the medium is busy this can be due to data frames or other control frames. During a contention phase several nodes try to access the medium.

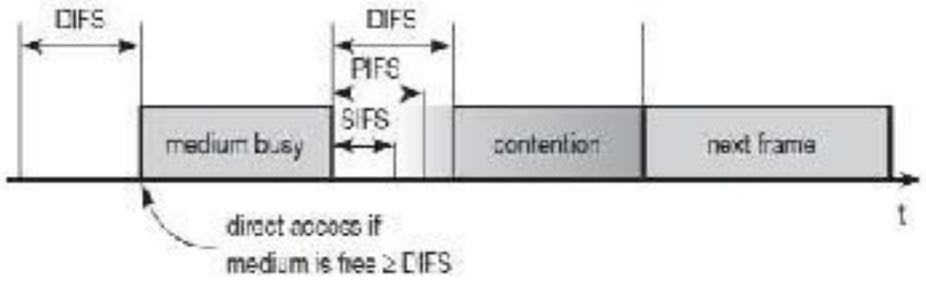


Figure 2.3 Medium Access and Inter Frame Spacing

Short inter-frame spacing (SIFS):

The shortest waiting time for medium access (so the highest priority) is defined for short control messages, such as acknowledgements of data packets or polling responses. For DSSS SIFS is 10 μs and for FHSS it is 28 μs.

PCF inter-frame spacing (PIFS):

A waiting time between DIFS and SIFS (and thus a medium priority) is used for a time-bounded service. An access point polling other nodes only has to wait PIFS for medium access. PIFS is defined as SIFS plus one slot time.

DCF inter-frame spacing (DIFS):

This parameter denotes the longest waiting time and has the lowest priority for medium access. This waiting time is used for asynchronous data service within a contention period. DIFS is defined as SIFS plus two slot times.

2.6 HIPERLAN

INTRODUCTION

HIPERLAN stands for **high performance local area network**. It is a wireless standard derived from traditional LAN environments and can support multimedia and asynchronous data effectively at high data rates of 23.5 Mbps. It is primarily a European standard alternative for the IEEE 802.11 standards and was published in 1996. It is defined by the European Telecommunications Standards

Institute (ETSI). It does not necessarily require any type of access point infrastructure for its operation, although a LAN extension via access points can be implemented.

Radio waves are used instead of a cable as a transmission medium to connect stations. Either, the radio transceiver is mounted to the movable station as an add-on and no base station has to be installed separately, or a base station is needed in addition per room. The stations may be moved during operation-pauses or even become mobile. The maximum data rate for the user depends on the distance of the communicating stations. With short distance(<50 m) and asynchronous transmission a data rate of 20 Mbit/s is achieved, with up to 800 m distance a data rate of 1 Mbit/s are provided. For connection-oriented services, e.g. video-telephony, at least 64 kbit/s are offered.

HIPERLAN uses cellular-based data networks to connect to an ATM backbone. The main idea behind HIPERLAN is to provide an infrastructure or ad-hoc wireless with low mobility and a small radius. HIPERLAN supports isochronous traffic with low latency. The HiperLAN standard family has four different versions.

The key feature of all four networks is their integration of time-sensitive data transfer services. Over time, names have changed and the former HIPERLANs 2,3, 1nd 4 are now called HiperLAN2, HIPERACCESS, and HIPERLINK.

Table 2.1: HIPERLAN protocol family

	HIPERLAN 1	HIPERLAN 2	HIPERLAN 3	HIPERLAN 4
Application	wireless LAN	access to ATM fixed networks	wireless local loop	point-to-point wireless ATM connections
Frequency	5.1-5.3GHz			17.2-17.3GHz
Topology	decentralized ad-hoc/infrastructure	cellular, centralized	point-to-multipoint	point-to-point
Antenna	omni-directional		directional	
Range	50 m	50-100 m	5000 m	150 m
QoS	statistical	ATM traffic classes (VBR, CBR, ABR, UBR)		
Mobility	<10m/s		stationary	
Interface	conventional LAN	ATM networks		
Data rate	23.5 Mbit/s	>20 Mbit/s		155 Mbit/s
Power conservation	yes		not necessary	

1. HIPERLAN 1

Planning for the first version of the standard, called HiperLAN/1, started 1991, when planning of 802.11 was already going on. The goal of the HiperLAN was the high data rate, higher than 802.11. The standard was approved in 1996. The functional specification is EN300652, the rest is in ETS300836.

The standard covers the Physical layer and the Media Access Control part of the Data link layer like 802.11. There is a new sub layer called Channel Access and Control sub layer (CAC). This sub layer deals with the access requests to the channels. The accomplishing of the request is dependent on the usage of the channel and the priority of the request.

CAC layer provides hierarchical independence with Elimination-Yield Non-Preemptive Multiple Access mechanism (EY-NPMA). EY-NPMA codes priority choices and other functions into one variable length radio pulse preceding the packet data. EY-NPMA enables the network to function with few collisions even though there would be a large number of users. Multimedia applications work in HiperLAN because of EY-NPMA priority mechanism. MAC layer defines protocols for routing, security and power saving and provides naturally data transfer to the upper layers.

On the physical layer FSK and GMSK modulations are used in HiperLAN/1.

HiperLAN features:

- range 50 m
- slow mobility (1.4 m/s)
- supports asynchronous and synchronous traffic
- sound 32 kbit/s, 10 ns latency
- video 2 Mbit/s, 100 ns latency
- data 10 Mbit/s

HiperLAN does not conflict with microwave and other kitchen appliances, which are on 2.4 GHz.

Elimination-yield non-preemptive priority multiple access (EY-NPMA)

EY-NPMA is a contention based protocol that has been standardized under ETSI's HIPERLAN, a standard for wireless LANs. Unlike other contention based protocols, EY-NPMA provides excellent support for different classes of traffic regarding quality of service and demonstrates very low collision rates. EY-NPMA is the medium access mechanism used by HIPERLAN Type 1. It uses active signaling.

Active signaling takes advantage of the fact that the current wireless technology

enables us to have a slot time very much smaller than the average packet size. Each node that wants to access the medium transmits a non-data preamble pattern consisting of slots. This pattern is made up of alternating idle and busy periods of different lengths (measured in slots). Conflict resolution and collision detection is done during this preamble. The main rule is that if a node detects a signal during one of its listening periods in its pattern, it aborts and defers until the next cycle. Otherwise, the node transmits its packet at the end of the pattern transmission.

With EYNPMA, each station may attempt to access the channel when a condition out of a group of three is met. The three conditions are:

- Channel free condition
- Synchronized channel condition
- Hidden elimination condition

The channel free condition occurs when the channel remains idle for at least a predefined time interval. A station willing to transmit senses the channel for this time interval, the station extends its period of sensing by a random number of slots (backoff). If the channel is still sensed as idle during the backoff period, the station commences transmitting. In both modes of operation unicast transmissions must get positively acknowledged or else the transmission is declared erroneous. Multicast and broadcast packets are not acknowledged.

The synchronized channel condition occurs when the channel is idle in the channel synchronization interval, which starts immediately after the end of the previous channel access cycle. The synchronized channel access cycle consists of three distinct phases:

- Prioritization
- Contention(Elimination and Yield)
- Transmission

Important features of the EY-NPMA

5. No preemption by frames with higher priority after the priority resolution possible.
6. Hierarchical independence of performance.
7. Fair contention resolution of frames with the same priority

In prioritization, EY-NPMA recognizes five distinct priorities from 0 to 4, with 0 being the highest priority. The cycle begins with each station having data to transmit sensing the channel for as many slots as the priority of the packet in its

buffer. All stations that successfully sense the channel as idle for the whole interval proceed to the next phase, the elimination phase.

During the elimination phase, each station transmits an energy burst of random length. These bursts ensure that only the stations having the highest priority data at a time proceed to the elimination phase. The length of the energy burst is a multiple of slots up to a predefined maximum. As soon as a station finishes bursting, it immediately senses the channel. If the channel is sensed as idle, the station proceeds to the next phase. Otherwise, it leaves the cycle.

During the yield phase, the station that survived the two previous ones, back off for a random number of slots. The station that backs off for the shortest interval eventually gets access of the channel for data transmission. All other station sense the beginning of the transmission and refrain from transmitting.

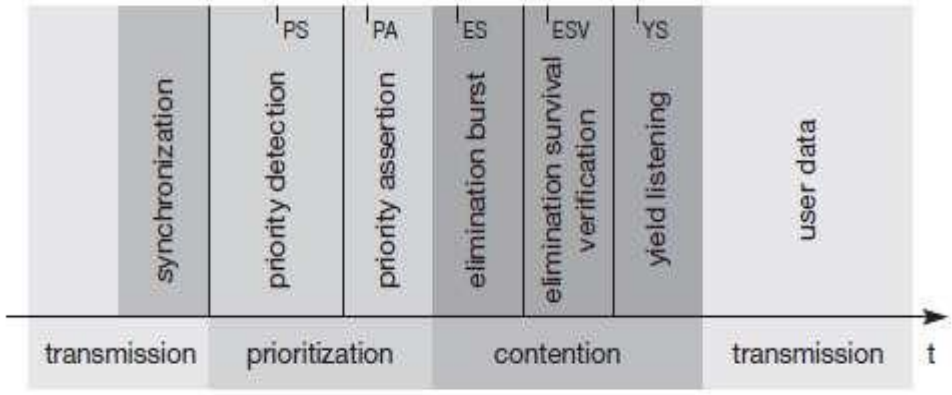


Fig 2.4 Phases of the HIPERLAN 1 EY-NPMA access scheme

a) Prioritization Phase

Prioritization Phase is the first attempt at reducing the number of contenders for the channel. Every contender calculates the number of idle slots according to the priority of its data, and senses the channel during those slots. Contenders with highest priority data will have no idle slots, while those with lowest priority data will have all idle slots If it detects a signal during those idle slots, it defers until the next cycle. This means that only the higher priority contenders survive. If it does not detect a signal during these idle slots, it sends the priority pulse and enters the elimination phase. In the first phase of the synchronized channel access cycle, known as the Prioritization Phase, every node allows a number of idle slots, where the default slot length is 168 high rate bit-periods.

The number of the idle slots is equal to the arithmetic value of the CAM

priority of the packet. Every contending node senses the channel, while it allows the idle slots. If it detects a signal transmission, it defers, that is it quits the effort to gain access to the channel and waits for the next channel access cycle to try to transmit. When a node detects no transmission during the Prioritization Phase, it transmits a pulse right after the idle slots, and proceeds to the next phase. This pulse is the one listened by every "defeated" node. The nodes that proceed to the next contention phase have a packet to send of the same highest CAM priority.

b) Elimination Phase

Elimination Phase is the second attempt at reducing the contenders. This phase consists of extending the priority pulse with a randomly calculated number of busy slots. The number of slots is independently calculated for each node. The probability of a larger than k-slot pulse is $1/2^k$. Therefore, the probability of a larger than 1-slot pulse is 1/2, larger than a 2-slot pulse is 1/4 and so on. Immediately after this pulse, the node senses the channel. If the channel is busy, it defers transmission until the next cycle. If the channel is idle, it enters the yield phase.

The nodes that "survive" the Prioritization Phase keep on trying to gain access to the channel. The objective of this medium access mechanism is to eliminate as more contending nodes as possible, but of course not all of them. During the Elimination Phase, a great percentage of the contending nodes is eliminated, but at least one of them survives. Every node that has not been defeated during the Prioritization Phase transmits an elimination pulse which is actually the lengthening of the priority pulse. Right after the end of this pulse, the nodes allow an idle slot, which is called survival verification slot, during which they sense the channel.

If a node detects a transmission during this time interval, this means that the specific node is "defeated", so it defers. Thus, the nodes that survive the Elimination Phase carry the packets of the highest priority and they have transmitted the longest elimination pulse.

Yield Phase

Yield Phase is the last phase of EY-NPMA, and is the last try to reduce collisions. Only the nodes that have survived elimination phase start the yield phase. The node selects a random number of idle slots uniformly distributed between 0 and 9. At the end of the yield phase, the node again senses the channel. If the channel is idle, it starts its transmission.

Yield Phase is the last phase before the transmission of a data packet and it is the last effort to reduce the number of the contending nodes. The nodes that have survived the Elimination Phase enter the Yield Phase allowing a number of idle

slots. Every node that detects transmission during these slots quits the current effort to gain access to the channel and waits till the next channel access cycle. If a node detects no transmission, it eventually transmits its data packet. Thus, a node "loses" in Yield Phase, when it listens some other node transmitting a data packet. The number of the idle slots is random and uniformly distributed between 0 and 9.

WATM

In the last decade of the twentieth century, technological improvements developed ways to achieve the objective of location and time independent communications. This objective has come into the light by the concept of personal communications networks and services. With the increasing role of multimedia and computer applications in communications, the main objective has become the extension of mobile communications and design a new generation of wireless personal communication networks, capable of supporting a variety voice, video and data traffic. The user demand for higher transmission speed and multimedia capability, as well as for mobile computing using portable computers becomes remarkable. These developments have motivated the studies on broadband wireless network technologies such as Wireless ATM (Asynchronous Transfer Mode) or WATM.

The concept of WATM was first proposed in 1992 as pointed out in and now it is actively considered as a potential framework for next-generation wireless communication networks capable of supporting integrated, quality-of-service (QoS) based multimedia services. The strength of wireless ATM technology is said to be its ability to provide support for different protocols, such as ISDN1 and Internet protocols. As the volume of wireless traffic is increasing, so is the role of QoS support, which will become very important when multiple services are multiplexed into the same radio access technology. As QoS support is a fundamental property of ATM technology, WATM promises a solution for this requirement. ATM is a very complex system and modifications for wireless communication and mobility management is going to make it more difficult.

Need for WATM

The area of wireless transmission systems has been increasing rapidly. Mobility raises a new set of questions, techniques, and solutions. This growth will occur in an environment characterized by rapid development of end-user applications and services towards the Internet and broadband multimedia delivery over the evolving fixed-wired infrastructure. Therefore, new developments of wireless networks are needed to enable wireless technologies to interwork with existing wired networks. Therefore, in order for ATM to be successful, it must offer a wireless extension. Otherwise it cannot participate in the rapidly growing field of mobile communications.

As ATM networks scale well from local area networks (LANs) to wide area networks (WANs), and there is a need for mobility in local and wide area applications, a mobile extension of ATM is required in order to have wireless access in local and wide environments. Many other wireless technologies, such as EEE 802.11, typically only offer best-effort services or to some extend time-bounded services. However, these services do not provide as many QoS parameters as ATM networks do. WATM could offer QoS for adequate support of multimedia data streams.

a. Reference Model

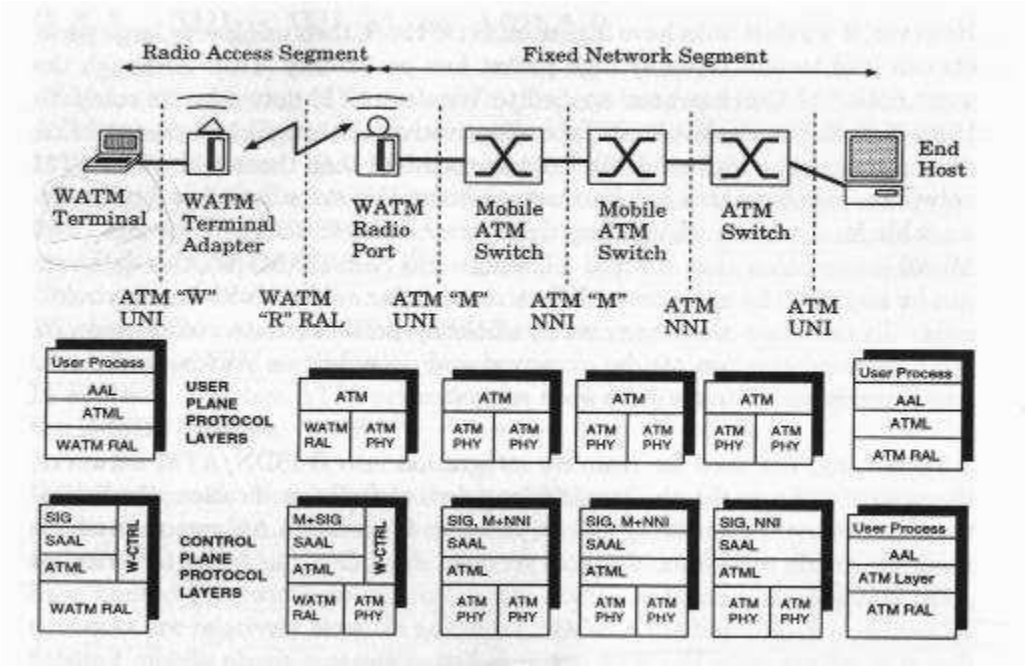


Figure 2 5 WATM Reference Model

The WATM system reference model, proposed by ATM Forum Wireless ATM (WATM) group, specifies the signaling interfaces among the mobile terminal, wireless terminal adapter, wireless radio port, mobile ATM switch and non-mobile ATM switch. It also specifies the user and control planes protocol layering architecture. This model is commonly advocated by many communication companies, such as NEC, Motorola, NTT, Nokia, Symbionics, and ORL.

The major components of a Wireless ATM system are: a) WATM terminal, b) WATM terminal adapter, c) WATM radio port, d) mobile ATM switch, e) standard ATM network and f) ATM host. The system reference model consists of a radio access segment and a fixed network segment. The fixed network is defined by "M (mobile ATM)" UNI and NNI interfaces while the wireless segment is defined by "R (Radio)" radio access layer (RAL) interface.

The "W" UNI is concerned with handover signaling, location management, wireless link and QoS control. The "R" RAL governs the signaling exchange between the WATM terminal adapter and the mobile base station. Hence, it concerns channel access, datalink control, meta-signaling, etc. The "M" NNI governs the signaling exchange between the WATM base station and a mobile capable ATM switch. It is also concerned with mobility-related signaling between the mobile capable ATM switches.

c) BRAN

The broadband radio access networks (BRAN), which have been standardized by the European Telecommunications Standards Institute (ETSI), could have been an RAL for WATM (ETSI, 2002b). The main motivation behind BRAN is the deregulation and privatization of the telecommunication sector in Europe. The primary market for BRAN includes private customers and small to medium-sized companies with Internet applications, multi-media conferencing, and virtual private networks. The BRAN standard and IEEE 802.16 (Broadband wireless access, IEEE, 2002b) have similar goals.

BRAN standardization has a rather large scope including indoor and campus mobility, transfer rates of 25–155 Mbit/s, and a transmission range of 50 m–5 km. Standardization efforts are coordinated with the ATM Forum, the IETF, other groups from ETSI, the IEEE etc. BRAN has specified four different network types (ETSI, 1998a):

4. **HIPERLAN 1:** This high-speed WLAN supports mobility at data rates above 20 Mbit/s. Range is 50 m, connections are multi-point-to-multi-point using ad-hoc or infrastructure networks
5. **HIPERLAN/2:** This technology can be used for wireless access to ATM or IP networks and supports up to 25 Mbit/s user data rate in a point-to-multi-point configuration.
6. **HIPERACCESS:** This technology could be used to cover the ‘last mile’ to a customer via a fixed radio link, so could be an alternative to cable modems or xDSL technologies (ETSI, 1998c).
7. **HIPERLINK:** To connect different HIPERLAN access points or HIPERACCESS nodes with a high-speed link, HIPERLINK technology can be chosen.
8. As an access network, BRAN technology is independent from the protocols of the fixed network. BRAN can be used for ATM and TCP/IP networks as illustrated in Figure. Based on possibly different physical layers, the DLC layer of BRAN offers a common interface to higher layers. To cover special

characteristics of wireless links and to adapt directly to different higher layer network technologies, BRAN provides a network convergence sub layer. This is the layer which can be used by a wireless ATM network, Ethernet, Fire wire, or an IP network. In the case of BRAN as the RAL for WATM, the core ATM network would use services of the BRAN network convergence sub layer.

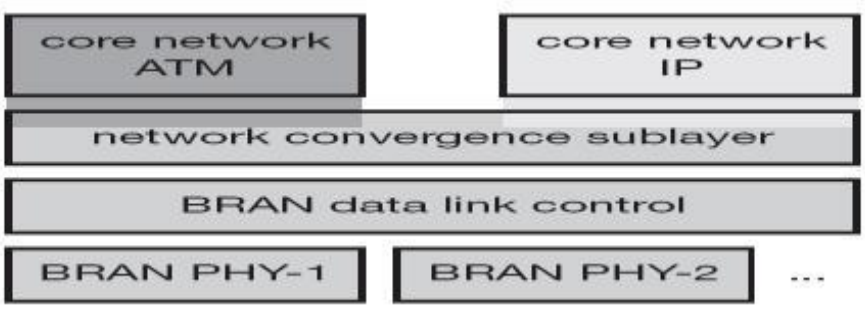


Fig 2.6 Layered model of BRAN wireless access Network

HiperLAN2 While HIPERLAN 1 did not succeed HiperLAN2 might have a better chance. HiperLAN2 offers more features in the mandatory parts of the standard (HiperLAN2, 2002).

- **High-throughput transmission:** Using OFDM in the physical layer and a dynamic TDMA/TDD-based MAC protocol, HiperLAN2 not only offers up to 54 Mbit/s at the physical layer but also about 35 Mbit/s at the network layer.
- **Connection-oriented:** Prior to data transmission HiperLAN2 networks establish logical connections between a sender and a receiver
- **Quality of service support:** support of QoS is much simpler. Each connection has its own set of QoS parameters (bandwidth, delay, jitter, bit error rate etc.).
- **Dynamic frequency selection:** HiperLAN2 does not require frequency planning of cellular networks or standard IEEE 802.11 networks.
- **Security support:** Authentication as well as encryption are supported by HiperLAN2.
- **Mobility support:** Mobile terminals can move around while transmission always takes place between the terminal and the access point with the best radio signal.

- **Application and network independence:** HiperLAN2 was not designed with a certain group of applications or networks in mind. Access points can connect to LANs running ethernet as well as IEEE 1394 (Firewire) systems used to connect home audio/video devices.
- **Power saves:** Mobile terminals can negotiate certain wake-up patterns to save power.

REFERENCE MODEL AND CONFIGURATIONS

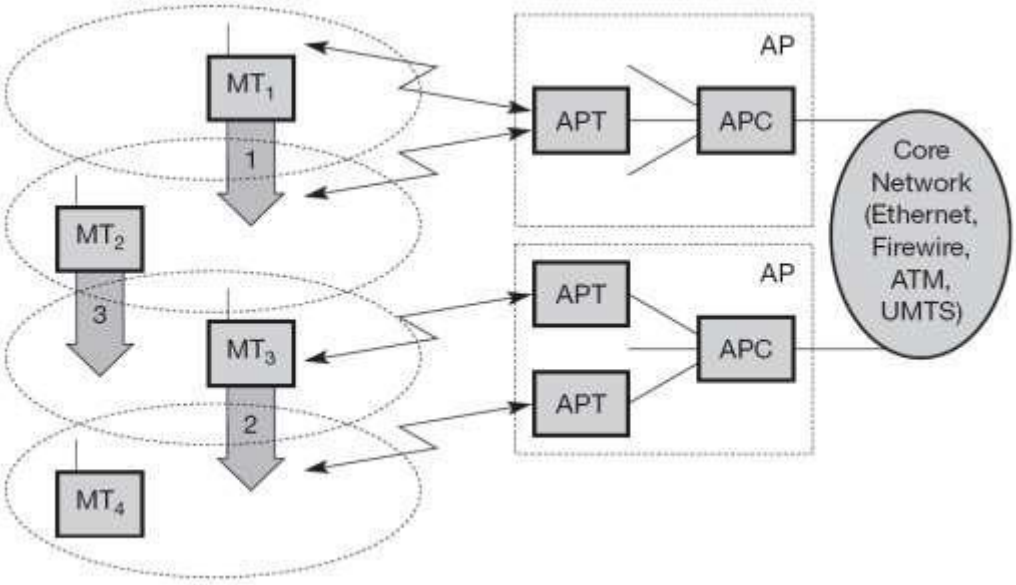


Fig 2.7 HiperLAN2 basic structure and handover scenarios

The above Figure shows the standard architecture of an infrastructure-based HiperLAN2 network. Here, two **access points** (AP) are attached to a core network. Core networks might be Ethernet LANs, Firewire (IEEE 1394) connections between audio and video equipment, ATM networks, UMTS 3G cellular phone networks etc. Each AP consists of an **access point controller** (APC) and one or more **access point transceivers** (APT).

An APT can comprise one or more sectors (shown as cell here). Finally, four **mobile terminals** (MT) are also shown. MTs can move around in the cell area as shown. No frequency planning is necessary as the APs automatically select the appropriate frequency via **dynamic frequency selection.** Three handover situations may occur:

7. **Sector handover** (Inter sector): If sector antennas are used for an AP, which is optional in the standard, the AP shall support sector handover. This type of handover is handled inside the DLC layer
8. **Radio handover** (Inter-APT/Intra-AP): As this handover type, too, is handled within the AP, no external interaction is needed.
9. **Network handover** (Inter-AP/Intra-network): This is the most complex situation: MT2 moves from one AP to another.

HiperLAN2 networks can operate in two different modes (which may be used simultaneously in the same network).

5. **Centralized mode** (CM): In infrastructure-based mode all APs are connected to a core network and MTs are associated with APs.
   - **Direct mode** (DM): The optional ad-hoc mode of HiperLAN2 directly

exchanged between MTs if they can receive each other, but the network still controlled.

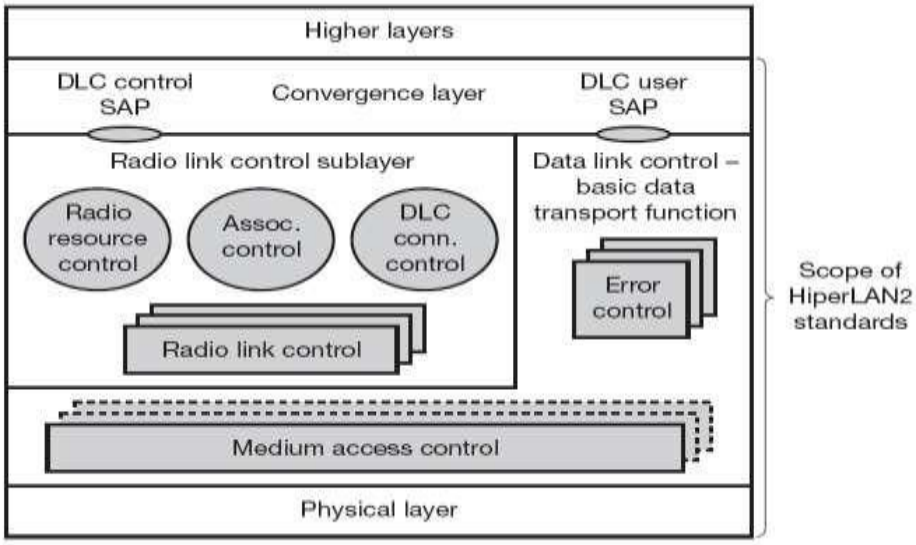


Fig 2.8 HiperLAN2 protocol stack

The above figure shows the HiperLAN2 protocol stack as used in access points. Protocol stacks in mobile terminals differ with respect to the number of MAC and RLC instances (only one of each). The lowest layer, the **physical layer**, handles as usual all functions related to modulation, forward error correction, signal detection, synchronization etc. The **data link control** (DLC) layer contains the MAC functions, the RLC sub layer and error control functions. The **MAC** of an AP assigns each MT a certain capacity to guarantee connection quality depending on available resources.

Above the MAC DLC is divided into a control and a user part. The user part contains **error control** mechanisms. HiperLAN2 offers reliable data transmission

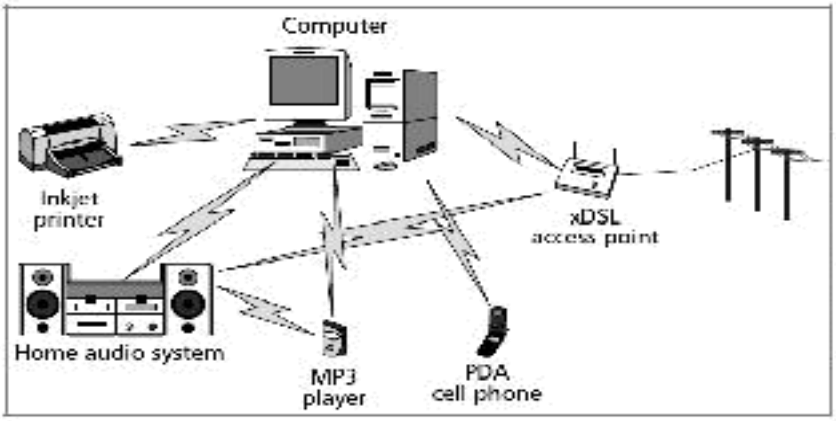
using acknowledgements and retransmissions. The **radio link control** (RLC) sub layer comprises most control functions in the DLC layer (the CC part of an AP). The **association control function** (ACF) controls association and authentication of new MTs as well as synchronization of the radio cell via beacons.

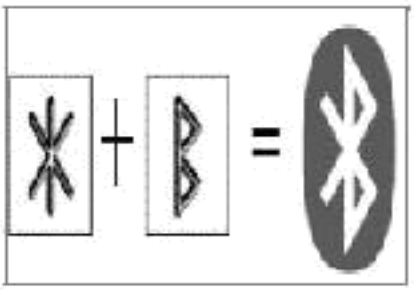
The **DLC user connection control** (DCC or DUCC) service controls connection setup, modification, and release. Finally, the **radio resource control** (RRC) handles handover between APs and within an AP. On top of the DLC layer there is the **convergence layer**. This highest layer of HiperLAN2 standardization may comprise segmentation and reassembly functions and adaptations to fixed LANs, 3G networks etc.

2.7 BLUETOOTH

INTRODUCTION

Bluetooth is a wireless technology standard for exchanging data over short distances (using short-wavelength radio transmissions in the ISM band from 2400–2480 MHz) from fixed and mobile devices, creating personal area networks (PANs) with high levels of security. Different type of network is needed to connect different small devices in close proximity (about 10 m) without expensive wiring or the need for a wireless infrastructure .Bluetooth is a new standard suggested by a group of electronics manufacturers that will allow any sort of electronic tools from computers and cell phones to keyboards and headphones to make its own connections, without wires, cables or any direct action from a user. A key distinction with other offered wireless technologies is that bluetooth enables combined usability models based on functions provided by different devices. Bluetooth was invented in1994 by L.M.Ericson of Sweden. The name is attributed to Harald Bluetooth was king of Denmark around the turn of the last millennium. Choosing this name for the standard indicates how important companies from the Baltic region (nations including Denmark, Sweden, Norway and Finland) are to the communications industry.





A BLUETOOTH NETWORK **SYMBOL OF BLUETOOH**

As famous as the name is the bluetooth symbol. Bluetooth icon can be recognized by all. The main strength of bluetooth is its ability to simultaneously handle both data and voice transmissions. It is capable of supporting one asynchronous data channel and up to three synchronous voice channels, or one channel sup-porting both voice and data. This ability combined with ad hoc device connection and automatic service discovery make it a superior solution for mobile devices and Internet applications. This grouping allows such novel solutions as a mobile hands-free headset for voice calls, print to fax capability, and automatically synchronizing PDA, laptop, and cell phone address book applications.

BLUETOOTH FEATURE:

- It is Wireless and automatic
- Bluetooth is inexpensive (< $5 per unit)
- It Handles both data and voice
- Signals are omni-directional and can pass through walls and briefcases
- Bluetooth uses frequency hopping at rate of 1600 Lops/sec
- It operates on 79 channels in 2.4GHZ band with 1MHZ carrier spacing
- Pi-conet is the important terminology

NETWORK TOPOLOGY

Piconet:

A set of bluetooth devices sharing a common channel is called piconet. A piconet is a collection of devices connected via Bluetooth technology in an ad hoc fashion. A piconet starts with two connected devices, and may grow to eight connected devices. All Bluetooth devices are peer units and have identical implementations. However, when establishing a piconet, one unit will act as a ***Master*** and the other(s) as ***slave***(s) for the duration of the piconet connection. Master is a Bluetooth device that sets the frequency hopping sequence. The Slave synchronizes to the Masters in time and frequency by following the Master's frequency hoping sequence. Every Bluetooth device has a unique

Bluetooth device address and a 28-bit Bluetooth clock. The baseband part of the Bluetooth System uses a special algorithm, which calculates the frequency hop sequence from the masters clock and device address. In addition to controlling the frequency hop sequence, the Master controls when Slaves are to transmit using Time Division Multiplexing (TDM).

When there is just one Master and one Slave the system is called a **Point to Point** connection. When many Slaves are connected to one Master, the system is called a **Point to Multipoint**. Both these types are referred to as a **Piconet** and all follow the frequency hopping sequence of the Master. The Slaves in the Piconet only have links to the Master and no direct links between Slaves.

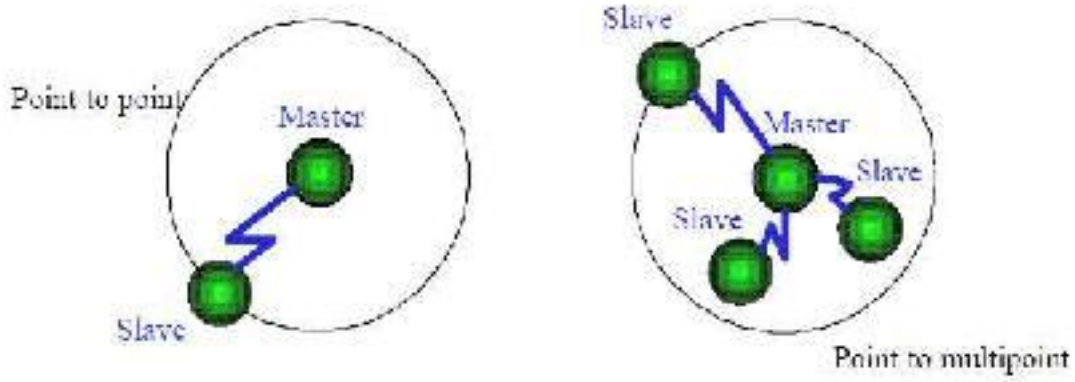


Fig 2.9 Piconet

Formation of piconet:

Two parameters are needed for the formation of piconet

- Hopping pattern of the radio it wishes to connect.
- Phase within the pattern i.e. the clock offset of the hops.

The global ID defines the hopping pattern. The master shares its global ID and its clock offset with the other radios which become slaves. The global ID and

the clock parameters are exchanged using a FHS (Frequency Hoping Synchronization) packet.

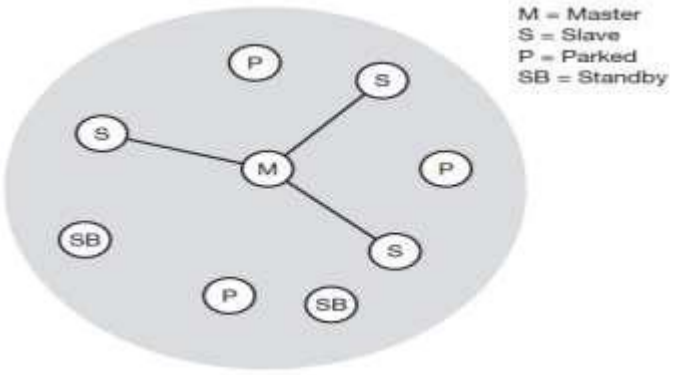


Fig 2.10 Simple Bluetooth piconet

There is no difference between terminals and base stations, two or more devices can form a piconet. The unit establishing the piconet repeatedly becomes the master, all other devices will be slaves. The hopping pattern is determined by the device ID, a 48-bit worldwide unique identifier. The phase in the hopping pattern is determined by the master's clock. After altering the interior clock according to the master a device may take part in the piconet. All active devices are assigned a 3-bit **active member address** (AMA). All parked devices use an 8-bit **parked member address** (PMA). Devices in stand-by do not need any address. All users within one piconet have the same hopping sequence and share the same 1 MHz channel. As more users join the piconet, the throughput per user drops quickly.

Scatternet :

Bluetooth defines a structure called scatternet to facilitate inter piconet communication. A scatternet is formed by interconnecting multiple piconet. A group of piconet is called **scatternet**.

If a device wants to take part in more than one piconet, it has to coordinate to the hopping sequence of the piconet it wants to take part in. If a device acts as slave in one piconet, it just starts to synchronize with the hopping sequence of the piconet it wants to join. After synchronization, it acts as a slave in this piconet and no longer participates in its former piconet. To permit synchronization, a slave has to know the uniqueness of the master that determines the hopping sequence of a piconet. Before leaving one piconet, a slave informs the current master that it will be unavailable for a certain amount of time.

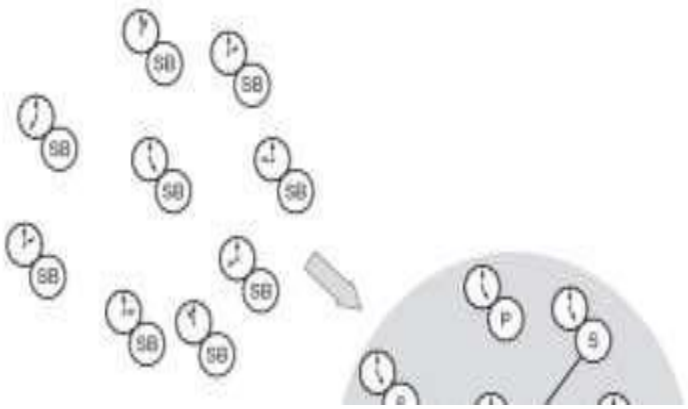


Fig 2.11 Formation of piconet

The left over devices in the piconet continue to communicate normal.

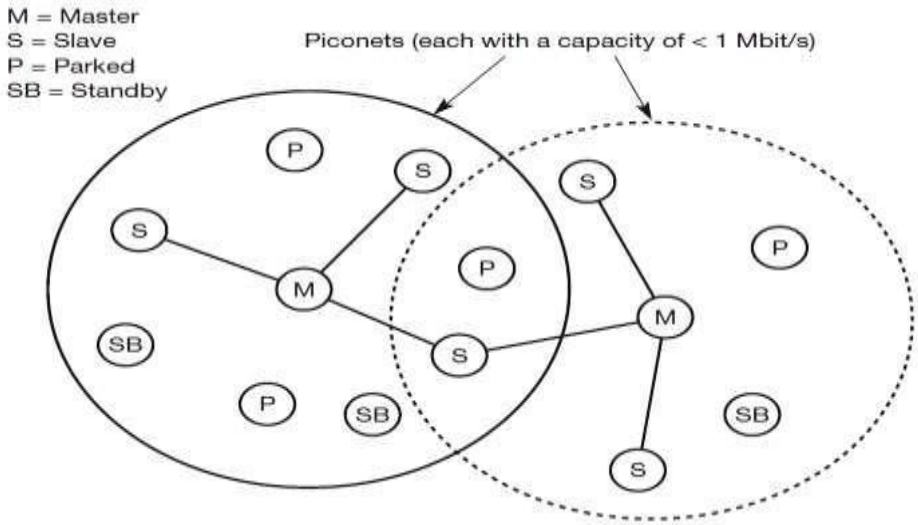


Fig 2.12 Bluetooth scatternet

A master can also go away from its piconet and act as a slave in another piconet. It is obviously not possible for a master of one piconet to act as the master of another piconet as this would direct to identical behavior. As soon as a master leaves a piconet, all traffic within this piconet is balanced until the master returns. Communication between different piconets takes place by devices jumping back and forth between these nets. If this is done occasionally, for instance, isochronous data streams can be forwarded from one piconet to another. On the other hand, scatternets are not yet supported by all piconet.

BLUETOOTH PROTOCOL STACK

:

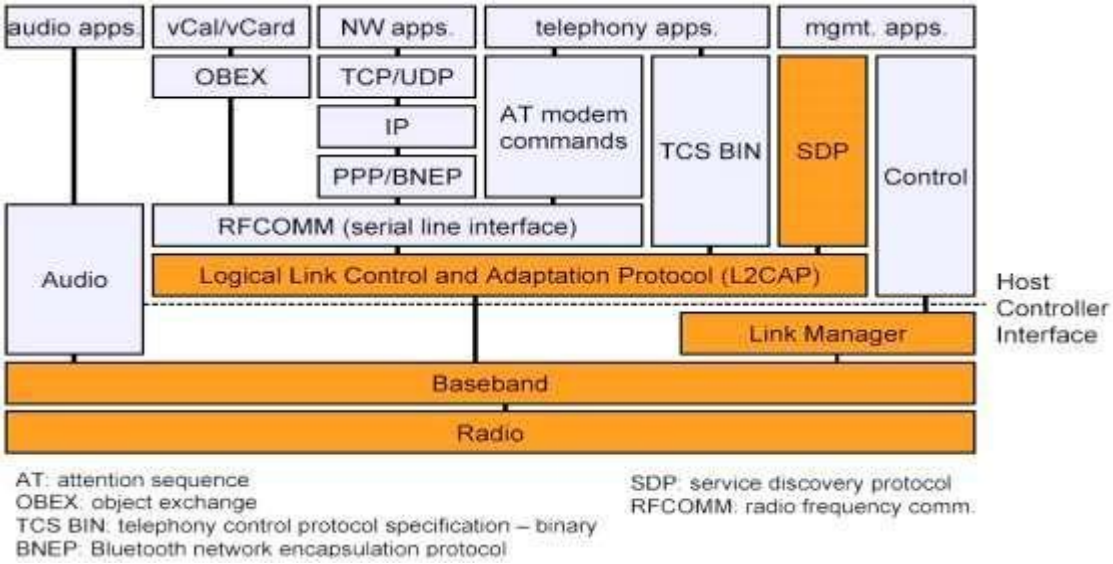


Fig 2.13 Bluetooth protocol stack

The Bluetooth protocol stack can be divided into:

- **Core Specification** -Deals with the lower layers of the architecture and describes how the technology works. It describe the protocol from physical to data link layer along with management functions.
- **Profile Specification** -Focuses on how to build interoperating devices using the core technology.
- **Bluetooth Radio** : specifics details of the air interface, including frequency, frequency hopping, modulation scheme, and transmission power.
- **Baseband**: concerned with connection establishment within a piconet, addressing, packet format, timing and power control.
- **Link manager protocol (LMP)**: establishes the link setup between Bluetooth devices and manages ongoing links, including security aspects (e.g. authentication and encryption), and control and negotiation of baseband packet size
- **Logical link control and adaptation protocol (L2CAP)**: adapts upper layer protocols to the baseband layer. Provides both connectionless and connection-oriented services.
- **Service discovery protocol (SDP)**: handles device information, services, and queries for service characteristics between two or more Bluetooth devices.
- **Host Controller Interface (HCI):** provides an interface method for accessing the Bluetooth hardware capabilities. It contains a command

interface, which acts between the Baseband controller and link manager

- **TCS BIN (Telephony Control Service)**: bit-oriented protocol that defines the call control signaling for the establishment of voice and data calls between Bluetooth devices.
- **OBEX(OBject EXchange)** : Session-layer protocol for the exchange of objects, providing a model for object and operation representation
- **RFCOMM**: a reliable transport protocol, which provides emulation of RS232 serial ports over the L2CAP protocol
- **WAE/WAP**: Bluetooth incorporates the wireless application environment and the wireless application protocol into its architecture.

Physical links

Different types of links can be established between master and slave. Two link types have been defined they are:

- Synchronous Connection-Oriented (SCO) link.
- Asynchronous Connection-Less (ACL) link.

1.Synchronous Connection Oriented (SCO): It Support symmetrical, circuit-switched, point-to-point connections . It is typically used for voice traffic. The Data rate is 64 kbit/s.

2.Asynchronous Connection-Less (ACL): It Support symmetrical and asymmetrical, packet-switched, point-to-multipoint connections. It is typically used for data transmission .Up to 433.9 kbit/s are used in symmetric or 723.2/57.6 kbit/s are used in asymmetric. The master uses polling. A slave may answer if it has used the preceeding slot.

Connection establishment states:

Standby : The State in which Bluetooth device is inactive, radio not switched on, enable low power operation.

Page : The Master enters page state and starts transmitting paging messages to Slave using earlier gained access code and timing information.

Page Scan : The Device periodically enters page state to allow paging devices to establish connections.

Inquiry: The State in which device tries to discover all Bluetooth enabled devices in the close vicinity.

Inquiry scan : Most devices periodically enter the inquiry scan state to make themselves available to inquiring devices.

Slave connection state modes:

Active –It participates in piconet It Listens, transmits and receives frames

Sniff – It only listens on specified slots
Hold –It does not support ACL frames. It has reduced power status. It May still participate in SCO exchanges

Park – It does not participate on piconet and it Still retained as part of piconet

Bluetooth security:

There are three modes of security for Bluetooth access between two devices.

- Non-secure
- Service level enforced security
- Link level enforced security

The following are the three basic security services specified in the Bluetooth standard:

Authentication : It verify the identity of communicating devices. User authentication is not provided natively by Bluetooth.

Confidentiality : It prevent information compromise caused by eavesdropping by ensuring that only authorized devices can access and view data.

Authorization :It allow the control of resources by ensuring that a device is authorized to use a service before permitting it to do so.

2.8 WI-FI-WIRELESS FIDELITY

Wireless Fidelity is commonly known as Wi-Fi, developed on IEEE 802.11 standards, is commonly used technology development in wireless communication. As the names indicate, WI-FI provides wireless access to applications and data across a radio network. WI-FI sets up many ways to build up a connection between the transmitter and the receiver such as DSSS, FHSS, IRInfrared and OFDM.

Wi-Fi provide its users with the authorization of connecting to the Internet from any place such as their home, office or a public place without the hassles of plugging in the wires. Wi-Fi is faster than the conventional modem for accessing information over a huge network. With the help of different amplifiers, the users

can easily alter their location without interference in their network access. Wi-Fi devices are yielding with each other to grant well-organized access of information to the user. Wi-Fi location, the users can attach to the wireless network is called a Wi-Fi hotspot. Through the Wi-Fi hotspot, the user can evenimprove their home business as accessing information through Wi-Fi is easy Accessing a wireless network through a hotspot in some cases is free of cost while in some it may carry extra charges. Many set of Wi-Fi devices such as PCI, miniPCI, USB, Cardbus and PC card, ExpressCard make the Wi-Fi experience suitable and enjoyable for the users. Distance from a wireless network can decrease the signal strength to quite an extent; some devices such as Ermanno Pietrosemoli and EsLaRed of Venezuela Distance are used for amplifying the signal power of the network. These devices create embedded systems that communicate with any other node on the Internet.

Wi-Fi uses radio networks to broadcast data between its nodes. Such networks are made up of cells that grant coverage across the network. The further the number of cells, the larger and stronger is the coverage on the radio network. The radio technology is a absolute package deal as it offers a secure and reliable connectivity. Radio bands such as 2.4GHz and 5GHz depend on wireless hardware such Ethernet protocol and CSMA. Originally, Phase Shift Keying (PSK), a modulation method for transmission of data was used, but now it has been replaced with CCK. Wi-Fi uses many spectrum such as FHSS and DSSS. The most accepted Wi-Fi technology such as 802.11b operates on the range of 2.40 GHz up to 2.4835 GHz band. This provides a complete platform for operating Bluetooth strategy, cellular phones, and other scientific equipments. While 802.11a technology has the range of 5.725 GHz to 5.850 GHz and provides up to 54 Mbps in speed. 802.11g technology is even enhanced as it cover three non-overlapping channels and permit PBCC. 802.11e technology takes a pale lead by providing outstanding streaming quality of video, audio, voice channels etc.

Fig 2.15 Scenario of wi-fi

Wi-Fi communication devices are extended forms of radios used for cell phones and walkie-talkies: they simultaneously transmit and receive radio waves and convert 1s to 0s into the radio waves along with reconverting the radio waves into 1s and 0s, however the Wi-Fi radios enjoy some exceptional features.

Advantages

Wi-Fi allows cheaper deployment of local area networks (LANs). Also spaces where cables cannot be run, such as outdoor areas and historical buildings, can host wireless LANs.

Manufacturers are building wireless network adapters into most laptops. The price of chipsets for Wi-Fi continues to drop, making it an economical networking option included in even more devices.

Different competitive brands of access points and client network-interfaces can inter-operate at a basic level of service. Products designated as "Wi-Fi Certified" by the Wi-Fi Alliance are backwards compatible. Unlike mobile phones, any standard Wi-Fi device will work anywhere in the world.

Wi-Fi Protected Access encryption (WPA2) is considered secure, provided a strong passphrase is used. New protocols for quality-of-service (WMM) make Wi-Fi more suitable for latency-sensitive applications (such as voice and video). Power saving mechanisms (WMM Power Save) extend battery life.

Limitations

Spectrum assignments and operational limitations are not consistent worldwide: most of Europe allows for an additional two channels beyond those permitted in the US for the 2.4 GHz band (1–13 vs. 1–11), while Japan has one more on top of that (1–14). As of 2007, Europe is essentially homogeneous in this respect.

A Wi-Fi signal occupies five channels in the 2.4 GHz band. Any two channel numbers that differ by five or more, such as 2 and 7, do not overlap. The oft-repeated adage that channels 1, 6, and 11 are the only non-overlapping channels is, therefore, not accurate. Channels 1, 6, and 11 are the only group of three non-overlapping channels in the U.S. In Europe and Japan using Channels 1, 5, 9, and 13 for 802.11g and 802.11n is recommended.

Equivalent isotropically radiated power (EIRP) in the EU is limited to 20 dBm (100 mW). The current 'fastest' norm, 802.11n, uses double the radio

spectrum/bandwidth (40 MHz) compared to 802.11a or 802.11g (20 MHz). This means there can be only one 802.11n network on the 2.4 GHz band at a given location, without interference to/from other WLAN traffic. 802.11n can also be set to use 20 MHz bandwidth only to prevent interference in dense community.

WIFI NETWORK SERVICES:

2. Distribution and integration
3. Association, re-association, and disassociation
4. Authentication and deauthentication
5. Providing privacy

Distribution:

This service is used by mobile stations in an infrastructure network every time they send data. Once a frame has been accepted by an access point, it uses the distribution service to deliver the frame to its destination. Any communication that uses an access point travels through the distribution service, including communications between two mobile stations associated with the same access point.

Integration:

Integration is a service provided by the distribution system; it allows the connection of the distribution system to a non-IEEE 802.11 network. The integration function is specific to the distribution system used and therefore is not specified by 802.11, except in terms of the services it must offer.

Association:

Delivery of frames to mobile stations is made possible because mobile stations register, or associate, with access points. The distribution system can then use the registration information to determine which access point to use for any mobile station.

Re-association:

When a mobile station moves between basic service areas within a single extended service area, it must evaluate signal strength and perhaps switch the access point with which it is associated. Reassociations are initiated by mobile stations when signal conditions indicate that a different association would be beneficial; they are never initiated by the access point. After the reassociation is complete, the distribution system updates its location records to reflect the reachability of the mobile station through a different access point.

Disassociation:

To terminate an existing association, stations may use the disassociation service. When stations invoke the disassociation service, any mobility data stored in the distribution system is removed. Once disassociation is complete, it is as if the station is no longer attached to the network. Disassociation is a polite task to do during the station shutdown process. The MAC is, however, designed to accommodate stations that leave the network without formally disassociating.

Authentication/deauthentication:

Physical security is a major component of a wired LAN security solution.

Wired network's equipment can be locked inside offices. Wireless networks cannot offer the same level of physical security, however, and therefore must depend on additional authentication routines to ensure that users accessing the network are authorized to do so. Authentication is a necessary prerequisite to association because only authenticated users are authorized to use the network. (In practice, though, many access points are configured for "open-system" mode and will authenticate any station.)

Deauthentication terminates an authenticated relationship. Because authentication is needed before network use is authorized, a side effect of deauthentication is termination of any current association.

WIFI SECURITY

WiFi hotspots can be open or secure. If a hotspot is open, then anyone with a WiFi card can access the hotspot. If it is secure, then the user needs to know a WEP key to connect. WEP stands for Wired Equivalent Privacy. WEP is an encryption system for the data that 802.11 sends through the air. Encryption system prevents any non-authorized party from reading or changing data. Specifically, it is the process of encoding bit stream in such a way that only the person (or computer) with the key (a digital sequence) can decode it.

2.9 WI-MAX

Wi-MAX (Worldwide Interoperability for Microwave Access) unites the technologies of wireless and broadband to provide high-speed internet access across long distances. The name was christened by WiMAX Forum that promotes interoperability and conformity of the standard. The forum defines the technology as "a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL". With the

guarantee of WiMAX Forum the vendors are authorized to sell their WiMAX certified products so they can enjoy operability with other products of same type. It is a telecommunication protocol capable of providing internet access to fixed and mobile users. For an outstanding performance like Wi-Fi networks along with QOS (Quality of Service) and coverage this Wireless Broadband Access (BAS) technology is assembled around IP (internet protocol). Currently it offers 40 Mbit/s but expected to offer 1 Gbit/s speed for fixed users.

WI-MAX ARCHITECTURE

There are three main components of WiMax network architecture.

7. The first component is the mobile stations which are used as a source of network connection for end user.
8. The second network is an access service network which is formed of more than two or three base stations. It also contains ASN gateways which build the radio access at the end.
9. The third component is connectivity service network which is responsible for providing IP functions. The base station provides the air interface for the mobile stations. The base stations also provide mobile management functions, triggering and tunnel establishment, radio resource management, dynamic host control protocol proxy, quality of service enforcement and multicast group management. ASN is responsible for radio resource management, encryption keys, routing to the selected network and client functionality. Connectivity service network is responsible for internet connections, corporate and public networks and many other user services.

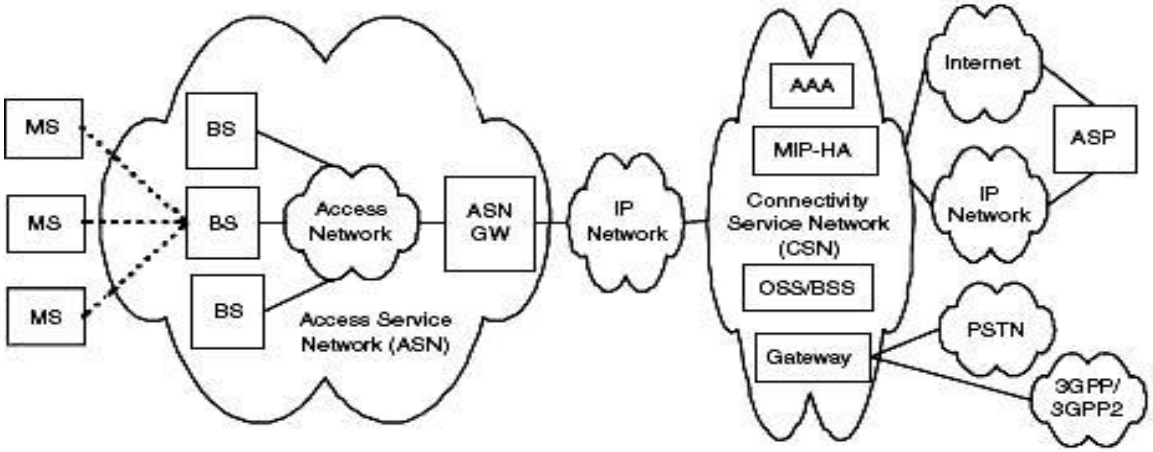


Fig 2.16 IP base WI-MAX architecture

Standard WiMax Architecture

The WiMax network is based on three four basic components they are:

3. AS gateway,
4. CSN and
5. MS.

The basic network has a inner IP core which is bounded by an ASN gateway, which is associated to service network or CSN. The main IP core is attach to the internet backbone for aid and coverage. The WiMax network which is also part of the ISP network is recognized as access service gateway. This ASN handles the micro and macro base stations, which offer WiMax access to end users. The connectivity examine network or CSN is an important part of WiMax architecture which provides the verification to the user devices.

CSN is in charge for providing roaming among the network service providers. It is CSN which is accountable for user security and quality for service for this reason it uses several protocols. The IP address management is also handled by CSN. IP core is in the middle of CSN and ASN. CSN provides the internet and telecommunications connectivity. ASP communicates to the base stations and the mobile stations. At the users end the WiMax architecture may additionally contain firewall for security. WiMax architecture provides discretion at user end to make possible amendments.

Two Dimensions of WiMax Network

WiMax network is composed of two parts the

1. WiMax tower 2. WiMax receiver.

WiMax tower is associated straightly to the internet backbone using a wired connection such as optical fiber. It can be linked to the WiMax tower using a line of sight link or a non line of sight link. The line of site communiqué involves the use of fixed antenna or dish. This antenna is unchanging or deployed on the roof top or the tower of the building. Line of sight connection is measured as more strong and stable connection. Thus it sends lot of error free data over the network. It uses a frequency range of 66Ghz. Higher frequency reduces the possibility of signal flaw and interference and provides extra bandwidth. On the other hand the non line of sight link provides you connectivity with the fixing of small antenna in your PC. This mode provides lower frequency range from 2 GHz to 11 GHz. The lower band signals are not prone to obstacles like trees and walls. Hence the signal

strength is more and the user receives the quality of service. For every WiMax connectivity and architecture it is significant to connect to an internet backbone via swift wired connection.

L2CAP-LOGICAL LINK CONTROL AND ADAPTION PROTOCOL:

The L2CAP is a data link control protocol.The L2CAP link layer operates over an ACL link provided by the baseband. A single ACL link, set up by the link manager using LMP, is always available between the master and any active slave. This provides a point-to-multipoint link supporting both asynchronous and isochronous data transfer. L2CAP provides services to upper-level protocols by transmitting data packets over L2CAP channels. Three types of L2CAP channels exist: bidirectional signaling channels that carry commands; connection-oriented channels for bidirectional point-to-point connections; and unidirectional connectionless channels that support point-to multipoint connections, allowing a local L2CAP entity to be connected to a group of remote devices.

Functions:

It Performs 4 major functions

2. Managing the creation and termination of logical links for each connection through "channel" structures
3. Enforcing and defining QoS requirements
4. Adapting Data, for each connection, between application (APIs) and Bluetooth Baseband formats through Segmentation and Reassembly (SAR)
5. Performing Multiplexing to support multiple concurrent connections over a single common radio interface.

Channels:

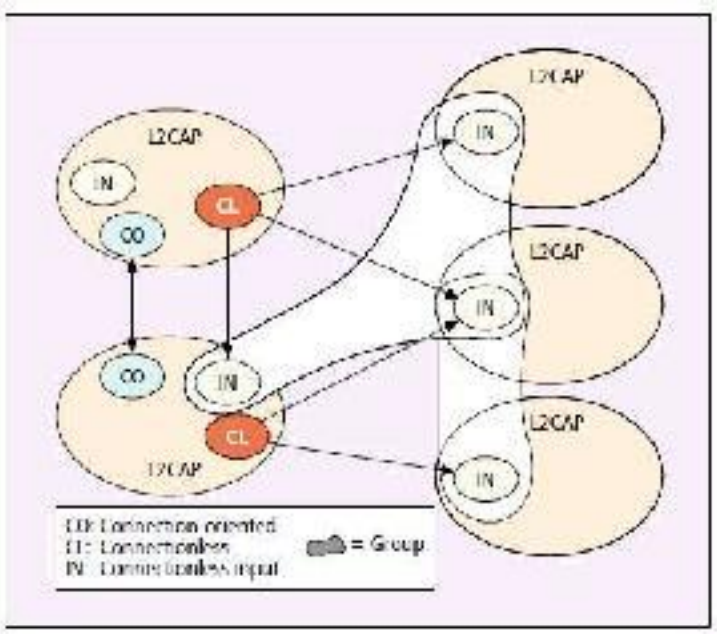


Fig 2.17 L2CAP Channels

L2CAP CHANNELS

The above figure shows L2CAP entities with various types of channels between them. Every L2CAP channel includes two endpoints referred to by a logical channel identifier (CID). Each CID may represent a channel endpoint for a connection oriented channel, a connectionless channel, or a signaling channel. Since a bi-directional signaling channel is required between any two L2CAP entities before communication can take place, every L2CAP entity will have one signaling channel endpoint with a reserved CID of 0x0001. All signal channels between the local L2CAP entity and any remote entities use this one endpoint. Each connection-oriented channel in an L2CAP entity will have a local CID that is dynamically allocated. All connection-oriented

CIDs must be connected to a single channel, and that channel must be configured before data

transfer can take place. Note that the channel will at that point be bound to a specific upper level

protocol. In addition, a quality of service (QoS) agreement for the channel will be established

between the two devices. QoS is negotiated for each channel during configuration and includes data flow parameters such as peak bandwidth, as well as the transmission type: best effort, guaranteed, or no traffic. Connectionless channels are unidirectional and used to form groups. A single outgoing connectionless CID on a local device may be logically connected to multiple remote devices.

The devices connected to this outgoing endpoint form a logical group. These outgoing CIDs are dynamically allocated. The incoming connectionless CID, however, is fixed at 0x0002. Although multiple outgoing CIDs may be created to form multiple logical groups, only one incoming connectionless CID is provided on each L2CAP entity. All incoming connectionless data arrives via this endpoint. These channels do not require connection or configuration. Therefore, any required configuration information, such as upper-level protocol, is passed as part of the data packet.

Functional requirement:

Protocol multiplexing distinguishes between upper-layer protocols like SDP, RFCOMM. It Segments larger packets from higher layers into smaller baseband packets. It allows QoS parameters to be exchanged during connection establishment and it also allows efficient mapping of protocol groups to piconets.

L2CAP Operation:

L2CAP channel end-points are represented by channel identifiers (CIDs). An L2CAP channel is uniquely defined by 2 CIDs and device addresses. Reserved CIDs

0x0001: Signaling channel 0x0002:
Connection-less reception
0x0003-0x003F: Reserved for future use

Operation between layers:

It transfers data between higher layer protocols and lower layer protocols. It Signal with peer L2CAP implementation. L2CA layer should be able to accept *events* from lower/upper layers. L2CA layer should be able to take appropriate *actions* in response to these events.

L2CAP Format

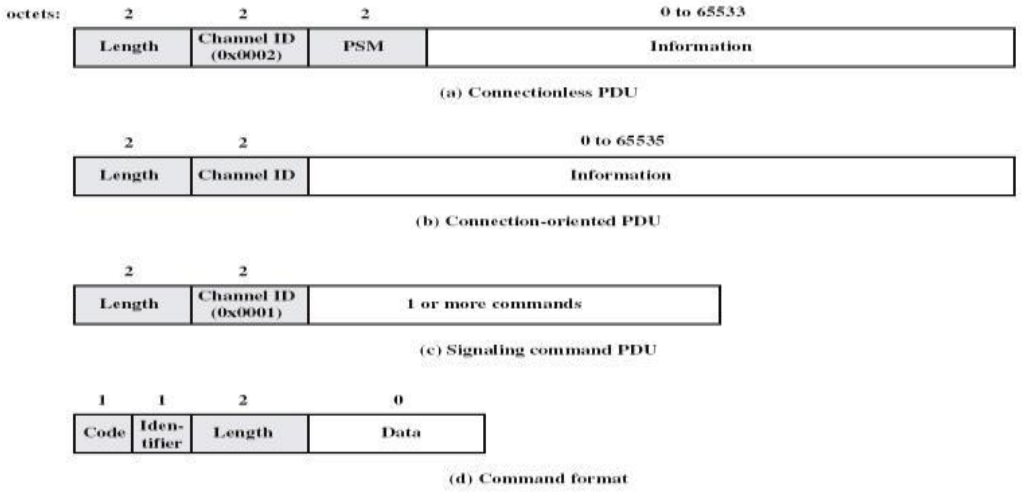


Fig 2.18 L2CAP Format

L2CAP Frame field for connectionless service:

Length – It indicates length of information payload, PSM fields
Channel ID – 2, indicating connectionless channel

Protocol/service multiplexer (PSM) – identifies higher-layer recipient for payload

Not included in connection-oriented frames
Information payload – higher-layer user data

Signaling frame payload:

It Consists of one or more L2CAP commands, each with four fields

Code – identifies type of command

Identifier – used to match request with reply

Length – length of data field for this command
Data – additional data for command, if necessary

L2CAP signaling command codes:

Code	Description	Parameters
0x01	Command reject	Reason
0x02	Connection request	PSM, Source CID
0x03	Connection response	Destination CID, Source CID, Result, Status
0x04	Configure request	Destination CID, Flags, Options
0x05	Configure response	Source CID, Flags, Result, Options
0x06	Disconnection request	Destination CID, Source CID
0x07	Disconnection response	Destination CID, Source CID
0x08	Echo request	Data (optional)
0x09	Echo response	Data (optional)
0x0A	Information request	InfoType
0x0B	Information response	InfoType, Result, Data (optional)

UNIT III ROUTING

Mobile IP – DHCP – AdHoc– Proactive and Reactive Routing Protocols – Multicast Routing.

PREQUISTIES DISCUSSION:

In this unit ,we presents several media access schemes and motivates why the standard schemes from fixed networks fail if used in a wireless environment.

3.ROUTING

3.1 MOBILE IP

While systems like GSM have been designed with mobility in mind, the internet started at a time when no one had thought of mobile computers. Today's internet lacks any mechanisms to support users traveling around the world. IP is the common base for thousands of applications and runs over dozens of different networks. This is the reason for supporting mobility at the IP layer; mobile phone systems, for example, cannot offer this type of mobility for heterogeneous networks. To merge the world of mobile phones with the internet and to support mobility in the small more efficiently, so-called micro mobility protocols have been developed.

a) Goals, Assumptions and Requirements

A host sends an IP packet with the header containing a destination address with other fields. The destination address not only determines the receiver of the packet, but also

the physical subnet of the receiver. For example, the destination address 129.13.42.99 shows that the receiver must be connected to the physical subnet with the network prefix 129.13.42. Routers in the internet now look at the destination addresses of incoming packets and forward them according to internal look-up tables. To avoid an explosion of routing tables, only prefixes are stored and further optimizations are applied. A router would otherwise have to store the addresses of all computers in the internet, which is obviously not feasible. As long as the receiver can be reached within its physical subnet, it gets the packets; as soon as it moves outside the subnet, a packet will not reach it.

A host needs a so-called **topologically correct address.** So moving to a new location would mean assigning a new IP address. The problem is that nobody knows about this new address. It is almost impossible to find a (mobile) host on the internet which has just changed its address. One could argue that with the help of dynamic DNS (DDNS, RFC 2136,Vixie, 1997) an update of the mapping logical name – IP address is possible. This is what many computer users do if they have a dynamic IP address and still want to be permanently reachable using the same logical computer name. It is important to note that these considerations, indeed most of mobile IP's motivation, are important if a user wants to offer services from a mobile node, i.e., the node should act as server. Typically, the IP address is of no special interest for service usage: in this case DHCP is sufficient. Another motivation for permanent IP addresses is emergency communication with permanent and quick reachability via the same IP address. So what about dynamically adapting the IP address with regard to the current location? The problem is that the domain name system (DNS) needs some time before it updates the internal tables necessary to map a logical name to an IP address. This approach does not work if the mobile node moves quite often.

The internet and DNS have not been built for frequent updates. Just imagine millions of nodes moving at the same time. DNS could never present a consistent view of names and addresses, as it uses caching to improve scalability. It is simply too expensive to update quickly.There is a severe problem with higher layer protocols like TCP which rely on IP addresses. Changing the IP address while still having a TCP connection open means breaking the connection. A TCP connection is identified by the tuple (source IP address, source port, destination IP address, destination port), also known as a **socket pair** (a socket consists of address and port). Therefore, a TCP connection cannot survive any address change. Breaking TCP connections is not an option, using even simple programs like telnet would be impossible. The mobile node would also have to notify all communication partners about the new address.

b) Requirements

Many field trials and proprietary systems finally led to mobile IP as a standard to enable mobility in the internet. Several requirements accompanied the development of the standard:

- **Compatibility:** The installed base of Internet computers, i.e., computers running TCP/IP and connected to the internet, is huge. A new standard cannot introduce changes for applications or network protocols already in use. People still want to use their favorite browser for www and do not want to change applications just for mobility, the same holds for operating systems.Mobile IP has to be integrated into existing operating systems or at least work with them (today it is available for many platforms). Routers within the internet should not necessarily require other software. While it is possible to enhance the capabilities of some routers to support mobility,

it is almost impossible to change all of them. Mobile IP has to remain compatible with all lower layers used for the standard, non-mobile, IP. Mobile IP must not require special media or MAC/LLC protocols, so it must use the same interfaces and mechanisms to access the lower layers as IP does. Finally, end-systems enhanced with a mobile IP implementation should still

be able to communicate with fixed systems without mobile IP. Mobile IP has to ensure that users can still access all the other servers and systems in the internet. But that implies using the same address format and routing mechanisms.

 Transparency: Mobility should remain 'invisible' for many higher layer protocols and applications. Besides maybe noticing a lower bandwidth and some interruption in service, higher layers should continue to work even if the mobile computer has changed its point of attachment to the network.For TCP this means that the computer must keep its IP address as explained above. If the interruption of the connectivity does not take too long, TCP connections survive the change of the attachment point. Problems related to the performance of TCP are discussed in chapter 9. Clearly, many of today's applications have not been designed for use in mobile environments, so the only effects of mobility should be a higher delay and lower bandwidth. However, there are some applications for which it is better to be 'mobility aware'. Examples are cost-based routing or video compression. Knowing that it is currently possible to use different networks, the software could choose the cheapest one. Or if a video application knows that only a low bandwidth connection is currently available, it could use a different compression scheme. Additional mechanisms are necessary to inform these applications about mobility .

 Scalability and efficiency: Introducing a new mechanism to the internet must not jeopardize its efficiency. Enhancing IP for mobility must not generate too many new messages flooding the whole network. Special care has to be taken considering the lower bandwidth of wireless links. Many mobile systems will have a wireless link to an attachment point, so only some additional packets should be necessary between a mobile system and a node in the network. Looking at the number of computers connected to the internet and at the growth rates of mobile communication, it is clear that myriad devices will participate in the internet as mobile components. Just think of cars, trucks, mobile phones, every seat in every plane around the world etc. – many of them will have some IP implementation inside and move between different networks and require mobile IP. It is crucial for a mobile IP to be scalable over a large number of participants in the whole internet, worldwide.

● **Security:** Mobility poses many security problems. The minimum requirement is that of all the messages related to the management of Mobile IP are authenticated. The IP layer must be sure that if it forwards a packet to a mobile host that this host receives the packet. The IP layer can only guarantee that the IP address of the receiver is correct. There are no ways of preventing fake IP addresses or other attacks. According to Internet philosophy, this is left to higher layers (keep the core of the internet simple, push more complex services to the edge).The goal of a mobile IP can be summarized as: 'supporting end-system mobility while maintaining scalability, efficiency, and compatibility in all respects with existing applications and Internet protocols'.

c) Entities and terminology

The following defines several entities and terms needed to understand mobile IP as defined in RFC 3344. Figure 3.1 illustrates an example scenario.

● **Mobile node (MN):** A mobile node is an end-system or router that can change its point of attachment to the internet using mobile IP. The MN keeps its IP address and can continuously communicate with any other system in the internet as long as link-layer connectivity is given. Mobile nodes are not necessarily small devices such as laptops with antennas or mobile phones; a router onboard an aircraft can be a powerful mobile node.

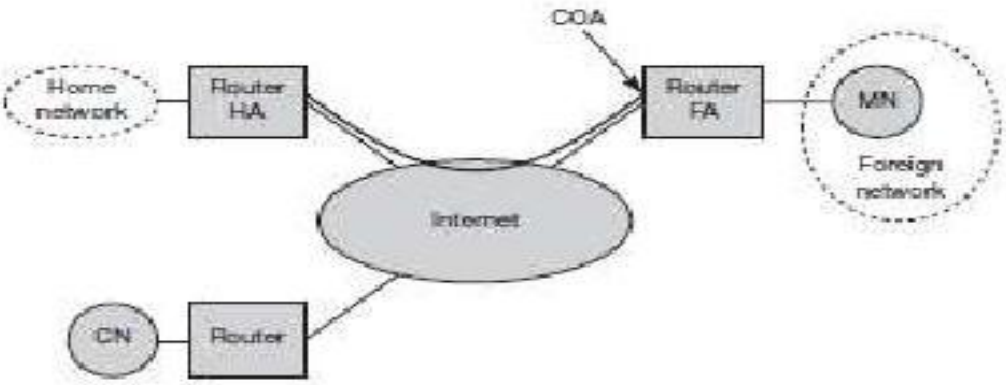


Fig 3.1 Entities and terminology

3.2 DYNAMIC HOST CONFIGURATION PROTOCOL

The dynamic host configuration protocol is mainly used to simplify the installation and maintenance of networked computers. If a new computer is connected to a network, DHCP can provide it with all the necessary information for full system integration into the network, e.g., addresses of a DNS server and the default router, the subnet mask, the domain name, and an IP address. Providing an IP address makes DHCP very attractive for mobile IP as a source of care-of-addresses. While the basic DHCP mechanisms are quite simple, many options are available as described in RFC 2132. DHCP is based on a client/server model as shown in Figure 8.17. DHCP clients send a request to a server (DHCPDISCOVER in the example) to which the server responds.

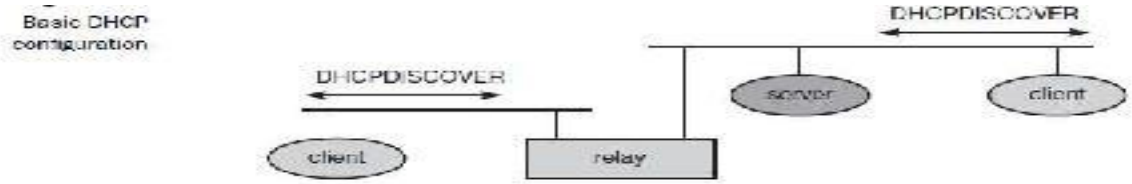


Fig 3.2 a) Client Initialization via DHCP

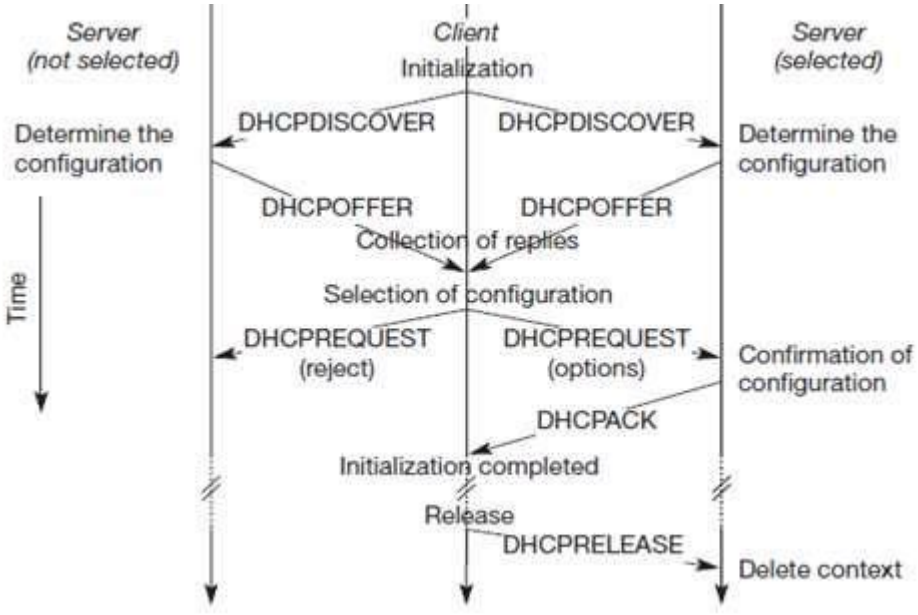


Fig 3.2 b) Dynamic Host Configuration Protocol

A typical initialization of a DHCP client is shown in the above Figure. The figure shows one client and two servers. As described above, the client broadcasts a DHCPDISCOVER into the subnet. There might be a relay to forward this broadcast. In the case shown, two servers receive this broadcast and determine the configuration they can offer to the client. One example for this could be the checking of available IP addresses and choosing one for the client. Servers reply to the client's request with DHCPOFFER and offer a list of configuration parameters. The client can now choose one of the configurations offered. The client in turn replies to the servers, accepting one of the configurations and rejecting the others using DHCPREQUEST. If a server receives a DHCPREQUEST with a rejection, it can free the reserved configuration for otherpossible clients. The server with the configuration accepted by the client now confirms the configuration with DHCPACK.

This completes the initialization phase. If a client leaves a subnet, it should release the configuration received by the server using DHCPRELEASE. Now the server can free the context stored for the client and offer the configuration again. The configuration a client gets from a server is only leased for a certain amount of time; it has to be reconfirmed from time to time. Otherwise the server will free the configuration. This timeout of configuration helps in the case of crashed nodes or nodes moved away without releasing the context. DHCP is a good candidate for supporting the acquisition of care-of addresses for mobile nodes. The same holds for all other parameters needed, such as addresses of the default router, DNS servers, the timeserver etc. A DHCP server should be located in the subnet of the access point of the mobile node, or at least a DHCP relay should provide forwarding of the messages. RFC 3118 specifies authentication for DHCP messages which is needed to protect mobile nodes from malicious DHCP servers. Without authentication, the mobile node cannot trust a DHCP server, and the DHCP server cannot trust the mobile node.

3.3 ADHOC– PROACTIVE AND REACTIVE ROUTING PROTOCOLS

0 ROUTING

Routing is the act of moving information across the network from a source to a destination. It is also referred as the process of choosing a path over which the packets are sent. The routing process usually directs forwarding on the basis of routing tables which maintain a record of the routes to various network destinations.

At least one intermediate node within the internetwork is encountered during the transfer of information. Basically two activities are involved in this concept: determining optimal routing paths and transferring the packets through an internetwork. The transferring of packets through an internetwork is called as packet switching which is straight forward, and the path determination could be very complex.

Routing protocols use several metrics as a standard measurement to calculate the best path for routing the packets to its destination that could be : **number of hops,** which are used by the routing algorithm to determine the optimal path for the packet to its destination. The process of path determination is that, routing algorithms find out and maintain routing tables, which contain the total route information for the packet. The information of route varies from one routing algorithm to another. The routing tables are filled with entries in the routing table are **ip-address prefix and the next hop**.

Routing is mainly classified into **static routing and dynamic routing**.

1. Static routing refers to the routing strategy being stated manually or statically, in the router. Static routing maintains a routing table usually written by a networks administrator. The routing table doesn‘t depend on the state of the network status, i.e., whether the destination is active or not.

2. Dynamic routing refers to the routing strategy that is being learnt by an interior or exterior routing protocol. This routing primarily depends on the state of the network i.e., the routing table is affected by the activeness of the destination.

b) **Routing in Mobile Ad-hoc Networks**

Mobile Ad-hoc networks are self-organizing and self-configuring multihop wireless networks, where the structure of the network changes dynamically. This is mainly due to the mobility of the nodes. Nodes in these networks utilize the same random access wireless channel, cooperating in an intimate manner to engaging themselves in multihop forwarding. The node in the network not only acts as hosts but also as routers that route data to / from other nodes in network.

In mobile ad-hoc networks there is no infrastructure support as is the case with wireless networks, and since a destination node might be out of range of a source node transferring packets; so there is need of a routing procedure. This is always ready to find a path so as to forward the packets appropriately between the source and the destination.

In infrastructure networks, within a cell, a base station can reach all mobile nodes without routing via broadcast in common wireless networks. In the case of ad-hoc networks, each node must be able to forward data for other nodes. This creates additional problems along with the problems of dynamic topology which is unpredictable connectivity changes.

c) Problems in routing with Mobile Ad hoc Networks

i) **Asymmetric links:** Most of the wired networks rely on the symmetric links which are always fixed. But this is not a case with ad-hoc networks as the nodes are mobile and constantly changing their position within network

ii) **Routing Overhead:** In wireless ad hoc networks, nodes often change their location within network. So, some stale routes are generated in the routing table which leads to unnecessary routing overhead.

iii) **Interference:** This is the major problem with mobile ad-hoc networks as links come and go depending on the transmission characteristics, one transmission might interfere with another one and node might overhear transmissions of other nodes and can corrupt the total transmission.

iv) **Dynamic Topology:** Since the topology is not constant; so the mobile node might move or medium characteristics might change. In ad-hoc networks, routing tables must somehow reflect these changes in topology and routing algorithms have to be adapted. For example in a fixed network routing table updating takes place for every 30sec. This updating frequency might be very low for ad-hoc networks.

d) **Classification of Routing Protocols**

Classification of routing protocols in mobile ad hoc network can be done in many ways, but most of these are done depending on routing strategy and network structure.

The routing protocols can be categorized as flat routing, hierarchical routing and geographic position assisted routing while depending on the network structure. According to the routing strategy routing protocols can be classified as Table-driven and source initiated. The classification of routing protocols is shown below.

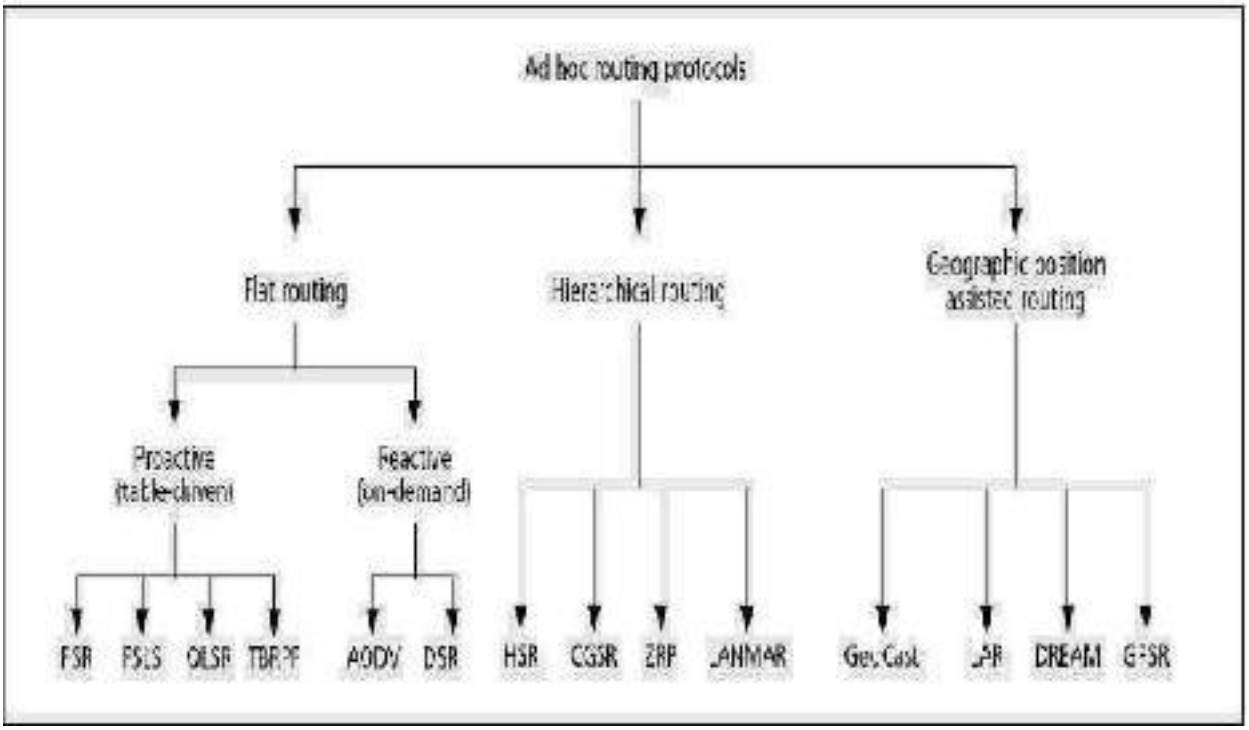


Fig 3.3 Classification of Routing Protocols

FSR – Fish Eye State Routing ;**FSLS** – Fuzzy Sighted Link state;

OLSR – Optimized Link State Routing; **DSR** – Dynamic Source Routing
TBRPF – Topology broadcast based on Reverse – Path Forwarding

AODV – Ad hoc On Demand Distance Vector; **HSR** – Hierarchical State Routing **CGSR** – Cluster Gateway Switch Rouing; **ZRP** – Zone Routing Protocol

LANMAR – Landmark Ad hoc Routing; **GeoCast** – Geographic Addressing and Routing
LAR – Location Aided Routing Protocol; **GPSR** – Greedy Perimeter Stateless Routing
DREAM – Distance Routing Effect Algorithm for mobility

Based on the Routing Information Update Mechanism :

Ad hoc wireless network routing protocols can be classified into three major categories based on the routing information update mechanism. They are

1. **Proactive or Table driven routing protocols** : In table drive routing protocols , every node maintains the network topology information in the form of routing tables by periodically exchanging routing network. Whenever a node requires a path to a

2. **Reactive or On-demand routing protocols** : Protocols that falls under this category do not maintain the network topology information. They obtain the necessary path when it is required by using connection establishment process. Hence these protocols do not exchange routing information periodically.
3. **Hybrid Routing Protocols** : Protocols belonging to this cateogory combine the best features of above two categories. Nodes within a certain distance from the node concerned or within a particular geographical region are said to be within the routing zone of the given node. For routing within this zone, a table- driven approach is used. For nodes that are located beyond this zone is on-demand approach is used

➢ **DESTINATION SEQUENCED DISTANCE VECTOR (DSDV)**

DSDV was one of the first proactive routing protocols available for Ad Hoc networks. It was developed by C. Perkins in 1994, 5 years before the informational RFC of the MANET group. It has not been standardised by any regulation authorities but is still a reference.

Algorithm

- DSDV is based on the Bellman-Ford algorithm.
- With DSDV, each routing table will contain all available destinations, with the associated next hop, the associated metric (numbers of hops), and a sequence number originated by the destination node.
- Tables are updated in the topology per exchange between nodes.
- Each node will broadcast to its neighbors entries in its table. This exchange of entries can be made by dumping the whole routing table, or by performing an incremental update, that means exchanging just recently updated routes.
- Nodes who receive this data can then update their tables if they received a better route, or a new one.
- Updates are performed on a regular basis, and are instantly scheduled if a new event is detected in the topology.
- If there are frequent changes in topology, full table exchange will be preferred whereas in a stable topology, incremental updates will cause less traffic.
- The route selection is performed on the metric and sequence number criteria. The sequence number is a time indication sent by the destination node. It allows the table update process, as if two identical routes are known, the one with the best sequence number is kept and used, while the other is destroyed (considered as a stale entry).

Illustration

Let us consider the two following topologies (figure 1 and figure 2). At t=0, the network is organized as shows figure 1. We suppose at this time the network is stable, each node has a correct routing table of all destinations.

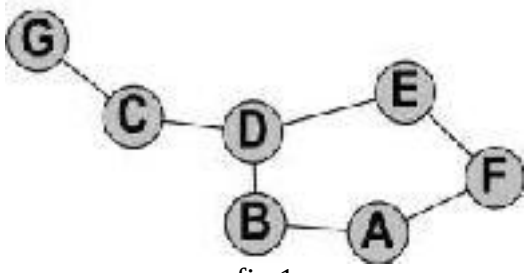


fig 1

Then, we suppose G is moving, and at t+1, the topology is as shown in figure 2.

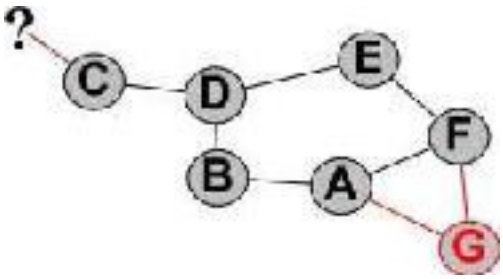


fig 2

At this stage, the following events are detected, and actions are taken:

- On node C: Link with G is broken, the route entry is deleted, and updates are sent to node D.
- On node A and F: A new link is detected, the new entry is added to the routing table and updates are sent to neighbors.
- On node G: Two new links are detected (to A and F), and one is broken (to C), the routing table is updated and a full dump is sent to neighbors (as the routing table is entirely changed, a full dump equals an incremental update).

Advantages

- DSDV was one of the early algorithms available. It is quite suitable for creating ad hoc networks with small number of nodes. Since no formal specification of this algorithm is present there is no commercial implementation of this algorithm.
- DSDV guarantees for loop free path.

Disadvantages

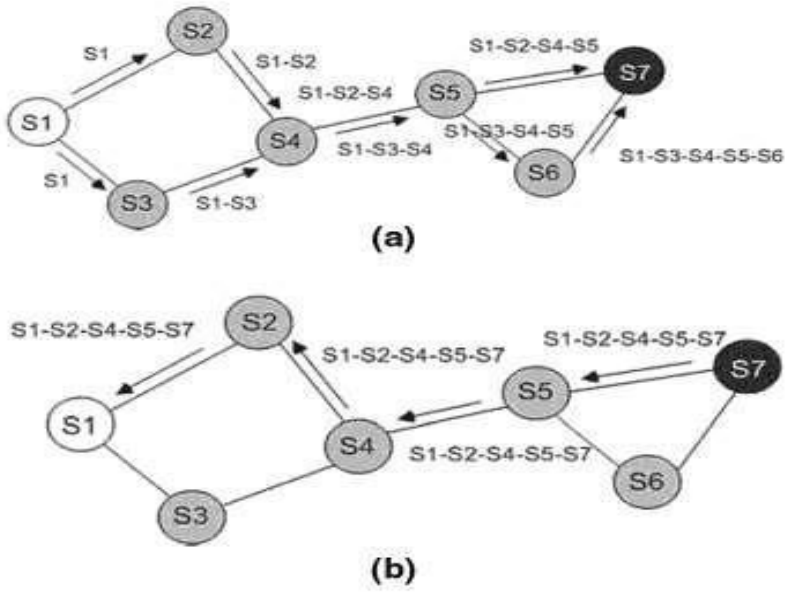
- DSDV requires a regular update of its routing tables, which uses up battery power and a small amount of bandwidth even when the network is idle.
- Whenever the topology of the network changes, a new sequence number is necessary before the network re-converges; thus, DSDV is not suitable for highly dynamic networks.

➢ **DYNAMIC SOURCE ROUTING (DSR)**

- The Dynamic Source Routing protocol (DSR) is a simple and efficient routing protocol designed specifically for use in multi-hop wireless ad hoc networks of mobile nodes.
- DSR allows the network to be completely self-organizing and self-configuring, without the need for any existing network infrastructure or administration.
- It is a reactive protocol and all aspects of the protocol operate entirely on-demand basis.
- It works on the concept of source routing. Source routing is a routing technique in which the sender of a packet determines the complete sequence of nodes through which, the packets are forwarded.
- The advantage of source routing is : intermediate nodes do not need to maintain up to date routing information in order to route the packets they forward.
- The protocol is composed of the two main mechanisms of "Route Discovery" and "Route Maintenance".
- DSR requires each node to maintain a route – cache of all known self – to – destination pairs. If a node has a packet to send, it attempts to use this cache to deliver the packet.
- If the destination does not exist in the cache, then a route discovery phase is initiated to discover a route to destination, by sending a route request.
- This request includes the destination address, source address and a unique identification number.
- If a route is available from the route – cache, but is not valid any more, a route maintenance procedure may be initiated.
- A node processes the route request packet only if it has not previously processes the packet and its address is not present in the route cache.
- A route reply is generated by the destination or by any of the intermediate nodes when it knows about how to reach the destination.

Example

In the following example, the route discovery procedure is shown where S1 is the source node and S7 is the destination node.



(a) Route Discovery **(b)** Using route record to send the route reply

In this example, the destination S7, gets the request through two paths. It chooses one path based on the route records in the incoming packet and sends a reply using the reverse path to the source node. At each hop, the best route with minimum hop is stored. In this example, it is shown the route record status ate each hop to reach the destination from the source node. Here, the chosen route is S1-S2-S4-S5-S7.

Advantages and Disadvantages:

a) DSR uses a reactive approach which eliminates the need to periodically flood the network with table update messages which are required in a table-driven approach. The intermediate nodes also utilize the route cache information efficiently to reduce the control overhead.

b) The disadvantage of DSR is that the route maintenance mechanism does not locally repair a broken down link. The connection setup delay is higher than in table-driven protocols. Even though the protocol performs well in static and low-mobility environments, the performance degrades rapidly with increasing mobility. Also, considerable routing overhead is involved due to the source-routing mechanism employed in DSR. This routing overhead is directly proportional to the path length.

➢ **AD HOC ON-DEMAND DISTANCE VECTOR (AODV)**

AODV was proposed to standardization by the RFC 3561 in July 2003. It was designed by the same people who designed DSDV. AODV is a distance vector routing protocol, which means routing decisions will be taken depending on the number of hops to destination. A particularity of this network is to support both multicast and unicast routing.

Algorithm

- The AODV algorithm is inspired from the Bellman-Ford algorithm like DSDV.
- The principle change is to be **On Demand.**
- The node will be silent while it does not have data to send. Then, if the upper layer is requesting a route for a packet, a "ROUTE REQUEST" packet will be sent to the direct neighborhood. If a neighbour has a route corresponding to the request, a packet "ROUTE REPLY" will be returned.
- Otherwise, each neighbour will forward the "ROUTE REQUEST" to their own neighborhood, except for the originator and increment the hop value in the packet data.
- They also use this packet for building a reverse route entry (to the originator). This process occurs until a route has been found.
- Another part of this algorithm is the **route maintenance**.
- While a neighbour is no longer available, if it was a hop for a route, this route is not valid anymore.
- AODV uses "HELLO" packets on a regular basis to check if they are active neighbours.

  Active neighbours are the ones used during a previous route discovery process. If there is no response to the "HELLO" packet sent to a node, then, the originator deletes all associated routes in its routing table. "HELLO" packets are similar to ping requests.

  While transmitting, if a link is broken (a station did not receive acknowledgment from the

layer 2), a “ROUTE ERROR” packet is unicast to all previous forwarders and to the sender of the packet.

Illustration

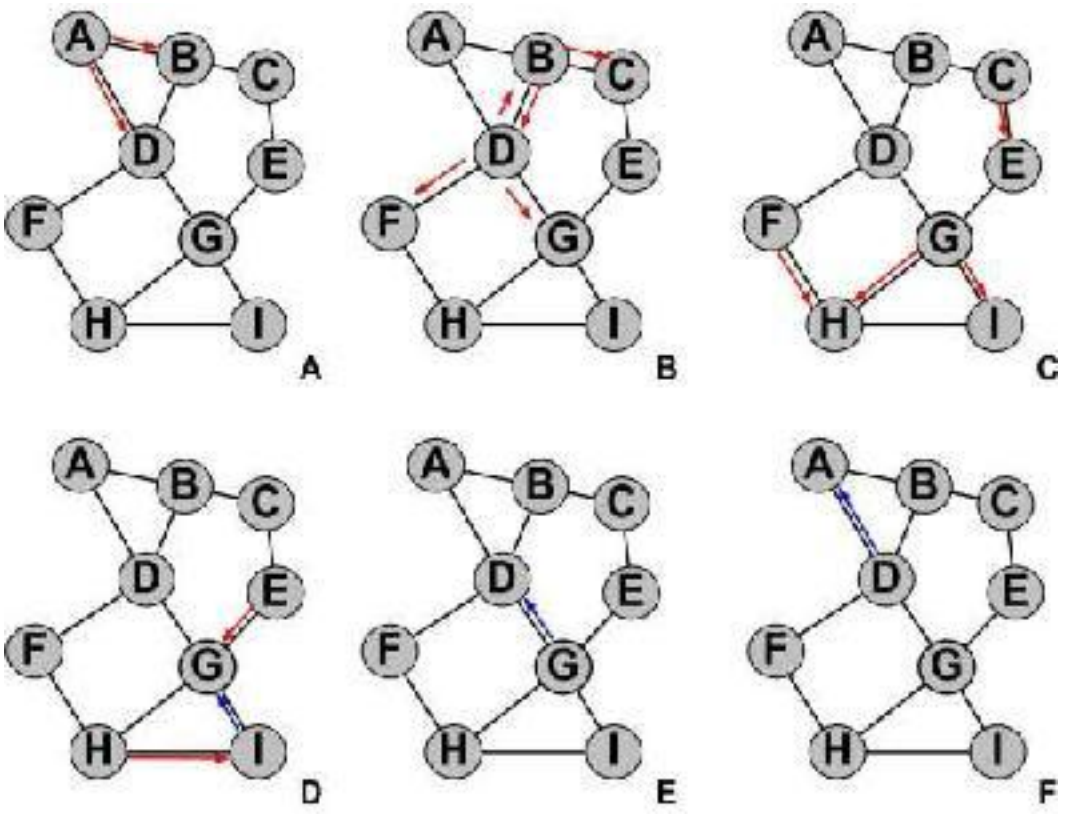


fig 1

In the example illustrated by figure 1, A needs to send a packet to I. A “ROUTE REQUEST” packet will be generated and sent to B and D (a). B and D add A in their routing table, as a reverse route, and forward the “ROUTE REQUEST” packet to their neighbours (b). B and D ignored the packet they exchanged each others (as duplicates). The forwarding process continues while no route is known (c). Once I receives the “ROUTE REQUEST” from G (d), it generates the “ROUTE REPLY” packet and sends it to the node it received from. Duplicate packets continue to be ignored while the “ROUTE REPLY” packet goes on the shortest way to A, using previously established reverse routes (e and f).

The reverse routes created by the other nodes that have not been used for the “ROUTE REPLY” are deleted after a delay. G and D will add the route to I once they receive the “ROUTE REPLY” packet.

Characteristics of AODV

- Unicast, Broadcast, and Multicast communication.
- On-demand route establishment with small delay.

- All routes are loop-free through use of sequence numbers.
- Use of Sequence numbers to track accuracy of information.
- Only keeps track of next hop for a route instead of the entire route.
- Use of periodic HELLO messages to track neighbors.

Advantages and Disadvantages

The main advantage of this protocol is that routes are established on demand and destination sequence numbers are used to find the latest route to the destination. The connection setup delay is lower.

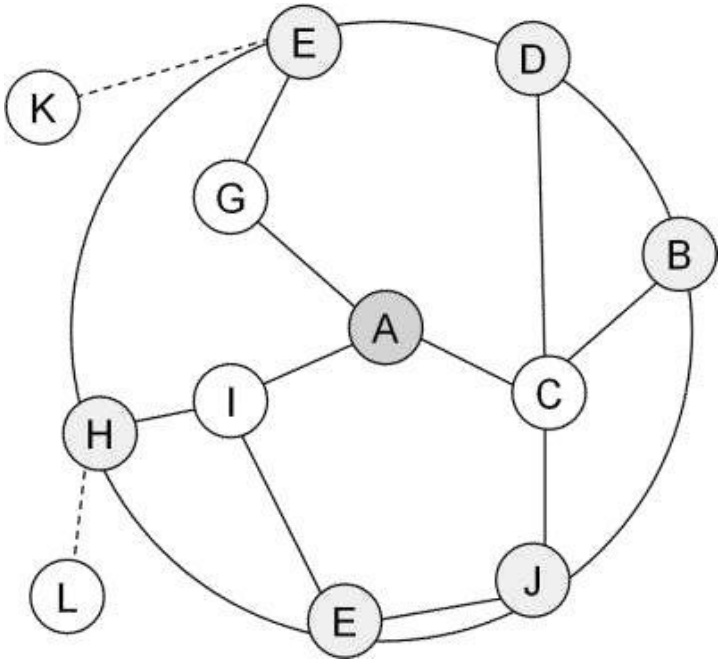
One of the disadvantages of this protocol is that intermediate nodes can lead to inconsistent routes if the source sequence number is very old and the intermediate nodes have a higher but not the latest destination sequence number, thereby having stale entries. Also multiple RouteReply packets in response to a single RouteRequest packet can lead to heavy control overhead.

Another disadvantage of AODV is that the periodic beaconing leads to unnecessary bandwidth consumption

➢ ZONE ROUTING PROTOCOL (ZRP)

- The Zone Routing Protocol (ZRP) was introduced in 1997 by Haas and Pearlman.
- It is either a proactive or reactive protocol. **It is a hybrid routing protocol.**
- It combines the advantages from proactive and reactive routing.
- It takes the advantage of pro-active discovery within a node's local neighborhood (**Intra zone Routing Protocol (IARP)**), and using a reactive protocol for communication between these neighborhoods (**Inter zone Routing Protocol(IERP)**).
- The **Broadcast Resolution Protocol (BRP)** is responsible for the forwarding of a route request.
- ZRP divides its network in different zones.
- Each node may be within multiple overlapping zones, and each zone may be of a different size.
- The size of a zone is not determined by geographical measurement. It is given by a radius of length, where the number of hops is the perimeter of the zone. Each node has its own zone.

Example



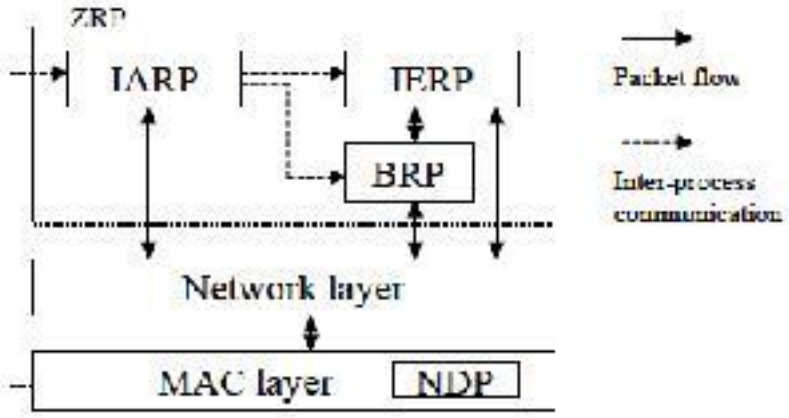
- radius=2-Hop; E, D, B, J, E and H are border-nodes

Illustration

- Before constructing a zone and determine border nodes , a node needs to know about its local neighbors.
- A node may use the media access control (MAC) protocols to learn about its direct neighbors. It also may require a **Neighbor Discovery Protocol (NDP).**
- ZRP does not strictly specify the protocol used but allows for local independent implementations.
- NDP relies on the transmission of hello messages by each node.
- When the node, for example node A, gets a response from a node B which has received the "Hello"-messages, the node A notice that it has a direct point-to-point connection with that node B.
- The NDP selects nodes on various criteria, e.g.:
    1. signal strength
    2. frequency/delay of beacons.
- Once the local routing information has been collected, the node periodically broadcasts discovery messages in order to keep its map of neighbours up to date.
- Sometimes the MAC layer of the nodes does not allow for such a NDP.

Components of ZRP

- The Zone Routing Protocol consists of several components, which only together provide the full routing benefit to ZRP
- Even though the hybrid nature of the ZRP seems to indicate that it is a hierarchical protocol, it is important to point out that the ZRP is in fact a flat protocol.
- ZRP is more efficiency for large networks



ZRP Components / Architecture

Advantage

6. Less control overhead as in a proactive protocol or an on demand protocol

Disadvantages

10. Short latency for finding new routes

➢ **ON DEMAND MULTICAST ROUTING PROTOCOL (ODMR)**

6. On- Demand Multicast routing protocol is a mesh architecture protocol, i.e, it has multiple paths from the sender to the receivers and uses a forwarding group concept.
7. It applies on-demand procedures to dynamically build route and maintain multicast group membership.

6. By maintaining a mesh instead of a tree, the drawbacks of multicast trees in ad hoc networks like frequent tree reconfiguration and non-shortest path in a shared tree are avoided.
7. In ODMRP, group membership and multicast routes are established by the source on demand when a multicast source has packets to send, but no route to the multicast group, it broadcasts a Join-Query control packets to the entire network.
8. This control packet is periodically broadcast to refresh the membership information and

updates routes.

9. When the Join-Query packet reaches a multicast receiver, it creates and broadcasts Join-Reply to its neighbours. When it has been received by the node, it checks if the next hop own id.
10. If it is does, the node realizes that it is on the path to the source and becomes the part of the forwarding group by setting the FG_FLAG (Forwarding Group flag).
11. When receiving a multicast data packet, a node forwards it only when it is not a duplicate, hence minimizing traffic overhead. Because the nodes maintain soft state, finding the optimal flooding interval is critical to ODMRP performance.
12. ODMRP uses location and movement information to predict the duration of time that routes will remain valid. With the predicted time of route disconnection, a "join data" packet is flooded when route breaks of ongoing data sessions are imminent.
13. It reveals that ODMRP is better suited for ad hoc networks in terms of bandwidth utilization

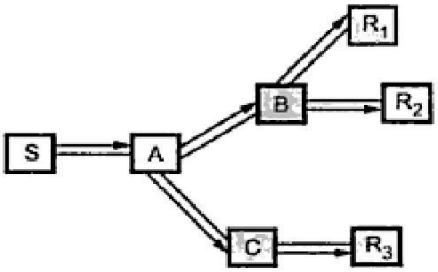
.

Example

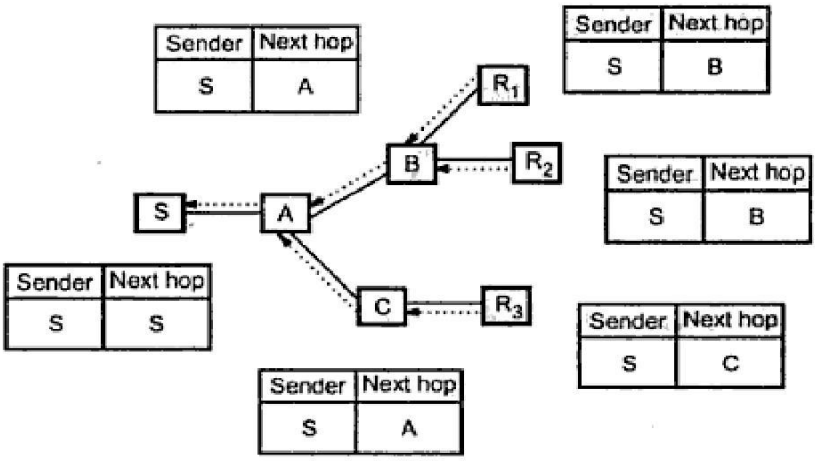
Consider the source node 'S'. It will flood the JOIN_DATA packets to all other nodes in the network. When a host node receives first JOIN_DATA packet it will rebroadcast it to form a reverse path with the previous host. Each host in the network acts as multicast receiver. It receives JOIN_DATA packet and replies in turn with a JOIN_TABLE packet to the upstream to establish reverse paths.

The process repeats until source host 'S' is reached. The method of packet forwarding is for Figure (a) is shown in Figure (b). As the JOIN_TABLE is received a host has to build a multicast table so as to facilitate future packet forwarding. For example the host B receives the R_1's JOIN_TABLE which is shown in the diagram.

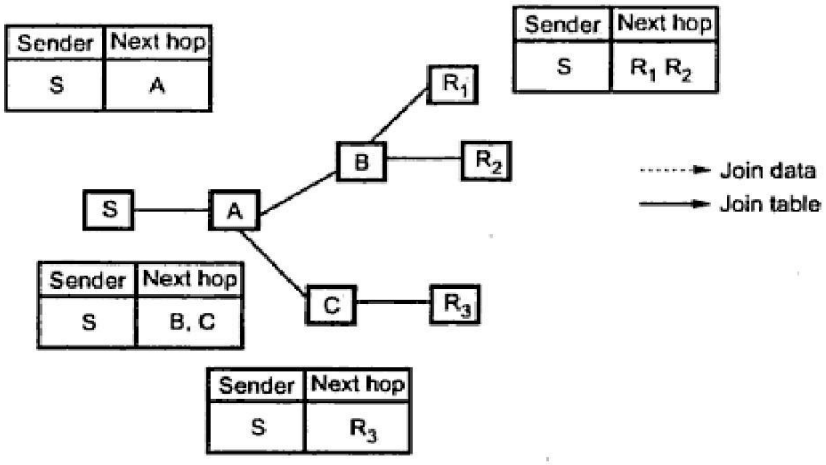
It will add R_1 as its next hop step. Assume B receives R_2's JOIN_TABLE. Now it will add R_2 as its next hop step. A simple final multicast table for each host is shown in the Figure (c) in propagation of data packets.



(a) Load_data packets propagation



(b) Load_table packets propagation



(c) Last multicast table

Advantages

1. Low channel and storage overhead

3. Usage of up-to-date shortest routes
4. Robustness to host mobility
5. Maintenance and exploitation of multiple redundant paths
6. Exploitation of the broadcast nature of the wireless environment
7. Unicast routing capability

Disadvantages

0 The main disadvantage of ODMRP is its excessive overhead, because broadcasting of the reply packets to many nodes.

1 It has a complex topology.

3.4 **MULTICAST IN THE INTERNET**

Some applications require data to be delivered from a sender to multiple receivers. Examples of such applications include audio and video broadcasts, real-time delivery of stock quotes, and teleconferencing applications. A service where data is delivered from a sender to multiple receivers is called multipoint communication or multicast, and applications that involve a multicast delivery service are called multicast applications.

Bellow Figure compares multicast to other communication paradigms. In unicast or point-to-point communication, data is sent to a single host. In broadcast or one-to-all communications, data is transmitted to all hosts with respect to a given scope, for example, all hosts in a LAN. Multicast can be thought of as a generalization of unicast and broadcast. In multicast, data is transmitted to a set of hosts that have indicated interest in receiving the data, referred to as a multicast group or host group.

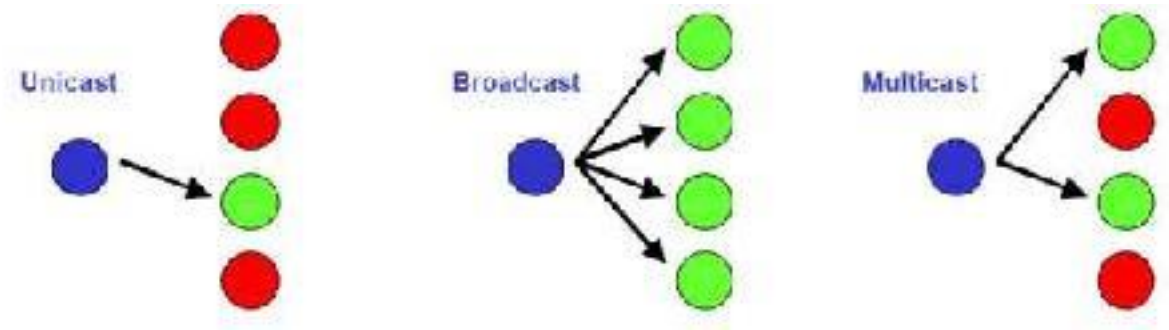


Fig 3.4 Multicast in Internet

In principle, it is feasible to implement multicast in a network using either unicast or broadcast. However, both solutions have shortcomings. In a unicast solution to multicast, the sender transmits one copy of the data separately to each host in the multicast group. This is viable for small multicast groups, but when the number of hosts is large, transmitting the same data multiple times wastes a lot of resources. In a broadcast solution to multicast, data is delivered to all hosts in a network; for example, and hosts drop the data if they are not members of the multicast group. This solutions works when the hosts of a multicast group are all located

on the same LAN and the LAN supports broadcast transmissions. Otherwise, sending data to a large number of hosts just to have it dropped by most hosts is not an economical use of network capacity.

Making multicast delivery efficient in a packet switching network requires a whole set of new protocols and mechanisms at the network layer. First, multicast addresses must be available that can designate a multicast group as the destination of a datagram. Second, there must be mechanisms that allow a host to join and leave a multicast group. Third, there is a need for multicast routing protocols that set up paths, called distribution tree, from the sender to the members of a multicast group. The issues related to setting up multicast distribution trees are referred to as multicast routing.

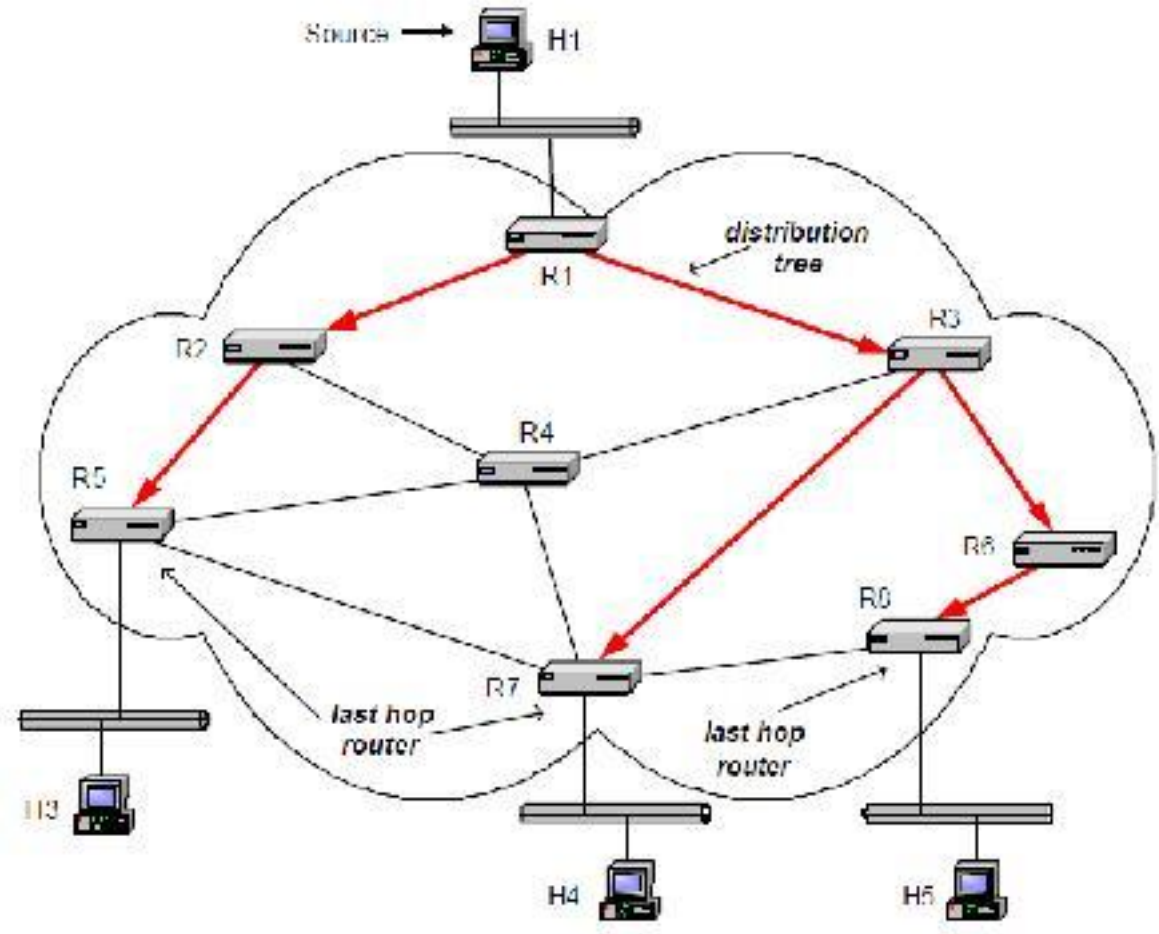


Fig 3.5 Multicast Delivery

The network-layer mechanisms in the Internet that support multicast are referred to as IP multicast. Figure depicts an example of multicast in an IP network. The figure shows four hosts and eight routers. Routers are connected by point-to-point links and hosts connect to routers via Ethernet LANs. Host H1 is a source of multicast data, and hosts H2, H3 and H4 are Multicast receivers. The distribution tree, indicated with arrows, is established by the routers using a multicast routing protocol. Data is delivered downstream in the distribution tree from the

source to the receivers.

IP multicast involves both hosts and routers. In IPv4, support of IP mult icast is optional, but almost all hosts and most routers support multicast. Hosts that are members of a multicast group exchange Internet Group Management Protocol (IGMP) messages with routers. Routers perform two main processes in IP multicast: multicast routing and multicast forwarding. Multicast routing sets up the distribution tree for a multicast group by setting the content of multicast routing tables. In multicast, a routing table may list multiple next hop addresses for a routing table entry. As in unicast, forwarding refers to the processing of an incoming datagram, the routing table lookup, and the transmission on an outgoing interface. When a multicast packet arrives at a router, the router performs a lookup in the multicast routing table for a matching entry. The router forwards one copy of the packet to each next hop address in the matching routing table entry.

3.4.1 MULTICAST ROUTING

Multicast routing is concerned with setting up distribution trees that provide a path from the multicast sources to multicast group members. This section discusses the objectives of multicast routing, and the protocol mechanisms used in routing protocols.

a) Multicast Routing Algorithms in a Graph

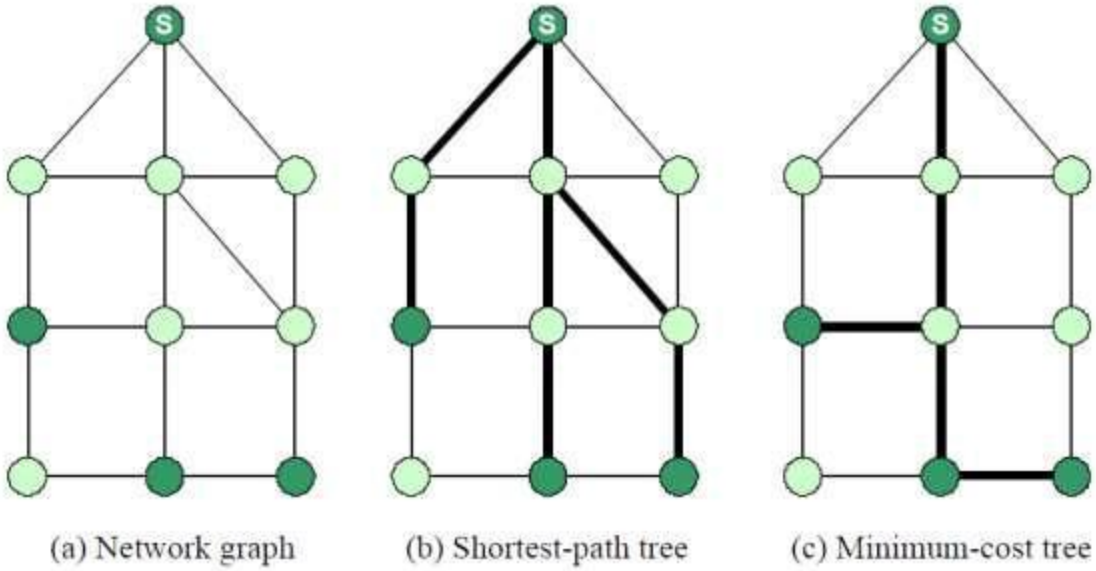


Fig 3.6 Multicast Routing Graph

In this graph, the task of multicast routing is the embedding of a tree into the graph such that all multicast group members are connected by the tree. So, how would an ideal embedding look like? One can think immediately of two objectives to build a good tree, which are shown. In the figures, the thick lines indicate the distribution trees. The first objective is to build a tree that minimizes the path cost from the source to each receiver. Such a tree is called a shortest-path tree or source-based tree. A shortest-path tree can be relatively quickly computed using a shortest-path algorithm, but has the drawback that the tree is dependent on the source of the multicast tree. Note in figure, that the shortest-path tree is different when a different node is

the source in the multicast group. Therefore, a distribution tree must be computed separately for each source. The second objective is to build a tree that minimizes the total cost of the edges, called a minimum-cost tree.

Different from a shortest-path tree, a minimum-cost tree does not change when a different node becomes the source. Thus, the same tree can be shared by all sources. The main drawback of a minimum-cost tree is that its computation is prohibitively expensive in most cases. In fact, the problem of calculating a minimum-cost tree is known to be an NP-complete problem, meaning that the computation of the tree is intractable unless the network is small. The two objectives for multicast routing, shortest-path tree and minimum-cost tree, represent a set of trade-off for multicast routing. Shortest-path trees minimize the cost of each receiver, whereas the minimum-cost tree minimizes the cost of the total tree. A single minimum-cost tree can be shared by all senders, whereas, a shortest-path trees must be built for each source. Lastly, a shortest-path tree can be computed relatively quickly, whereas the computation of a minimum cost-tree is not tractable in large networks. Some of the above trade-offs of multicast routing are reflected in the multicast routing protocols for the Internet. However, routing protocols have to satisfy additional constraints, which are not expressed in the formulation of a graph problem. For example, routing protocols must be able to adapt the distribution tree quickly when hosts join and leave a multicast group. Additionally, most multicast routing protocols require computing the distribution tree in a distributed fashion without any central coordination. Finally, most multicast routing protocols try to leverage off unicast routing protocols, by constructing the distribution tree using information from the unicast routing tables. This imposes additional constraints on a multicast routing protocol.

b) Reverse Path Forwarding

The majority of routing protocols in the Internet built either source-based trees or core-based trees. A source-based tree is essentially a shortest-path tree, with a multicast source at the root of the distributions tree. With source-based trees, one distribution tree must be built for each multicast source. Protocols that built core-based trees avoid the need for multiple distribution trees by building a single distribution tree that is shared by all sources. Core-based tree routing protocols do not attempt to construct a minimum-cost tree. Instead, one router is selected as core or rendezvous-point, and a shortest-path tree is constructed with the core router as the root of the tree. Thus, both source-based trees and core-based trees build shortest path trees. For reasons that will become clear in a moment, routing protocols generally minimize the paths from the receivers to the source, as opposed to minimizing the path from the source to the receiver.

A reverse shortest path tree can be built by applying a concept called reverse path forwarding (RPF). RPF is a mechanism to build a shortest path tree in a distributed fashion by taking of the unicast routing tables. The idea of RPF is simple. Given the address of the root of the tree, a router selects as its upstream neighbor in the tree of the router which is the next-hop neighbor for forwarding unicast packets to the root. The network interface which is used to reach this upstream neighbor is called the RPF interface and the neighbor router is called the RPF neighbor. An illustration of the basic concept of reverse path forwarding is given in Figure 10.12. Here, host H1 is the root of the distribution tree. Router R3 determines that R2 is its RPF neighbor for H1 from a lookup in its routing table. The interface that connects to R2 is the RPF interface.

If all routers determine their upstream neighbor using reverse path forwarding, the resulting distribution tree is such that packets from the root to a receiver are forwarded on the reverse path taken by unicast packets sent from the receiver to the root. This is where the name

reverse path forwarding comes from. Since unicast routing tables are set so that the path from the receiver to the root is minimal, the tree generated by reverse path forwarding is a reverse shortest path. Reverse path forwarding is applied to construct source-based, as well as core-based trees. In the former case, the root of the tree is a multicast source. In the latter case, the root of the tree is the core.

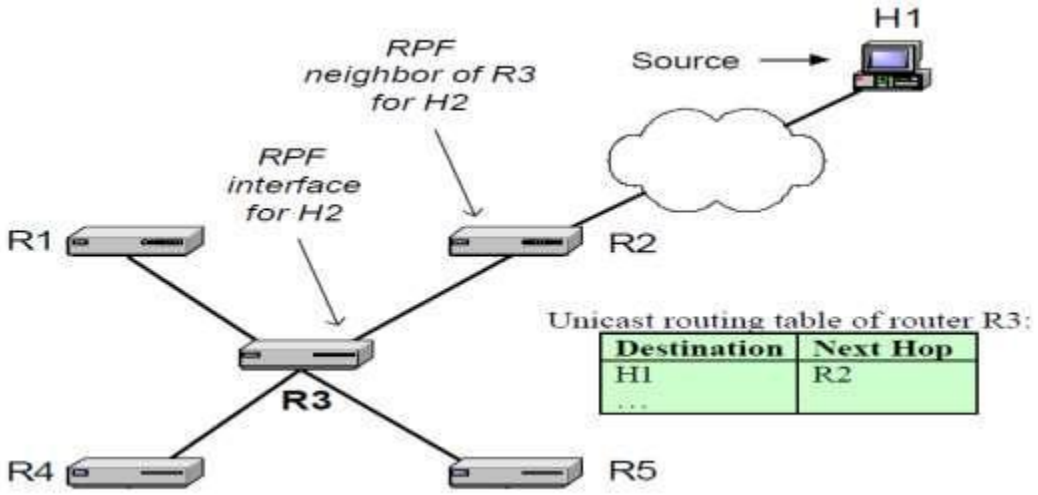


Fig 3.7 Multicast Routing

The general structure of a multicast routing table is shown in Figure 10.13. There are columns for a source IP address, a multicast address, an incoming interface and a list of outgoing interfaces. Routing table entries for source-based trees and for core-based trees are different. In a source based tree, a routing table entry contains a source address as well as a group address. This is often called a (Source, Group) or (S, G) entry. Since a source-based tree has one distribution tree for each source there may be multiple routing entries for the same multicast group. Routing table entries for a core-based tree do not list a source IP address, which is indicated with a '*' (star) in the source IP address column. The corresponding entry is called a (*, G) entry. In an (S, G) entry, the incoming interface entry is the RPF interface for address S. In an (*, G) entry, the incoming interface is set to the RPF interface with respect to the core.

An arriving multicast packet must match the source address and the group address in an (S, G) entry, and the group address in an (*, G) entry. If there are matches for both (S, G) and (*, G) entries, the (S, G) entries take preference. When a match is found in the routing table, the router verifies that the packet arrived on the incoming interface listed in the incoming interface column of the routing table. This is called an RPF check. If an RPF check is successful, one copy of the datagram is forwarded on each interface listed in the outgoing interface list. If there is no match in the routing table or if the RPF check failed, the packet is discarded.

Source IP address	Multicast group	Incoming interface (RPF interface)	Outgoing interface list
S1	G1	I1	I2, I3
*	G2	I2	I1, I3

Fig 3.8 General Structure of Multicast Routing Table

Source-based trees

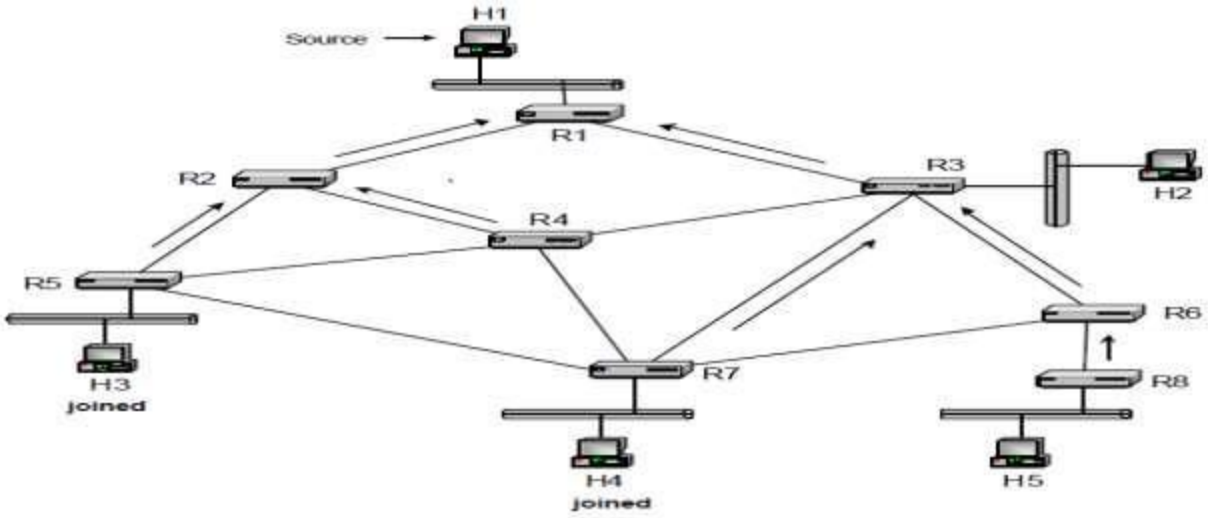


Fig 3.9 Multicast Group with RPF forwarding tree

We illustrate the construction of a source-based tree for the network shown in Figure. Each host is connected to a router via an Ethernet network and routers are connected by point to point links. H1 is the source of a multicast group and hosts H3 and H4 are receivers that have joined the multicast group. The arrows in the figure indicate the reverse shortest paths for all routers in the network. Data transmissions in the source-based distribution tree for H1 follow the opposite direction of the reverse shortest paths. The paths for H3 and H4 are H1à R1à R2 à R5 à H3 and H1à R1à R3 à R7 à H4, respectively.

Since the RPF interface tells each router the upstream neighbor in the distribution tree, but not the downstream neighbors in a distribution tree, additional protocol mechanisms are needed to determine the outgoing interfaces in the multicast routing tables. One method to achieve this is flood-and-prune, which starts out by forwarding multicast packets on all its interfaces, and then deletes interfaces which are not part of the distribution tree. Another method, called explicit join, requires that multicast receivers initiate the process of getting connected to the distribution tree. We will describe both methods.

In flood-and-prune, the outgoing interface list in a routing table entry initially lists all interfaces other than the RPF interface. When a router receives a multicast packet from source address S for a multicast group G, on the RPF interface for source S, it forwards the packet on all

interfaces, with exception of the RPF interface. When a router receives a packet on an interface that is different from the RPF interface for S, it discards the packet. Figure 10.15 shows how multicast packets from H1 are forwarded with this method. Arrows indicate the transmission of packets and crosses show where packets are discarded. Figure 10.15 show that due the flooding of messages, many routers receive multiple copies of the same packet. All but the copy arriving on the RPF interface are discarded. A router connected to a LAN, forwards a datagram to the LAN only if some hosts on the LAN are multicast group members. In Figure 10.15, these are the LANs where the multicast group members H3 and H4 reside. A router knows about group members on a LAN from IGMP messages.

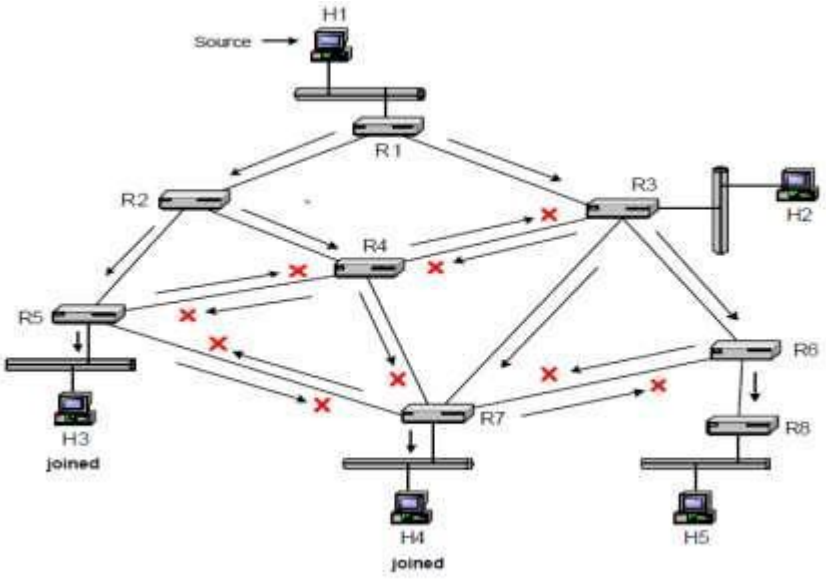


Fig 3.10 Flood and Prune

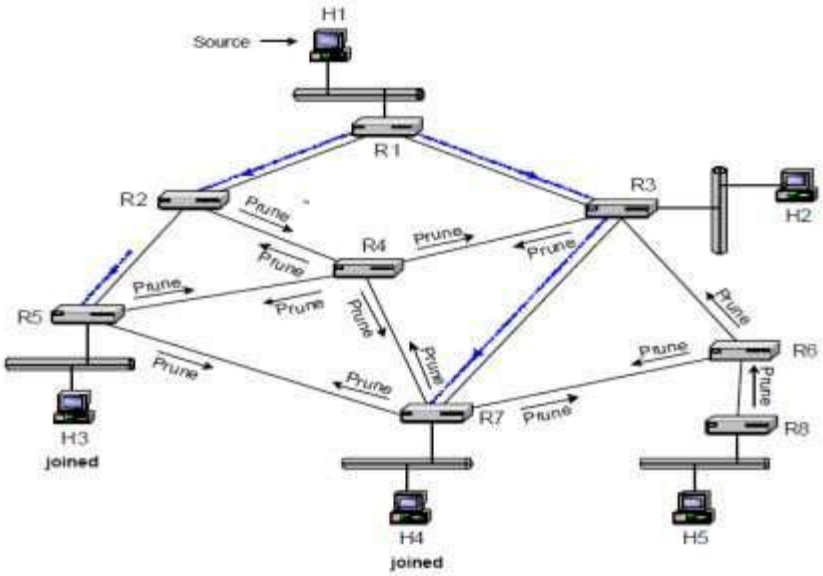


Fig 3.11 Prune Messages

In Figure when hosts H3 and H4 want to receive multicast packets from source S for group G, the connected routers, R5 and R7, send a join message to the router on their RPF interfaces. When a router receives a join message, it adds the interface where the join message was received to the list of outgoing interfaces of the (S, G) routing table entry. If an (S, G) routing entry does not exist, it will be created when the join message arrives. In Figure 10.18, when R2 receives a join message from R5, it adds the interface that connects to R5 to the outgoing interface list. As with graft messages, a router forwards a join message on its RPF interface only if the router is not part of the distribution tree.

When a router in the distribution tree receives a join message, it does not forward the join message. This method of explicit join messages avoids the transmission of duplicate packets as in flood-and-prune. On the other hand, building source-based trees with explicit join messages assume that routers know the identity of hosts that transmit to a multicast group.

Routing protocols that built core-based trees designate a router, called the core, and built a reverse shortest path tree with the core as the root of the tree. The core becomes the central hub for disseminating multicast packets sent to the group. When a source transmits a packet to a multicast group, the packet is sent to the core. When the packet reaches the core, it is forwarded using the reverse shortest path tree.

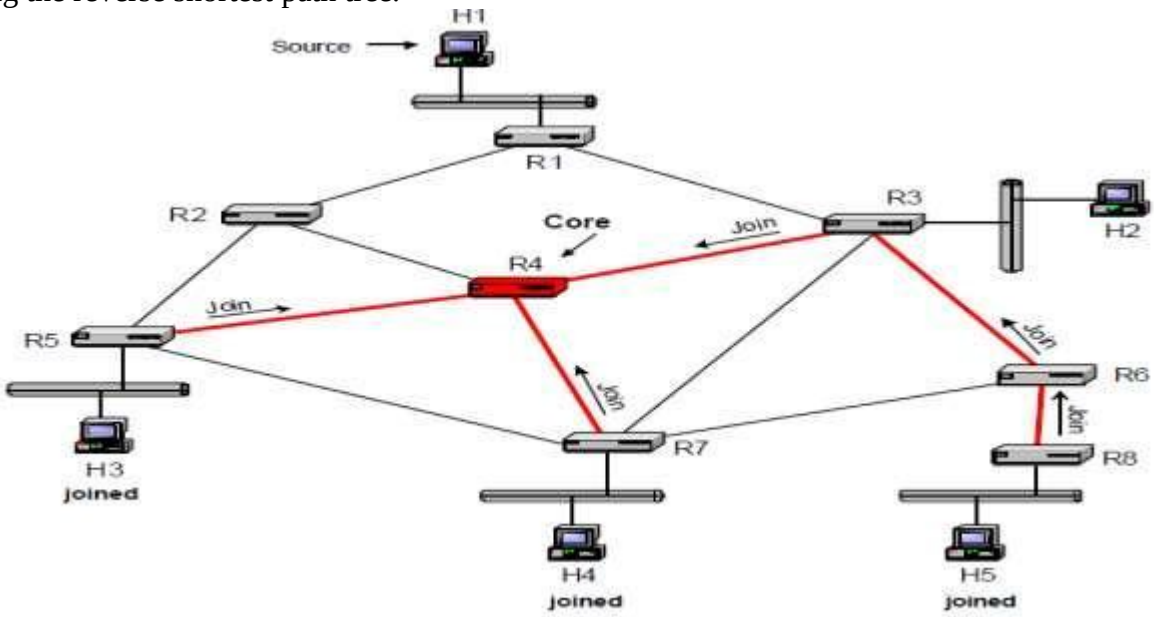


Fig 3.12 Construction of Core Based Trees

The construction of a core-based tree for a multicast group G is illustrated in Figure. Each receiver that joins a multicast group sends a join message to the core router of the group. The message is sent on the RPF interface with respect to the core. If the join message reaches a router that is not part of the tree, the router adds an (*, G) entry to the multicast routing table, adds the interface where the join message arrived to the outgoing interface list, and sets the incoming interface to the RPF interface. Then, the router forwards the join message on its RPF interface in the direction of the core. If the router is already part of the shared tree for group G,

the router, upon receiving a join message, adds the interface where the join message arrived to the outgoing interface list of the corresponding (*, G) entry, but does not forward the join message. Since join messages are sent on the reverse shortest path, the resulting distribution tree is a reverse shortest-path tree rooted at the core. The Figure illustrates the transmission of join messages with router R4 as core, and hosts H3, H4, and H5 has members of the multicast group.

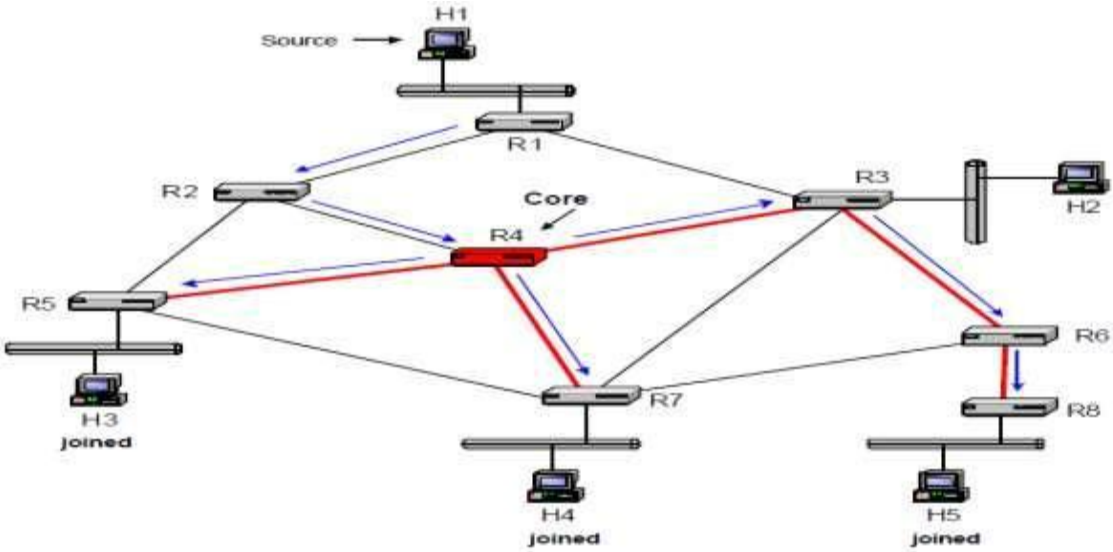


Fig 3.13 Multicast Forwarding in core based tree

Forwarding of multicast packets in core-based trees is illustrated in Figure. When H1 transmits a multicast packet, the packet is forwarded on the unicast shortest path to the core. A possible route is H1à R2 à R4. When the packet reaches the tree at the core, it is sent downstream in the core-based tree. Sending packets always to the core and inserting packets into the distribution tree only at the core can be inefficient. For example, suppose the path from R1 to R4 is through R3. Then, a packet from H1 to H5 takes the route H1à R1 à R3 à R4 à R3 à à R6 à R8. Here, it would be better to have R3, instead of forwarding the packet to the core, forward the datagram using the core-based tree. R3 should forward the datagram downstream the tree to R6, as well as upstream the tree to R4. However, such bidirectional trees, where packets are forwarded both upstream and downstream a core-based tree, are prone to routing loops, and special precautions are necessary to avoid such loops.

The main advantage of using core-based trees is that only one distribution tree is required for each multicast group. This reduces the complexity of the routing tables, as well as the volume of multicast routing protocol messages. The main disadvantage of core-based trees is that, dependent on the placement of the core, the paths from the sender to the receivers through the core can be much longer than the direct path between a source and a receiver.

Overview of Multicast Routing Protocols

Here is a brief overview of the multicast routing protocols that have been developed for the Internet. **Distance Vector Multicast Routing Protocol (DVMRP):5** DVMRP was the first multicast routing protocol developed for the Internet. DVMRP can operate in an environment where some, but not all routers in the network are capable of multicast forwarding and routing. This is achieved by having DVMRP run a separate unicast routing algorithm, similar to RIP, to determine the shortest-paths between all multicast-capable routers. DVMRP uses flood-and prune to set up source-based trees. DVMRP messages are encapsulated in IGMP messages, where the type field is set to 3.

DVMRP played an important role in the early deployment of IP multicast. IP multicast deployment in the Internet began in the early 1990s with the creation of the Multicast Backbone (MBONE). The multicast routing algorithm in the MBONE is DVMRP. The MBONE solved the problem of wide-area IP multicast routing on the Internet where only few routers were capable of IP multicast routing, by setting up a virtual network of multicast routers that are connected by unicast path. These multicast routers exchanged multicast IP datagram that were encapsulated in IP unicast datagrams, using the IP-in-IP option in the IP header.

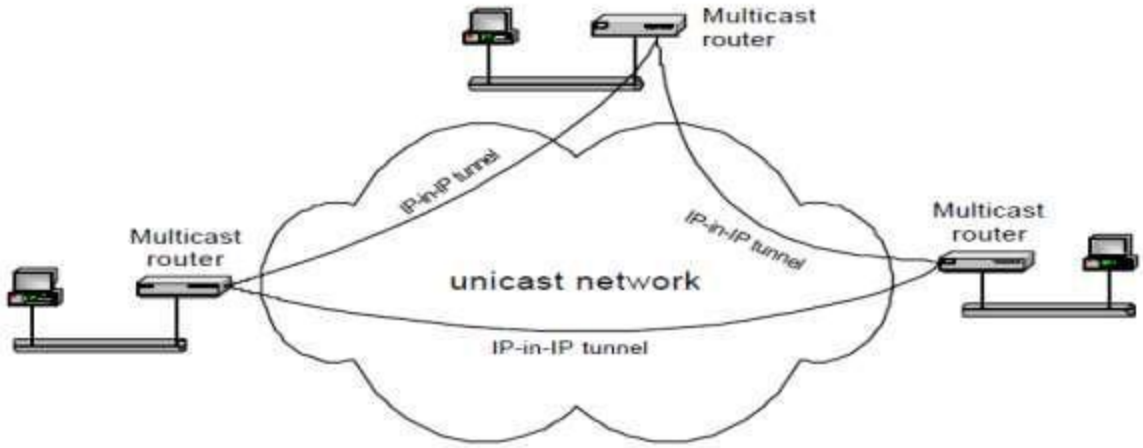


Fig 3.14 Tunnel based topology in MBONE

As a result of the encapsulation, the MBONE is a virtual network, where each link between two multicast routers consists of a complete unicast path. As more and more routers provide native support for IP multicast, meaning that they are capable of forwarding IP multicast traffic and running a multicast routing protocol, the need for a virtual multicast network has all but disappeared. **Multicast Open Shortest Path First (MOSPF):6** MOSPF consists of multicast extensions to the unicast routing protocol OSPF, and requires that OSPF is used for unicast routing. In MOSPF, multicast routers broadcast link state advertisements (LSAs) to all other multicast routers. Then, as in unicast OSPF, each multicast router calculates routes independently.

MOSPF computes shortest-path trees for each sender in the multicast group. A router computes a shortest-path tree for a source only if there is traffic from that sender. **Core Based Tree** (**CBT)**:7 CBT was the first routing protocol for the Internet that took a corebased tree approach. CBT builds a shared tree using reverse-path forwarding, without making assumptions on the unicast routing protocol used. The core of a group is either statically configured, or determined as the outcome of a selection process from a candidate set. Different multicast groups may use different core-bases trees. Distribution trees in CBT are bidirectional, that is, routers are capable of forwarding multicast packets downstream away from the core as well as upstream towards the core.

Protocol Independent Multicast (PIM):8 Protocol independent multicast consists of two multicast routing protocols: PIM Dense Mode (PIM-DM) and PIM Sparse Mode (PIM-SM). PIM-DM builds source-based trees using flood-and-prune, and is intended for large multicastence of this is that PIM must assume that all routers in the network are multicast enabled. An important difference between the core-based trees of PIM and CBT is that the trees in PIM are unidirectional, that is, sources always forward packets to the core, and the core transmits packets downstream the corebased tree.

All of the above multicast routing protocols follow the service model of IP multicast where any host can transmit to all multicast groups. Since this service model makes multicast routing complex, recently there has been support for a source-specific multicast service model, where hosts join a multicast group separately for each source. This service model is called Source- Specific Multicast (SSM). SSM requires that a host, when it joins a multicast group G, also specifies the source S from which it wishes to receive multicast packets. This can be accommodated by IGMPv39, a recently completed new revision IGMP, which allows hosts to join a multicast group for specific sources. SSM does not require new multicast routing protocols. In fact, routing for SSM can be realized with a subset of most existing protocols.

UNIT IV TRANSPORT AND APPLICATION LAYERS

Mobile TCP– WAP – Architecture – WWW Programming Model– WDP – WTLS – WTP – WSP – WAE – WTA Architecture – WML – WMLScripts.

PREREQUISTIES DISCUSSION:

In this unit,we comprises the global system for mobile communications (GSM) as today's most Preface
successful public mobile phone system, cordless phone technology, trunked
radios, and the future development with the universal mobile telecommunication
system (UMTS)

4.TRANSPORT AND APPLICATION LAYERS

4.1 MOBILE TCP

The **M-TCP (mobile TCP)**1 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 (Brown, 1997). 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 supervisory host is responsible for exchanging data between both parts similar to the proxy in ITCP (see Figure 9.1).

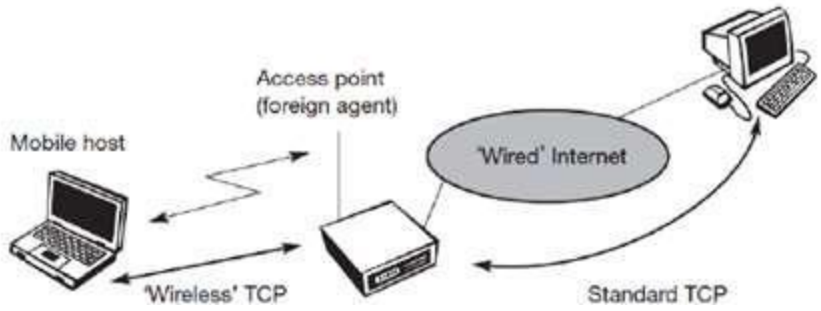


Fig 4.1 Indirect TCP segments a TCP Connection into two parts

The M-TCP approach assumes a relatively low bit error rate on the wireless link. Therefore, it does not perform caching/retransmission of data via the SH. If a packet is lost on the wireless link, it has to be retransmitted by the original sender. This maintains the TCP end-to-end semantics. 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. The wireless side uses an adapted TCP that can recover from packet loss much faster. This modified TCP does not use slow start, thus, M-TCP needs a **bandwidth manager** to implement fair sharing over the wireless link.

The **advantages** of M-TCP are the following:

- It maintains the TCP end-to-end semantics. The SH does not send any ACK itself but forwards the ACKs from the MH.

0 If the MH is disconnected, it avoids useless retransmissions, slow starts or breaking connections by simply shrinking the sender's window to 0.Since it does not buffer data in the SH as I-TCP does, it is not necessary to forward buffers to a new SH. Lost packets will be automatically retransmitted to the new SH.

The lack of buffers and changing TCP on the wireless part also has some **disadvantages:**

1 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.

2 A modified TCP on the wireless link not only requires modifications to the MH protocol software but also new network elements like the bandwidth manager.

WIRELESS APPLICATION PROTOCOL

The growth of the internet, internet applications, and mobile communications led to many early proprietary solutions providing internet services for mobile, wireless devices. Some of the problems these partial solutions face were discussed in section 10.2 because the World Wide Web is the most important and fastest growing internet application. To avoid many islands of incompatible solutions, e.g. special solutions for GSM, IS-136, or certain manufacturers, the **wireless application protocol forum (WAP Forum)** was founded in June 1997 by Ericsson, Motorola, Nokia, and Unwired Planet.

The basic objectives of the WAP Forum and now of the OMA are to bring diverse internet content (e.g., web pages, push services) and other data services (e.g., stock quotes) to digital cellular phones and other wireless, mobile terminals (e.g., PDAs, laptops). Moreover, a protocol suite should enable global wireless communication across different wireless network technologies, e.g., GSM, CDPD, UMTS etc. The forum is embracing and extending existing standards and technologies of the internet wherever possible and is creating a framework for the development of contents and applications that scale across a very wide range of wireless bearer networks and wireless device types.

- **Interoperable,** i.e., allowing terminals and software from different vendors to communicate with networks from different providers;
- **Scaleale,** i.e., protocols and services should scale with customer needs and number of customers;
- **Efficient,** i.e., provision of QoS suited to the characteristics of the wireless and mobile networks;

4.2 WAP-ARCHITECTURE

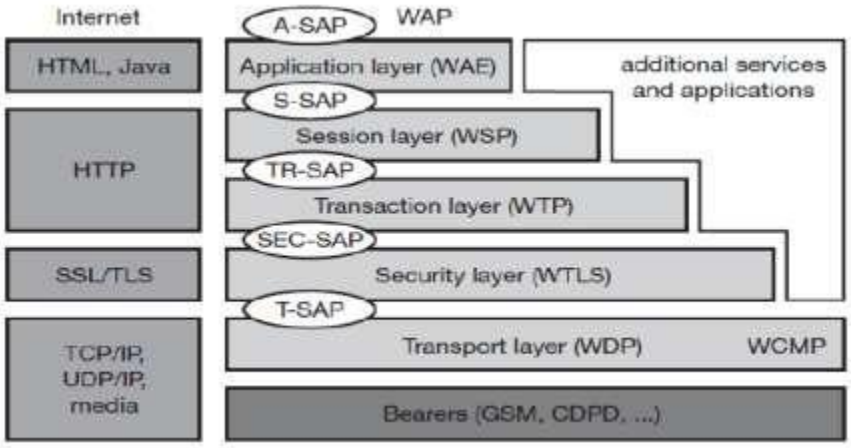


Fig 4.1 Components and Interface of WAP Architecture

The above Figure gives an overview of the WAP architecture, its protocols and components, and compares this architecture with the typical internet architecture when using the World Wide Web. This comparison is often cited by the WAP Forum and it helps to understand the architecture (WAP Forum, 2000a). This comparison can be misleading as not all components and protocols shown at the same layer are comparable. For consistency reasons with the existing specification, the following stays with the model as shown in Figure 10.9.The basis for transmission of data is formed by different **bearer services**. WAP does not specify bearer services, but uses existing data services and will integrate further services. Examples are message services, such as short message service (SMS) of GSM, circuit-switched data, such as high-speed circuit switched data (HSCSD) in GSM, or packet switched data, such as general packet radio service (GPRS) in GSM. Many other bearers are supported, such as CDPD, IS-136, PHS. No special interface has been specified between the bearer service and the next higher layer, the **transport layer** with its **wireless datagram protocol (WDP)** and the additional **wireless control message protocol (WCMP),** because the adaptation of these protocols are bearer-specific.

The transport layer offers a bearer independent, consistent datagram-oriented service to the higher layers of the WAP architecture. Communication is done transparently over one of the available bearer services. The **transport layer service access point (T-SAP)** is the common interface to be used by higher layers independent of the underlying network. The next higher layer, the **security layer** with its **wireless transport layer security** protocol **WTLS** offers its service at the **security SAP (SEC-SAP)**. WTLS is based on the transport layer security (TLS, formerly SSL, secure sockets layer) already known from the www. WTLS has been optimized for use in wireless networks with narrow-band channels. It can offer data integrity, privacy, authentication, and (some) denial-of-service protection.

The WAP **transaction layer** with its **wireless transaction protocol (WTP)** offers a lightweight transaction service at the **transaction SAP (TR-SAP)**. This service efficiently provides reliable or unreliable requests and asynchronous transactions as explained in the above section. Tightly coupled to this layer is the next higher layer, if used for connection-oriented service as described in section 10.3.5. The **session layer** with the **wireless session protocol (WSP)** currently offers two services at the **session-SAP (S-SAP)**, one connection-oriented and one connectionless if used directly on top of WDP. A special service for browsing the web (WSP/B) has been defined that offers HTTP/1.1 functionality, long-lived session state, session suspend and resume, session migration and other features needed for wireless mobile access to the web. Finally the **application layer** with the **wireless application environment (WAE)** offers a framework for the integration of different www and mobile telephony applications. It offers many protocols and services with special service access.

The main issues here are scripting languages, special markup languages, interfaces to telephony applications, and many content formats adapted to the special requirements of small, handheld, wireless devices.Figure 10.9 not only shows the overall WAP architecture, but also its relation to the traditional internet architecture for www applications. The WAP transport layer together with the bearers can be (roughly) compared to the services offered by TCP or UDP over IP and different media in the internet. If a bearer in the WAP architecture already offers IP services (e.g., GPRS, CDPD) then UDP is used as WDP. The TLS/SSL layer of the internet has also been adopted for the WAP architecture with some changes required for optimization. The functionality of the session and transaction layer can roughly be compared with the role of HTTP in the web architecture. However, HTTP does not offer all the additional mechanisms needed for efficient wireless, mobile access (e.g., session migration, suspend/resume).

Finally, the application layer offers similar features as HTML and Java. Again, special formats and features optimized for the wireless scenario have been defined and telephony access has been added.WAP does not always force all applications to use the whole protocol architecture. Applications can use only a part of the architecture as shown in Figure 10.9. For example, this means that, if an application does not require security but needs the reliable transport of data, it can **directly** use a service of the transaction layer. Simple applications can directly use WDP. Different scenarios are possible for the integration of WAP components into existing wireless and fixed networks (see Figure 10.10). On the left side, different fixed networks, such as the traditional internet and the public switched telephone network (PSTN), are shown. One cannot change protocols and services of these existing networks so several new elements will be implemented between these networks and the WAP-enabled wireless, mobile devices in a wireless network on the right-hand side.

The current www in the internet offers web pages with the help of HTML and web servers. To be able to browse these pages or additional pages with handheld devices, a wireless markup language (WML) has been defined in WAP. Special filters within the fixed network can now translate HTML into WML, web servers can already provide pages in WML, or the gateways between the fixed and wireless network can translate HTML into WML. These gateways not only filter pages but also act as proxies for web access, as explained in the following sections.WML is additionally converted into binary WML for more efficient transmission. In a similar way, a special gateway can be implemented to access traditional

telephone services via binary WML. This wireless telephony application (WTA) server translates, e.g., signaling of the telephone network (incoming call etc.) into WML events displayed at the handheld device. It is important to notice the integrated view for the wireless client of all different services; telephony and web, via the WAE.

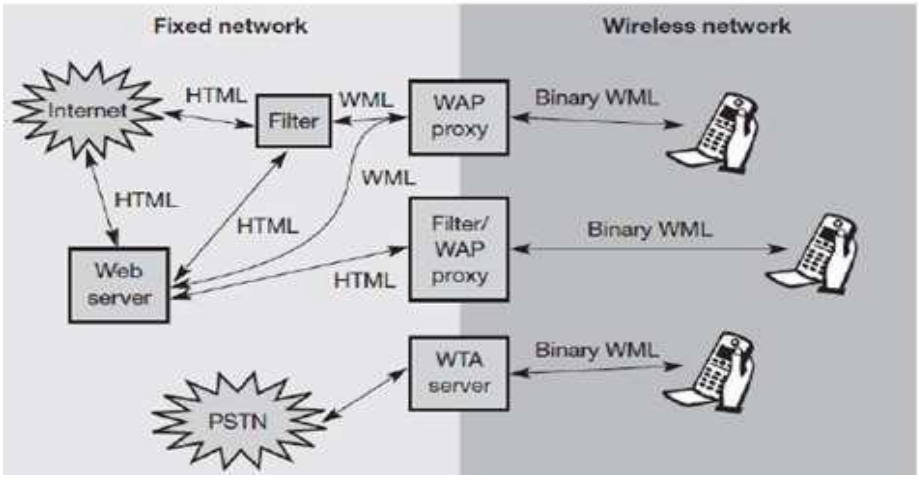


Fig 4.2 Examples of Integration of WAP Components

4.3 WWW Programming Model

4.3.1 Wireless Datagram Protocol (WDP)

The W**ireless Datagram Protocol (WDP)** operates on top of many different bearer services capable of carrying data. At the T-SAP WDP offers a consistent datagram transport service independent of the underlying bearer. To offer this consistent service, the adaptation needed in the transport layer can differ depending on the services of the bearer. The closer the bearer service is to IP, the smaller the adaptation can be. If the bearer already offers IP services, UDP is used as WDP. WDP offers more or less the same services as UDP.WDP offers **source** and **destination port numbers** used for multiplexing and demultiplexing of data respectively. The service primitive to send a datagram is **TDUnitdata.req** with the **destination address (DA), destination port (DP), Source address (SA), source port (SP)**, and **user data (UD)** as mandatory parameters (see Figure 10.11). Destination and source address are unique addresses for the receiver and sender of the user data. These could be MSISDNs (i.e., a telephone number), IP addresses, or any other unique identifiers. The **T-DUnitdata.ind** service primitive indicates the reception of data. Here destination address and port are only optional parameters.

If a higher layer requests a service the WDP cannot fulfill, this error is indicated with the **T-DError.ind** service primitive as shown in Figure 10.11. An **error code (EC)** is returned indicating the reason for the error to the higher layer. WDP is not allowed to use this primitive to indicate problems with the bearer service. It is only allowed to use the primitive to indicate local problems, such as a user data size that is too large. If any errors happen when WDP datagram's are sent from one WDP entity to another (e.g. the destination is unreachable, no application is

listening to the specified destination port etc.), the **wireless control message protocol (WCMP)** provides error handling mechanisms for WDP and should therefore be implemented. WCMP contains control messages that resemble the internet control message protocol messages and can also be used for diagnostic and informational purposes.

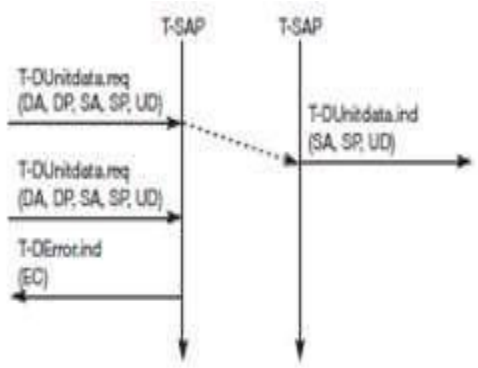


Fig 4.3 WAP Service Primitives

WCMP can be used by WDP nodes and gateways to report errors. However, WCMP error messages must not be sent as response to other WCMP error messages. In IP-based networks, ICMP will be used as WCMP (e.g., CDPD, GPRS). Typical WCMP messages are **destination unreachable** (route, port, address unreachable), **parameter problem** (errors in the packet header), **message too big, reassembly failure**, or **echo request/reply**. An additional

WDP management entity supports WDP and provides information about changes in the environment, which may influence the correct operation of WDP. Important information is the current configuration of the device, currently available bearer services, processing and memory resources etc. Design and implementation of this management component is considered vendor-specific and is outside the scope of WAP.

If the bearer already offers IP transmission, WDP (i.e., UDP in this case) relies on the segmentation (called fragmentation in the IP context) and reassembly capabilities of the IP layer as specified in (Postal, 1981a). Otherwise, WDP has to include these capabilities, which is, e.g., necessary for the GSM SMS. The WAP specification provides many more adaptations to almost all bearer services currently available or planned for the future (WAP Forum, 2000q), (WAP Forum, 2000b).

4.4 Wireless Transport Layer Security (WTLS)

If requested by an application, a security service, the **wireless transport layer security (WTLS)**, can be integrated into the WAP architecture on top of WDP as specified in (WAP Forum, 2000c). WTLS can provide different levels of security (for privacy, data integrity, and authentication) and has been optimized for low bandwidth, high-delay bearer networks. WTLS takes into account the low processing power and very limited memory capacity of the mobile devices for cryptographic algorithms. WTLS supports datagram and connection-oriented transport layer protocols. New compared to, e.g. GSM, is the security relation between two peers and not only between the mobile device and the base station . WTLS took over many features

and mechanisms from TLS (formerly SSL, secure sockets layer, but it has an optimized handshaking between the peers.

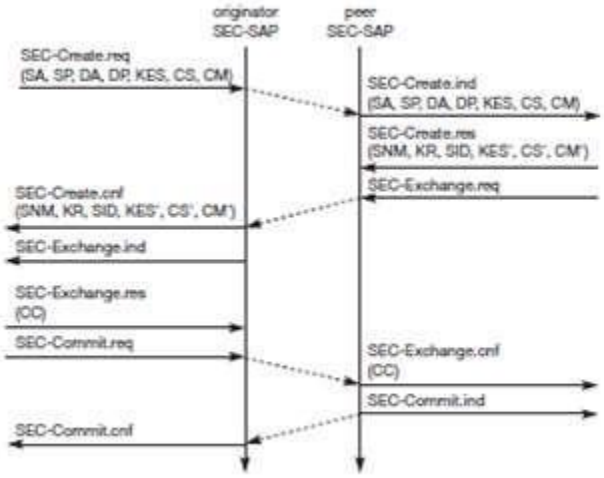


Fig 4.4 WTLS establishing a Secure Session

Before data can be exchanged via WTLS, a secure session has to be established. This session establishment consists of several steps: Figure illustrates the sequence of service primitives needed for a so-called 'full handshake' (several optimizations are possible). The originator and the peer of the secure session can both interrupt session establishment any time, e.g., if the parameters proposed are not acceptable.

The first step is to initiate the session with the SEC-Create primitive. Parameters are source address (SA), source port (SP) of the originator, destination address (DA), destination port (DP) of the peer. The originator proposes a key exchange suite (KES) (e.g., RSA, DH, ECC, a cipher suite (CS) (e.g., DES, IDEA, and a compression method (CM) (currently not further specified). The peer answers with parameters for the sequence number mode (SNM), the key refresh cycle (KR) (i.e., how often keys are refreshed within this secure session), the session identifier (SID) (which is unique with each peer), and the selected key exchange suite (KES'), cipher suite (CS'), compression method (CM'). The peer also issues a SEC-Exchange primitive. This indicates that the peer wishes to perform public-key authentication with the client, i.e., the peer requests a client certificate (CC) from the originator.

The first step of the secure session creation, the negotiation of the security parameters and suites, is indicated on the originator's side, followed by the request for a certificate. The originator answers with its certificate and issues a SEC-Commit.req primitive. This primitive indicates that the handshake is completed for the originator's side and that the originator now wants to switch into the newly negotiated connection state. The certificate is delivered to the peer side and the SEC-Commit is indicated. The WTLS layer of the peer sends back a confirmation to the originator. This concludes the full handshake for secure session setup.

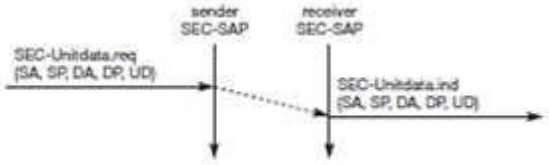


Fig 4.5 WTLS Datagram Transfer

After setting up a secure connection between two peers, user data can be exchanged. This is done using the simple SEC-Unit data primitive as shown in Figure 10.13. SEC-Unit data has exactly the same function as T-D Unit data on the WDP layer, namely it transfers a datagram between a sender and a receiver. This data transfer is still unreliable, but is now secure. This shows that WTLS can be easily plugged into the protocol stack on top of WDP. The higher layers simply use SEC-Unit data instead of T-D Unit data. The parameters are the same here: source address (SA), source port (SP), destination address (DA), destination port (DP), and user data (UD).

This section will not discuss the security-related features of WTLS or the pros and cons of different encryption algorithms. The reader is referred to the specification and excellent cryptography literature. Although WTLS allows for different encryption mechanisms with different key lengths, it is quite clear that due to computing power on the handheld devices the encryption provided cannot be very strong. If applications require stronger security, it is up to an application or a user to apply stronger encryption on top of the whole protocol stack and use WTLS as a basic security level only. Many programs are available for this purpose. It is important to note that the security association in WTLS exists between the mobile WAP-enabled devices and a WAP server or WAP gateway only. If an application accesses another server via the gateway, additional mechanisms are needed for end-to-end security. If for example a user accesses his or her bank account using WAP, the WTLS security association typically ends at the WAP gateway inside the network operator's domain. The bank and user will want to apply additional security mechanisms in this scenario.

Future work in the WTLS layer comprises consistent support for application level security (e.g. digital signatures) and different implementation classes with different capabilities to select from.

4.5 Wireless transaction protocol (WTP)

The wireless transaction protocol (WTP) is on top of either WDP or, if security is required, WTLS (WAP Forum, 2000d). WTP has been designed to run on very thin clients, such as mobile phones. WTP offers several advantages to higher layers, including an improved reliability over datagram services, improved efficiency over connection-oriented services, and support for transaction-oriented services such as web browsing. In this context, a transaction is defined as a request with its response, e.g. for a web page. WTP offers many features to the higher layers. The basis is formed from three classes of transaction service as explained in the following paragraphs. Class 0 provides unreliable message transfer without any result message. Classes 1 and 2 provide reliable message transfer, class 1 without, class 2 with, exactly one reliable result message (the typical request/response case).

WTP achieves reliability using duplicate removal, retransmission, acknowledgements and unique transaction identifiers. No WTP-class requires any connection set-up or tear-down phase. This avoids unnecessary overhead on the communication link. WTP allows for asynchronous transactions, abort of transactions, concatenation of messages, and can report success or failure of reliable messages (e.g., a server cannot handle the request). To be consistent with the specification, in the following the term initiator is used for a WTP entity initiating a transaction (aka client), and the term responder for the WTP entity responding to a transaction (aka server). The three service primitives offered by WTP are TR-Invoke to initiate a new transaction, TR-Result to send back the result of a previously initiated transaction, and TR-Abort to abort an existing transaction. The PDUs exchanged between two WTP entities for normal transactions are the invoke PDU, ack PDU, and result PDU.

A special feature of WTP is its ability to provide a user acknowledgement or, alternatively, an automatic acknowledgement by the WTP entity. If user acknowledgement is required, a WTP user has to confirm every message received by a WTP entity. A user acknowledgement provides a stronger version of a confirmed service because it guarantees that the response comes from the user of the WTP and not the WTP entity itself. WTP class 0 Class 0 offers an unreliable transaction service without a result message. The transaction is stateless and cannot be aborted. The service is requested with the TR-Invoke.req primitive as shown in Figure 10.14. Parameters are the source address (SA), source port (SP), destination address (DA), destination port (DP) as already explained in section 10.3.2. Additionally, with the A flag the user of this service can determine, if the responder WTP entity should generate an acknowledgement or if a user acknowledgement should be used. The WTP layer will transmit the user data (UD) transparently to its destination. The class type C indicates here class 0. Finally, the transaction handle H provides a simple index to uniquely identify the transaction and is an alias for the tuple (SA, SP, DA, DP), i.e., a socket pair, with only local significance.

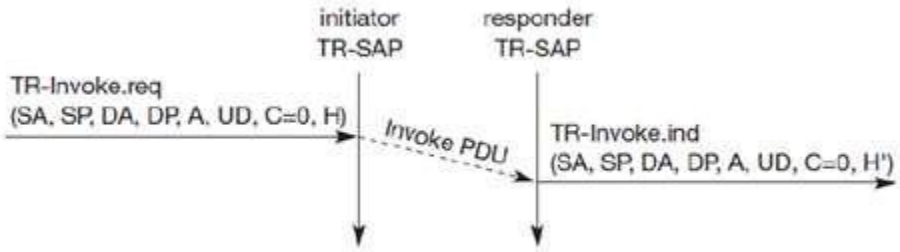


Fig 4.6 Basic Transaction , WTP Class 0

The WTP entity at the initiator sends an invoke PDU which the responder receives. The WTP entity at the responder then generates a TR-Invoke.ind primitive with the same parameters as on the initiators side, except for which is now the local handle for the transaction on the responders side. In this class, the responder does not acknowledge the message and the initiator does not perform any retransmission. Although this resembles a simple datagram service, it is recommended to use WDP if only a datagram service is required. WTP class 0

augments the transaction service with a simple datagram like service for occasional use by higher layers.

WTP class 1 Class 1 offers a reliable transaction service but without a result message. Again, the initiator sends an invoke PDU after a TR-Invoke.req from a higher layer. This time, class equals „1, and no user acknowledgement has been selected as shown in Figure 10.15. The responder signals the incoming invoke PDU via the TR-Invoke.ind primitive to the higher layer and acknowledges automatically without user intervention. The specification also allows the user on the responders side to acknowledge, but this acknowledgement is not required. For the initiator the transaction ends with the reception of the acknowledgement. The responder keeps the transaction state for some time to be able to retransmit the acknowledgement if it receives the same invoke PDU again indicating a loss of the acknowledgement.

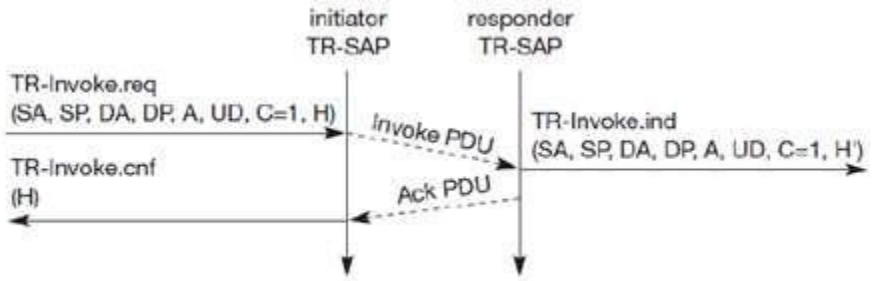


Fig 4.7 Basic Transaction , WTP Class 1, no user Acknowledgement

If a user of the WTP class 1 service on the initiators side requests a user acknowledgement on the responders side, the sequence diagram looks like Figure. Now the WTP entity on the responders side does not send an acknowledgement automatically, but waits for the TR-Invoke.res service primitive from the user. This service primitive must have the appropriate local handle H for identification of the right transaction. The WTP entity can now send the ack PDU. Typical uses for this transaction class are reliable push services.

WTP class 2 Finally, class 2 transaction service provides the classic reliable request/response transaction known from many client/server scenarios. Depending on user requirements, many different scenarios are possible for initiator/responder interaction. Three examples are presented below. Figure shows the basic transaction of class 2 without-user acknowledgement. Here, a user on the initiators side requests the service and the WTP entity sends the invoke PDU to the responder. The WTP entity on the responders side indicates the request with the TR-Invoke.ind primitive to a user. The responder now waits for the processing of the request, the user on the responders side can finally give the result UD* to the WTP entity on the responder side using TR-Result.req. The result PDU can now be sent back to the initiator, which implicitly acknowledges the invoke PDU. The initiator can indicate the successful transmission of the invoke message and the result with the two service primitives TR-Invoke.cnf and TR-Result.ind. A user may respond to this result with TR-Result.res. An acknowledgement PDU is then generated which finally triggers the TR-Result.cnf primitive on the responder‟s side. This example clearly shows the combination of two reliable services (TR-Invoke and TR-Result) with an efficient data transmission/acknowledgement.

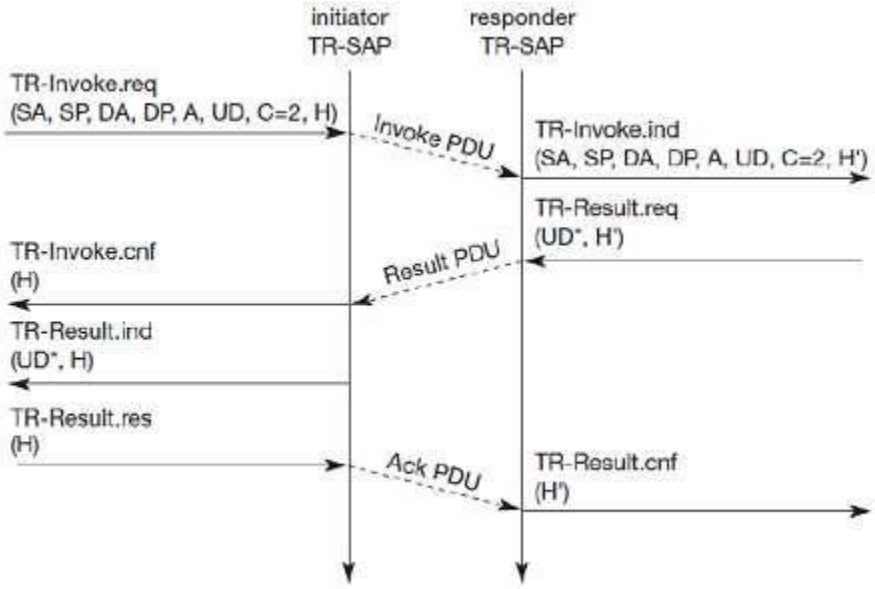


Fig 4.8 Basic Transaction , WTP Class 2, no user Acknowledgement

An even more reliable service can be provided by user acknowledgement as explained above. The time-sequence diagram looks different (see Figure 10.18). The user on the responder's side now explicitly responds to the Invoke PDU using the TR-Invoke.res primitive, which triggers the TR-Invoke.cnf on the initiator's side via an ack PDU. The transmission of the result is also a confirmed service, as indicated by the next four service primitives. This service will likely be the most common in standard request/response scenarios as, e.g., distributed computing.

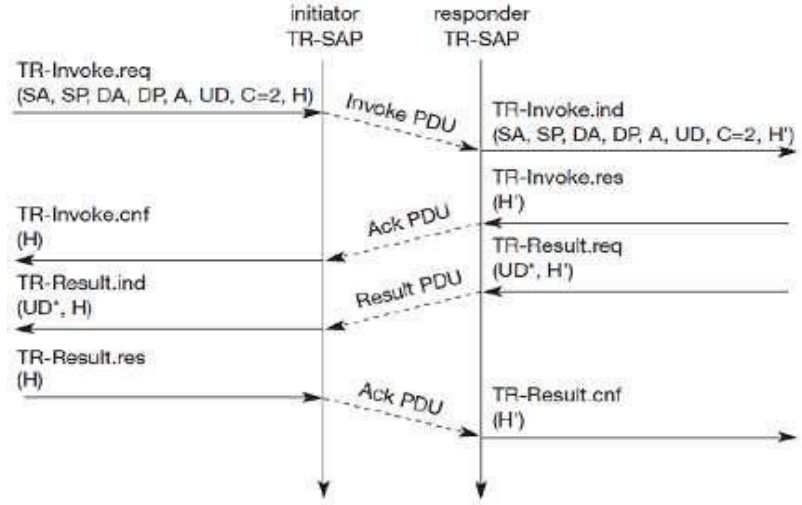


Fig 4.9 Basic Transaction , WTP Class 2,with user Acknowledgement

If the calculation of the result takes some time, the responder can put the initiator on "hold on" to prevent a retransmission of the invoke PDU as the initiator might assume packet loss if no result is sent back within a certain timeframe. This is shown in the Figure.

After a time-out, the responder automatically generates an acknowledgement for the Invoke PDU. This shows the initiator that the responder is still alive and currently busy processing the request. WTP provides many more features not explained here, such as concatenation and separation of messages, asynchronous transactions with up to 215 transactions outstanding, i.e., requested but without result up to now, and segmentation/ reassembly of messages.

4.6 Wireless session protocol (WSP)

The **wireless session protocol (WSP)** has been designed to operate on top of the datagram service WDP or the transaction service WTP (WAP Forum, 2000e).For both types, security can be inserted using the WTLS security layer if required. WSP provides a shared state between a client and a server to optimize content transfer. HTTP, a protocol WSP tries to replace within the wireless domain, is stateless, which already causes many problems in fixed networks. Many web content providers therefore use cookies to store some state on a client machine, which is not an elegant solution. State is needed in web browsing, for example, to resume browsing in exactly the same context in which browsing has been suspended. This is an important feature for clients and servers. Client users can continue to work where they left the browser or when the network was interrupted, or users can get their customized environment every time they start the browser. Content providers can customize their pages to clients' needs and do not have to retransmit the same pages over and over again. WSP offers the following general features needed for content exchange between cooperating clients and servers:

- **Session management:** WSP introduces sessions that can be **established** from a client to a server and may be long lived. Sessions can also be **released** in an orderly manner. The capabilities of **suspending** and **resuming** a session are important to mobile applications. Assume a mobile device is being switched off – it would be useful for a user to be able to continue operation at exactly the point where the device was switched off. Session lifetime is independent of transport connection lifetime or continuous operation of a bearer network.

- **Capability negotiation:** Clients and servers can agree upon a common level of protocol functionality during session establishment. Example parameters to negotiate are maximum client SDU size, maximum outstanding requests, protocol options, and server SDU size.

c) **Content encoding:** WSP also defines the efficient binary encoding for the content it transfers. WSP offers content typing and composite objects, as explained for web browsing. While WSP is a general-purpose session protocol, WAP has specified the **wireless session protocol/browsing (WSP/B)** which comprises protocols and services most suited for browsing-type applications. In addition to the general
features of WSP, WSP/B offers the following features adapted to web browsing:

d) **HTTP/1.1 functionality:** WSP/B supports the functions HTTP/1.1 offers, such as extensible request/reply methods, composite objects, and content type negotiation. WSP/B is a binary form

of HTTP/1.1. HTTP/1.1 content headers are used to define content type, character set encoding, languages etc., but binary encodings are defined for well-known headers to reduce protocol overheads.

e) **Exchange of session headers:** Client and server can exchange request/reply headers that remain constant over the lifetime of the session. These headers may include content types, character sets, languages, device capabilities, and other static parameters. WSP/B will not interpret header information
but passes all headers directly to service users.

f) **Push and pull data transfer:** Pulling data from a server is the traditional mechanism of the web. This is also supported by WSP/B using the request/response mechanism from HTTP/1.1. Additionally, WSP/B supports three push mechanisms for data transfer: a confirmed data push within an existing session context, a non-confirmed data push within an existing session context, and a non-confirmed data push without an existing session context.

g) **Asynchronous requests:** Optionally, WSP/B supports a client that can send multiple requests to a server simultaneously. This improves efficiency for the requests and replies can now be coalesced into fewer messages. Latency is also improved, as each result can be sent to the client as soon as it is available.

As already mentioned, WSP/B can run over the transaction service WTP or the datagram service WDP. The following shows several protocol sequences typical for session management, method invocation, and push services.

4.7 Wireless Application Environment

The main idea behind the wireless application environment (WAE) is to create a general-purpose application environment based mainly on existing technologies and philosophies of the world wide web. This environment should allow service providers, software manufacturers, or hardware vendors to integrate their applications so they can reach a wide variety of different wireless platforms in an efficient way. However, WAE does not dictate or assume any specific man-machine-interface model, but allows for a variety of devices, each with its own capabilities and probably vendor-specific extras (i.e., each vendor can have its own look and feel). WAE has already integrated the following technologies and adapted them for use in a wireless environment with low power handheld devices.

HTML, JavaScript, and the handheld device markup language HDML form the basis of the wireless markup language (WML) and the scripting language WML script. The exchange formats for business cards and phone books vCard and for calendars vCalendar have been included. URLs from the web can be used. A wide range of mobile telecommunication technologies have been adopted and integrated into the wireless telephony application (WTA).

Besides relying on mature and established technology, WAE focuses on devices with very limited capabilities, narrow-band environments, and special security and access control features. The first phase of the WAE specification developed a whole application suite, especially for wireless clients as presented in the following sections. Future developments for the WAE will include extensions for more content formats, integration of further existing or

emerging technologies, more server-side aspects, and the integration of intelligent telephone networks.

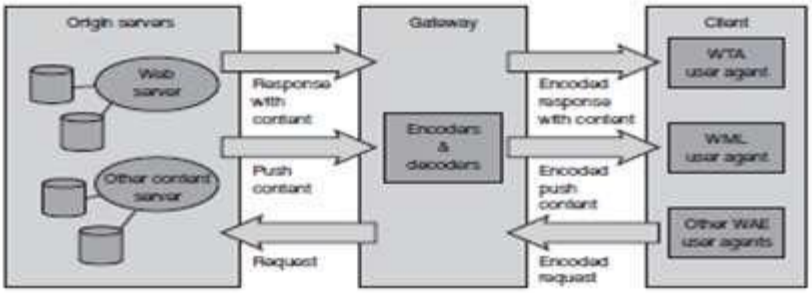


Fig 4.10 WAE Logical Model

One global goal of the WAE is to minimize over-the-air traffic and resource consumption on the handheld device. This goal is also reflected in the logical model underlying WAE (Figure 10.29) showing some more detail than the general overview in Figure 10.10. WAE adopts a model that closely follows the www model, but assumes additional gateways that can enhance transmission efficiency.

A client issues an encoded request for an operation on a remote server. Encoding is necessary to minimize data sent over the air and to save resources on the handheld device as explained together with the languages WML and WMLscript. Decoders in a gateway now translate this encoded request into a standard request as understood by the origin servers. This could be a request to get a web page to set up a call. The gateway transfers this request to the appropriate origin server as if it came from a standard client. Origin servers could be standard web servers running HTTP and generating content using scripts, providing pages using a database, or applying any other (proprietary) technology. WAE does not specify any standard content generator or server, but assumes that the majority will follow the standard technology used in today's www.

The origin servers will respond to the request. The gateway now encodes the response and its content (if there is any) and transfers the encoded response with the content to the client. The WAE logical model not only includes this standard request/response scheme, but it also includes push services. Then an origin server pushes content to the gateway. The gateway encodes the pushed content and transmits the encoded push content to the client.

Several user agents can reside within a client. User agents include such items as: browsers, phonebooks, message editors etc. WAE does not specify the number of user agents or their functionality, but assumes a basic WML user agent that supports WML, WML script, or both (i.e., a 'WML browser'). Further domain specific user agents with varying architectures can be implemented. Again, this is left to vendors. However, one more user agent has been specified with its fundamental services, the WTA user agent. This user agent handles access to, and interaction with, mobile telephone features (such as call control). As over time many vendor dependent user agents may develop, the standard defines a user agent profile (UAProf), which describes the capabilities of a user agent. Capabilities may be related to hardware or software.

Examples are: display size, operating system, browser version, processor, memory size, audio/video codec, or supported network types. The basic languages WML and WML Script , and the WTA will be described in the following three sections.

4.8 WTA Architecture

Browsing the web using the WML browser is only one application for a handheld device user. Say a user still wants to make phone calls and access all the features of the mobile phone network as with a traditional mobile phone. This is where the wireless telephony application (WTA), the WTA user agent (as shown in Figure), and the wireless telephony application interface WTAI come in. WTA is a collection of telephony specific extensions for call and feature control mechanisms, merging data networks and voice networks.

The WTA framework integrates advanced telephony services using a consistent user interface (e.g., the WML browser) and allows network operators to increase accessibility for various special services in their network. A network operator can reach more end-devices using WTA because this is integrated in the wireless application environment (WAE) which handles device-specific characteristics and environments. WTA should enable third-party developers as well as network operators to create network-independent content that accesses the basic features of the bearer network. However, most of the WTA functionality is reserved for the network operators for security and stability reasons.

WTA extends the basic WAE application model in several ways:

● **Content push:** A WTA origin server can push content, i.e., WML decks or WMLScript, to the client. A push can take place without prior client request. The content can enable, e.g., the client to handle new network events that were unknown before.

● **Access to telephony functions:** The **wireless telephony application interface** (**WTAI**, WAP Forum, 2000m) provides many functions to handle telephony events (call accept, call setup, change of phone book entries etc.).

● **Repository for event handlers:** The repository represents a persistent storage on the client for content required to offer WTA services. Content are either channels or resources. Examples for resources are WML decks, WMLScript objects, or WBMP pictures. Resources are loaded using WSP or are pre-installed. A channel comprises references to resources and is associated with a lifetime. Within this lifetime, it is guaranteed that all resources the channel points to are locally available in the repository. The motivation behind the repository is the necessity to react very quickly for time-critical events (e.g., call accept). It would take too long to load content from a server for this purpose.

● **Security model**: Mandatory for WTA is a security model as many frauds happen with wrong phone numbers or faked services. WTA allows the client to only connect to trustworthy gateways, which then have to check if the servers providing content are authorized to send this content to the client. Obviously, it is not easy to define trustworthy in this context. In the beginning, the network operator‟s gateway may be the only trusted gateway and the network operator may decide which servers are allowed to provide content. Figure 10.30 gives an overview of the WTA logical architecture.

The components shown are not all mandatory in this architecture; however, firewalls or other origin servers may be useful. A minimal configuration could be a single server from the network operator serving all clients. The **client** is connected via a mobile network with a **WTA server**, other telephone networks (e.g., fixed PSTN), and a **WAP gateway**. A WML user agent running on the client or on other user agents is not shown here.

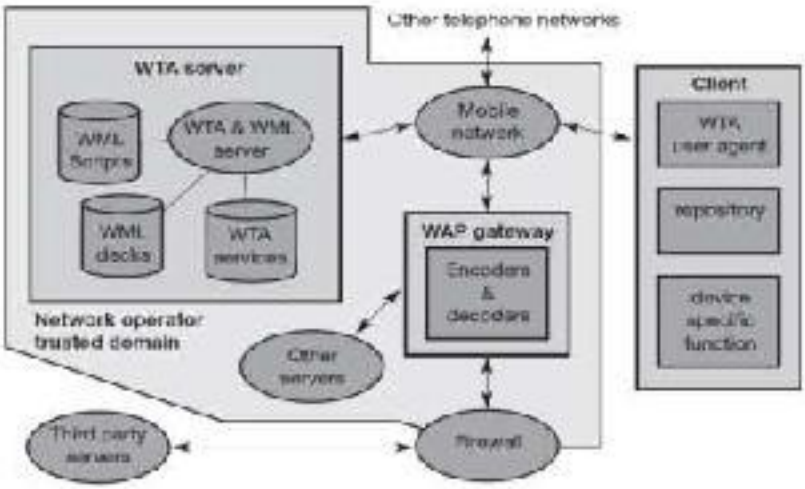


Fig 4.11 WTA Architecture

The client may have voice and data connections over the mobile network. Other origin servers within the trusted domain may be connected via the WAP gateway. A firewall is useful to connect third-party origin servers outside the trusted domain. One difference between WTA servers and other servers besides security is the tighter control of QoS. A network operator knows (more or less precisely) the latency, reliability, and capacity of its mobile network and can have more control over the behavior of the services. Other servers, probably located in the internet, may not be able to give as good QoS guarantees as the network operator.

Similarly, the WTA user agent has a very rigid and real-time context management for browsing the web compared to the standard WML user agent. Figure shows an exemplary interaction between a WTA client, a WTA gateway, a WTA server, the mobile network (with probably many more servers) and a voice box server. Someone might leave a message on a voice box server as indicated. Without WAP, the network operator then typically generates an SMS indicating the new message on the voice box via a little symbol on the mobile phone. However, it is typically not indicated who left a message, what messages are stored etc. Users have to call the voice box to check and cannot choose a particular message. In a WAP scenario, the voice box can induce the WTA server to generate new content for pushing to the client. An example could be a WML deck containing a list of callers plus length and priority of the calls. The server does not push this deck immediately to the client, but sends a push message containing a single URL to the client. A short note, e.g., “5 new calls are stored", could accompany the push message. The WTA gateway translates the push URL into a service indication and codes it into a more compact binary format. The WTA user agent then indicates that new messages are stored. If the user wants to listen to the stored messages, he or she can request a list of the messages.

This is done with the help of the URL. A WSP get requests the content the URL points to. The gateway translates this WSP get into an HTTP get and the server responds with the prepared list of callers.

After displaying the content, the user can select a voice message from the list. Each voice message in this example has an associated URL, which can request a certain WML card from the server. The purpose of this card is to prepare the client for an incoming call. As soon as the client receives the card, it waits for the incoming call. The call is then automatically accepted. The WTA server also signals the voice box system to set up a (traditional) voice connection to play the selected voice message. Setting up the call and accepting the call is shown using dashed lines, as these are standard interactions from the mobile phone network, which are not controlled by WAP.

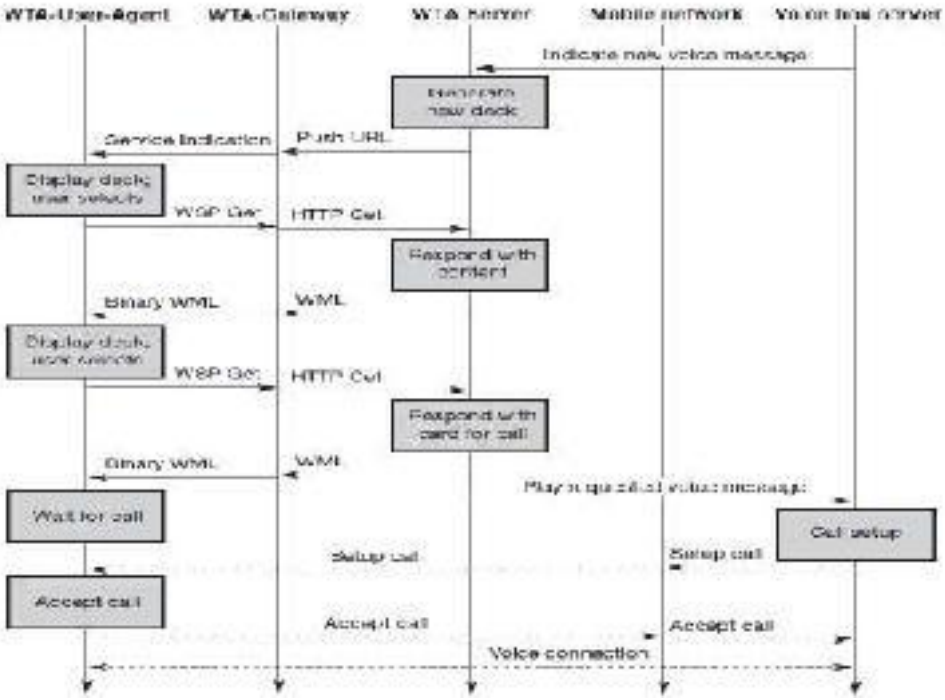


Fig 4.12 Exemplary Interaction Between A WTA Client, A WTA Gateway, A WTA Server, The Mobile Network

4.9 Wireless markup language (WML)

The **wireless markup language (WML)** (WAP Forum, 2000j) is based on the standard HTML known from the www and on HDML (King, 1997). WML is specified as an XML (W3C, 1998a) document type. When designing WML, several constraints of wireless handheld devices had to be taken into account. First of all, the wireless link will always have only a very limited capacity compared to a wire. Current handheld devices have small displays, limited user input facilities, limited memory, and only low performance computational resources. While the bandwidth argument will remain for many years, it currently seems that the gap between mobile and fixed devices regarding processing
power is getting narrower.

Today's CPUs in PDAs have performance figures close to desktop CPUs just a few years ago. WML follows a deck and card metaphor. A WML document is made up of multiple **cards**. Cards can be grouped together into a **deck**. A WML deck is similar to an HTML page, in that it is identified by a URL and is the unit of content transmission. A user navigates with the WML browser through a series of WML

cards, reviews the contents, enters requested data, makes choices etc. The WML browser fetches decks as required from origin servers. Either these decks can be static files on the server or they can be dynamically generated.

It is important to note that WML does not specify how the implementation of a WML browser has to interact with a user. Instead, WML describes the intent of interaction in an abstract manner. The user agent on a handheld device has to decide how to best present all elements of a card. This presentation depends much on the capabilities of the device.

WML includes several basic features:

● **Text and images:** WML gives, as do other mark-up languages, hints how text and images can be presented to a user. However, the exact presentation of data to a user is up to the user agent running on the handheld device. WML only provides a set of mark-up elements, such as emphasis elements (bold, italic, etc.) for text, or tab columns for tabbing alignment.

● **User interaction:** WML supports different elements for user input.Examples are: text entry controls for text or password entry, option selections or controls for task invocation. Again, the user agent is free to choose how these inputs are implemented. They could be bound to, e.g., physical keys, soft keys, or voice input.

• **Navigation:** As with HTML browsers, WML offers a history mechanism with navigation through the browsing history, hyperlinks and other intercard navigation elements.

● **Context management:** WML allows for saving the state between different decks without server interaction, i.e., variable state can last longer than a single deck, and so state can be shared across different decks. Cards can have parameters defined by using this state without access to the server

over the narrow-band wireless channel.

The following paragraph gives a simple example of WML; the reader is referred to the standard or Singhal (2001) for a full reference and in-depth discussion of the language.

First, a reference to XML is given where WML was derived from. Then, after the keyword wml the first card is defined. This first card of the deck "displays" a text after loading ("displaying" could also mean voice output etc.). As soon as a user activates the do element (a button or voice command), the user agent displays the second card. On this second card, the user can select one out of three pizza options. Depending on the choice of the user, PIZZA can have one of the values Mar, Fun, or Vul. If the user proceeds to the third card without choosing a pizza, the value Mar is used as default. Again, describing these options with WML does not automatically mean that these options are displayed as text. It could also be possible that the user

agent reads the options through a voice output and the user answers through a voice input. WML only describes the intention of a choice. The third card finally outputs the value of PIZZA.

```
<?xml version="1.0"?>
<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD WML
1.1//EN" "http://www.wapforum.org/DTD/wml_1.1.xml">
<wml>
<card id="card_one" title="Simple example">
<do type="accept">
<go href="#card_two"/>
</do>
<p>
This is a simple first card!
<br/>
On the next one you can choose ...
</p>
</card>
<card id="card_two" title="Pizza selection">
<do type="accept" label="cont">
<go href="#card_three"/>
</do>
<p>
... your favourite pizza!
<select value="Mar" name="PIZZA">
<option value="Mar">Margherita</option>
<option value="Fun">Funghi</option>
<option value="Vul">Vulcano</option>
</select>
</p>
</card>
<card id="card_three" title="Your
Pizza!"> <p>
Your personal pizza parameter is <b>$(PIZZA)</b>!
</p>
</card>
</wml>
```

WML may be encoded using a compact binary representation to save bandwidth on the wireless link. This compact representation is based on the binary XML content format as specified in WAP Forum (2000k). The binary coding of WML is only one special version of this format; the compact representation is

valid in general for XML content. The compact format allows for transmission without loss of functionality or of semantic information. For example, the URL prefix href=_http://, which is very common in URLs, will be coded as 4B. The code for the select keyword is 37 and option is 35. These single byte codes are much more efficient than the plain ASCII text used in HTML and today’s www.

4.10 WMLScript

WMLScript complements to WML and provides a general scripting capability in the WAP architecture (WAP Forum, 2000h). While all WML content is static (after loading on the client), WMLScript offers several capabilities not supported by WML:

● **Validity check of user input:** before user input is sent to a server, WMLScript can check the validity and save bandwidth and latency in case of an error. Otherwise, the server has to perform all the checks, which always includes at least one round-trip if problems occur.

● **Access to device facilities:** WMLScript offers functions to access hardware components and software functions of the device. On a phone a user could, e.g., make a phone call, access the address book, or send a message via the message service of the mobile phone.

● **Local user interaction:** Without introducing round-trip delays, WMLScript can directly and locally interact with a user, show messages or prompt for input. Only, for example, the result of several interactions could be transmitted to a server.

● **Extensions to the device software:** With the help of WMLScript a device can be configured and new functionality can be added even after deployment. Users can download new software from vendors and, thus, upgrade their device easily.

The following six libraries have been defined so far:

● **Lang:** This library provides functions closely related to WMLScipt itself.Examples are isInt to check if a value could be converted into an integer or float to check if floating-point operations are supported.

● **Float:** Many typical arithmetic floating-point operations are in this library (which is optional as mentioned before). Example functions are round for rounding a number and sqrt for calculating the square root of a given value.

● **String**: Many string manipulation functions are in this library. Examples are well-known functions such as length to return the length of a string or subString to return a substring of a given string. Nevertheless, this library also provides more advanced functions such as find to find a substring within a string or squeeze to replace several consecutive whitespaces with only one.

● **URL:** This library provides many functions for handling URLs with the syntax defined in Fielding :

<scheme>://<host>:<port>/<path>;<parameters>?<query>#<fragment>
for example: http://www.xyz.int:8080/mypages;5;2?j=2&p=1#crd.
The function getPath could now extract the path of this URL, i.e., "mypages", getQuery has the query part j=2&p=1" as return value, and getFragment delivers the fragment used in the URL, i.e., "crd".

UNIT V PERVASIVE COMPUTING

Pervasive computing infrastructure-applications- Device Technology - Hardware, Human-machine Interfaces, Biometrics, and Operating systems– Device Connectivity –Protocols, Security, and Device Management- Pervasive Web Application architecture-Access from PCs and PDAs - Access via WAP

PREQUISTIES DISCUSSION:

The aim of ubiquitous computing is to design computing infrastructures in such a manner that they integrate seamlessly with the environment and become almost invisible. The essence of that vision was the creation of environments saturated with computing and communication capability, yet gracefully integrated with human users.

5. PERVASIVE COMPUTING

Introduction:

- Our life in the future should be very carefree with little to no hassle.
- Less searching, faster and accurate access to information, when needed.
- Time and location boundaries will eventually be eliminated, resulting in a true information age style of civilization.
- Future devices will become more and more intelligent; they will start to talk among themselves to serve us better.

5.1 PERVASIVE COMPUTING INFRASTRUCTURE

The aim of ubiquitous computing is to design computing infrastructures in such a manner that they integrate seamlessly with the environment and become almost invisible. The essence of that vision was the creation of environments saturated with computing and communication capability, yet gracefully integrated with human users.

Pervasive – all around us

- Should be there where we need them
- Not go and get them

Human Centred

- Computers should adapt to the humans
- Computations enter our world

Must be unobtrusive and minimize user distraction

- Computers as we know it will disappear

Better ways of Computer-Human interaction

- The computers need to be aware of humans – Context

Pervasive computing integrates computation into the environment, rather than having computers which are distinct objects.

Other terms for pervasive computing:

- Ubiquitous computing
- Calm technology
- Things that think
- Pervasive internet
- Ambient intelligence
- Proactive computing
- Augmented reality
- Sentient Computing
- Urban Computing

Ubiquitous Computing

- Ubiquitous computing (ubicomp) integrates computation into the environment, rather than having computers which are distinct objects.
- Promoters of this idea hope that embedding computation into the environment
- Everyday objects would enable people to interact with information-processing devices more naturally and casually than they currently do, and in ways that suit whatever location or context they find themselves in.
- Ubiquitous computing encompasses wide range of research topics, including distributed computing,

Mobile computing
Sensor networks
Human-computer interaction
3
Artificial intelligence.

Pervasive – diffused among us

- It will make information available everywhere

Ubiquitous – State of being everywhere

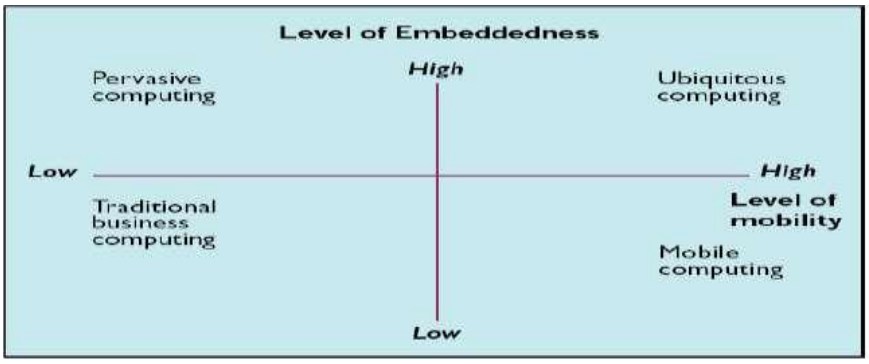


Fig.5.1 Level of Embeddedness

The most important characteristics of pervasive environments are:

- **Heterogeneity:** Computing will be carried out on a wide spectrum of client devices, each with different configurations and functionalities.
- **Prevalence of "Small" Devices:** Many devices will be small, not only in size but also in computing power, memory size, etc.
- **Limited Network Capabilities:** Most of the devices would have some form of connection. However, even with the new networking standards such as GPRS, Bluetooth, 802.11x, etc., the bandwidth is still relatively limited compared to wired network technologies. Besides, the connections are usually unstable.
- **High Mobility:** Users can carry devices from one place to another without stopping the services.

5.2 APPLICATION

5.2.1 Retail

Fig.5.2 Retail

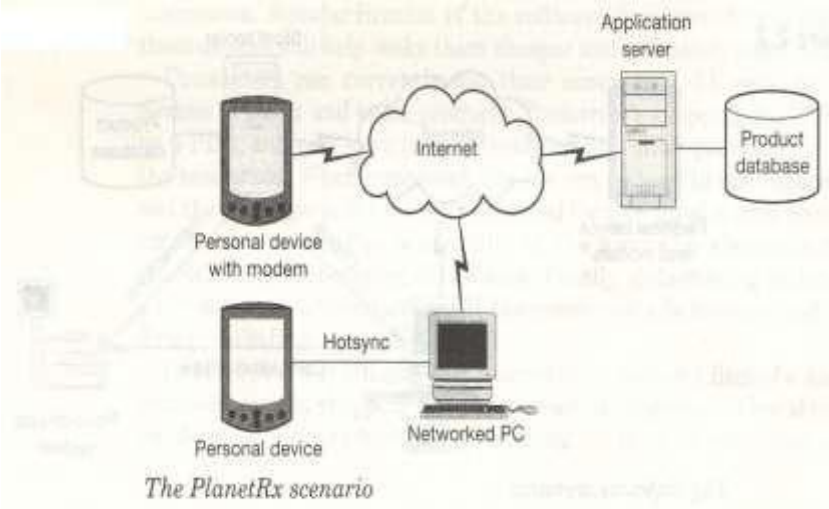


Fig.5.3 Air Line in and Booking

c) **Sales Force Automation:**

- Mobile workers relied on their portable computers in order to access and process data on the road
- Availability of wireless modems has enabled them to travel and allow them to access to enterprise data
- Mobile professionals to use the phone book and calendar while working out of office and stay in contact via email.
- Used to access mission-critical data – contracts, technical descriptions, small database and graphics etc at anytime & anywhere.
- Used to control the delivery of goods, update work assignment, submit orders, even billing info while on road.

Healthcare:

- Modern medicine already depends on a wide range of computerized devices, sensors, actors.
- Clinical professional learn about new methods and how to use them
- Access to laboratory results and surgical reports as well as ordering processes and physician directory look ups can be improved
- With respect to security, integrity of data and privacy assured though hardware, software and system design.

- Patient and clinical data must be exchangeable and accessible whenever needed but with restriction

- Modern healthcare systems are now using smart cards for patients and professionals.

Tracking:

1. Use of barcodes has revolutionized processes in many industries.

2. Provide fast and accurate identification of goods during transportation.

3. Barcodes on all products.

4. One dimensional barcode – encodes only few characters.

5. Two dimensional barcode – several hundreds of characters of information to be stored..

6. Allows tracking of goods – eg- airline luggage.

7. Tags are cheaper attached to each luggage and detected at certain points on the journey – enables airline to track individual pieces of luggage from check-in to baggage claim.

Car Information System:

Car manufacturers are very interested in using pervasive computing technology. Nowadays cars are equipped with more than 30 microprocessors. To connect the car with outside world, OGSI (open Service Gateway Initiative) is an industry group which defines features such as downloading software, application lifecycle management and gateway circuitry etc, which is implemented in JAVA,

Email Access via WAP and Voice:

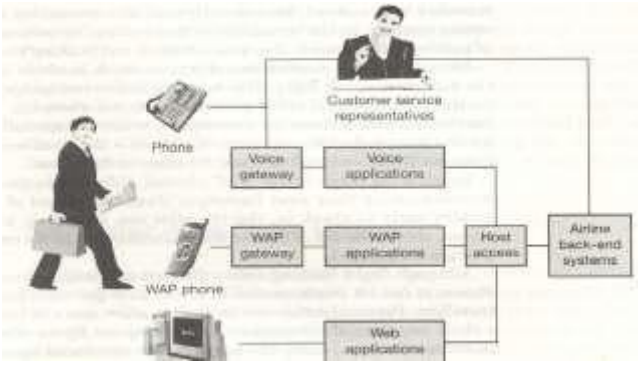


Fig.5.4 Email Access via WAP and Voice

5.3 DEVICE TECHNOLOGY

7. **Hardware and Battery**

Today, lithium ion (Li ion) batteries can be found in all sorts of electronic equipment. These batteries are lighter and have better energy density, resulting in more power delivered. The weight of a NiCad battery for a five-year-old mobile phone is often higher than the total weight of a modern mobile phone, including the Li ion battery. While the latter does have a lower capacity, it still offers longer talk time because of the reduced power requirements of modern devices.

Table. 1 gives an estimate of the expected standby and talk time for a mobile phone when used with typical batteries available today. The data is taken from the specification of three batteries of comparable size. The latest in battery technology is the emergence of lithium polymer cells, which use a gel material for the electrolyte. The batteries are made from a few thin and flexible layers, and do not require a leak-proof casing. This means the batteries can be made in almost any shape or size.

Expected lifetime for NiCad, NiMH, and Li ion batteries

Chemistry	**Standby time (h)**	**Talk time (m)**
NiCad	12-27	85-160
NiMH	**16-37**	**110-210**
Li ion	**21-50**	**170-225**

Table. 5.1

11. **Displays**

LCDs are already replacing the bulky cathode ray tubes.

Larger and more readable

Dramatic weight, size, and power consumption benefits of LCD technology outweigh their relatively high cost.

Today's PDAs usually feature dual-scan (DSTN) displays that control individual display elements via passive matrix addressing.

This technology consumes considerably less power than the thin-film transistor (TFT) active matrix technology.

This latter technology is more expensive, but is capable of significantly superior display performance and thus is generally used in portable computers.

6 Better and thinner displays will be available in the future based on the light-emitting organic diode (OLED) or light-emitting polymer (LEP) technologies.

7 OLED technology was invented about 15 years ago.

- It only recently became commercially attractive when the initial problems with the expected life and efficiency were solved.
- Instead of crystalline semiconductor material, organic compounds are used.
- The simplified manufacturing process of smaller structures and a rich selection of organic compounds enable OLEDs to be built in almost any size and colour.
- This will eventually allow manufacturers to create extremely thin displays that are flexible enough to be bent and shaped as required.
- Other new display technologies, such as chip-on-glass (CoG) and liquid-crystal-on-glass (LCoG), integrate the picture elements with transistors on a layer of glass.
- This allows manufacturing of extremely small displays, with a pixel size of only 10 micrometers.
- In contrast to regular small displays like those on the back of a camcorder, the microdisplays usually require some form of magnification.
- They can be found, for example, in projection systems and in head-mounted displays used with wearable computers.

5.3.3 Memory

- Memory is becoming cheaper, while the demand from applications is growing.
- Development is driven in part by smart phones, digital cameras, MP3 players and PDAs.
- For these mobile devices, the currently available technologies and their associated costs have reached a point where it is now feasible to integrate several megabytes of memory into a mobile device with an acceptable form factor.
- On PCs, permanent data can be stored on hard disk drives.
- For mobile devices, this is often not an option because neither the space nor the power supply is available.
- Recently, extremely small removable disk drives like the IBM Microdrive became available.
- Their capacity ranges between 340 MB and 1 GB, and is sufficient to store, for example, several hundred pictures when used in a digital camera.
- Other devices such as smart phones and PDAs store their operating system

code and application data in non-volatile Flash memory and battery-backed random-access memory (RAM) instead.

- These semiconductor-based technologies require less power and offer faster access than disk drives.

- The typical capacity of built-in memory in mobile devices ranges from 2 to 16 MB.

- Expansion slots allow additional memory modules to be plugged into the device, which in turn allow data exchange and replace removable media such as diskettes and CD-ROM for a PC.

5.3.4 **Processors**

- During the last couple of years, the clock rate of microprocessors and the processing power available from them has increased steadily.

- Rapid improvements in the CMOS manufacturing process have created ever-smaller structures and delivered higher and higher numbers of transistors per chip.

  0 At the same time, the processor core voltage was lowered from the industry standard 3.3 V in 1995 to 1.35 V in 2000.

  1 This means lower heat emissions, which in turn paves the way for new improvements like larger on-die caches.

  2 This, together with advances in packaging technologies, delivers the modern Central Processing Units (CPUs) found in mobile computers and PDAs today.

- **Intel's Speed Step technology**

  0 Recent processors include improvements in power management.

  1 These processors are capable of changing their internal clock frequencies and core voltage to adapt to changes in power supply.

  2 Newer designs are even capable of switching parts of the CPU on or off depending on whether the current calculations require them to be available.

  3 One such design is the Speed Step technology from Intel.

  4 While the system is connected to an external power supply, the full clock rate and core voltage are available to the processor, resulting in the maximum performance.

- When running on batteries, the clock rate and core voltage of the processor are reduced, resulting in significant power savings.

- The transition between both modes is very fast and completely transparent to the user.
- During the boot cycle, the Crusoe processor loads its software into a section of the main memory.
- Frequently used code parts are optimized during run-time and kept in a separate cache.
- A technology called LongRun promises to reduce the power consumption even more by reducing the processor's voltage on the fly when the processor is idle.
- The big advantage of this approach is that the Crusoe processor can be used to emulate almost any other processor and uses only a few watts, even with high clock rates.

5.4 HUMAN-MACHINE INTERFACES

- When reaching a haptic mark, the user feels a resistance generated by the motor against the turning direction.
- This force increases until a specific position is reached.
- When the knob passes that position, the force gets smaller again.
- This can be used to create the impression of a knob that can be put into a programmable number of positions.
- It allows a single knob to be used for navigating through a menu structure where each menu choice is represented by one position.

Navigation:

- In order to operate applications in mobile devices, the user navigates through a menu structure, often using special navigation keys.
- An example is the integrated cursor key that delivers signals for all four directions by pressing or moving it up, down, left, or right.
- Buttons that can be operated with the thumb while holding the device are especially suited for selecting entries from a menu list.
- These buttons can usually be turned or pressed.

Haptic Interfaces:

- The programmable rotating actuator with haptic feedback is available from VDO.1
- It is basically a rotating control with force-feedback and a push button integrated into one.

- Sensors detect the position of the knob and an integrated motor produces feedback of torque when rotated.

- The way in which the motor responds when turning the knob is programmable.

- Haptic marks define positions of specific feedback force changes.

5.4.1 Keyboards

- Depending on the size of the mobile device, keyboards offer either the full set of keys or a limited set of keys for data input.

- Adding a full keyboard with a typewriter layout to a mobile device inevitably makes these devices larger.

- On the other hand, limiting the number of keys will automatically make the operation of the device more complex.

- Sometimes keyboards cannot be used at all because the form factor of the device simply does not offer the space for it, or the device is used in environmental conditions where a keyboard wouldn't work.

- Therefore, some devices completely omit keyboards in favor of other input technologies, such as handwriting or voice recognition.

On Screen Keyboards

- Devices with a reasonably large touch-sensitive display often make a compromise by replacing the mechanical keyboard with a virtual on-screen keyboard.

- This does not allow touch typing but still offers a convenient method for text entry.

- Numbers and special characters can be entered after switching into another mode, which alters the keyboard layout accordingly.

5.4.2 Handwriting Recognition:

- With the availability of sufficient processing power and touch-sensitive displays, handwriting recognition became feasible.

0 The technologies available today differ widely in the amount of processing power and input precision they require.

1 Recognition of cursive handwriting is much more complex than recognition of individually printed letters.

3. Character Recognition:

0 Other methods limit the recognition to separated characters, and require the stylus to be lifted between letters.

1 These technologies usually achieve a very high recognition rate but require some cooperation from the user.

2 Usually there is a limited number of ways how an individual letter has to be drawn in order to be recognized by the device.

4. Speech Recognition:

0 Speech recognition has the advantage of being the most natural input method with only a minimum of requirements in terms of space required to integrate it into mobile devices.

1 However, it is also the most expensive technology in terms of computing power, and the most vulnerable in extreme environments.

2 Recognition of continuous speech is available in computers today, and will certainly become available in mobile devices too.

3 The most obvious devices for the integration of speech recognition are telephones.

4 Some mobile phones already allow the selection of an entry from the address book by just speaking the name.

5 In the future they will be operated entirely by voice, understand complex queries, and we may even be able to translate speech into other languages.

5.5 BIOMETRICS

Biometric authentication system capture the user's characteristics with a sensor, derive characteristic values, and compare this with a known reference. The result of the comparison is either 0, if the authentication was not successfully performed, or 1, if authentication was successfully performed. The image system gets the bifurcation points of finger line and compare with the existing one stored in the system. Concept is shown in the Fig.

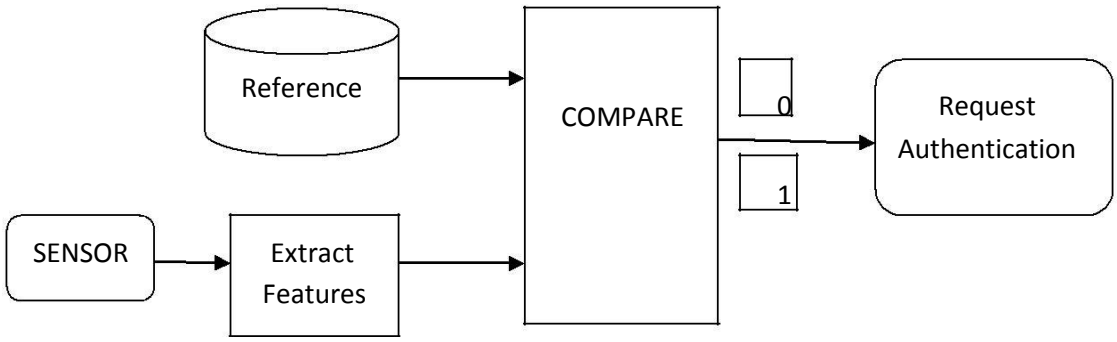


Fig. 5.5 Biometrics

5.5.1 Operating system

- The core functionality of every pervasive computing device is determined by its operating system.
- The major differences of operating systems for pervasive devices from the user's point of view are the human-machine interface, and the speed with which a task can be performed.
- For pervasive devices, there will likely be no equivalent to the Windows/Intel monopoly in the near future because pervasive devices do have a wide range of usages (from mobile phones to set-top boxes) with very constrained hardware

5.5.2. Palm OS:

Suitable and easy to use operating system for PDAs, optimized restricted features are available which leads to lower memory and CPU usage which results in longer battery life.**Features:** Enhanced Communication Support and Multimedia with Mobile Phones

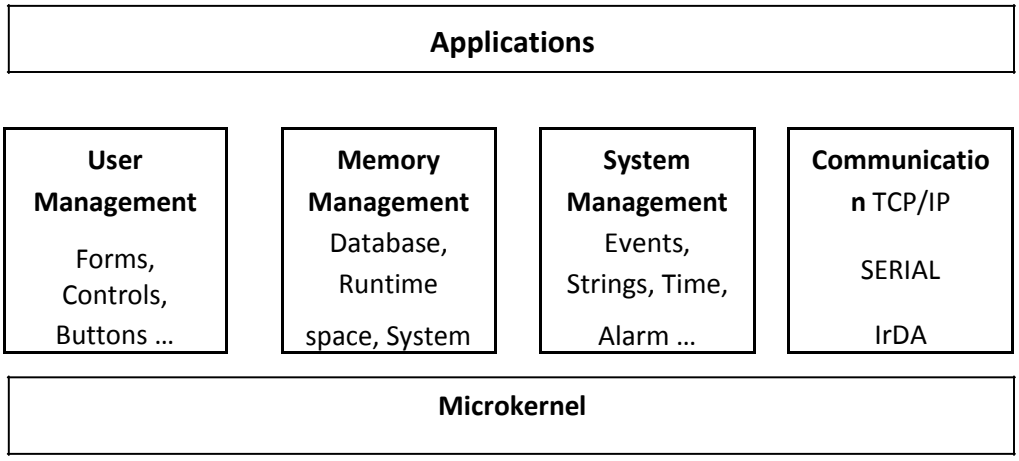


Fig. 5.6 Palm OS

1.User Management:

Single user operating system

2.Task Management:

One application runs at a time and can call other applications.

3.Power Management:

Power modes (sleep, doze and running)

4.OS size:

OS 3.5 is about 1.4 MB.

User Interface:

It recognizes only the palm handwriting alphabets, one button access to applications; minimize taps for often used operations.

Memory Management:

Applications should be well tested since if one application crashes then the system crashes. Thus memory is divided into dynamic heap which is execution based and clears on reset and storage is designed to hold permanent data.

Software can be developed with both C and C++ in Palm OS.

5.5.3 EPOC:

The EPOC operating system was designed specifically for phones. There are two versions: EPOC16 for 16-bit processors and EPOC32 for 32-bit processors.

Core operating system functionality:

- Heavily Multitasking.
- The base layer provides the fundamental APIs.
- The middleware layer provides the graphics, data, and other components to support the graphical user interface and applications.
- EIKON is the system graphical user interface framework.
- User Management:

  Single user operating system

- Task management

  Provides multitasking with a pre-emptive, priority-driven scheduler.

- User interface

  The EPOC user interface supports display, keyboard, and sound. . It is also responsible for handling the data and command input. Figure shows the EPOC user interface of an Ericsson device with a map application.

- Memory management

  EPOC has a memory management unit (MMU) concept to provide separate address spaces for each application. These tools include design patterns, stack clean-up heap failure, and heap-checking tools.

Programming languages supported by EPOC are C++, Java and OPL. C++ used to develop system development and high performance application programming.

5.5.4 Window CE:

Windows CE is an embedded operating system developed by Microsoft.

Windows CE 3.0 offers real-time support, a smart card subsystem for PC/SC compliant readers, is Unicode based, and supports grayscale and color graphics up to 32-bit depth. Windows CE is a modular operating system that can be configured by the device manufacturer. This is a result of the read-only memory (ROM)-based design of Windows CE, in contrast to more desktop-oriented, disk-based operating systems like Linux or BeOS. It can even be configured at runtime.

- The kernel provides memory management, task scheduling, and interrupt handling.
- The graphics/window/event manager (GWE) integrates the user interface functions of graphical output and user input.
- The object store is the persistent memory of Windows CE and includes files, the registry, and a database.
- Finally, the communication interfaces include infrared communication via IrDA, TCP/IP, and serial drivers.
- User management:

  Because Windows CE is designed for PDAs, it supports only one user.

- `Task management:

  The task manager supports 32 simultaneous processes and an unlimited number of threads.

- Operating system size:

  The Windows CE footprint can be as small as 400 kb for the kernel, up to 3 MB with all modules, and up to 8 MB including Pocket Word and Internet Explorer.

- User interface:

  Windows CE provides menu controls, dialog boxes, an: icons, and supports sound.

- Memory management:

  A protected virtual memory system that supports up to 32 MB memory per process protects applications against each other. There exists a special heap for the file system, registry, and object store that has a transaction service for ensuring data integrity. The object store can have a size up to 256 MB.

- Security*:* Windows CE has support for cryptography with a cryptographic library (Cryptographic Application Programming Interface, CAPI) to securely store information in memory. The kernel-loader authentication program can use public-key signatures to prevent unauthorized applications from running. Access to the data, however, will be slower because of the electrically erasable and programmable read-only memory (EEPROM) memory used instead of battery-backed RAM.

- Software development for Windows CE

Since Windows CE is based on the Win32 API development tools, such as Visual C++ or Visual Basic, available for this API.

5.5.5 QNX Neutrino:

QNX is a real time operating system consisting of microkernel surrounded by a collection of optimal processes that provides UNIX based system services. Due to microkernel architecture even if the file system driver or network driver crashes, still the system will work which leads to stable system. QNX is very well suitable for car devices.

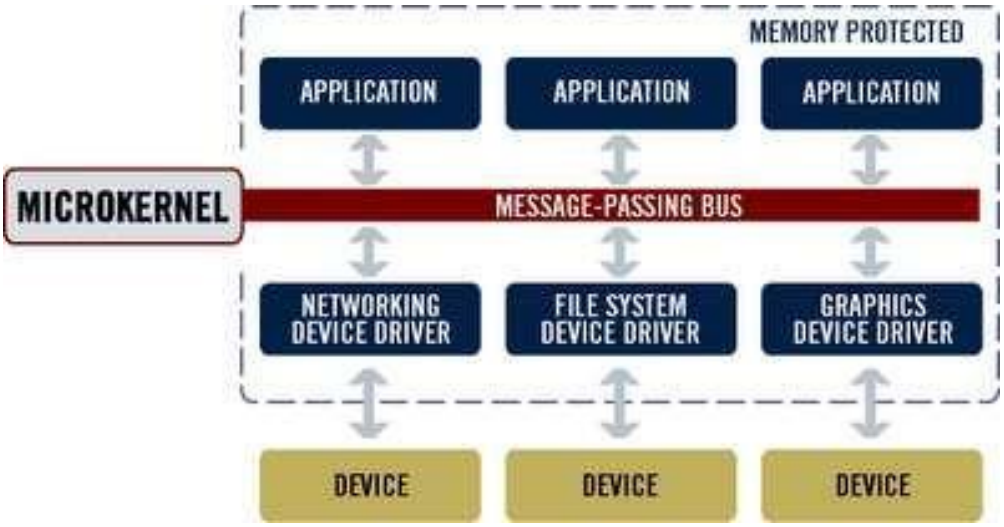


Fig. 5.7 QNX Neutrino

- User management:

  QNX supports single user.

- Task management:

  Supports real multitasking.

- User interface:

  Consist of micro graphical user interfaces and widgets for easy interface.

- Memory management:

  QNX has an MMU concept for separation of address space of applications. Different application runs on different threads.

Software development for QNX is in C language.

5.5.6 Be OS:

Be OS is highly optimized for multimedia application. It posses sound and graphic processor. It deals with 64 bit file system

The architecture is based on a symmetric multi processor model, allowing each

processor full access to resources and also it provides pre-emptive multitasking and pervasive mulltithreading (rapind switching between several task).

- User management:

  It supports multiuser as like standard operating system.

- Task management:

  Pre-emptive multitasking by pervasive threads enables the task management much speeder.

- Memory management:

  Provides memory protection between applications and virtual memory support.

Software development is done using C/C++.

3.2.6. Embedded Linux:

Embedded Linux is a stripped down operating system with special support to pervasive devices. Mainly used for handheld devices.

The core features are Configurable kernel, Scalability and Networking.

- User management:

  It supports multiuser as like standard operating system.

- Task management:

  Preemptive multitasking with optional real time scheduler is implemented.

- Operating System Size:

  Depending on the configuration, the size of the kernel can range from 200 KB to several megabytes.

- User Interface:

  x-Window system for user interface and have striped down version to save memory.

- Memory management:

  Supports MMUs to provide memory protection between applications and virtual memory for paging memory to hard disc.

Software development is done using C/C++ and JAVA.

5.6 DEVICE CONNECTIVITY:

5.6.1 PROTOCOLS

Standardized protocols are basic requirements for pervasive computing devices.

Wireless protocols supporting IP (Mobile IP) is needed for existing pervasive devices. Another issue is regarding the consistency of database and their data. Connected pervasive computing devices to the distributed environment and their services are important in pervasive domain. Data delivery, in time and integrity is maintained by transaction protocols. Protocols such as WAP and Bluetooth having role in pervasive domain.

5.6.1 Wireless Protocols:

Wireless protocols are enhanced mainly for PDAs and Mobile phones. The protocols such as WAP, Bluetooth, IrDA, Object Exchange Protocol (OBEX) and other mobile phone technologies are concentrated.

WAP/WML:

- WAP is a technology which mainly pronounces on rapid access to internet.
- WAP works for Ericsson, Nokia, Motorola.
- Integrates telephony services with browser technology.
- WAP application includes e-commerce, online banking and messaging.
- WAP is similar to HTTP and then optimized for narrow band channel.
- WML is used for textual format and WBXML is used for compressed binary format.

OBEX:

- OBEX was originally designed for IrDA later due to creation of Bluetooth it was emerged as high level protocol.
- OBEX simplifies communication enabled application by using push and pull commands.
- Two models- The Session Model and The Object Model.
- Session Model: Constructs the dialog between two devices by packet based client/server request/response model.
- Object Model: It contains information about the object being sent with header and the header has name, length, descriptive text and the object body.
- Security is provided using challenge response scheme.

Bluetooth:

- It is a short range communication and data exchange.
- The various characteristics are
- Frequency Band: operates at 2.45 GHz.
- Security: Authentication based on private key and encryption.
- Transmitting Capability: It is omnidirectional and range upto 10m.
- Bandwidth: Data Transfer Rate of 1 Mbps.
- Speech: Support digital speech channel.
- Cost: Reasonable Price.

IrDA:

- Pervasive Computing device are IrDA-Data and Infrared Mobile Communication (IrMC)**.**
- Frequency Band: It is a physical transport medium.
- Security: Higher level protocol service.
- Transmitting Capabilities: Point-point connection with narrow angle between sender and receiver. Range of communication is about 0-30 cm.
- Bandwidth: Data rate-4 Mbps
- Speech: one digital speech channel.
- Cost: very cheap.

5.6.2 **Mobile Phone Technologies:**

Cellular system for mobile communication:

Mobile environment is in need of electromagnetic waves at frequency 1Ghz. Small antennas emit only limited power and thus mobile and cellular environment is designed as such the transmitters are placed in small grid like space where the frequency in one grid is reused in another grid which is placed far away and thus reusability is enhanced.

The grid is called as cell. Base station within the cell and the mobile station should be connected with stronger signal.

Features	1G	2G	2+G	3G
Protocols	AMPS, C-Net	GSM, TDMA, CDMA	GPRS, HSCSD, EDGE	UMTS,W-CDMA
Technology	Analog Circuit Switched	Digital Circuit Switched	Digital or Packet Switched	Digital Packet Switched
Speech Quality	Poor	High	High	High
Bandwidth	Low	Low	Medium	High
Security	None	Low to High	High	High

Table. 5.2

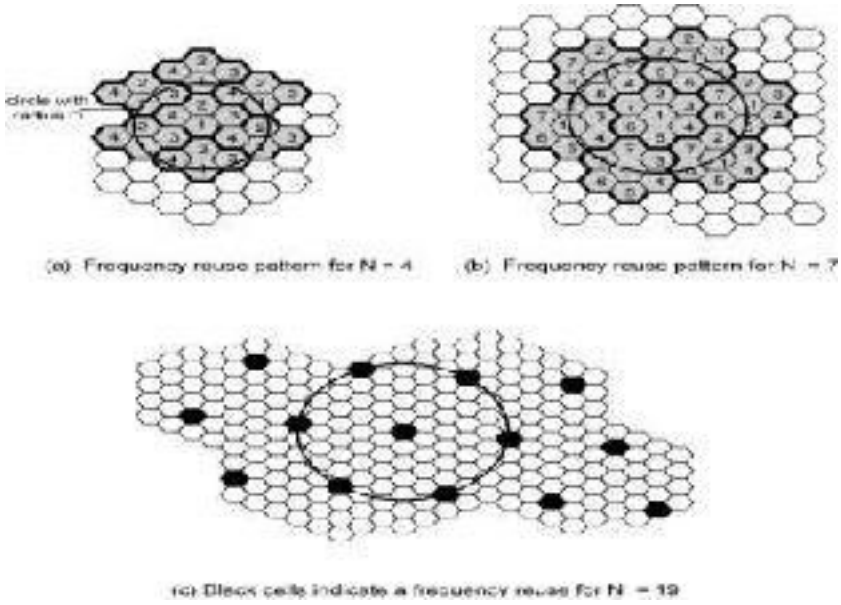


Fig. 5.8 Cells indicating Frequency

5.6.3 Mobile Internet Protocol:

Once the mobile connection is available, the next step is to access internet and thus in mobile environment we are moving to IPv6 working on Mobile IP.

Standard internet protocol and mobile devices:

The standard protocol used to access internet is IPv4 which provides with unique address to nodes. Similar to postal addresses IP addressing is provided to each node. The data will be in the form of packets and the packet contains source address, source port, and destination address and destination port. Then in order to reach the destination, it is traversed through intermediate nodes and this is achieved by routing.

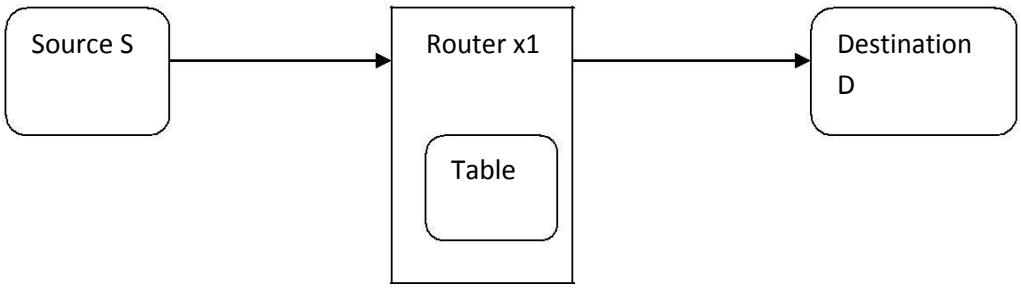


Fig. 5.9 Standard Internet Protocol and Mobile device

Steps to have connection to a mobile node:

- Discover a care-of-address
- Register the care-of address
- Tunnel the care-of address

Changes to Internet Protocol Version 6:

IPv6 includes the features of IPv4 but the significance is address configuration and neighbour discovery. Ipv6 expects all nodes to implement strong encryption and authentication features. Supports mobility at greater extent.

5.6.4 Synchronization and replication Protocols:

Synchronization also called as replication in concern with database. For instance, if the calendar application is stored in the PC and PDAs. The updation in one device related to the application should be reflected in the other device also. Software updation also comes under the same category.

Steps in Synchronization Protocol:

- Presynchronization: Before the actual synchronization authentication (authenticates client and server), authorization (allowed to perform action) and determination of device capabilities (maximum buffer size) should be done.

 Synchronization: Data exchanged here, al; local Ids are mapped with global Ids. Only updated entries are allowed to exchange. if both the partners are updating the same entry then the conflict is resolved by deleting the duplicate one.

SyncML

Scope of SyncML

 XML based framework for data synchronization
 Message oriented data exchange protocol
 Transport agnostic
 Universal deployment
 Extension for device management

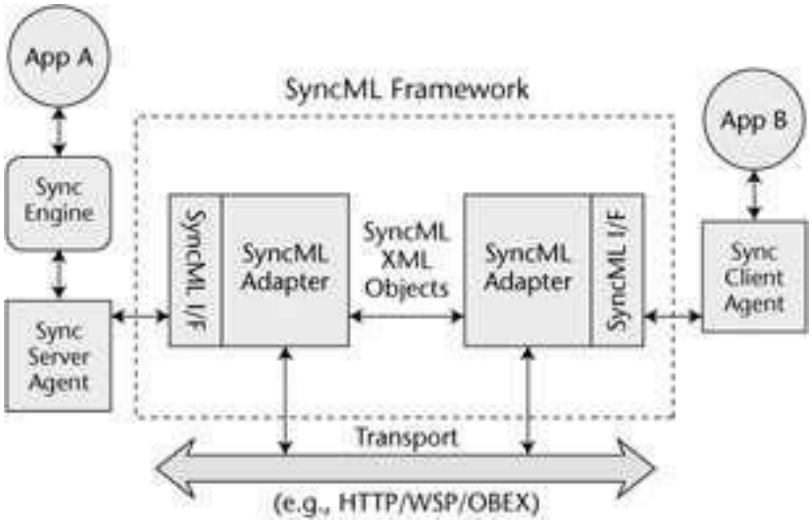


Fig. 5.10 SyncML Framework

5.6.5 Distributed Services:

Jini:

A Jini system consists of the following parts:

 A set of components that provides an infrastructure for federating services in a distributed system.
 A programming model that supports and encourages the production of reliable distributed services.
 Services that can be made part of a federated Jini system and that offer functionality to any other member of the federation.
 Discovery: finds other members in the jinni community and joins to enable networking capability.
 Look-up: Search based on object type.
 Leasing: Stable and self healing. Resources are only leased since each have some timestamp.

 Remote Events: Sending notification to the participants.
 Transactions: similar to database transaction.

Jini scenario for accessing a remote service- first step

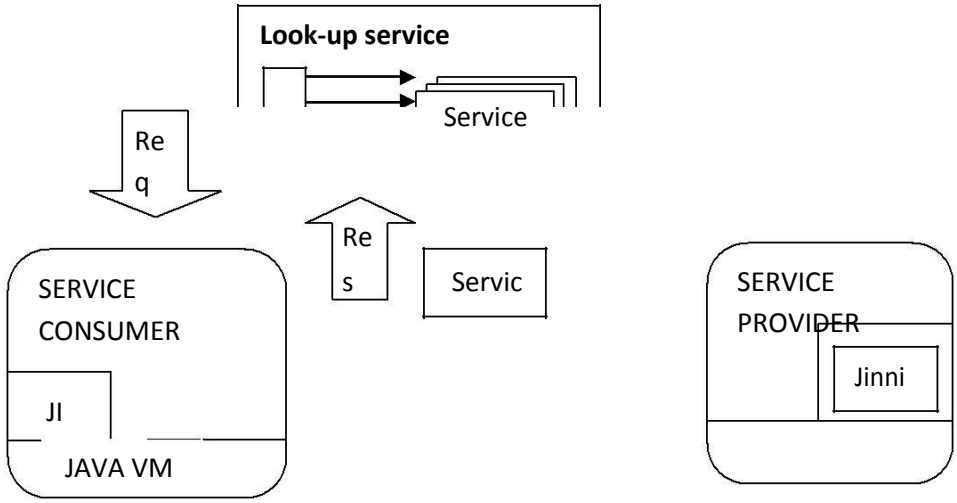


Fig. 5.11 Jini scenario for accessing a remote service- first step

Jini scenario for accessing a remote service- second step

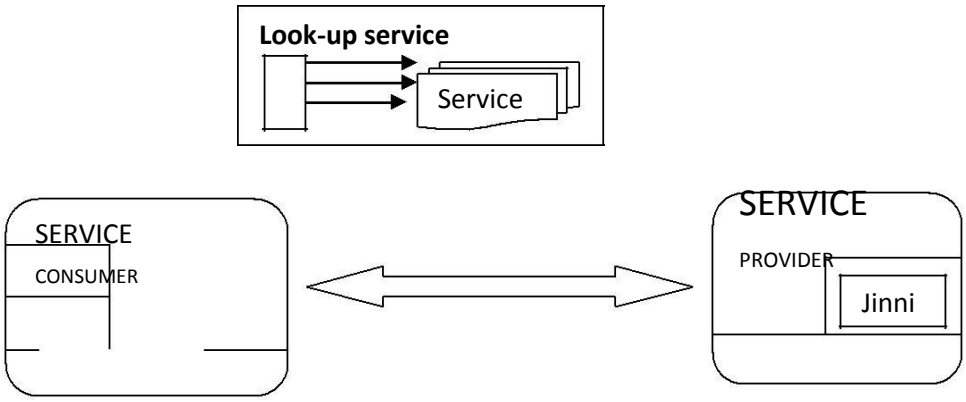


Fig. 5.12 Jini scenario for accessing a remote service- Second step

5.6.6 Message and Transaction Based Protocol:

Messaging and Transaction Technology:

- Assured Message delivery and atomic operation are the two different technologies. Standards are needed in order to implement these technologies.
- SOAP provides message exchanging structure and sends information between peers.

- Message queuing is another concept which emphasis on storing of messages and as soon as the connection established, the message will be delivered. This is suitable for asynchronous network.
- Topology can be build with multiple queues. One to one, one to many, many to one.

5.7 SECURITY:

5.7.1. **Security Concept**

Identification:

Various identification methods are used in pervasive devices. One is user have to type the one which is stored in the device otherwise the user can be identified using certificates.

Authentication:

The most common way of authenticating peers is by using user name and password. For a client server connection authentication is ensured by establishing secure socket layer. WTLS is used in WAP phones. Wireless Authentication Module (WIM) is for smart cards. SIMs will be authenticating the mobile terminal in GSM network.

When smart card at the client and authentication software is installed in the server then the server will produce a challenge to the client and in return the client gives the signature and the certificate which will be checked by the server.

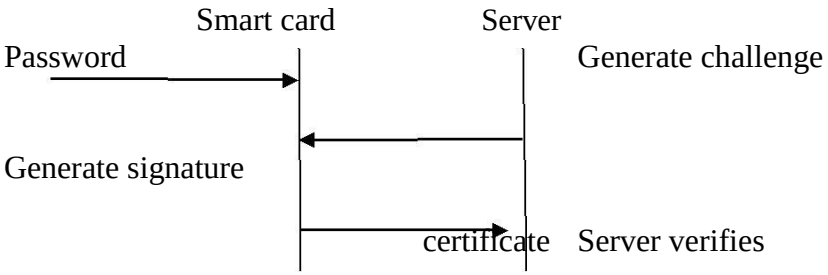


Fig. 5.13 Authentication

Authorization:

Authorization is based on from which device the user is accessing. In a banking transaction, If the user access from PC using user identification and password and this will be the same for user from WAP phone, PDAs but when user is from normal telephone then there is no entry for user identification and password.

Transaction Authorization:

The applications such as home-banking, placing orders, brokerage should authorize every individual transaction.

Digital Signature Endorsed by a Password:

Digital signature ensures that the particular request or message comes from the authorized user since the user will be having the unique key from which a token will be

generated and the token generates the required signature which will be passed to the server for verifying the transaction.

Transaction Authorization Number:

TAN (in blocks) is a secret number generated for legitimate users in an organization. The users have to acknowledge immediately after receiving the TAN and the user should ensure that it will be kept secret. Whenever the user contacts the server, TAN should be entered which will be compared with the stored one.

Non-Repudiation:

User A denying user B and user B denying user A is called non-repudiation. This can be avoided by using efficient digital signature.

5.7.2. Device Security

Trojan Horses

A **Trojan horse**, or **Trojan**, is a non-self-replicating type of malware which gains privileged access to the operating system while appearing to perform a desirable function but instead drops a malicious payload, often including a backdoor allowing unauthorized access to the target's computer. These backdoors tend to be invisible to average users, but may cause the computer to run slow. Trojans do not attempt to inject themselves into other files like a computer virus. Trojan horses may steal information, or harm their host computer systems. Trojans may use drive-by downloads or install via online games or internet-driven applications in order to reach target computers.

Security in WAP phones

WTLS is the layer that provides most of the security functionalities for WAP applications. These functionalities include client-server mutual authentication, privacy, data integrity, and non-repudiation.

WTLS and TLS: The design of WTLS is based upon TLS (Transport Layer Security) that is in turn built upon SSL (Secure Socket Layer). TLS has become de facto security protocol for ensuring end-to-end security for Internet communications. Similar to TLS, WTLS requires the client and the server negotiate and agree on a set of security parameters during the handshake before the communicate channel can be established. Once handshake succeeds, the client and the server can exchange information using the secrets known to both ends of the channel. Since WTLS resembles TLS so much, one could consider that the WTLS provides the same level of security as TLS does. However, due to the limitations of wireless communications and the modifications WTLS made to accommodate to these limitations, it has been shown that WTLS is vulnerable to a variety of known attacks such as plaintext recovery attacks and datagram truncation attacks.

Security in PDAs

Downloading and installing arbitrary software makes prone Trojan horse attacks.

5.7.3. Server Side Security

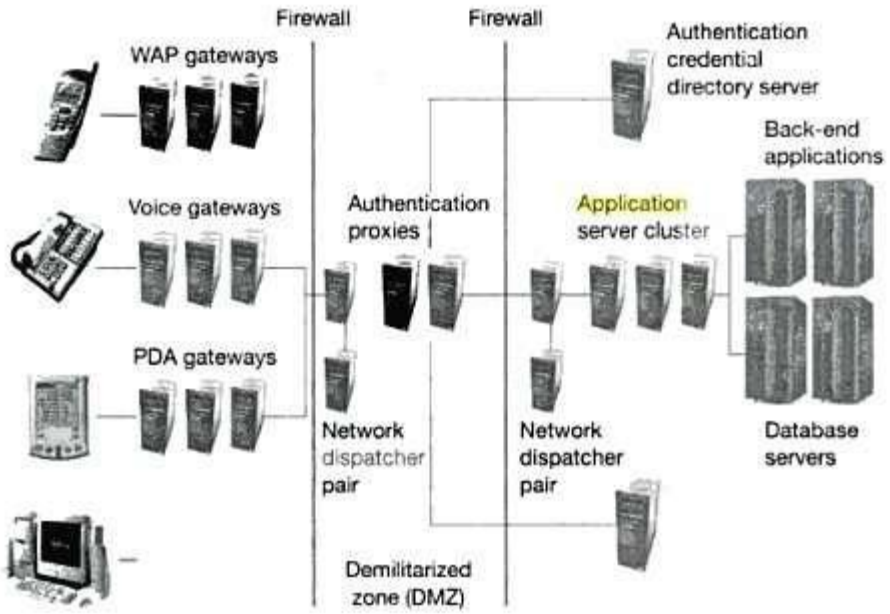


Fig.5.14 Server Side Security

5.7.4. Cryptographic Algorithms:

A cryptographic algorithm, or cipher, is a mathematical function used in the encryption and decryption process. A cryptographic algorithm works in combination with a key—a word, number, or phrase—to encrypt the plaintext. The same plaintext encrypts to different ciphertext with different keys. The security of encrypted data is entirely dependent on two things: the strength of the cryptographic algorithm and the secrecy of the key. A cryptographic algorithm, plus all possible keys and all the protocols that make it work comprise a cryptosystem

Symmetric cryptographic algorithm:

In secret key cryptography, a single key is used for both encryption and decryption. The sender uses the key (or some set of rules) to encrypt the plaintext and sends the cipher text to the receiver. The receiver applies the same key to decrypt the message and recover the plaintext. Because a single key is used for both functions, secret key cryptography is also called symmetric encryption. With this form of cryptography, it is obvious that the key must be known to both the sender and the receiver; that, in fact, is the secret. The biggest difficulty with this approach, of course, is the distribution of the key.

Asymmetric cryptographic algorithm:

Public or asymmetric key cryptography involves the use of key pairs: one private key and one public key. Both are required to encrypt and decrypt a message or transmission. The private key, not to be confused with the key utilized in private key

cryptography, is just that, private. It is not to be shared with anyone. The owner of the key is responsible for securing it in such a manner that it will not be lost or compromised. On the other hand, the public key is just that, public. Public key cryptography intends for public keys to be accessible to all users. In fact, this is what makes the system strong. If a person can access anyone public key easily, usually via some form of directory service, then the two parties can communicate securely and with little effort, i.e. without a prior key distribution arrangement.

Data Encryption Standard (DES):

DES is the archetypal block cipher — an algorithm that takes a fixed-length string of plaintext bits and transforms it through a series of complicated operations into another cipher text bit string of the same length. In the case of DES, the block size is 64 bits. DES also uses a key to customize the transformation, so that decryption can supposedly only be performed by those who know the particular key used to encrypt. The key ostensibly consists of 64 bits; however, only 56 of these are actually used by the algorithm. Eight bits are used solely for checking parity, and are thereafter discarded. Hence the effective key length is 56 bits

Triple Data Encryption Standard (Triple DES):

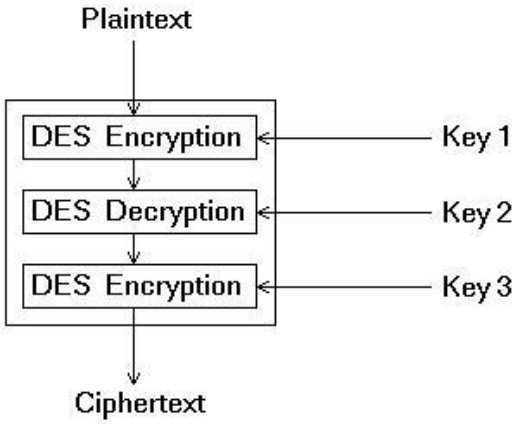


Fig.5.15 Triple Data Encryption

AES:

AES is based on a design principle known as a substitution-permutation network, and is fast in both software and hardware. Unlike its predecessor DES, AES does not use a Feistel network. AES is a variant of Rijndael which has a fixed block size of 128 bits, and a key size of 128, 192, or 256 bits.

RSA:

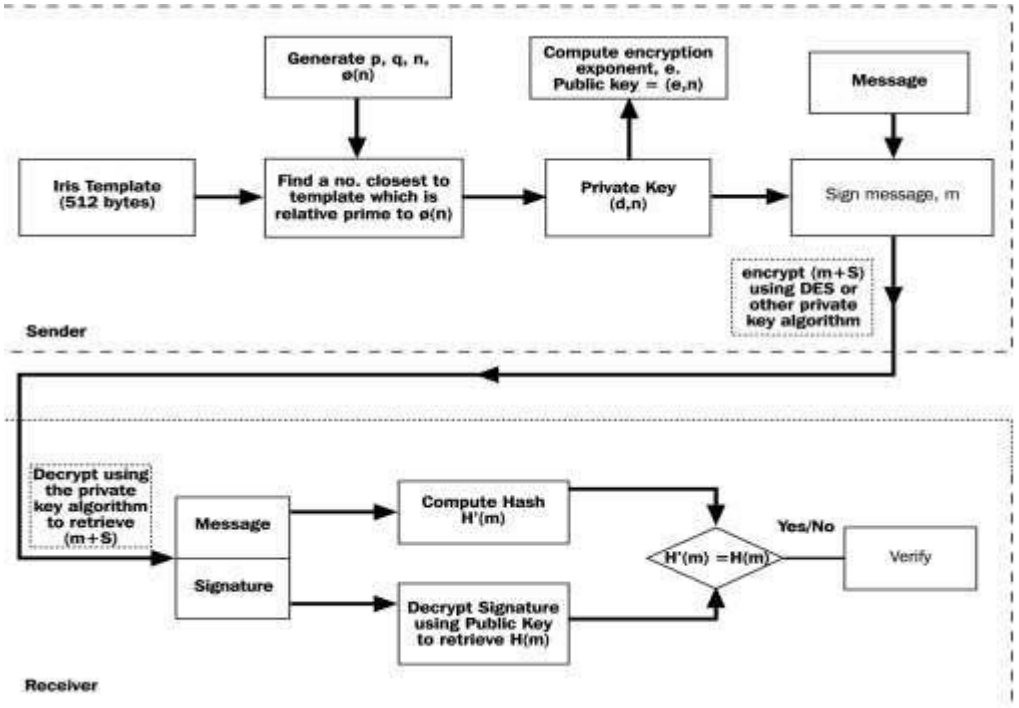


Fig. 5.16 RSA

Digital Signature:

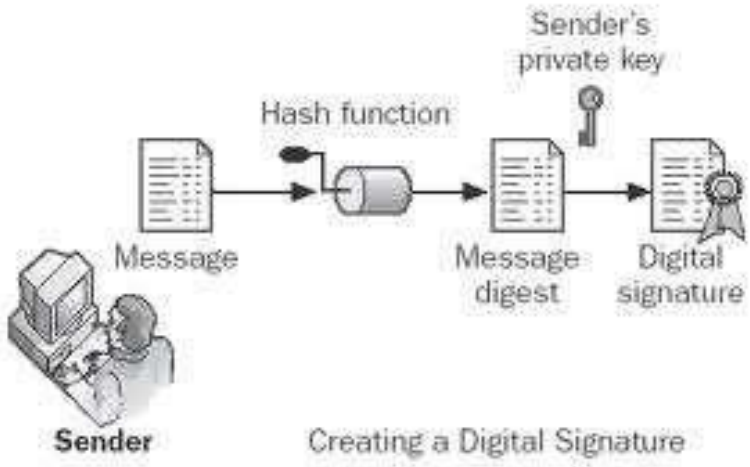


Fig.5.17 Digital Signature

Elliptic Curve Cryptography:

Public-key cryptography is based on the intractability of certain mathematical problems. Early public-key systems are secure assuming that it is difficult to factor a large integer composed of two or more large prime factors. For elliptic-curve-based protocols, it is assumed that finding the discrete logarithm of a random elliptic curve element with respect to a publicly known base point is infeasible. The size of the elliptic curve determines the difficulty of the problem. The primary benefit promised by ECC is a

smaller key size, reducing storage and transmission requirements—i.e., that an elliptic curve group could provide the same level of security afforded by an RSA-based system with a large modulus and correspondingly larger key—e.g., a 256-bit ECC public key should provide comparable security to a 3072-bit RSA public key.

For current cryptographic purposes, an *elliptic curve* is a plane curve which consists of the points satisfying the equation

$$y^2 = x^3 + ax + b,$$

along with a distinguished point at infinity, denoted ∞. (The coordinates here are to be chosen from a fixed finite field of characteristic not equal to 2 or 3, or the curve equation will be somewhat more complicated.)

5.8 DEVICE MANAGEMENT:

This section covers device management- Software Distribution and Solving Device Management Task.

Device Management Challenges:

- Tracking device location
- Device-user relationship
- Version control for software and hardware
- Software updates
- Installing new software
- Providing secure access

Software Distribution:

Issues when new software is installed and when updating the older one,

- Hardware Capabilities: Whether the processor type, memory and features are compatible to install.
- Hardware Version Management: Software downloading should be compatible with hardware version.
- Software Version Management: Whether the new version software runs in older version of OS.
- Library Management: Tracking all libraries needed for each program version.
- Devices are not always connected: Tracking of updates received.
- Insecure Connection: Distributed Software and data needs to be protected.
- Unstable Connection: Connection should not break during updates.
- Operating System Updates: Device should be operational after the updates.

5.9 PERVASIVE WEB APPLICATION ARCHITECTURE:

Requirements of computational infrastructure:

- Failure management.
- Security.
- Performance.
- Dependability.

The architecture for pervasive computing applications that support multiple devices, such as PCs, WAP phones, PDA and voice-only phones enabled to access Web servers through voice gate-ways. The architecture addresses the special problems associated with pervasive computing, including diversity of devices, markup language and authentication methods shows how pervasive computing applications based on this architecture can be secured. Users have many different devices that look and behave in very different ways. Examples of several kinds of pervasive computing devices include WAP phones, PDAs, and voice-recognition devices. These devices proving different user interfaces, use different markup languages, use different communication protocols, and have different ways of authenticating themselves to servers.

Scalability and Availability are the two important issues because pervasive architecture should as scalable to meet the large amount of user who are subscribing for applications. According to scalability, when a user accessing the application and if it is not available then the user will assume that it does not works and the user will switch on to next service. Thus the service should be available whenever needed.

A scalable topology for pervasive computing is shown in the Fig. This shows several gateways used for accessing the server.

- WAP gateway which executes WTLS protocol in the direction of clients and SSL in the direction of servers.
- Voice gateways use voice recognition engine which consumes more power.
- PDA gateway for the PDAs.
- Network Dispatcher which routes incoming request to the approporiate server. Support handling of HTTP request from a particular client is always sent to the same server to avoid repeated SSL handshakes.
- Two dispatchers available to increase availability.
- Two firewalls are placed to perform secure connection.

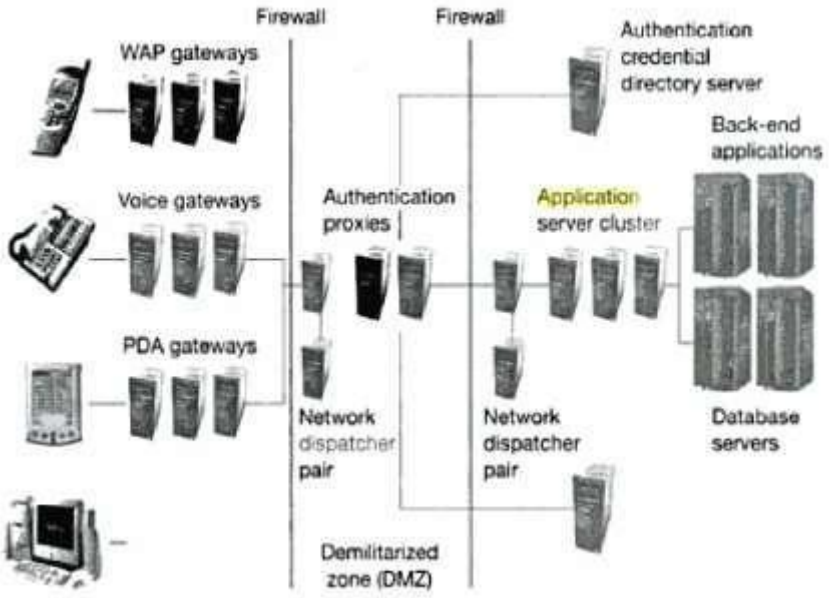


Fig. 5.18 A Scalable Topology for pervasive computing web applications

- Authentication proxy is placed which checks all incoming request according to security policy defined and use the credentials given by the client in order to secure transactions. Authentication proxy consumes significant computing power.

- Cluster of application servers are arranged in order to add additional machines if load increases

Development of Pervasive Computing Web Application

- Business Logic Designers, User Interface Designers, Application Programmers and experts for existing database system are the roles assigned to implement web application**.**

- Application flow is designed by Business Logic Designers.

- Look and feel of the system is given by User Interface Designers.

- Programmers are concerned with technologies such as HTML and JSP and implementing the application logic.

- Those experts who monitor the gateways are responsible in knowing the technologies such as WML, Voice XML.

Pervasive Application Architecture:

- The model-view-controller (MVC) pattern is a good choice when implementing Web applications.

- Standard mapping of the pattern to servlets, JSPs, and EJBs, where controller is implemented as a servlet, the model implemented as a secure EJBs, and the views as JSPs..
- As devices are very different from each other, we can assume that one controller will fit all device classes. In the MVC pattern the controller encapsulates the dialog flow of an application.
- This flow will be different for different classes of devices, such as WAP phone, voice-only phones, PCs, or PDAs.
- Thus, we need different controller for different classes of devices.
- To support multiple controllers, we replace the servlet's role to that of a simple dispatcher that invokes the appropriate controller depending on the type of device being used.

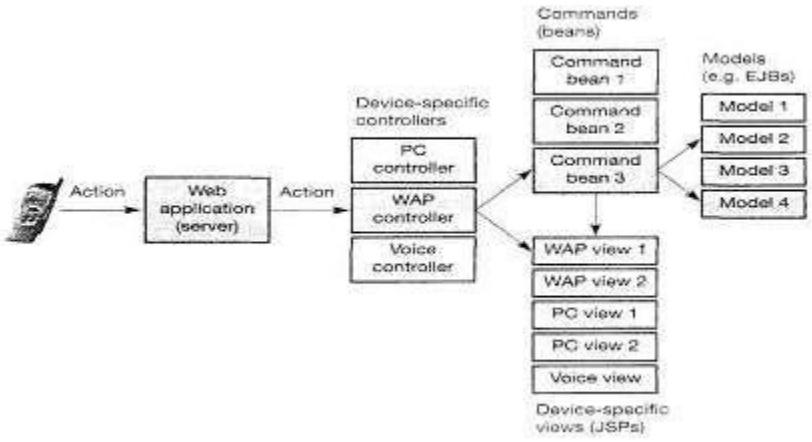


Fig.5.19 MVC Pattern applied to Pervasive Computing Applications

Securing Pervasive Computing Application

- Web applications have to be secured by appropriate encryption, authentication, using authorization mechanisms.
- The secure pervasive access architecture is designed to process client requests on the application server in a secure and efficient way.
- It addresses user identification, authentication, and authorization of invocation of application depending on configurable security policies.

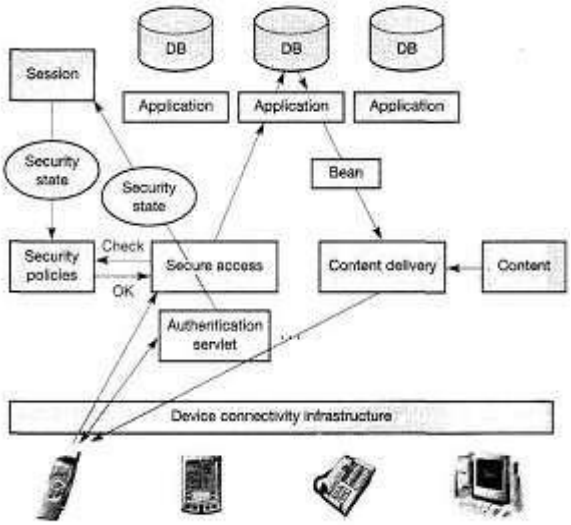


Fig 5.20 Secure Pervasive Access Architecture

All incoming requests originate from the device connectivity infrastructure. This infrastructure may include different kinds of gateways that convert device specific requests to a canonical form, i.e. HTTP request that may carry information about the device type, the desired language and the desired reply content type, e.g. HTML, WML, or Voice XML. Examples of gateways in the device connectivity layer are voice gateways with remote Voice XML browsers, WAP gateways, and gateways for con-necting PDAs.An important function that the device connectivity layer must provide is support of session cookies to allow the application server to associate a session with the device.The secure access component is the only system component allowed to invoke application functions. It checks all incoming requests and calls application functions according to security policies stored in a database or directory.A particular security state - part of the session state – is reached by authentication of the client using user-ID and password, public-key client authentication, or authentication with a smart card, for example. If the requirements for permissions defined in the security policy are met by the current security state of a request's session, then the secure access layer invokes the requested application function, e.g. a function that accesses a database and returns a bean. Otherwise, the secure access component can redirect the user to the appropriate authentication page.Typically, the secure access component will be implemented as an authentication proxy within a demilitarized zone as shown earlier.Finally, the output generated by the application logic is delivered back to the user in a form appropriate for the device him or her issuing. In the Figure, the information to be displayed is prepared by the application logic and passed to the content-delivery module encapsulated in beans.The content-delivery module then extracts the relevant part of the information from the bean and renders it into content that depends on the device type and desired reply content type, for example by calling appropriate JSPs.The content-delivery module delivers the content generated in the previous step via the device connectivity infrastructure that converts canonical responses (HTTP responses) to device-specific responses, using-appropriate gateways.For example, if a user accesses the system via a telephone, the voice gateway receives the HTTP response with Voice XML content and leads an appropriate 'conversation' with the user, finally resulting in a new request being sent to the server.

5.9 ACCESS FROM PCs:

Smart card based authentication via the internet:

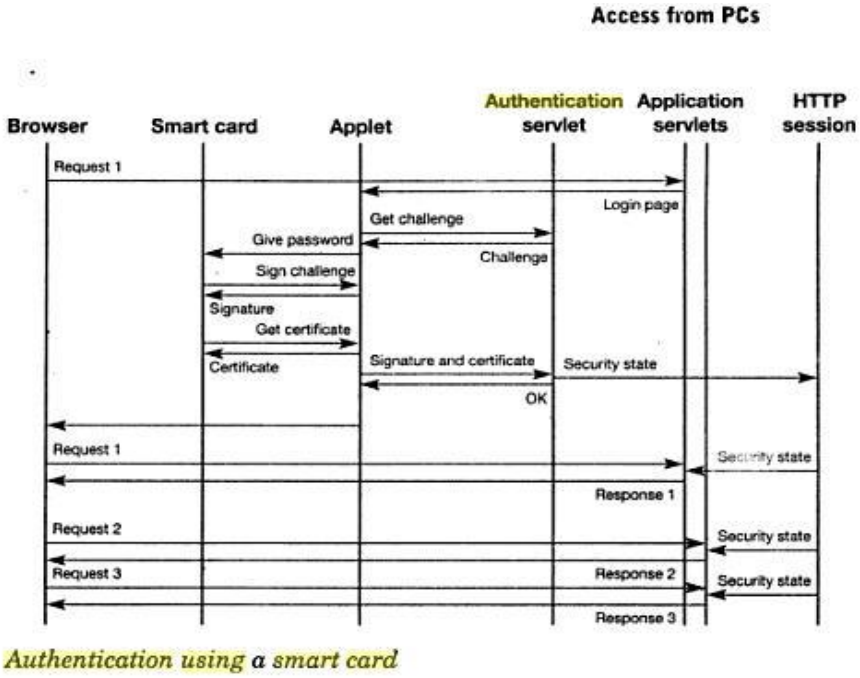


Fig.5.21 Smart card based authentication via the internet 5.9.1 Implementation:

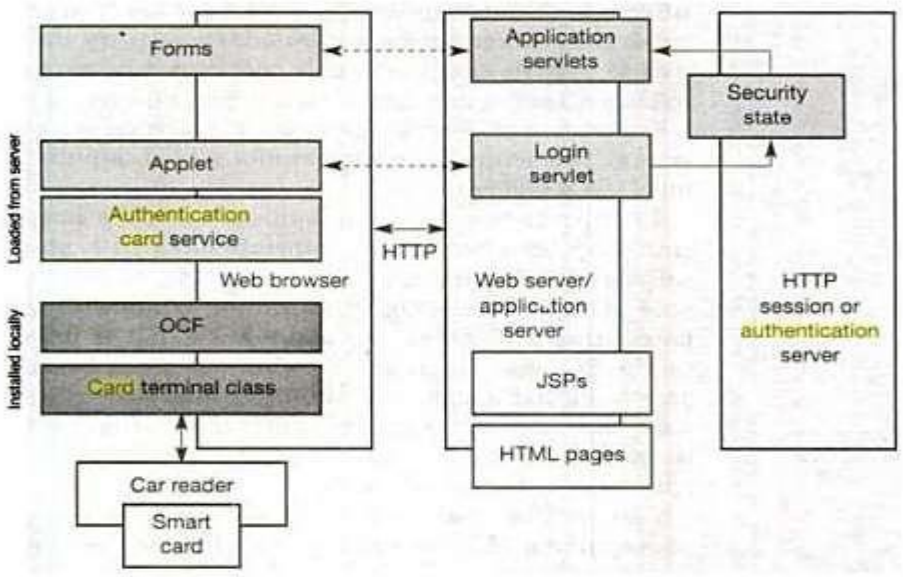


Fig.5.22 Architecture Overview

Smart Card Login Controller:

- Get Logon Page
- Get Challenge
- Logon allowing authentication

Authentication Applet:

- Authentication Servlet URI
- Success Page
- Failure Page

Authentication Card Service:

Verifying the PIN of smart card.

Designing goods ordering application via pc:

- Login Page
- Menu
- Items List
- Purchase Confirmation
- Purchase History

5.10 ACCESS VIA WAP:

WAP Functionality:

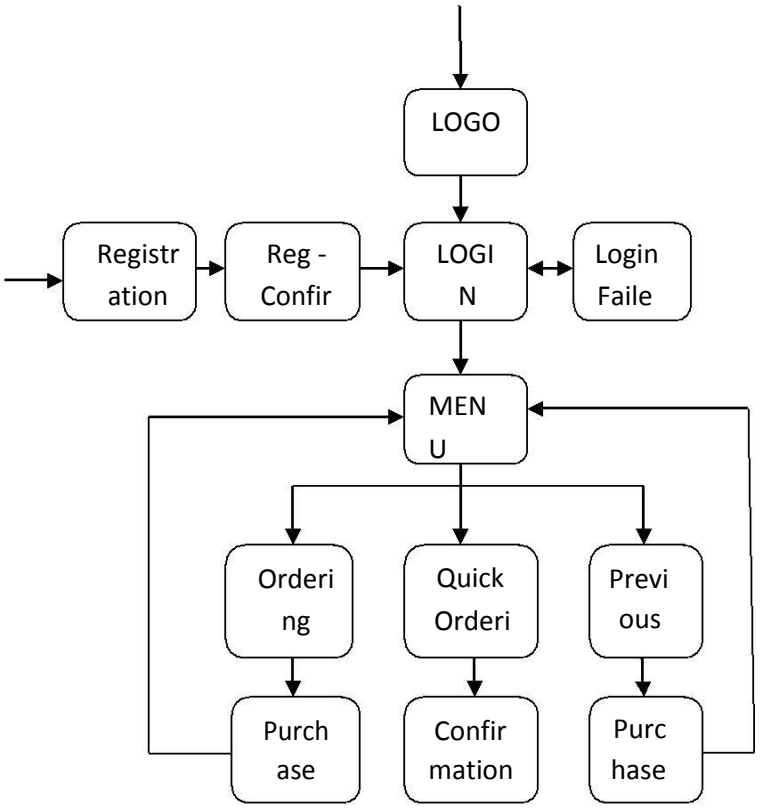


Fig. 5.23 Overview of Infrastructure accessing from WAP phone:

Customer calls originating from mobile device, it crosses WAP gateway-dial in hardware and finally reaches the application server. Three Servlet are implemented such as Registration, Login and Shop. The Controllers are Registration controller, 2 Login controllers, one is PC specific and another one is device specific. The whole process is very similar to MVC pattern.

5.10.1 ACCESS FROM PDAs:

The same order processing is done from PDAs which will follow the same MVC pattern and secure pervasive infrastructure. The device specific capabilities of PDAs demand special attention to the layout and design of web pages, or the tool and formats used by client application. Three implementation are layered. Necessary support will be given for accessing from PDAs.

SIGNIFICANCE:

Pervasive computing refers to the ubiquitous presence of computing in both mobile and embedded environments, with the ability to access and update information anywhere, anyplace and anytime. This idea has been around for a long time, but only now is pervasive computing truly taking root.