EC6802 NOTES - REGULATION 2013

UNIT – I WIRELESS LANs

INTRODUCTION

1.1WIRELESS LAN

A wireless LAN (or WLAN, for wireless local area network, sometimes referred to as LAWN, for local area wireless network) is one in which a mobile user can connect to a local area network (LAN) through a wireless (radio) connection. The IEEE802.11 group of standards specify the technologies for wireless LANs. 802.11 standards use the Ethernetprotocol and CSMA/CA protocol.

There are three main ways by which WLANs transmit information: microwave, spread spectrum and infrared.WLANs have data transfer speeds ranging from 1 to 54Mbps, with some manufacturers offering proprietary 108Mbps solutions. The 802.11n standard can reach 300 to 600Mbps.

1.1.1 Types of Wireless LAN

There are two types of wireless LAN : "ad-hoc" and "infrastructred" networks.

1.1.1.aAd-hoc Networks

This network can be set up by a number **of**mobile users meeting in a small room. It does not need any support from a wired/wireless backbone. There are two ways to implement this network.

- **Broadcasting/Flooding**
  Suppose that a mobile user A wants to send data to another user B in the same area. When the packets containing the data are ready, user A broadcasts the packets. On receiving the packets, the receiver checks the identification on the packet. If that receiver was not the correct destination, then it rebroadcasts the packets. This process is repeated until user B gets the data.
- **TemporaryInfrastructure**
  In this method, the mobile users set up a temporary infrastructure. But this method is complicated and it introduces overheads. It is useful only when there is a small number of mobile users.

  Ad-hoc wireless networks, however, do not need any infrastructure to work. Each node can communicate directly with other nodes, so no access point controllingmedium access is necessary. Nodes within an ad-hoc network can only communicate ifthey can reach each other physically, i.e., if they are within each other's radiorange or if other nodes can forward the message.

In ad-hoc networks, the complexity of each node is higher because everynode has to implement medium access mechanisms, mechanisms to handlehidden or exposed terminal problems, and perhaps priority mechanisms, to providea certain quality of service. This type of wireless network exhibits thegreatest possible flexibility as it is, for example, needed for unexpected meetings,quick replacements of infrastructure or communication scenarios far awayfrom any infrastructure.

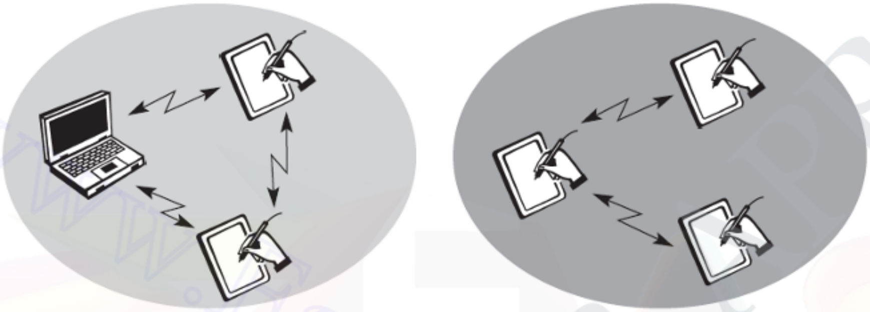


Fig. 1.1 Two adhoc wireless networks

1.1.1.bInfrastructure Networks

The design of infrastructure-based wireless networks is simplerbecause most of the network functionality lies within the access point, whereasthe wireless clients can remain quite simple. This structure is reminiscent ofswitched Ethernet or other star-based networks, where a central element (e.g., aswitch) controls network flow. This type of network can use different accessschemes with or without collision.

Collisions may occur if medium access of thewireless nodes and the access point is not coordinated. However, if only the accesspoint controls medium access, no collisions are possible. This setting may beuseful for quality of service guarantees such as minimum bandwidth for certainnodes. The access point may poll the single wireless nodes to ensure the data rate.

This type of network allows users to move in a building while they are connected to computerresources. The IEEE Project 802.11 specified the components in a wireless LAN architecture. In an infrastructure network, a cell is also known as a Basic Service Area (BSA). It contains a number of wireless stations. The size of a BSA depends on the power of the transmitter and receiver units; it also depends on the environment. A number of BSAs are connected to each other and to a distribution system by Access Points (APs). A group of stations belonging to an AP is called a Basic Service Set (BSS).

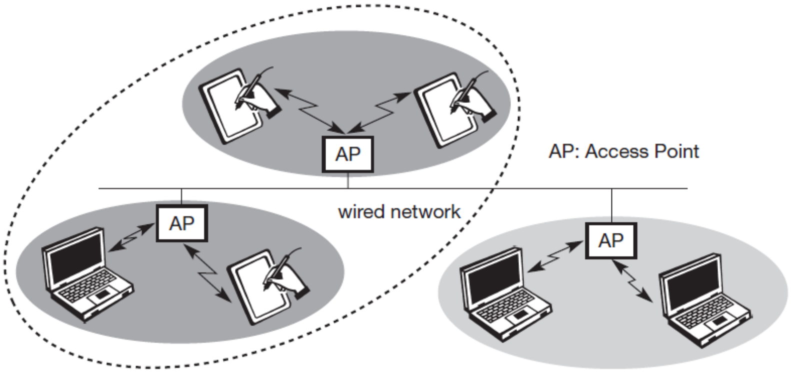


Fig. 1.2 Three infrastructure based wireless networks

1.2 WLAN TECHNOLOGIES

1.2.1 INFRARED (IR)

Infrared is an invisible band of radiation that exists at the lower end of the visible electromagnetic spectrum. This type of transmissionis most effective when a clear line-of-sight exists between the transmitter and the receiver.

Two types of infrared WLAN solutions are available: diffused-beam and direct-beam (or line-of-sight). Currently, direct-beam WLANs offer a faster data rate than diffused-beam networks, but is more directional since diffused-beam technology uses reflected rays to transmit/receive a data signal, it achieves lower data rates in the 1-2 Mbps range.

Infrared optical signals are often used in remote control device applications. The users connect to the local wired network via an infrared device for retrieving information or using fax and print functions on a server. A group of users may also set up a peer-to peer infrared network while on location to share printer, fax, or other server facilities within their own LAN environment. When used indoors, it can be limited by solid objects such as doors, walls, merchandise,or racking. In addition, the lighting environment can affect signal quality.

For example, loss of communications may occur because of the large amount of sunlight or background light in an environment. Fluorescent lights also maycontain large amounts of infrared. This problem maybe solved by using high signal power and an optical bandwidth filter, which reduces the infrared signals coming from outside sources

Advantages

- No government regulations controlling use
- Immunity to electromagnetic and RF interference

Dis – Advantages

- A short range technology (30 to 50 ft. radius)
- Signals cannot penetrate solid objects
- Signal affected by light, fog, snow, etc.
- Dirt can interfere with infrared

1.2.2 UHF (Narrowband)

UHF wireless data communication systems normally transmit in the 430 to 470 MHz frequency range, with rare systems using segments of the 800 MHz range. The lower portion of this band 430-450 MHz is often referenced as unprotected(unlicensed) and 450-470 MHz is referred to as the protected (licensed) band.

In the unprotected band, RF licenses are not granted for specific frequencies and anyone is allowed to use any frequency in the band.In the protected band, RF licenses are granted for specific frequencies, giving customers some assurance that they will have complete use of that frequency.

Other terms for UHF include narrowband and 400 MHz RF. Because independent narrowband RF systems cannot coexist on the same frequency, government agencies allocate specific radio frequencies to users through RF site licenses. A limited amount of unlicensed spectrum is also available in some countries. In order to have many frequencies that can be allocated to users, the bandwidth given to a specific user is very small.

The term “narrowband” is used to describe this technology because the RF signal is sent in a very narrow bandwidth, typically 12.5 kHz or 25 kHz. Power levels range from 1 to 2 watts for narrowband RF data systems. This narrow bandwidth combined with high power results in largetransmission distances than are available from 900 MHz or 2.4 GHz spread spectrum systems, which have lower power levels and wider bandwidths.

1.3 IEEE 802.11

The IEEE standard 802.11 (IEEE, 1999) is 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. The standard specifies the physical and medium access layer adapted to the special requirements of wirelessLANs.

The primary goal of the standard was the specification of a simple and robust WLAN which offers time-bounded and asynchronous services. Additional features of the WLAN should include the support of power management to save battery power, the handling of hidden nodes, and the ability to operate worldwide. The 2.4 GHz ISM band, which is available in most countries around the world, was chosen for the original standard. Data rates envisaged for the standard were 1 Mbit/s mandatory and 2 Mbit/s optional.

1.3.1 System architecture

Wireless networks can exhibit two different basic system architectures as : infrastructure-based or ad-hoc. 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. Thedistribution system connects the wireless networks via the APs with a portal, which forms the interworking unit to other LANs.

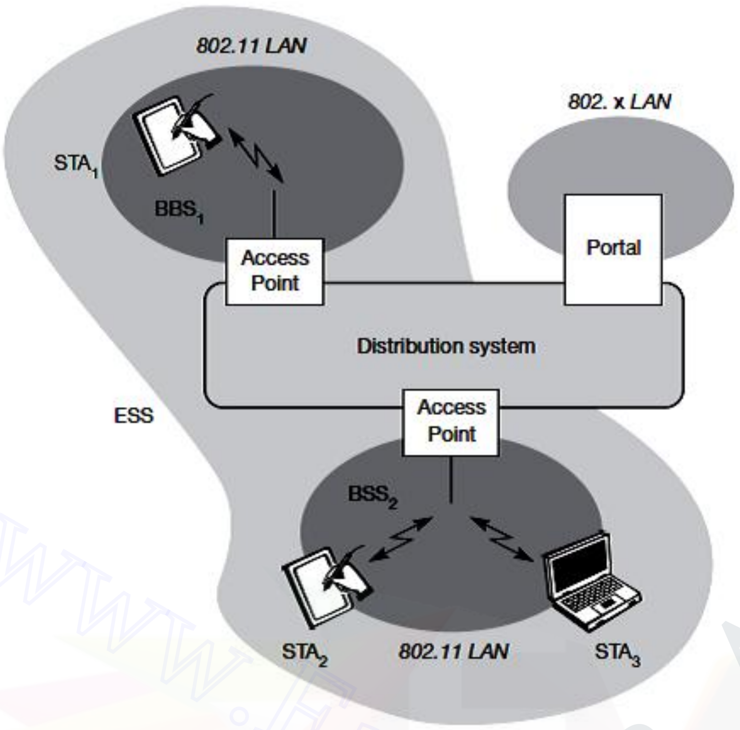


Fig. 1.3 Architecture of Infrastructure IEEE 802.11

The architecture of the distribution system consists of bridged IEEE LANs, wireless links, or any other networks. 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 IEEE 802.11 allows the building of ad-hoc networks between stations, thus forming one or more independentBSSs (IBSS). 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.

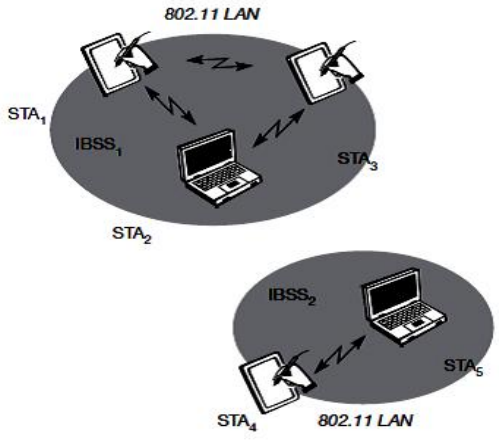


Fig. 1.4 Architecture of Adhoc IEEE 802.11

1.3.2 Protocol architecture

IEEE 802.11 fits into the other 802.x standards for wired LANs. In the most common scenario: an IEEE 802.11 wireless LAN connected to a switched IEEE 802.3 Ethernet via a bridge. The WLAN behaves like a slow wired LAN. Consequently, thehigher layers (application, TCP, IP) look the same for wireless nodes as for wirednodes. The upper part of the data link control layer, the logical link control(LLC), covers the differences of the medium access control layers needed for thedifferent media.

The IEEE 802.11 standard only covers the physical layer PHY and mediumaccess layer MAC like the other 802.x LANs do. The physical layer is subdividedinto the physical layer convergence protocol (PLCP) and the physicalmedium dependent sublayer PMD. The basic tasks of the MAClayer comprise medium access, fragmentation of user data, and encryption. ThePLCP sublayer provides a carrier sense signal, called clear channel assessment(CCA), and provides a common PHY service access point (SAP) independent ofthe transmission technology. Finally, the PMD sublayer handles modulationand encoding/decoding of signals.

Apart from the protocol sublayers, the standard specifies managementlayers and the station management. The MAC management supports the associationand re-association of a station to an access point and roaming betweendifferent access points. It also controls authentication mechanisms, encryption, synchronization of a station with regard to an access point, and power managementto save battery power. MAC management also maintains the MACmanagement information base (MIB).

The main tasks of the PHY management include channel tuning and PHYMIB maintenance. Finally, station management interacts with both managementlayers and is responsible for additional higher layer functions (e.g., control of bridgingand interaction with the distribution system in the case of an access point).

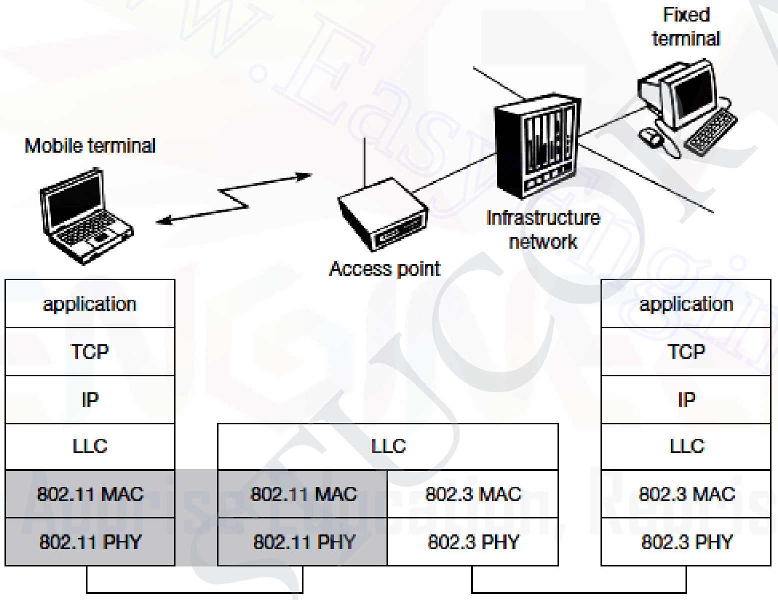
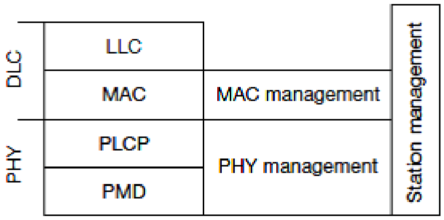


Fig. 1.5 IEEE 802.11 Protocol architecture and bridging



1.3.3 Physical layer

IEEE 802.11 supports three different physical layers: one layer based on infraredand two layers based on radio transmission (primarily in the ISM band at 2.4GHz, which is available worldwide). All PHY variants include the provision ofthe clear channel assessment signal (CCA). This is needed for the MAC mechanismscontrolling medium access and indicates if the medium is currently idle.

The transmission technology determines exactlyhow this signal is obtained.

The PHY layer offers a service access point (SAP) with 1 or 2 Mbit/s transferrate to the MAC layer (basic version of the standard).

1.3.4 Frequency hopping spread spectrum

Frequency hopping spread spectrum (FHSS) is a spread spectrum techniquewhich allows for the coexistence of multiple networks in the same area by separatingdifferent networks using different hopping sequences. The original standard defines 79 hopping channels for North America andEurope, and 23 hopping channels for Japan (each with a bandwidth of 1 MHzin the 2.4 GHz ISM band). The selection of a particular channel is achieved byusing a pseudo-random hopping pattern.

The standard specifies Gaussian shaped FSK (frequency shift keying), GFSK,as modulation for the FHSS PHY. For 1 Mbit/s a 2 level GFSK is used (i.e., 1 bit ismapped to one frequency), a 4 level GFSK for 2 Mbit/s (i.e., 2 bitsare mapped to one frequency). While sending and receiving at 1 Mbit/s ismandatory for all devices, operation at 2 Mbit/s is optional. This facilitated theproduction of low-cost devices for the lower rate only and more powerfuldevices for both transmission rates in the early days of 802.11.

The physical layer used with FHSS has the frame that consists of two basic parts, the PLCP part (preamble and header) and the payloadpart. While the PLCP part is always transmitted at 1 Mbit/s, payload, i.e.MAC data, can use 1 or 2 Mbit/s.

The fields of the frame fulfill the following functions:

Synchronization: The PLCP preamble starts with 80 bit synchronization,which is a 010101... bit pattern. This pattern is used for synchronization ofpotential receivers and signal detection by the CCA.

Start frame delimiter (SFD): The following 16 bits indicate the start of theframe and provide frame synchronization. The SFD pattern is0000110010111101.

PLCP_PDU length word (PLW): This first field of the PLCP header indicatesthe length of the payload in bytes including the 32 bit CRC at the endof the payload. PLW can range between 0 and 4,095.

PLCP signalling field (PSF): This 4 bit field indicates the data rate of thepayload following. All bits set to zero (0000) indicate the lowest data rateof 1 Mbit/s. The granularity is 500 kbit/s, thus 2 Mbit/s is indicated by 0010and the maximum is 8.5 Mbit/s (1111). This system obviously does notaccommodate today's higher data rates.

Header error check (HEC): Finally, the PLCP header is protected by a16 bit checksum with the standard ITU-T generator polynomial
G(x) = x16 + x12 + x5 + 1.

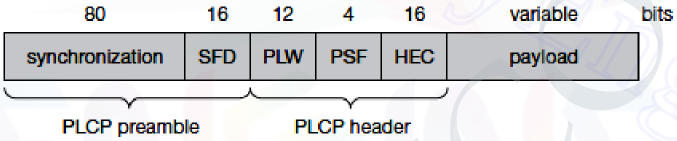


Fig. 1.6 Frame Format of IEEE 802.11 using FHSS

1.3.5 Direct sequence spread spectrum

Direct sequence spread spectrum (DSSS) is the alternative spread spectrummethod separating by code and not by frequency. In the case of IEEE 802.11DSSS, spreading is achieved using the 11-chip Barker sequence (+1, –1, +1, +1, –1,+1, +1, +1, –1, –1, –1). The key characteristics of this method are its robustnessagainst interference and its insensitivity to multipath propagation (time delayspread). However, the implementation is more complex compared to FHSS.IEEE 802.11 DSSS PHY also uses the 2.4 GHz ISM band and offers both 1 and2 Mbit/s data rates. The system uses differential binary phase shift keying (DBPSK)for 1 Mbit/s transmission and differential quadrature phase shift keying (DQPSK)for 2 Mbit/s as modulation schemes. The symbol rate is1 MHz, resulting in a chipping rate of 11 MHz. All bits transmitted by the DSSSPHY are scrambled with the polynomial s(z) = z7 + z4 + 1 for DC blocking andwhitening of the spectrum. Many of today's products offering 11 Mbit/s accordingto 802.11b are still backward compatible to these lower data rates.

The frame of the physical layer using DSSSconsistsof two basic parts, the PLCP part (preamble and header) and the payloadpart. While the PLCP part is always transmitted at 1 Mbit/s, payload, i.e., MACdata, can use 1 or 2 Mbit/s.

The fields of the frame have the following functions:

Synchronization: The first 128 bits are not only used for synchronization,but also gain setting, energy detection (for the CCA), and frequency offsetcompensation. The synchronization field only consists of scrambled 1 bits.

Start frame delimiter (SFD): This 16 bit field is used for synchronization atthe beginning of a frame and consists of the pattern 1111001110100000.

Signal: Originally, only two values have been defined for this field to indicatethe data rate of the payload. The value 0x0A indicates 1 Mbit/s (andthus DBPSK), 0x14 indicates 2 Mbit/s (and thus DQPSK). Other values havebeen reserved for future use, i.e., higher bit rates.

Service: This field is reserved for future use; however, 0x00 indicates anIEEE 802.11 compliant frame.

Length: 16 bits are used in this case for length indication of the payloadin microseconds.

Header error check (HEC): Signal, service, and length fields are protectedby this checksum using the ITU-T CRC-16 standard polynomial.

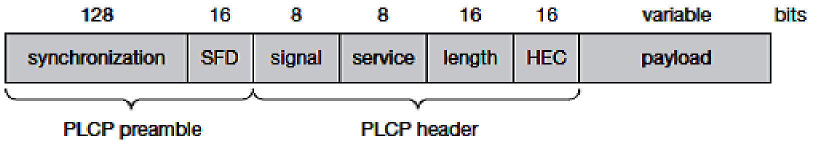


Fig. 1.7 Frame Format of IEEE 802.11 using DSSS

1.3.6 Medium access control layer (MAC LAYER)

The MAC layer has to control mediumaccess, but it can also offer support for roaming, authentication, and power conservation.The basic services provided by the MAC layer are the mandatoryasynchronous data service and an optional time-bounded service. While802.11 only offer the asynchronous service in ad-hoc network mode, both servicetypes can be offered using an infrastructure-based network together withthe access point coordinating medium access. The asynchronous service supportsbroadcast and multi-cast packets, and packet exchange is based on a 'besteffort' model, i.e., no delay bounds can be given for transmission.

The following three basic access mechanisms have been defined for IEEE802.11:

- 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 arealso summarized as distributed coordination function (DCF), the thirdmethod is called point coordination function (PCF). DCF only offers asynchronousservice, while PCF offers both asynchronous and time-boundedservice but needs an access point to control medium access and to avoid contention.The MAC mechanisms are also called distributed foundation wirelessmedium access control (DFWMAC).

For all access methods, several parameters for controlling the waiting timebefore medium access are important. The three different parametersthat define the priorities of medium access. The values of the parametersdepend on the PHY and are defined in relation to a slot time. Slot time isderived from the medium propagation delay, transmitter delay, and other PHYdependent 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.

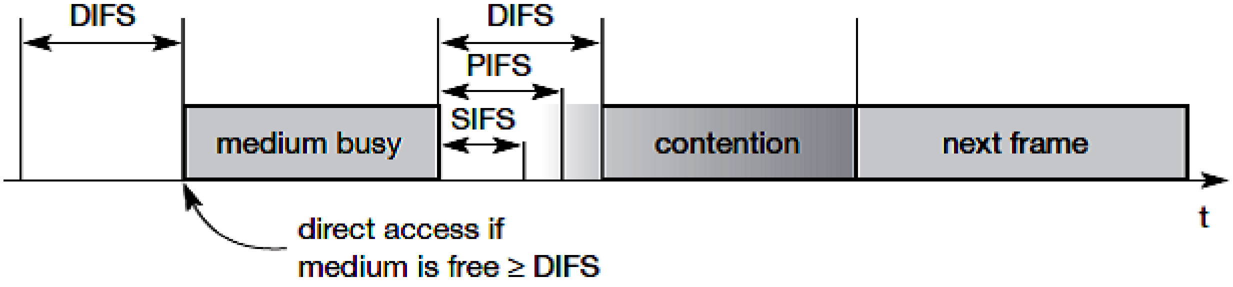


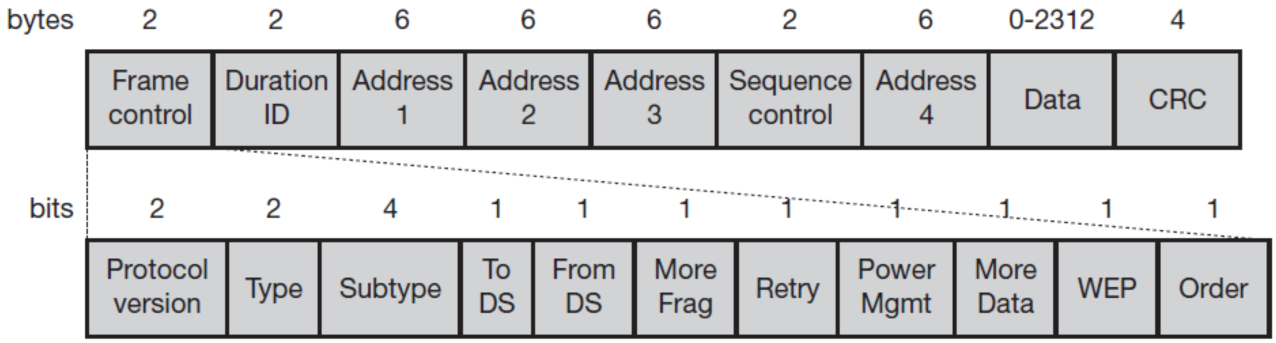
Fig. 1.8 Medium access and inter frame spacing

During a contention phase several nodes try to access the medium.

Short inter-frame spacing (SIFS): The shortest waiting time for mediumaccess (so the highest priority) is defined for short control messages, such asacknowledgements of data packets or polling responses. For DSSS SIFS is10 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 accesspoint polling other nodes only has to wait PIFS for medium. PIFS is defined as SIFS plus one slot time.

DCF inter-frame spacing (DIFS): This parameter denotes the longest waitingtime and has the lowest priority for medium access. This waiting time isused for asynchronous data service within a contention period.DIFS isdefined as SIFS plus two slot times.



bytes	Frame control	Duration ID	Address 1	Address 2	Address 3	Sequence control	Address 4	Data	CRC
	2	2	6	6	6	2	6	0-2312	4

bits	Protocol version	Type	Subtype	To DS	From DS	More Frag	Retry	Power Mgmt	More Data	WEP	Order
	2	2	4	1	1	1	1	1	1	1	1

Fig. 1.9 Frame Control

1.3.7 MAC frames

The following figure shows the basic structure of an IEEE 802.11 MAC data rametogether with the content of the frame control field.

● Frame control: The first 2 bytes serve several purposes. They contain several sub-fields.

● Duration/ID: If the field value is less than 32,768, the duration field containsthe value indicating the period of time in which the medium isoccupied (in μs). This field is used for setting the NAV for the virtual reservationmechanism using RTS/CTS and during fragmentation. Certain valuesabove 32,768 are reserved for identifiers.

● Address 1 to 4: The four address fields contain standard IEEE 802 MACaddresses (48 bit each), as they are known from other 802.x LANs. Themeaning of each address depends on the DS bits in the frame control field.

● Sequence control: Due to the acknowledgement mechanism frames may beduplicated. Therefore a sequence number is used to filter duplicates.

● Data: The MAC frame may contain arbitrary data (max. 2,312 byte), whichis transferred transparently from a sender to the receiver(s).

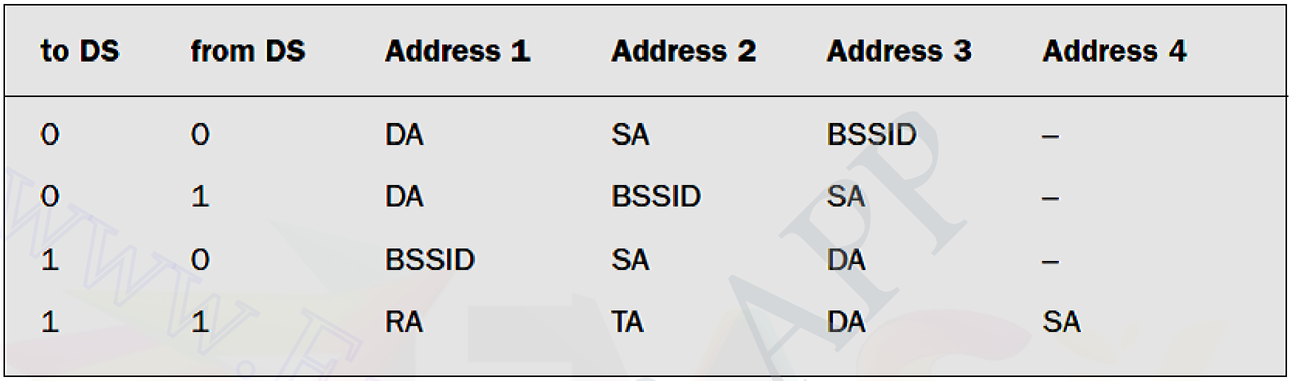
● Checksum (CRC): Finally, a 32 bit checksum is used to protect the frame asit is common practice in all 802.x networks.

MAC frames can be transmitted

- Between mobile stations;

- Between mobile stations and
- An access point and between access points over a DS.

Two bits within the Frame Control field, ‘to DS’ and ‘from DS’, differentiatethese cases and control the meaning of the four addresses used. The following Table will gives an overview of the four possible bit values of the DS bits and the associated interpretationof the four address fields.



to DS	from DS	Address 1	Address 2	Address 3	Address 4
0	0	DA	SA	BSSID	–
0	1	DA	BSSID	SA	–
1	0	BSSID	SA	DA	–
1	1	RA	TA	DA	SA

Fig. 1.10 Interpretation of the MAC addresses in an 802.11 MAC frame

Every station, access point filters on address 1. This addressidentifies the physical receiver(s) of the frame. Based on this address, a station candecide whether the frame is relevant or not. The second address, address 2, representsthe physical transmitter of a frame. This information is important because thisparticular sender is also the recipient of the MAC layer acknowledgement. If apacket from a transmitter (address 2) is received by the receiver with address 1, thisreceiver in turn acknowledges the data packet using address 2 as receiver address asshown in the Figure. The remaining two addresses address 3and address 4, are mainly necessary for the logical assignment of frames (logicalsender, BSS identifier, logical receiver). If address 4 is not needed the field is omitted.

For addressing, the following four scenarios are possible:

- **Ad-hoc network:** If both DS bits are zero, the MAC frame organizes a packet which is exchanged between two wireless nodes without a distribution system. DA indicates the destination address, SA is the source address of the frame, which is identical to the physical receiver and sender addresses respectively. The third address identifies the basic service set (BSSID, the fourth address is unused.
- **Infrastructure network, from AP:** If the bit only from DSis set, the framephysically originates from an access point. DA is the logical and physicalreceiver, the second

address identifies the BSS, and the third address specifies thelogical sender, the source address of the MAC frame.

- **Infrastructure network, to AP:** If a station sends a packet to another station through the access point, only the 'to DS' bit is set. Now the first addressrepresents the physical receiver of the frame, the access point, via the BSSidentifier. The second address is the logical and physical sender of theframe, while the third address indicates the logical receiver.

- **Infrastructure network, within DS:** For packets transmitted between twoaccess points over the distribution system, both bits are set. The firstreceiver address (RA), represents the MAC address of the receiving accesspoint. Similarly, the second address transmitter address (TA), identifies thesending access point within the distribution system. Now two moreaddresses are needed to identify the original destination DA of the frameand the original source of the frame SA.

1.3.8 MAC management

MAC management plays a vital role in an IEEE 802.11 station as it controls all the functions related to integration of awireless station into a BSS, formation of an ESS, synchronization of stations etc.

- Synchronization: It is used to support finding a wireless LAN, synchronization of internal clocks, and generation of beacon signals.
- Power management: It is used to control transmitter activity for power conservation, e.g., periodic sleep, buffering, without missing a frame.
- Roaming: Functions for joining a network (association), changing access points, scanning for access points.
- Management information base (MIB): All parameters representing the current state of a wireless station and an access point are stored within a MIB for internal and external access. A MIB can be accessed via standardized protocols such as the simple network management protocol (SNMP).

1.3.9 Synchronization

An internal clock is maintained by each node of an 802.11 network. Timing synchronization function is specified by the IEEE 802.11 to synchronizethe clocks of all nodes.

In power management synchronized clocks are needed, but also for coordination of the PCF and forsynchronization of the hopping sequence in an FHSS system. The start of a super frame can be predicted by the local timer of the node. FHSS physical layers need the same hoppingsequences so that all nodes can communicate within a BSS.

A beacon contains a timestamp and other management informationused for power management and roaming (e.g., identification of the BSS).The timestamp is used by a node to adjust its local clock. The node is notrequired to hear every beacon to stay synchronized; however, from time to timeinternal clocks should be adjusted.The transmission of a beacon frame is notalways periodic because the beacon frame is also delayed if the medium is busy.

Within infrastructure-based networks, the access point performs synchronizationby transmitting the (quasi)periodic beacon signal. However, the access pointalways tries to schedule transmissions according to the target beacon interval, i.e., beacon intervals are not shifted if onebeacon is delayed.The timestamp of a beacon always reflects the real transmittime, not the scheduled time.

Ad-hoc networks, does not have an access point for beacon transmission. In this case, each node maintainsits own synchronization timer and starts the transmission of a beaconframe after the beacon interval. All otherstations now adjust their internal clocks according to the received beacon andsuppress their beacons for this cycle. If collision occurs, the beacon is lost.

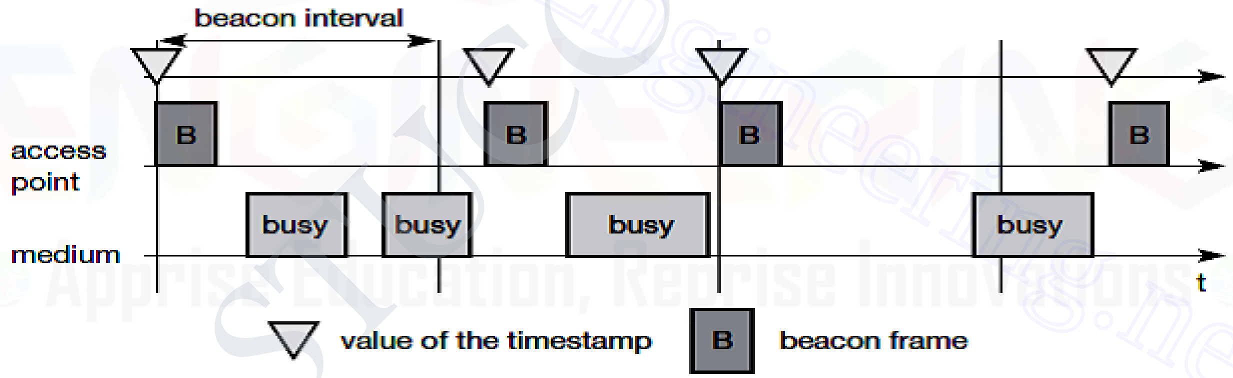


Fig. 1.11 Beacon transmission in a 802.11 network

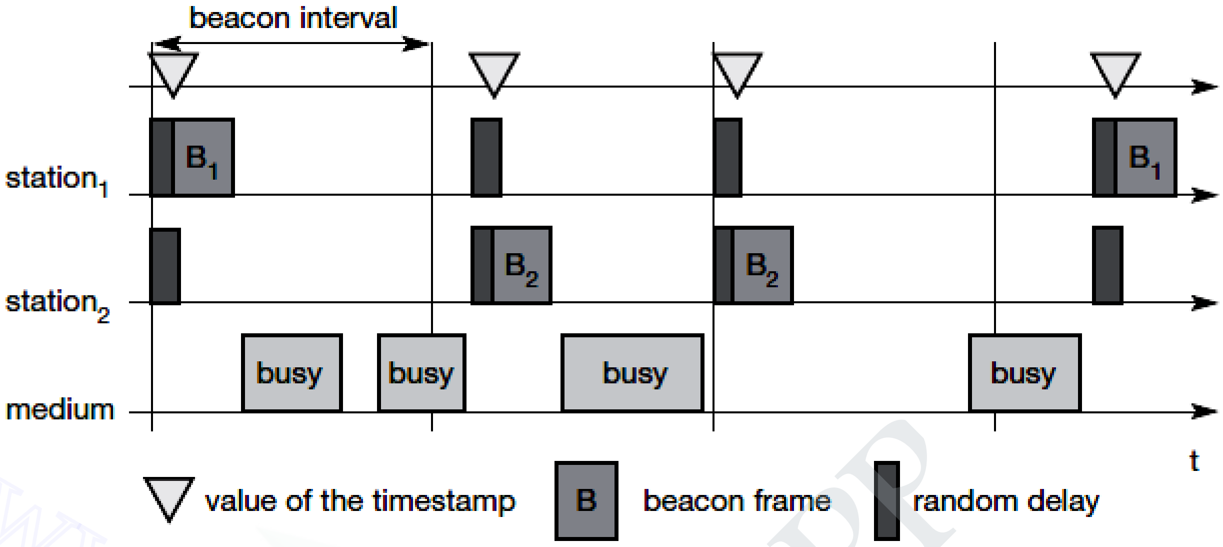


Fig. 1.11 Beacon transmission in a adhoc network

1.3.10 Power management

Wireless devices are battery powered. Hence power-saving mechanisms are critical for suchdevices. Standard LAN protocols are always ready to receivedata, although receivers are idle most of the time in lightly loaded networks.

In IEEE 802.11 power management is to switch off the transceiverwhenever it is not needed. This is simple for the sending device toachieve as the transfer is generated by the device itself. However, since the powermanagement of a receiver cannot know in advance when the transceiver has tobe active for a specific packet, it has to 'wake up' the transceiver periodically.

Switching off the transceiver should be transparent to present protocols and able to support different applications. However, throughputcan be traded-off for battery life. Longer off-periods save battery life butaverage throughput will be reduce and vice versa.

The basic idea of power saving includes two states for a station: sleep andawake, and buffering of data in senders. If a sender aims to communicate with a power-saving station, if the station is asleepit needs to buffer data. On the other handthe sleeping stationhas to wake up periodically and stay awake for a certaintime. During this time, all senders can reveal the destinations of their buffereddata frames. If a station detects that it is a destination of a buffered packet it has tostay awake until the transmission takes place. All stations have to wake up or be awake at the same time.

Comparedto ad-hoc networksinfrastructure-based networkshas a simpler power management. The access point buffers all frames of stations operating in power-save mode. With every beacon sent by the accesspoint, a traffic indication map (TIM) is transmitted. The TIM contains a list ofstations for which unicast data frames are buffered in the access point.

If the TIM indicates a unicast frame forthe station, the station stays awake for transmission. Stations will always stay awake for multi-cast/broadcast transmission.

A sleepingstation still has the TSF timer running.

The following figure shows an example with an access point and one station.

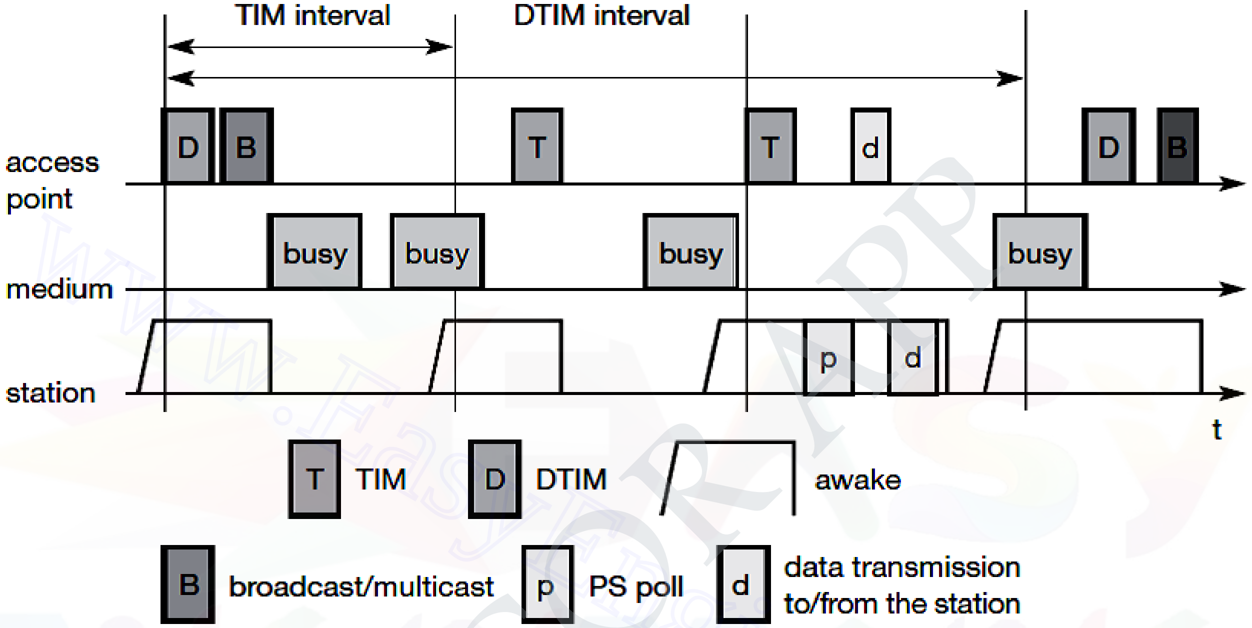


Fig. 1.12 Access Point with one station

1.3.11 Power management in IEEE 802.11 Network

Thestate of the medium is indicated. The access point transmits a beaconframe each beacon interval. This interval is now the same as the TIM interval. For sending broadcast/multicast frames, the access point maintains a delivery traffic indication map(DTIM) interval. The DTIM interval isalways a multiple of the TIM interval.

In the first case, the access point has to transmit abroadcast frame and the station stays awake to receive it. After receiving thebroadcast frame, the station returns to sleeping mode. Before the next TIM transmission starts,the station wakes up. This time the TIM is delayed due toa busy medium so, the station stays awake. The access point has nothing to send and the station goes back to sleep.

At the next TIM interval, the station is thedestination for a buffered frame indicated by the access point. The access point then transmits the datafor the station; the station acknowledges the receipt and may also send somedata and it is acknowledged by the access point,afterwards, the station switchesto sleep mode again.Finally, the access point has more broadcast data to send during the next DTIMinterval, which is again delayed by a busy medium. A station may stay awake if the sleeping period would be too short.

This mechanism clearly shows the trade-off between short delays in stationaccess and saving battery power. The shorter the TIM interval, the shorter thedelay, but the lower the power-saving effect.

The power management for ad-hoc networks is much more complicated than ininfrastructure networks. In this case, there is no access point to buffer data inone location but each station wants to buffer data if it wants to communicatewith a power-saving station. Buffered frames list are announced by the stationsduring a period when they are all awake. Destinations areannounced using ad-hoc traffic indication map (ATIMs) – the announcementperiod is called the ATIM window.

All stations stay awake for the ATIMinterval as shown in the first two steps and go to sleep again if no frame isbuffered for them. In the third step, station1 has data buffered for station2. Thisis indicated in an ATIM transmitted by station1.

Station2 acknowledges thisATIM and stays awake for the transmission. After the ATIM window, station1can transmit the data frame, and station2 acknowledges its receipt. In this case,the stations stay awake for the next beacon.More ATIM transmissions take place, morecollisions happen and more stations are delayed. The access delay of large networksis difficult to predict. QoS guarantees cannot be given under heavy load.

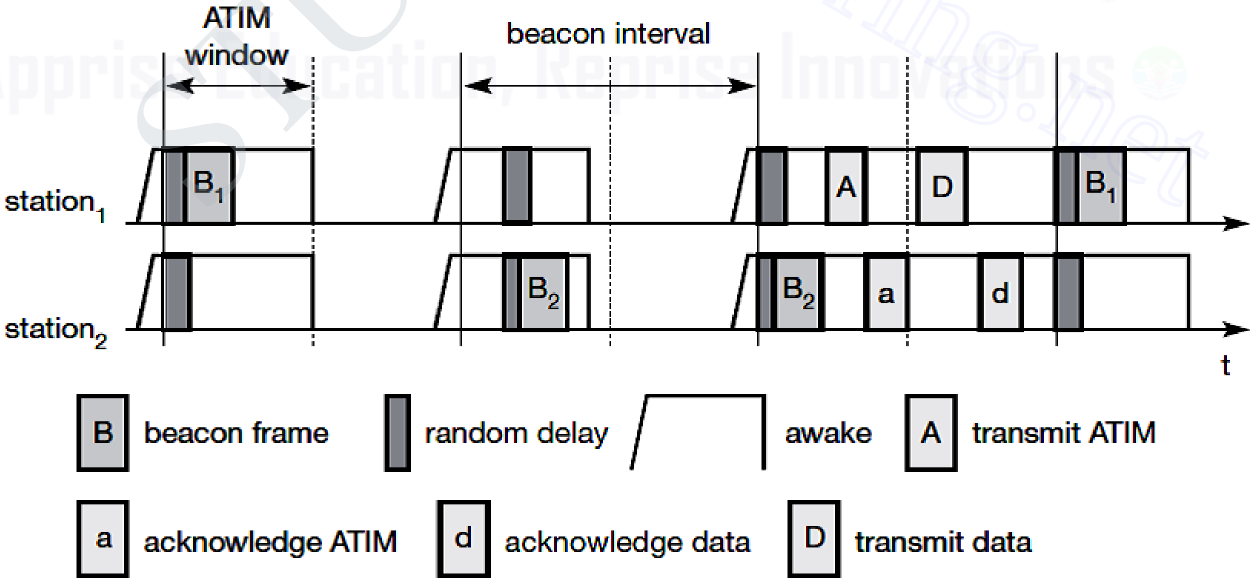


Fig. 1.13 Power management in IEEE 802.11 ad-hoc network

1.3.12 Roaming

Accesspoint more than one is required to cover all rooms when the wireless networks are within the buildings. Depending on the structure of the walls,one access point has a transmission range of 10–20 m. Each storey of a building needs its own access point(s) as quite often walls are thinner than floors. If a user walks around within a wireless station,the station has to change from one access point to another to provideuninterrupted service. Moving between access points is refers to roaming.

The steps for roaming between access points are:

- When the current link quality to its access point AP1 is toopoor. The station then starts scanning for another access point.

- Scanning will search for another BSS and can also be used forsetting up a new BSS in case of ad-hoc networks. IEEE 802.11 specifies scanningon single or multiple channels anddifferentiates between passive scanning and active scanning. Passive scanningmeans listening into the medium to find other networks, i.e.,receiving the beacon of another network issued by the synchronizationfunction within an access point. Active scanning comprises sending aprobe on each channel and waiting for a response. Beacon and proberesponses contain the information necessary to join the new BSS.
- The station then selects the best access point for roaming based on, signal strength, and sends an association request to the selected accesspoint AP2.
- The new access point AP2 answers with an association response. If theresponse is successful, the station has roamed to the new access point AP2.

The access point accepting an association request indicates the new stationin its BSS to the distribution system (DS). The DS then updates its database,which contains the current location of the wireless stations. This database isneeded for forwarding frames between different BSSs, i.e. between the differentaccess points controlling the BSSs, which combine to form an ESS.

The standard IEEE 802.11f should provide a compatible solution for all vendors.This also includes load-balancing between access points and keygeneration for security algorithms based on IEEE 802.1x (IEEE, 2001).

1.4 IEEE 802.11b

IEEE 802.11b was the first wireless LAN standard to be widely adopted and built in to many laptop computers and other forms of equipment. It was only after 802.11 was ratified and products became available that W-Fi took off in a large way. This standard describes a new PHY layerand is by far the most successful version of IEEE 802.11 available today. Thestandards are named according to the order in which the respective studygroups have been established.Although the IEEE 802.11a standard was introduced at the same time, it did not catch on in the same way even though it was capable of higher speeds.

1.4.1 802.11b specification

It is able to transfer data with raw data rates up to 11 Mbps, and has a good range, although not when operating at its full data rate.

Table 1.1 Wi-Fi Standard Specifications

PARAMETER	VALUE
Date of standard approval	July 1999
Maximum data rate (Mbps)	11
Typical data rate (Mbps)	5
Typical range indoors (Metres)	~30
Modulation	CCK (DSSS)
RF Band (GHz)	2.4
Channel width (MHz)	20

When transmitting data 802.11b uses the CSMA/CA technique that was defined in the original 802.11 base standard and retained for 802.11b. Using this technique, when a node wants to make a transmission it listens for a clear channel and then transmits. With 802.11b WLANs, mobile users can get Ethernet levels of performance, throughput, and availability. The standards-based technology allows administrators to build networks that seamlessly combine more than one LAN technology to best fit their business and user needs. The basic architecture features, and services of 802.11b are defined by the original 802.11 standard. The 802.11b specification affects only the physical layer, adding higher data rates and more robust connectivity.

1.4.2 802.11b Enhancements to the PHY Layer

The key contribution of the 802.11b addition to the wireless LAN standard was to standardize the physical layer support of two new speeds, 5.5 Mbps and 11 Mbps. To accomplish this, DSSS had to be selected as the sole physical layer technique for the standard since, as noted above, frequency hopping cannot support the higher speeds without violating current FCC regulations. The implication is that 802.11b systems will interoperate with 1 Mbps and 2 Mbps 802.11 DSSS systems, but will not work with 1 Mbps and 2 Mbps 802.11 FHSS systems.

The original 802.11 DSSS standard specifies an 11-bit chipping—called a Barker sequence—to encode all data sent over the air. Each 11-chip sequence represents a single data bit (1 or 0), and is converted to a waveform, called a symbol, that can be sent over the air. These symbols are transmitted at a 1 MSps (1 million symbols per second) symbol rate using a technique called Binary Phase Shift Keying (BPSK).

In the case of 2 Mbps, a more sophisticated implementation called Quadrature Phase Shift Keying (QPSK) is used; it doubles the data rate available in BPSK, via improved efficiency in the use of the radio bandwidth. To increase the data rate in the 802.11b standard, advanced coding techniques are employed.

Rather than the two 11-bit Barker sequences, 802.11b specifies Complementary Code Keying (CCK), which consists of a set of 64 8-bit code words. As a set, these code words have unique mathematical properties that allow them to be correctly distinguished from one another by a receiver even in the presence of substantial noise and multipath interference (e.g., interference caused by receiving multiple radio reflections within a building).

To support very noisy environments as well as extended range, 802.11b WLANs use dynamic rate shifting, allowing data rates to be automatically adjusted to compensate for the changing nature of the radio channel. Ideally, users connect at the full 11 Mbps rate. However when devices move beyond the optimal range for 11 Mbps operation, or if substantial interference is present, 802.11b devices will transmit at lower speeds, falling back to 5.5, 2, and 1 Mbps. Likewise, if the device moves back within the range of a higher-speed transmission, the connection will automatically speed up again. Rate shifting is a physicallayer mechanism transparent to the user and the upper layers of the protocol stack.

Table 1.2 802.11b data rate specifications

Data Rate	Modulation	Modulation Rate	Chip Size	Symbol Rate	Bits/Symbol
1Mbps	BPSK	11.000.000	11	1.000.000	1

2Mbps	QPSK	11.000.000	11	1.000.000	2
5.5 Mbps	QPSK	11.000.000	8	1.375.000	4
11 Mbps	QPSK	11.000.000	8	1.375.000	8

The following figure shows two packet formats standardized for 802.11b. The mandatory format is called long PLCP PPDU and is similar to the format illustrated in figure. One difference is the rate encoded in the signal field this is encodedin multiples of 100 kbit/s. Thus, 0x0A represents 1 Mbit/s, 0x14 is used for2 Mbit/s, 0x37 for 5.5 Mbit/s and 0x6E for 11 Mbit/s. Note that the preambleand the header are transmitted at 1 Mbit/s using DBPSK. The optional shortPLCP PPDU format differs in several ways.

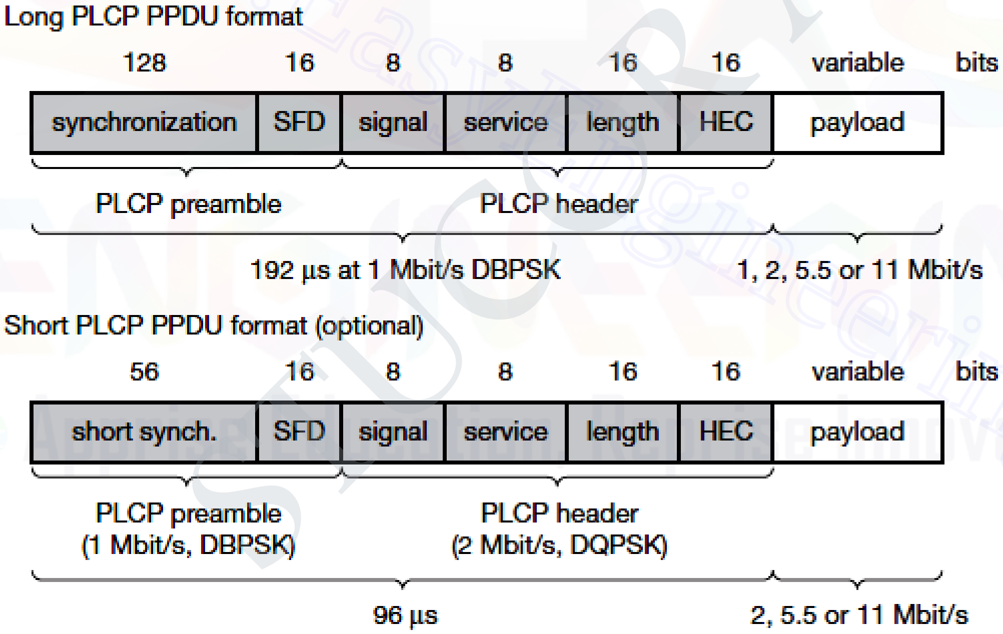


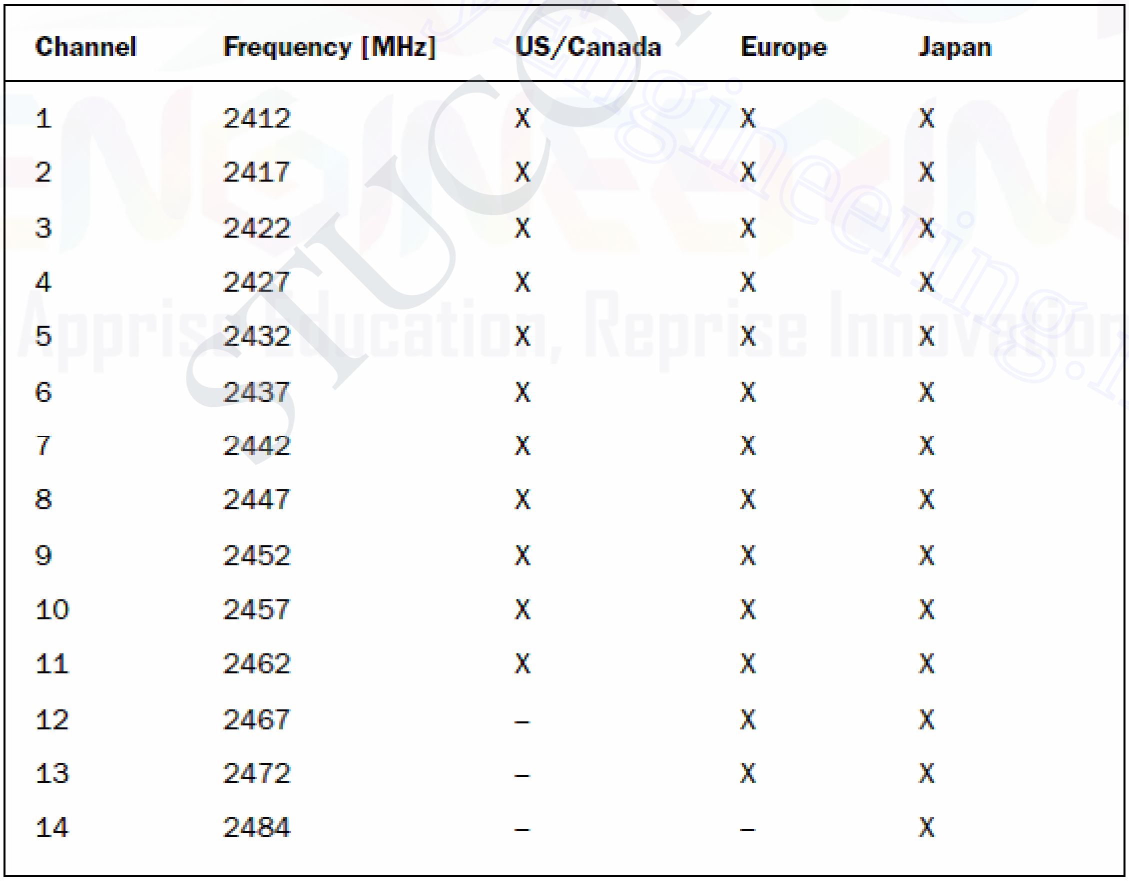
Fig. 1.14 IEEE 802.11b Phypacket formats

The short synchronization field consistsof 56 scrambled zeros instead of scrambled ones. The short start framedelimiter SFD consists of a mirrored bit pattern compared to the SFD of the longformat: 0000 0101 1100 1111 is used for the short PLCP PDU instead of 1111 0011 1010 0000 for the long PLCP PPDU. Receivers that are unable to receivethe short format will not detect the start of a frame (but will sense the mediumis busy). Only the preamble is transmitted at 1 Mbit/s, DBPSK. The followingheader is already transmitted at 2 Mbit/s, DQPSK, which is also the lowest availabledata rate.

As IEEE 802.11b is the most widespread version, some more information isgiven for practical usage. The standards operates (like the DSSS version of802.11) on certain frequencies in the 2.4 GHz ISM band.

The following figure illustrates the non-overlapping usage of channels for anIEEE 802.11b installation with minimal interference in the US/Canada andEurope. The spacing between the center frequencies should be at least 25 MHz(the occupied bandwidth of the main lobe of the signal is 22 MHz). This results in the channels 1, 6, and 11 for the US/Canada or 1, 7, 13 for Europe, respectively.It may be the case that, e.g., travellers from the US cannot use theadditional channels (12 and 13) in Europe as their hardware is limited to 11channels. Some European installations use channel 13 to minimize interference.
Users can install overlapping cells for WLANs using the three non-overlappingchannels to provide seamless coverage. This is similar to the cell planning formobile phone systems.

Table 1.3 IEEE 802.11b channel plan



Channel	Frequency [MHz]	US/Canada	Europe	Japan
1	2412	X	X	X
2	2417	X	X	X
3	2422	X	X	X
4	2427	X	X	X
5	2432	X	X	X
6	2437	X	X	X
7	2442	X	X	X
8	2447	X	X	X
9	2452	X	X	X
10	2457	X	X	X
11	2462	X	X	X
12	2467	–	X	X
13	2472	–	X	X
14	2484	–	–	X

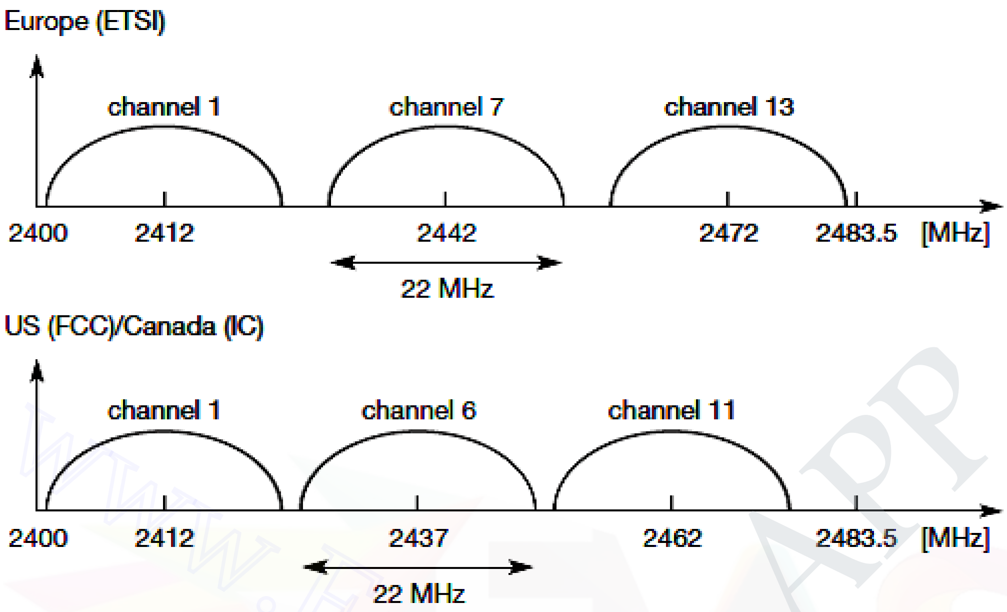


Fig. 1.15 IEEE 802.11b non overlapping channel selection

1.5 802.11a

An 802.11a wireless network supports a maximum theoretical bandwidth of 54Mbps, substantially better than the 11 Mbps of 802.11b and on par with what 802.11g would start to offer a few years later. The performance of 802.11a made it an attractive technology, but achieving that level of performance required using relatively higher cost hardware.

The IEEE 802.11a is an Orthogonal Frequency Division Multiplexing (OFDM) system very similar to Asymmetrical Digital Subscriber Loop (ADSL) Discrete Multi Tone (DMT) modems sending several sub-carriers in parallel using the Inverse Fast Fourier Transform (IFFT), and receiving those subcarriers using the Fast Fourier Transform (FFT).

In 802.11a the transmission medium is wireless and the operating frequency band is 5 GHz. The OFDM of the 802.11a system provides a Wireless LAN with data payload communication capabilities of 6, 9, 12, 18, 24, 36, 48 and 54 Mbps. The support of transmitting and receiving at data rates of 6, 12, and 24 Mbps is mandatory in the standard. The 802.11a system uses 52 subcarriers that are modulated using binary or quadrature phase shift keying (BPSK/QPSK), 16 Quadrature Amplitude Modulation

(QAM), or 64 QAM. Forward Error Correction (FEC) coding (convolutional coding) is used with a coding rate of 1/2, 2/3, or 3/4.

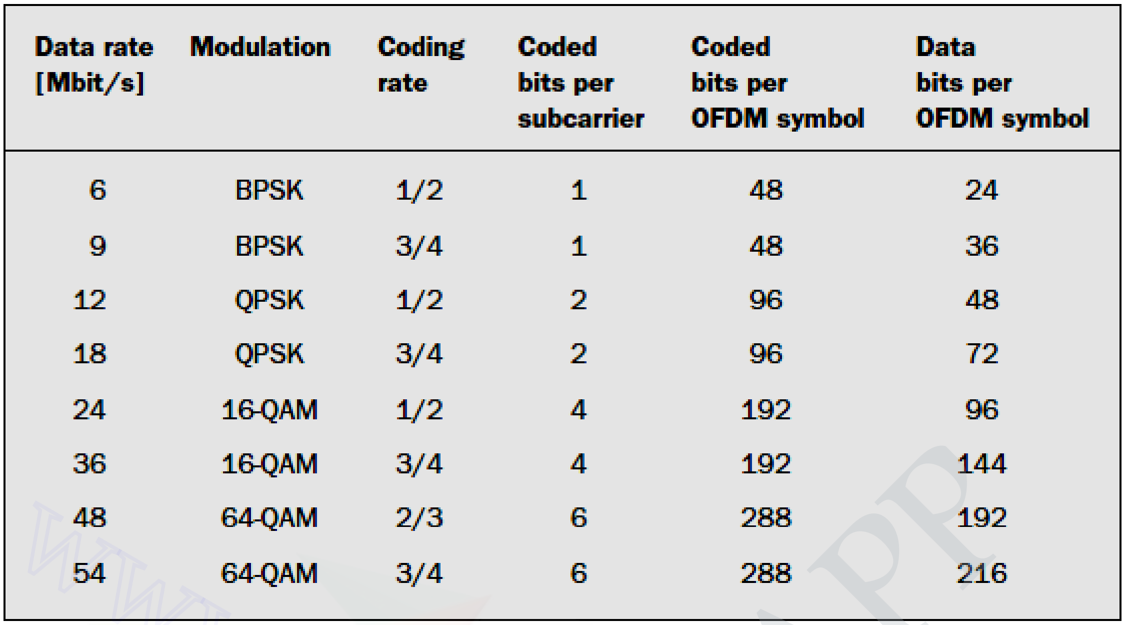
The OFDM PHY layer consists of two protocol functions: first a PHY convergence functions, which adapts the capabilities of the Physical Medium Dependent (PMD) system to the PHY service. This function is supported by the Physical Layer Convergence Procedure (PLCP), which defines a method of mapping the IEEE 802.11 PHY Sublayer Service Data Units (PSDU) into a framing format suitable for sending and receiving user data and management information between two or more stations using the associated PMD system.

Second a PMD system whose function defines the characteristics and method of transmitting and receiving data through a wireless medium between two or more stations, each using the OFDM system.

IEEE 802.11a uses the same MAC layer as all 802.11 physical layers. IEEE 802.11a uses many different technologiesto offer data rates up to 54 Mbit/s.The system uses 52 subcarriers (48 data + 4 pilot) that are modulatedusing BPSK, QPSK, 16-QAM, or 64-QAM. To mitigate transmission errors, FEC isapplied using coding rates of 1/2, 2/3, or 3/4.

The following table gives an overview of thestandardized combinations of modulation and coding schemes together withthe resulting data rates. To offer a data rate of 12 Mbit/s, 96 bits are coded intoone OFDM symbol. These 96 bits are distributed over 48 subcarriers and2 bits are modulated per sub-carrier using QPSK (2 bits per point in the constellationdiagram). Using a coding rate of 1/2 only 48 data bits can be transmitted.

Table 1.4 Rate dependent parameters for IEEE 802.11a



Data rate [Mbit/s]	Modulation	Coding rate	Coded bits per subcarrier	Coded bits per OFDM symbol	Data bits per OFDM symbol
6	BPSK	1/2	1	48	24
9	BPSK	3/4	1	48	36
12	QPSK	1/2	2	96	48
18	QPSK	3/4	2	96	72
24	16-QAM	1/2	4	192	96
36	16-QAM	3/4	4	192	144
48	64-QAM	2/3	6	288	192
54	64-QAM	3/4	6	288	216

1.5.1 OFDM PLCP sublayer of the 802.11a

The PHY Sublayer Service Data Units (PSDU) of the 802.11a is converted to a PLCP Protocol Data Unit (PPDU). The PSDU of the 802.11a is provided with a PLCP preamble and header to create the PPDU. At the receiver of the 802.11a, the PLCP preamble and header are processed to aid in demodulation and delivery of the PSDU. The PPDU is unique to the OFDM PHY. The PPDU format of the standard 802.11a is shown in figure and it includes: "PLCP preamble. This field is used to acquire the incoming OFDM signal and train and synchronize the demodulator.

The PLCP preamble consists of 12 symbols, 10 of which are short symbols and 2 long symbols. The short symbols are used to train the receiver's AGC and to estimate a coarse estimate of the carrier frequency and the channel. The long symbols are used to fine-tune the frequency and the channel estimates. Twelve subcarriers are used for the sort symbols and 53 for the long. The training of an OFDM is accomplished in 16 μs. The PLCP preamble is BPSK-OFDM modulated at 6 Mbps using convolutional encoding rate R=1/2. !"SIGNAL. This is a 24 bits field, which contains information about the rate and length of the PSDU.

The PLCP preamble is BPSK-OFDM modulated at 6 Mbps using convolutional encoding rate R=1/2. The first 4 bits (R1-R4) are used to encode the rate. The next bit is 1 reserved bit. A continuation they are 12 bits used for the length that indicated the number of octets in the PSDU. A continuation is a parity bit and 6 tail bits.This field contains 16 bits for the service field, the PSDU, tails bits and pad bits. A total of 6 tail bits containing 0s are appended to the PPDU to ensure that the convolutional encoder is brought back to zero state. The data portion of the packet is transmitted at the data rate indicated in the signal field.

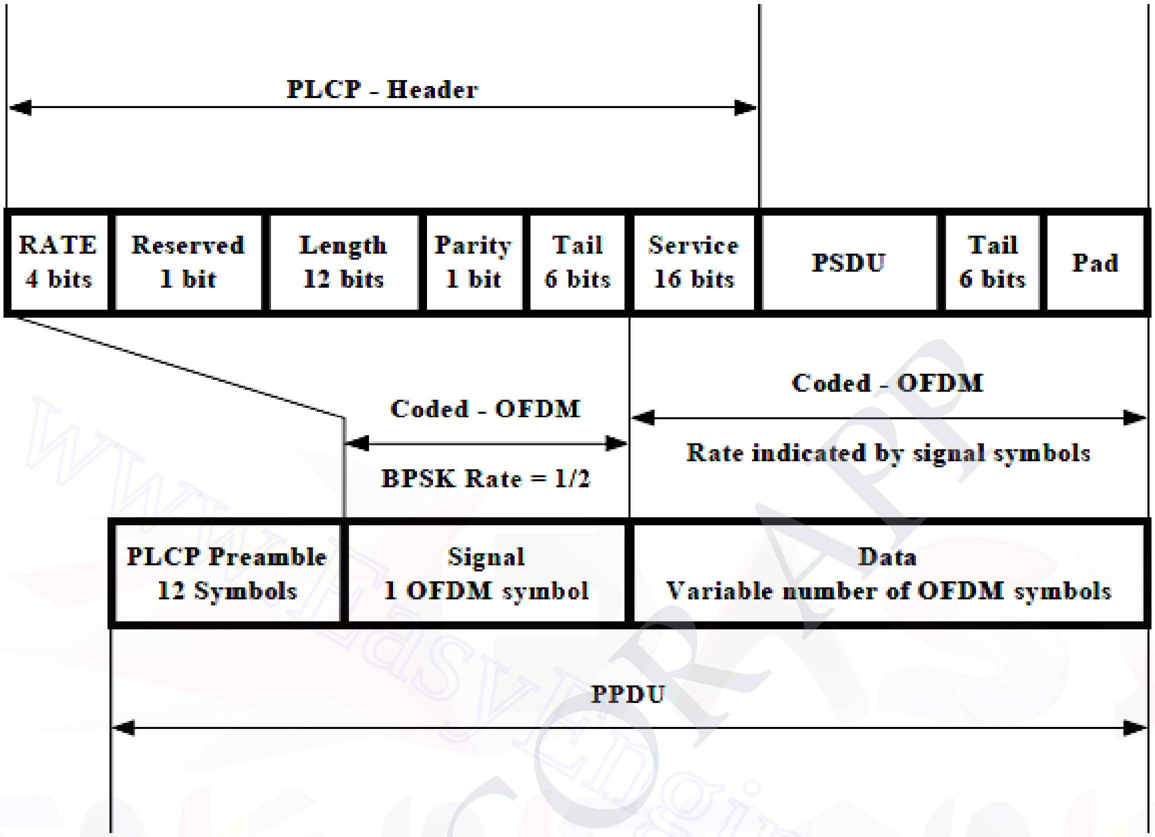


Fig. 1.16 OFDM PLCP Preamble, Header, PSDU of 802.11a

The PLCP header of 802.11a contains:

- 4 bits for the rate
- 1 reserved bit
- 12 bits for length
- 1 bit for parity
- 6 bits for tail
- 16 bits for service

Compared to IEEE 802.11b working at 2.4 GHz IEEE 802.11a at 5 GHz offersmuch higher data rates. However, shading at 5 GHz is much more severe comparedto 2.4 GHz and depending on the SNR, propagation conditions and thedistance between sender and receiver, data rates may drop fast (e.g., 54 Mbit/smay be available only in an LOS or near LOS condition). Additionally, the MAClayer of IEEE 802.11 adds overheads. User data rates are therefore much lowerthan the data rates listed above. Typical user rates in Mbit/s are (transmissionrates in brackets) 5.3 (6), 18 (24), 24 (36), and 32 (54).

1.5.2 Data Scrambler

All the bits transmitted by the 802.11a OFDM PMD in the data portion are scrambled using a frame synchronous 127 bits sequence generator. Scrambling is used to randomize the service, PSDU, pad and data patterns, which may contain long strings of binary 1s or 0s. The tail bits are not scrambled. The octets of the PSDU are placed in the transmitted serial bit stream, bit 0 first and bit 7 last.

The frame synchronous scrambler uses the generator polynomial S(x) as follows:

$$\mathbf{S(x) = x^7 + x^4 +1}$$

The 127bit sequence generated repeatedly by the scrambler is (leftmost used first), 00001110 11110010 11001001 00000010 00100110 00101110 10110110 00001100 11010100 11100111 10110100 00101010 11111010 01010001 10111000 1111111, when the "all ones" initial state is used. The same scrambler is used to scramble transmit data and to de-scramble receive data.

When transmitting, the initial state of the 802.11a scrambler will be set to a pseudo random non-zero state. The seven LSBs of the SERVICE field will be set to all zeros prior to scrambling to enable estimation of the initial state of the scrambler in the receiver. The contents of the SIGNAL field of the 802.11a are not scrambled. The PLCP length field of the 802.11a is an unsigned 12 bits integer that indicates the number of octets in the PSDU that the MAC is currently requesting the PHY to transmit. This value is used by the PHY to determine the number of octet transfers that will occur between the MAC and the PHY after receiving a request to start transmission.

The bits from 0-6 of the SERVICE field, which are transmitted first, are set to zeros and are used to synchronize the descrambler in the receiver. The remaining 9 bits (7-15) of the SERVICE field is reserved for future use. All reserved bits are set to zero.

1.5.3 Data Interleaving

In the 802.11a standard blocks inter - leaver interleaves all encoded data bits. The block size corresponds to the number of bits in a single OFDM symbol, NCBPS. The inter - leaver is defined by a two steps permutation. The first permutation ensures that adjacent coded bits are mapped onto nonadjacent subcarriers. The second ensures that adjacent coded bits are mapped alternately onto less and more significant bits of the constellation and, thereby, long runs of low reliability (LSB) bits are avoided.

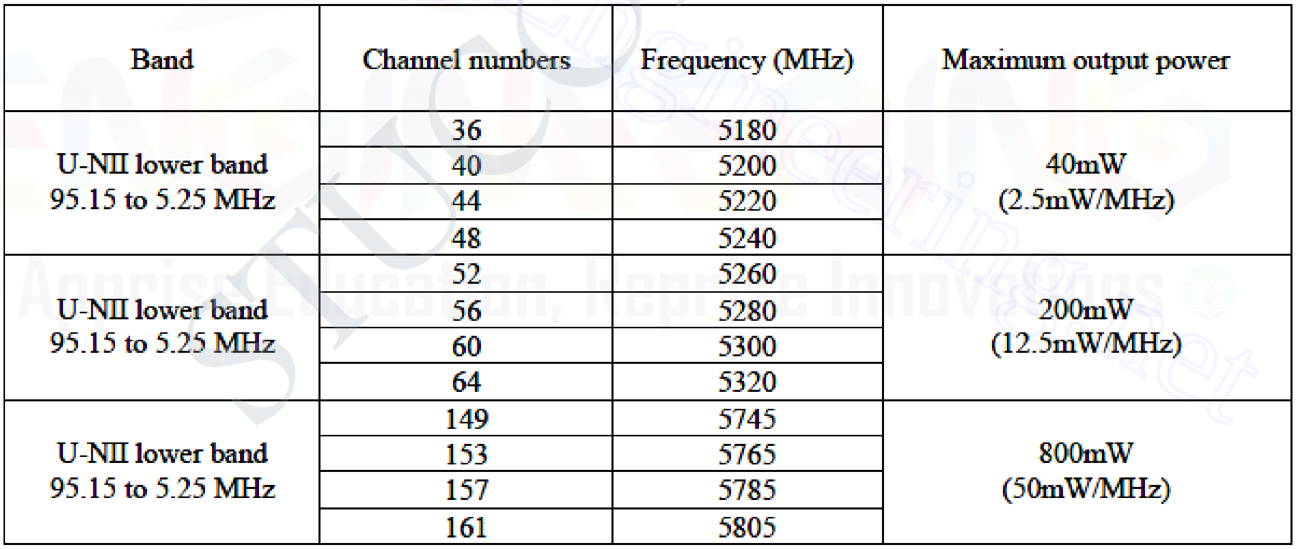
1.5.4 Operating frequency and maximum power of the 802.11a

For the 802.11a standard the 5 GHz U-NII frequency bans is segmented into three 100 MHz bands for operation in the US. The lower band ranges from 5.15 –5.25 GHz, the middle band ranges from 5.25-5.35 GHz and the upper band ranges from 5.725-5.825 GHz. The lower and middle band, accommodate 8 channels in a total bandwidth of 200 MHz and the upper band accommodates 4 channels in a 100 MHz bandwidth.

The frequency channel center frequencies are spaced 20 MHz apart. The outermost channels of the lower and middle bands are centered 30 MHz from the outer edges. In the upper band the outermost channel centers are 20 MHz from the outer edges. In addition to the frequency and channel allocations, transmit power is a key parameter regulated in the 5 GHz U-NII band. Three transmit power levels are specified: 40 mW, 200 mW and 800 mW. The upper band defines RF transmit power levels suitable for bridging applications while the lower band specifies a transmit power level suitable for short-range indoor home and small office environments.

The following table shows the operating frequency and maximum power of the 802.11a standard.

Table 1.5Operating frequency and maximum power of the 802.11a standard.



Band	Channel numbers	Frequency (MHz)	Maximum output power
U-NII lower band 95.15 to 5.25 MHz	36	5180	40mW (2.5mW/MHz)
	40	5200	
	44	5220	
	48	5240	
U-NII lower band 95.15 to 5.25 MHz	52	5260	200mW (12.5mW/MHz)
	56	5280	
	60	5300	
	64	5320	
U-NII lower band 95.15 to 5.25 MHz	149	5745	800mW (50mW/MHz)
	153	5765	
	157	5785	
	161	5805	

1.6 HIPERLAN

HIPERLAN is a European (ETSI) standardization initiative for a High Performance wireless Local Area Network. 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 distances (<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.

1.6.1 HiperLAN features:

- Range 50 m
- Slow mobility (1.4 m/s)
- Supports asynchronous and synchronous traffic
- Bit rate - 23.2 Mbit/s
- Description- wireless Ethernet
- Frequency range- 5 GHz

1.6.2 HIPERLAN 1

HIPERLAN1 was originally one out of four HIPERLANs envisaged, as ETSI decided tohave different types of networks for different purposes. The key feature of allfour networks is their integration of time-sensitive data transfer services. Overtime, names have changed and the former HIPERLANs 2, 3, and 4 are nowcalled HiperLAN2, HIPERACCESS, and HIPERLINK. The current focus is onHiperLAN2, a standard that comprises many elements from ETSI's BRAN (broadbandradio access networks) and wireless ATM activities.

ETSI describes HIPERLAN 1 as a wireless LAN supporting priorities andpacket life time for data transfer at 23.5 Mbit/s, including forwarding mechanisms,topology discovery, user data encryption, network identification andpower conservation mechanisms.

HIPERLAN 1 should operate at 5.1–5.3 GHzwith a range of 50 m in buildings at 1 W transmit power.The service offered by a HIPERLAN 1 is compatible with the standard MACservices known from IEEE 802.x LANs. Addressing is based on standard 48 bit MAC addresses. Confidentiality is ensured by an encryption/decryptionalgorithm that requires the identical keys and initialization vectors for successfuldecryption of a data stream encrypted by a sender.An innovative feature of HIPERLAN 1, which many other wireless networksdo not offer, is its ability to forward data packets using several relays. Relays canextend the communication on the MAC layer beyond the radio range.

Forpower conservation, a node may set up a specific wake-up pattern. This patterndetermines at what time the node is ready to receive, so that at other times, thenode can turn off its receiver and save energy. These nodes are called p-saversand need so-called p-supporters that contain information about the wake-uppatterns of all the p-savers they are responsible for. A p-supporter only forwardsdata to a p-saver at the moment the p-saver is awake. This action also requiresbuffering mechanisms for packets on p-supporting forwarders.

Elimination-yield non-preemptive priority multiple access (EY-NPMA) isnot only a complex acronym, but also the heart of the channel access providingpriorities and different access schemes. EY-NPMA divides the medium access ofdifferent competing nodes into three phases:

- Prioritization: Determine the highest priority of a data packet ready to be sent by competing nodes.
- Contention: Eliminate all but one of the contenders, if more than one sender has the highest current priority.
- Transmission: Finally, transmit the packet of the remaining node.

The dynamic extension is randomly chosenbetween 0 and 3 times 200 high bit rate. This extension further minimizes the probability of collisions accessing a free channelif stations are synchronized on higher layers and try to access the freechannel at the same time. HIPERLAN 1 also supports ‘channel access in thehidden elimination condition’ to handle the problem of hidden terminals as described in ETSI.
The contention phase is further subdivided into an elimination phase and ayield phase. The elimination phase is to eliminate as many contending nodes as possible. The result of the elimination phaseis a more or less constant number of remaining nodes, almost independent ofthe initial number of competing nodes. Finally, the yield phase completes thework of the elimination phase with the goal of only one remaining node.

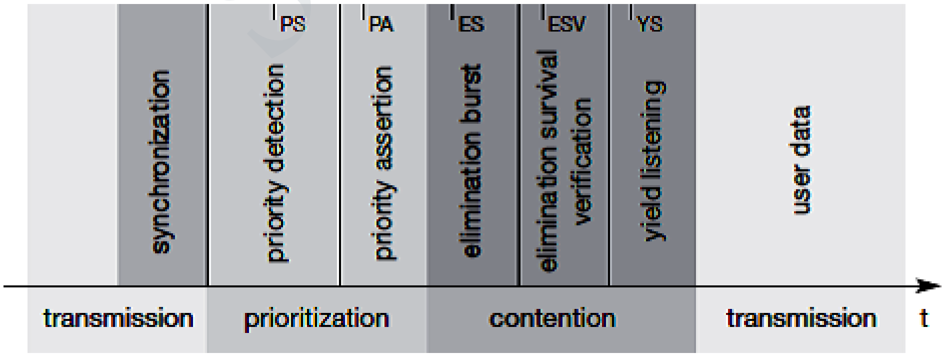


Fig. 1.17 Phases of HIPERLAN1

The above figure gives an overview of the three main phases and some moredetails which will be explained in the following sections. For every node readyto send data, the access cycle starts with synchronization to the current sender.

The first phase, prioritization, follows. After that, the elimination and yield partof the contention phase follow. Finally, the remaining node can transmit itsdata. Every phase has a certain duration which is measured in numbers of slotsand is determined by the variables IPS, IPA, IES, IESV, and IYS.

1.6.3 Prioritization phase

HIPERLAN 1 offers five different priorities for data packets ready to be sent.After one node has finished sending, many other nodes can compete for theright to send. The first objective of the prioritization phase is to make sure thatno node with a lower priority gains access to the medium while packets withhigher priority are waiting at other nodes. This mechanism always grants nodeswith higher priority access to the medium, no matter how high the load onlower priorities.

In the first step of the prioritization phase, the priority detection, time isdivided into five slots, slot 0 (highest priority) to slot 4 (lowest priority). Eachslot has a duration of IPS = 168 high rate bit-periods. If a node has the accesspriority p, it has to listen into the medium for p slots (priority detection).

Consider for example, that there are three nodes with data ready to besent, the packets of node 1 and node 2 having the priority 2, the packet of node3 having the priority 4. Then nodes 1, 2 and 3 listen into the medium and sense
slots 0 and 1 are idle. Nodes 1 and 2 both send a burst in slot 2 as priority assertion.
Node 3 stops its attempt to transmit its packet. In this example, theprioritization phase has taken three slots.

After this first phase at least one of the contending nodes will survive, thesurviving nodes being all nodes with the highest priority of this cycle.

1.6.4 Elimination Phase

Several nodes may now enter the elimination phase. Again, time is divided intoslots, using the elimination slot interval IES = 212 high rate bit periods. Thelength of an individual elimination burst is 0 to 12 slot intervals long, the probabilityof bursting within a slot is 0.5. The probability PE(n) of an eliminationburst to be in elimination slot intervals long is given by:

- ➢ $PE(n) = 0.5n+1$ for $0 \leq n < 12$
- ➢ $PE(n) = 0.512$ for $n = 12$

The elimination phase now resolves contention by means of elimination burstingand elimination survival verification. Each contending node sends anelimination burst with length n as determined via the probabilities and then listensto the channel during the survival verification interval IESV = 256 high ratebit periods. The burst sent is the same as for the priority assertion.

A contendingnode survives this elimination phase if, and only if, it senses the channel is idleduring its survival verification period. Otherwise, the node is eliminated andstops its attempt to send data during this transmission cycle.

1.6.5 Yield phase

During the yield phase, the remaining nodes only listen into the medium withoutsending any additional bursts. Again, time is divided into slots, this time called yieldslots with a duration of IYS = 168 high rate bit-periods. The length of an individualyield listening period can be 0 to 9 slots with equal likelihood. The probability PY(n)for a yield listening period to be n slots long is 0.1 for all n, $0 \leq n \leq 9$.

Each node now listens for its yield listening period. If it senses the channelis idle during the whole period, it has survived the yield listening. At least one node will survive this phase and can start totransmit data. This is what the other nodes with longer yield listening periodcan sense. It is important to note that at this point there can still be more thanone surviving node so a collision is still possible.

1.6.6 Transmission phase

A node that has survived the prioritization and contention phase can now sendits data, called a low bit-rate high bit-rate HIPERLAN 1 CAC protocol data unit(LBR-HBR HCPDU). This PDU can either be multicast or unicast. In case of aunicast transmission, the sender expects to receive an immediate acknowledgementfrom the destination, called an acknowledgement HCPDU (AK-HCPDU),which is an LBR HCPDU containing only an LBR part.

1.7 WATM(WIRELESS ATM)

Wireless ATM also called as wireless, mobile ATM, wmATM. It describes a transmission technology to specify a completecommunication system. IEEE WLAN originates from the data communication community whereas WLAN arise from the tele communication industry.

1.7.1 Development of WATM

- The wireless terminals are integrated into an ATM network for supporting different types of traffic streams as ATM does in fixed network.

- ATM network will scale well from LANs to WANs & mobility is needed in local and wide applications.
- WATM offers QOS for adequate support of multimedia data streams.
- For telecommunication service providers, merging of mobile wireless communication & ATM technology will leads to wireless ATM.

1.7.2 Standardization of WATM

WATM is a specific broadband wireless solution which is significantly meets the architectural and performance goals needed. WATM has been driven by the wide acceptance of ATM switching technology as a basis for broadband networks which support integrated services with QoS control. ATM signaling protocol (e.g.,Q2931) for connection establishment and QoS control also provide a suitable basis for mobility extensions such as handover and location management.

1.7.3 Extension of ATM

The following more general extensions of the ATM system also need to beconsidered for a mobile ATM:

•**Location management:** Similar to other cellular networks, WATM networksmust be able to locate a wireless terminal or a mobile user, i.e., tofind the current access point of the terminal to the network.

•**Mobile routing:** Even if the location of a terminal is known to the system,it still has to route the traffic through the network to the access point currentlyresponsible for the wireless terminal. Each time a user moves to anew access point, the system must reroute traffic.

•**Handover signalling:** The network must provide mechanisms which searchfor new access points, set up new connections between intermediate systemsand signal the actual change of the access point.

•**QoS and traffic control:** In contrast to wireless networks offering only besteffort traffic, and to cellular networks offering only a few different types of traffic,WATM should be able to offer many QoS parameters. To maintain theseparameters, all actions such as rerouting, handover etc. have to be controlled.The network must pay attention to the incoming traffic (and check if it conformsto some traffic contract) in a similar way to today's ATM (policing).

•**Network management:** All extensions of protocols or other mechanisms alsorequire an extension of the management functions to control the network.

1.7.4 Frame Format for WATM

Wireless header	ATM header (5 bytes)	ATM payload (48 bytes)	Wireless trailer

ATM packets has a fixed size of 53 bytes with a 48 – byte payload to facilitates fast switching in a multimedia environment.. The ATM cells were operating on reliable optical channels that do not need acknowledgement. When the same packet format is used in a wireless environment, another additional 16 bytes for the PLCP header is used and a few more for a wireless MAC layer that makes the overhead so large that a 48 – byte payload length makes the transmission inefficient.

1.7.6 WIRELESS ATM ARCHITECTURE

Wireless ATM architecture is obtained by incorporating new wireless protocols at the access level and extensions into the standard ATM protocol stack which is shown in Figure. At the access level, new protocols are needed for:

- Physical layer radio channels between the mobile terminals and base stations,
- Medium access control (MAC) to arbitrate the shared use of the radio channels by the mobile terminals,
- Data/logical link control (DLC/LLC) to detect and/or correct the radio channel errors and maintain end-to-end QoS.
- Wireless control to support such functions as radio resource management at the physical, MAC and DLC layers, as well as mobility management.

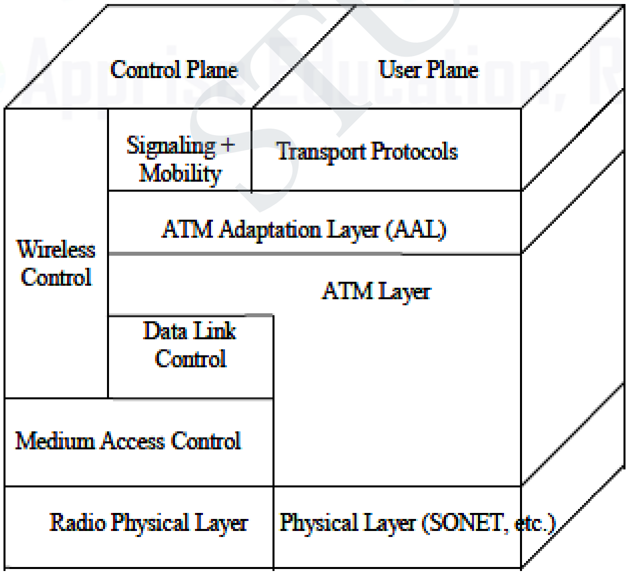


Fig. 1.18 Wireless ATM Architecture

WATM protocol architecture is based on integration of radio access and mobilityfeatures capabilities within the standard ATM protocol stack. A WATM system may be partitioned into two relativelyindependent parts: a mobile ATM infrastructure and a radio access segment, each ofwhich can be designed and specified separately. This facilitates standardization bymultiple organizations and allows for gradual evolution of radio access technologieswithout having to modify the core mobile ATM network specification.

The WATM radio access structure consists of Radio Physical Layer (PHY),Medium Access Control (MAC), Data Link Control (DLC) and wireless control. The WATM DLC layer interfaces each ATM virtual circuit (both service data andsignaling) with the ATM network layer above. An additional wireless control interface isprovided within the control plane to deal with radio link specific control functions such asinitial registration, resource allocation, and power control.

1.7.7 Handover

As a mobile terminal moves from one place to another, it becomes necessary to hand over its ongoing connections from the old radio port to the new one. The decision to change the radio port is made either by the mobile terminal or the base station based on signal strength measurements.

There are three handover scenarios. In the first scenario, the old andnew radio ports belong to the same base station. This case can be handled completely bythe radio-level protocols. In the second scenario, the target radio port belongs to differentbase stations, with the old and new base stations connected to (and supported by) thesame ATM switch.

This latter switch controls the rerouting of the connections from theold to the new base stations, and is called the crossover switch. In the third case, each ofthe two base stations is connected to its own access switch. An intermediate switch actsas the crossover switch for rerouting connections from the old to the new access switch.

The discovery and selection of the crossover switch is an important issue in handover. Unless handover occurs within the same base station, ATM-level protocols are said to be required for discovery of crossover switch, path rerouting, etc.

There are two typesof handovers:

1. Soft hand over
2. Hard hand over

In soft handover, the mobile terminal connections are passed to the new base station without interrupting communication with the old base station.

In hard handover, the connections are interrupted at the old base station and reestablished at the new base station. Only hard handover is supported in the current WATM specification.

1.7.8 Location Management

Location management is required to maintainthe association between the mobile's physical location at a foreign switch and itspermanent address at the home switch. To achieve this, a mobile terminal must registerwith the base station of every new service area it may enter. The main purpose of locationmanagement in wireless ATM networks is to allow a mobile terminal to use itspermanent address in connection set-up messages regardless of its attachment to thenetwork. In addition, location management incorporates features for access control,privacy, accounting and inter-provider roaming.

The functions of location management are handled by mobility-enhancedswitches, location servers, authentication servers, and mobile terminals. The locationserver is a database of associations between the permanent and temporary addresses ofmobile terminals. The temporary address identifies the location of the mobile terminalaway from its permanent home address (switch). This database is queried and updatedaccording to specific protocols.

On the other hand, the authentication server is a databasecontaining secure information relating to the privacy and identification of each mobileterminal.
Location management is required in local and wide area WATM networks. Locallocation management enables any host connected to a switch to establish a virtual circuitwith any mobile terminal moving between the base stations within a local network. Widearea location management permits mobile terminals attached to one local network toestablish connections with hosts (or other mobile terminals) attached to remote networkgroups.

1.7.9 Mobile quality of service

Quality of service (QoS) guarantees are one of the main advantages predictedfor WATM networks compared to, e.g., mobile IP working over packet radio networks.
While the internet protocol IP does not guarantee QoS, ATM networksdo (at the cost of higher complexity). WATM networks should provide mobile

QoS (M-QoS). M-QoS is composed of three different parts:

- Wired QoS: The infrastructure network needed for WATM has the sameQoS properties as any wired ATM network. Typical traditional QoS parametersare link delay, cell delay variation, bandwidth, cell error rate etc.
- Wireless QoS: The QoS properties of the wireless part of a WATM networkdiffer from those of the wired part. Again, link delay and error rate can bespecified, but

now error rate is typically some order of magnitude that ishigher than, e.g., fiber optics. Channel reservation and multiplexing mechanismsat the air interface strongly influence cell delay variation.

- Handover QoS: A new set of QoS parameters are introduced by handover.For example, handover blocking due to limited resources at target accesspoints, cell loss during handover, or the speed of the whole handover procedurerepresent critical factors for QoS.

1.8 BRAN(BROADBAND RADIO ACCESS NETWORK)

The broadband radio access networks (BRAN), which have been standardized bythe European Telecommunications Standards Institute (ETSI).

Many service providers experience problemsgetting access to customers because the telephone infrastructure belongs toa few big companies. One possible technology to provide network access for customersis radio. The advantages of radio access are high flexibility and quickinstallation. Different types of traffic are supported, one can multiplex traffic forhigher efficiency, and the connection can be asymmetrical.

Radio access allows for economicgrowth of access bandwidth. If more bandwidth is needed, additional transceiversystems can be installed easily. For wired transmission this would involve theinstallation of additional wires. The primary market for BRAN includes privatecustomers and small to medium-sized companies with Internet applications,multi-media conferencing, and virtual private networks. The BRAN standard andIEEE 802.16 (Broadband wireless access, IEEE, 2002b) have similar goals.

BRAN has specified four different network types

- 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.
- 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. Transmission range is 50 m with support of slow (< 10 m/s) mobility.
- 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. Transmission range is up to 5 km, data rates of up to 25 Mbit/s are supported. However, many proprietary products already offer 155 Mbit/s and more, plus QoS.

- HIPERLINK: To connect different HIPERLAN access points or HIPERACCESSnodes with a high-speed link, HIPERLINK technology can be chosen.HIPERLINK provides a fixed point-to-point connection with up to 155 Mbit/s.

BRAN technology is independent from the protocols ofthe fixed network. BRAN can be used for ATM and TCP/IP networks. Based on possibly differentphysical layers, the DLC layer of BRAN offers a common interface to higherlayers. To cover special characteristics of wireless links and to adapt directly to differenthigher layer network technologies, BRAN provides a network convergencesublayer. This is the layer which can be used by a wireless ATM network, Ethernet,Firewire, or an IP network. In the case of BRAN as the RAL for WATM, the core ATMnetwork would use services of the BRAN network convergence sublayer.

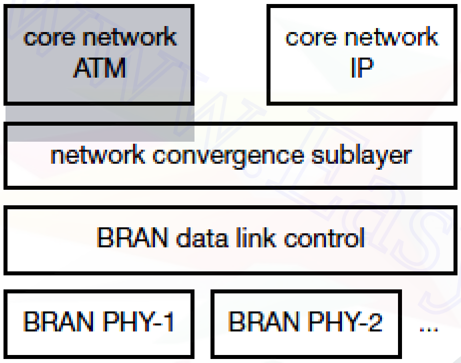


Fig. 1.19 Layered model of BRAN wireless access networks

1.9 HiperLAN2

It is a mobile short-range access network specified in the Broadband Radio Access Networks (BRAN) project chartered by the European Telecommunications Standards Institute (ETSI). HiperLAN/2, a competes directly with IEEE 802.11g/n, aka Wi-Fi.HiperLAN2 supports both asynchronous data and time critical services (e.g. packetized voice and video) that are bounded by specific time delays to achieve an acceptable Quality of Service (QoS) being developed under the auspices of the ETSI’s Project BRAN (Broadband Radio Access Networks).

The HiperLAN2 standard is nearly identical to 802.11 in terms of its physical layers – both use OFDM technology to achieve their data rates, for instance – but is very different at the MAC (Media Access Control) level and in the way the data packets are formed and devices are addressed. On a technical level, whereas 802.11 can be viewed as true wireless Ethernet, HiperLAN2 is more akin to wireless Asynchronous Transfer Mode (ATM). It operates by sharing the 20MHz channels in the 5GHz spectrum in time, using Time Division Multiple Access (TDMA) to provide QoS through ATM-like mechanisms.

It supports two basic modes of operation: centralized mode and direct mode. The centralized mode is used in the cellular networking topology where each radio cell is controlled by an access point covering a certain geographical area. In this mode, a mobile terminal communicates with other mobile terminals or with the core network via an access point. It is mainly used in business applications – both indoors and outdoors – where an area much larger than a radio cell has to be covered. The direct mode is used in the ad-hoc networking topology – mainly in typical private home environments – where a radio cell covers the whole serving area.

1.9.1 Features of HiperLAN/2:

- High-speed transmission
- Connection-oriented
- Quality-of-Service (QoS) support ·
- Automatic frequency allocation
- Security support
- Mobility support
- Network & application independent
- Power save

1.9.2 High-speed transmission

HiperLAN/2 has a very high transmission rate, which at the physical layer extends up to 54 Mbit/s and on layer 3 up to 25 Mbit/s. To achieve this, HiperLAN/2 makes use of Orthogonal Frequency Digital Multiplexing (OFDM) to transmit the analogue signals. OFDM is very efficient in time-dispersive environments, e.g within offices, where the transmitted radio signals are reflected from many points, leading to different propagation times before they eventually reach the receiver. Above the physical layer, the Medium Access Control (MAC) protocol is all new which implements a form of dynamic time-division duplex to allow for most efficient utilization of radio resources.

1.9.3 Connection-oriented

In a HiperLAN/2 network, data is transmitted between the MT and the AP that have been established prior to the transmission using signaling functions of the HiperLAN/2 control plane. Connections are time-division-multiplexed over the air interface. There are two types of connections, point to-point and point-to-multipoint.

Point-to-point connections are bidirectional whereas point-to-multipoint is unidirectional in the direction towards the Mobile Terminal. In addition, there is also a dedicated broadcast channel through which traffic reaches all terminals transmitted from one AP.

1.9.4 Quality of service support:

With the help of connections, support of QoSis much simpler. Each connection has its own set of QoS parameters (bandwidth,delay, jitter, bit error rate etc.). A more simplistic scheme usingpriorities only is available.

1.9.5Dynamic frequency selection:

In a HiperLAN/2 network, there is no need for manual frequency planning as in cellular networks like GSM. The radio base stations, which are called Access Points in HiperLAN/2, have a built-in support for automatically selecting an appropriate radio channel for transmission within each AP's coverage area. An AP listens to neighboring APs as well as to other radio sources in the environment, and selects an appropriate radio channel based on both what radio channels are already in use by those other APs and to minimize interference with the environment.

1.9.6 Security support

The HiperLAN/2 network has support for both authentication and encryption. With authentication both the AP and the MT can authenticate each other to ensure authorized access to the network (from the AP's point of view) or to ensure access to a valid network operator (from the MT's point of view). Authentication relies on the existence of a supporting function, such as a directory service, but which is outside the scope of HiperLAN/2. The user traffic on established connections can be encrypted to protect against for instance eaves-dropping and man-in-middle attacks.

1.9.7 Mobility support

Mobile terminals can move around while transmissionalways takes place between the terminal and the access point with the bestradio signal. Handover between access points is performed automatically. Ifenough resources are available, all connections including their QoS parameterswill be supported by a new access point after handover. However,some data packets may be lost during handover.

1.9.8 Application and network independence:

HiperLAN2 was not designedwith a certain group of applications or networks in mind. Access points canconnect to LANs running Ethernet as well as IEEE 1394 (Firewire) systemsused to connect home audio/video devices. Interoperation with 3G networks is also supported, so not only best effort data is supported but alsothe wireless connection of, e.g., a digital camera with a TV set for livestreaming of video data.

1.9.9 Power save

Mobile terminals can negotiate certain wake-up patterns tosave power. Depending on the sleep periods either short latency requirementsor low power requirements can be supported.

1.9.10 Protocol architecture & the layers

In the protocol reference model for the HiperLAN/2 radio interface is depicted. The protocol stack is divided into a control plane part and a user plane part i.e. user plane includes functions for transmission of traffic over established connections, and the control plane includes functions for the control of connection establishment, release, and supervision. The HiperLAN/2 protocol has three basic layers; Physical layer (PHY), Data Link Control layer (DLC), and the Convergence layer (CL). At the moment, there is only control plane functionality defined within DLC.

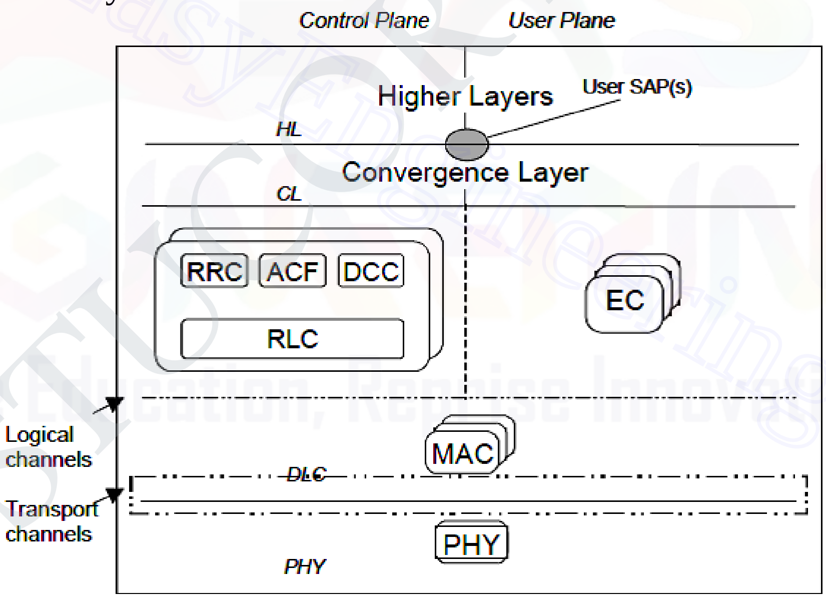


Fig. 1.20 HiperLAN/2 protocol reference model

1.9.11 Physical Layer

The transmission format on the physical layer is a burst, which consists of a preamble part and a data part, where the latter could originate from each of the transport channels within DLC. The channel spacing is 20 MHz, which allows high bit rates per channel but still has a reasonable number of channels in the allocated spectrum.52 subcarriers are used per channel, where 48 subcarriers carry actual data and 4 subcarriers are pilots which facilitate phase tracking for coherent demodulation. The duration of the guard interval is equal to

800 ns, which is sufficient to enable good performance on channels with delay spread of up to 250 ns.

1.9.12 Data Link Control Layer

The Data Link Control (DLC) layer constitutes the logical link between an AP and the MTs. The DLC includes functions for medium access and transmission (user plane) as well as terminal/user and connection handling (controlplane). Thus, the DLC layer consists of a set of sublayers: - Medium Access Control (MAC) protocol. - Error Control (EC) protocol - Radio Link Control (RLC) protocol with theassociated signalingentities DLC Connection Control (DCC), the Radio Resource Control (RRC) and the Association Control Function (ACF)

Each MAC frame is further sub-divided into four phases withvariable boundaries:

- Broadcast phase: The AP of a cell broadcasts the content of the current frame plus information about the cell (identification, status, resources).
- Downlink phase: Transmission of user data from an AP to the MTs.
- Uplink phase: Transmission of user data from MTs to an AP.
- Random access phase: Capacity requests from already registered MTs and access requests from non-registered MTs (slotted Aloha).

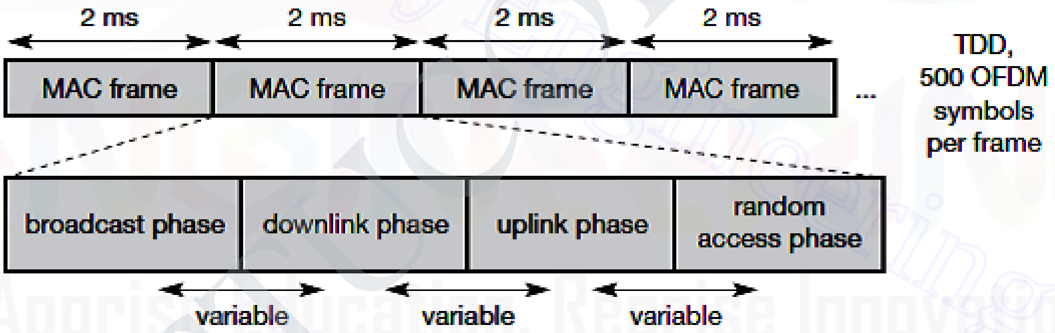


Fig. 1.21 HiperLAN2 MAC Frames

1.9.13 Convergence Layer

The convergence layer (CL) has two main functions: adapting service request from higher layers to the service offered by the DLC and to convert the higher layer packets (SDUs) with variable or possibly fixed size into a fixed size that is used within the DLC. The padding, segmentation and reassembly function of the fixed size DLC SDUs is one key issue that makes it possible to standardize and implement a DLC and PHY that is independent of the fixed network to which the HiperLAN/2 network is connected. The generic architecture of the CL makes HiperLAN/2 suitable as a radio access network for a diversity of fixed networks, e.g. Ethernet, IP, ATM, UMTS, etc. There are currently two different types of CLs defined; cell-based and packet-based. The former is intended for interconnection to ATM networks, whereas the latter can be used in a variety of configurations depending on fixed network type and how the interworking is specified.

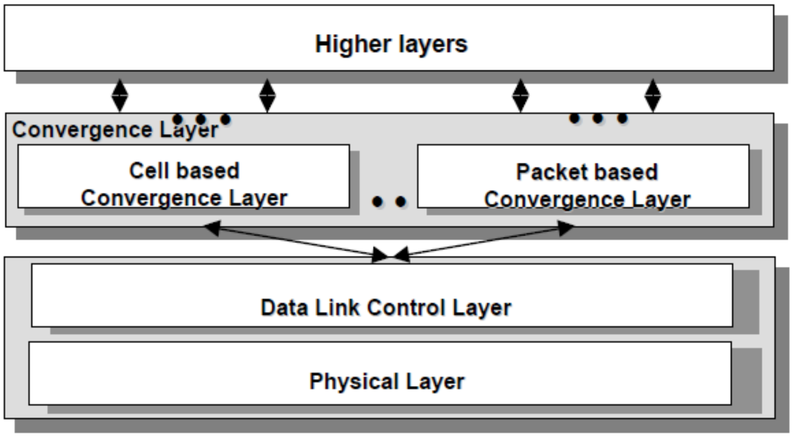


Fig. 1.22 Convergence layer

1.9.14 Modes of Operation

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

- Centralized mode (CM): All APs are connected to a corenetwork and MTs are associated with APs. Even if two MTs share the samecell, all data is transferred via the AP. In this mode the AP will takescomplete control of everything.
- Direct mode (DM): Data is directly exchanged between MTs ifthey can receive each other, but the network still has to be controlled. Thiscan be done via an AP that contains a central controller (CC) anyway or viaan MT that contains the CC functionality. There is no real differencebetween an AP and a CC besides the fact that APs are always connected toan infrastructure but here only the CC functionality is needed. This is whythe standard coined two different names. IEEE 802.11, too, offers an ad-hocmode, but not the CC functionality for QoS support.

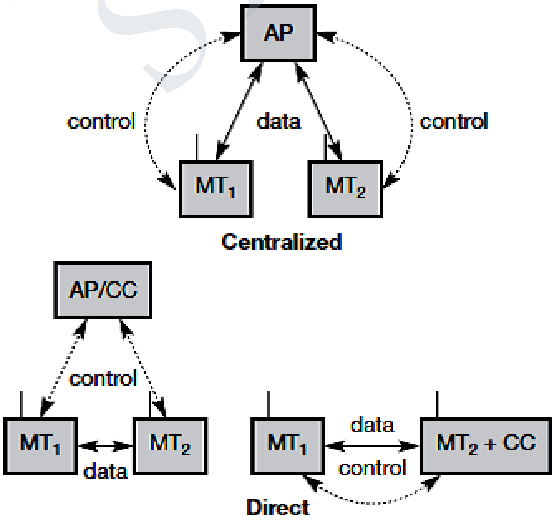


Fig. 1.23 HiperLAN2 Modes

HiperLAN2 defines six channels for data transfer.

- Broadcast channel (BCH): This channel conveys basic information for the radio cell to all MTs. This comprises the identification and current transmission power of the AP. Furthermore, the channel contains pointers to the FCH and RCH which allows for a flexible structure of the MAC frame. The length is 15 bytes.
- Frame channel (FCH): This channel contains a directory of the downlink and uplink phases (LCHs, SCHs, and empty parts). The length is a multiple of 27 bytes.
- Access feedback channel (ACH): This channel gives feedback to MTs regarding the random access during the RCH of the previous frame. The length is 9 bytes.
- Long transport channel (LCH): This channel transports user and control data for downlinks and uplinks. The length is 54 bytes.
- Short transport channel (SCH): This channel transports control data for downlinks and uplinks. The length is 9 bytes.
- Random channel (RCH): This channel is needed to give an MT the opportunity to send information to the AP/CC even without a granted SCH.The length is 9 bytes. A maximum number of 31 RCHs is currently supported.

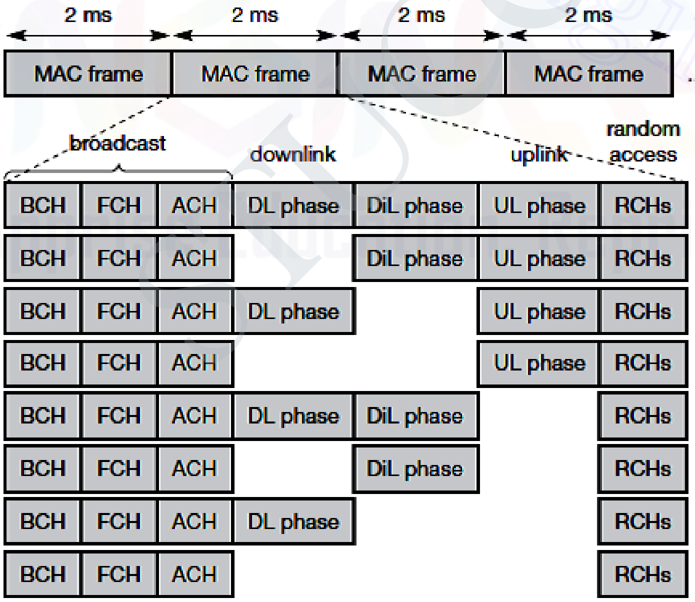

2 ms
2 ms
2 ms
2 ms
MAC frame
MAC frame
MAC frame
MAC frame
...
broadcast
downlink
uplink
random access
BCH FCH ACH DL phase DiL phase UL phase RCHs
BCH FCH ACH DiL phase UL phase RCHs
BCH FCH ACH DL phase UL phase RCHs
BCH FCH ACH UL phase RCHs
BCH FCH ACH DL phase DiL phase RCHs
BCH FCH ACH DiL phase RCHs
BCH FCH ACH DL phase RCHs
BCH FCH ACH RCHs

Fig. 1.24 MAC frames Configurations

1.10 Bluetooth

Bluetooth is a standard for short range, low power, low cost wireless communication that uses radio technology. Although originally envisioned as a cable-replacement technology.Bluetooth technology can be used at home, in the office, in the car, etc. This technology allows to the users instantaneous connections of voice and information between several devices in real time. The way of transmission used assures protection against interferences and safety in the sending of information.

The Bluetooth is a small microchip that operates in a band of available frequency throughout the world. Communications can realize point to point and point multipoint.The standard Bluetooth operates in the band of 2,4 GHz. Though worldwide, this band is available, the width of the band can differ in different countries. This is the frequency of band of the scientific and medical industries 2.45 GHz (ISM*). The ranges of the bandwidth in The United States and Europe are between 2.400 to 2.483,5 MHz and it covers part of France and Spain. The ranges of the bandwidth in Japan are between 2.471 to 2.497 MHz.

1.10.1 User scenarios

Many different user scenarios can be imagined for wireless piconets or WPANs:

●Connection of peripheral devices: Most of the devices are connected to adesktop computer via wires (e.g., keyboard, mouse, joystick, headset, speakers).This type of connection has several disadvantages: each device has itsown type of cable, different plugs are needed, and wires block office space. In awireless network, no wires are needed for data transmission. However, batteriesnow have to replace the power supply, as the wires not only transferdata but also supply the peripheral devices with power.

● Support of ad-hoc networking: Imagine several people coming together,
Wireless networks can support interactive exchange of data as a group. Small devices might not have WLAN adapters followingthe IEEE 802.11 standard, but cheaper Bluetooth chips built in.

●Bridging of networks: Using wireless piconets, a mobile phone can be connectedto a PDA or laptop in a simple way. Mobile phones will not have fullWLAN adapters built in, but could have a Bluetooth chip. The mobile phonecan then act as a bridge between the local piconet and, e.g., the global GSM network.

1.10.2 ARCHITECTURE OVERVIEW

Bluetooth link control hardware, integrated as either one chip or a radio module and a baseband module, implements the RF, baseband, and link manager portions of the Bluetooth specification. This hardware handles radio transmission and reception as well as required digital signal processing for the baseband protocol. Its functions include establishing connections, support for asynchronous (data) and synchronous (voice) links, error correction, and authentication. The link manager firmware provided with the

baseband CPU performs low-level device discovery, link setup, authentication, and link configuration.
Bluetooth operates on 79 channels in the 2.4 GHzband with 1 MHz carrier spacing. Each device performs frequency hopping with1,600 hops/s in a pseudo random fashion. A piconet isa collection of Bluetooth devices which are synchronized to the same hopping sequence. Onedevice in the piconet can act as master (M), all other devices connected to themaster must act as slaves (S).

The master determines the hopping pattern in thepiconet and the slaves have to synchronize to this pattern. Each piconet has aunique hopping pattern. If a device wants to participate it has to synchronize tothis. Two additional types of devices are shown: parked devices (P) cannotactively participate in the piconet (i.e., they do not have a connection), but areknown and can be reactivated within some milliseconds.

Devices in stand-by (SB) do not participate in the piconet. Each piconet hasexactly one master and up to seven simultaneous slaves. More than 200 devicescan be parked. The reason for the upper limit of eight active devices is the 3-bitaddress used in Bluetooth. If a parked device wants to communicate and thereare already seven active slaves, one slave has to switch to park mode to allowthe parked device to switch to active mode.

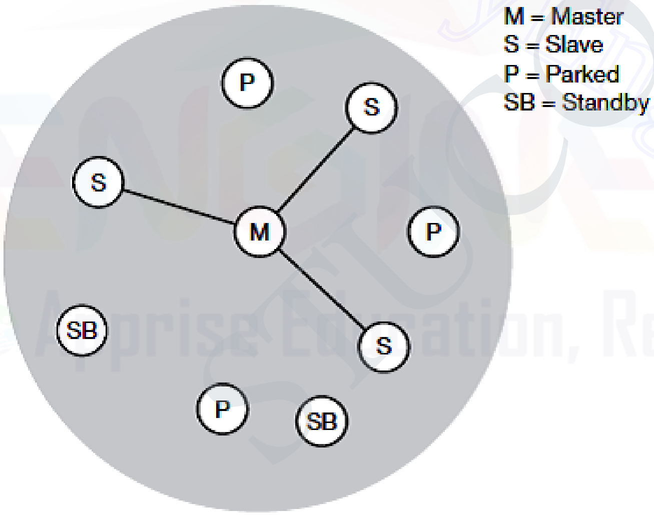


Fig. 1.25 Bluetooth piconet

The Piconetare several devices that are in the same radio of coverage where they share the same channel and that is constituted between two and eight of these units. Every device has the unique direction of 48 bits, based on the standard IEEE 802.11 for WLAN, whereas the Scatternet formed by the connection of a Piconet to other one, with a maximum of interconnections of ten Piconets.

As all activedevices have to use the same hopping sequence they must be synchronized. Thefirst step involves a master sending its clock and device ID. All Bluetooth deviceshave the same networking capabilities, i.e., they can be master or slave. There is nodistinction between terminals and base stations, any two or more devices can forma piconet.

The unit establishing the piconet automatically becomes the master, all other devices will be slaves.The phase in the hopping pattern is determinedby the master's clock. After adjusting the internal clock according to themaster a device may participate in the piconet. All active devices are assigned a3-bit active member address (AMA). All parked devices use an 8-bit parkedmember address (PMA). Devices in stand-by do not need an address.

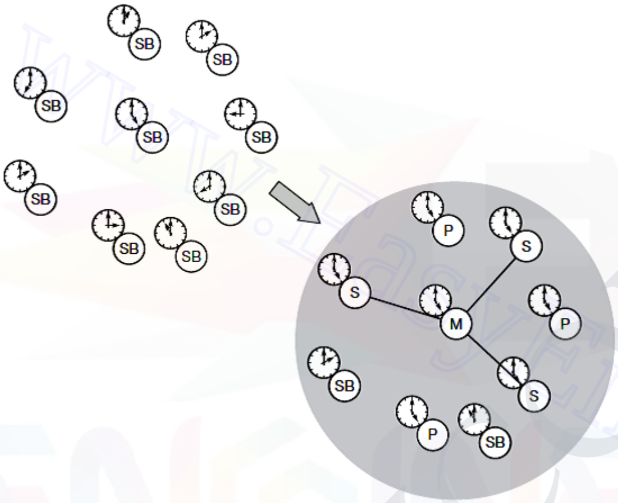


Fig. 1.26 Forming a Bluetooth piconet

All users within one piconet have the same hopping sequence and share thesame 1 MHz channel. As more users join the piconet, the throughput per userdrops quickly (a single piconet offers less than 1 Mbit/s gross data rate).Thisled to the idea of forming groups of piconets calledscatternet. If a device wants to participate in more than one piconet, it has to synchronizeto the hopping sequence of the piconet it wants to take part in.

If adevice acts as slave in one piconet, it simply starts to synchronize with the hoppingsequence of the piconet it wants to join. After synchronization, it acts as aslave in this piconet and no longer participates in its former piconet. To enablesynchronization, a slave has to know the identity of the master that determinesthe hopping sequence of a piconet. Before leaving one piconet, a slave informsthe current master that it will be unavailable for a certain amount of time. Theremaining devices in the piconet continue to communicate as usual.

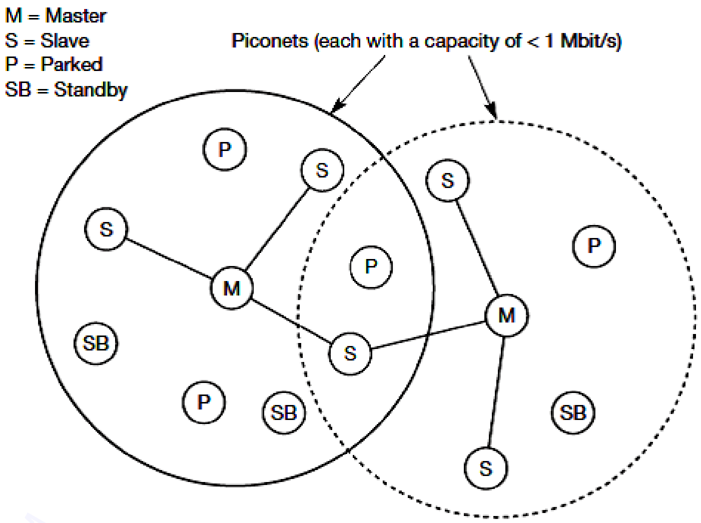


Fig. 1.27 Bluetooth scatternet

1.10.3 Protocols Stack

The protocol architecture of the Bluetooth consists of following in a Bluetooth protocol stack:

- Core protocols consisting 5 layer protocol stack viz. radio,baseband,link manager protocol, logical link control and adaptation protocol, service discovery protocol.

- Cable replacement protocol,RFCOMM

- Telephony Control Protocols

- Adopted protocols viz. PPP,TCP/UDP/IP,OBEX and WAE/WAPCore protocols

Radio: This protocol specification defines air interface, frequency bands, frequency hopping specifications, modulation technique used and transmits power classes.

Baseband: Addressing scheme, packet frame format, timing and power control algorithms required for establishing connection between Bluetooth devices within piconet defined in this part of protocol specification.

Link Manager Protocol: It is responsible to establish link between Bluetooth devices and to maintain the link between them. This protocol also includes authentication and encryption specifications. Negotiation of packet sizes between devices can be taken care by this.

Logical link control and adaptation protocol: This L2CAP protocol adapts upper layer frame to baseband layer frame format and vice versa. L2CAP take care of both connections oriented and connectionless services.

Service discovery protocol: Service related queries including device information can be taken care at this protocol so that connection can be established between Bluetooth devices.

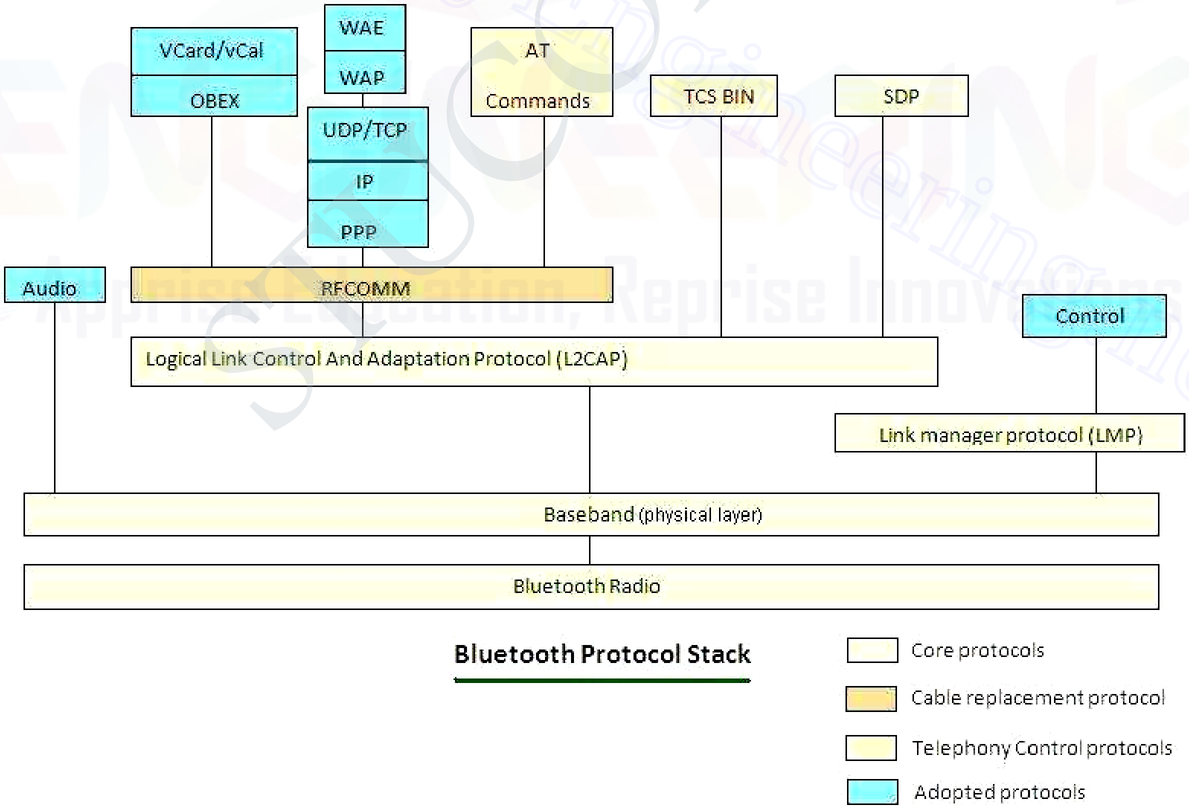

WAE
WAP
VCard/vCal
OBEX
UDP/TCP
IP
PPP
AT
Commands
TCS BIN
SDP
Audio
RFCOMM
Control
Logical Link Control And Adaptation Protocol (L2CAP)
Link manager protocol (LMP)
Baseband (physical layer)
Bluetooth Radio
Bluetooth Protocol Stack
Core protocols
Cable replacement protocol
Telephony Control protocols
Adopted protocols

Fig.1.28 Bluetooth Protocol Stack

1.10.4 Radio Layer

• The Bluetooth radio layer corresponds to the physical layer of OSI model. It deals with radio transmission and modulation.The radio layer moves data from master to slave or vice versa. It is a low power system that uses 2.4 GHz ISM band in a range of 10 meters.

• This band is divided into 79 channels of 1MHz each. Bluetooth uses the Frequency Hopping Spread Spectrum (FHSS) method in the physical layer to avoid interference from other devices or networks.

The radio specification defines the carrier frequencies and output power. Bluetoothdevices will be integrated into typical mobile devices and rely on battery power.This requires small, low power chips which can be built into handheld devices.The combined use for data and voice transmission has to be reflected in thedesign, i.e., Bluetooth has to support multi-media data.

Bluetooth uses the license-free frequency band at 2.4 GHz allowing forworldwide operation with some minor adaptations to national restrictions. Afrequency-hopping/time-division duplex scheme is used for transmission, witha fast hopping rate of 1,600 hops per second. The time between two hops iscalled a slot, which is an interval of 625 µs. Each slot uses a different frequency.In order to change bits into a signal, it uses a FSK with Gaussian bandwidth filtering.

Bluetooth transceivers use Gaussian FSK for modulation and are available inthree classes:

- Power class 1: Maximum power is 100 mW and minimum is 1 mW (typ. 100 m range without obstacles). Power control is mandatory.
- Power class 2: Maximum power is 2.5 mW, nominal power is 1 mW, andminimum power is 0.25 mW (typ. 10 m range without obstacles). Power control is optional.
- Power class 3: Maximum power is 1 mW.

1.10.5 Baseband Layer

Baseband layer is equivalent to the MAC sublayer in LANs.

The baseband layer controls transmission of frames in association with frequency hopping.Master and slave stations communicate with each other using time slots.The master in a piconet takes the channel to transmit in even-numbered hops, and slavestransmit in odd-numbered hops, reflecting a time-division duplex for all devices in a piconet.

A single frame can be transmitted in the duration of one, three, or five hops. Depending onthe nature of the logical link between a slave and the master, two types of links are offered.Bluetooth uses a form of TDMA called TDD-TDMA (time division duplex TDMA).The master in each piconet defines the time slot of 625 µsec.

In TDD- TDMA, communication is half duplex in which receiver can send and receive data but not at the same time.If the piconet has only no slave; the master uses even numbered slots (0, 2, 4, ...) and the slave uses odd-numbered slots (1, 3, 5, ...). Both master and slave communicate in half duplex mode. In slot 0, master sends & secondary receives; in slot 1, secondary sends and primary receives.If piconet has more than one slave, the master uses even numbered slots. The slave sends in the next odd-numbered slot if the packet in the previous slot was addressed to it.

The baseband layer has defined some types of frames that correspond to various purposes ofthe baseband frames. Different types of frames can carry different sizes of payload data anderror-correction schemes. In particular, the access code field in a baseband frame indicatesthe purpose of the frame in a special state. For example, a frame with the inquiry accesscode (IAC) will be sent when a device elects to scan for other devices within the radio rangein a series of 32 frequency hops.

Bluetooth devices can be configured to periodically hopaccording to the inquiry scan hopping sequence to scan inquires. When an inquiry is detected,the device, now the slave, will reply with its address and timing information to the master,and then the master and the slave begin the paging process to determine a common hoppingsequence to establish a connection. Eventually, both the master and the slave will hop on thesame sequence of channels for the duration of the connection.

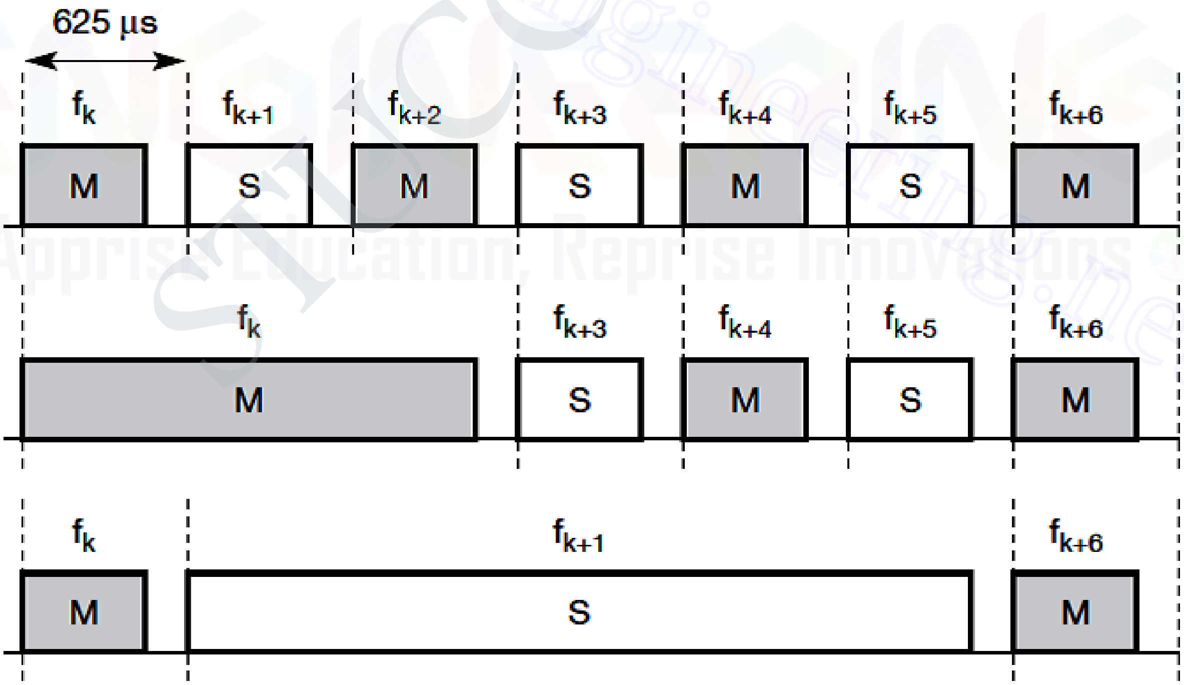


Fig. 1.29 Frequency selection during data transmission using 1, 3, 5 packet slots

1.10.6 Link manager protocol

The Link Manager (LM) translates the commands into operations at the Baseband level, managing the following operations.

1) Attaching slaves to piconets, and allocating their active member addresses.
2) Breaking connections to detach Slaves from a piconet.
3) Configuring the link including Master/Slave switches
4) Establishing ACL and SCO links.
5) Putting connections into Low Power modes: Hold, Sniff and Park.
6) Controlling test modes.

A bluetooth Link Manager communicates with Link Managers on other Bluetooth devices using the Link Management protocol (LMP).

The link can be configured at any time, including at mode changes, quality of service changes, packet type changes and any power level changes. Finally, information about an active link can be retrieved at any time.When the connection is no longer required, LMP can cause disconnection.

The link manager protocol (LMP) manages various aspects of the radio linkbetween a master and a slave and the current parameter setting of the devices.LMP enhances baseband functionality, but higher layers can still directly accessthe baseband. The following groups of functions are covered by the LMP:

• Authentication, pairing, and encryption: Although basic authentication ishandled in the baseband, LMP has to control the exchange of random numbersand signed responses. The pairing service is needed to establish an initialtrust relationship between two devices that have never communicated before.

The result of pairing is a link key. This may be changed, accepted or rejected.LMP is not directly involved in the encryption process, but sets the encryptionmode (no encryption, point-to-point, or broadcast), key size, and random speed.

• Synchronization: Precise synchronization is of major importance within aBluetooth network. The clock offset is updated each time a packet isreceived from the master. Additionally, special synchronization packets canbe received. Devices can also exchange timing information related to thetime differences (slot boundaries) between two adjacent piconets.

• Capability negotiation: Not only the version of the LMP can be exchangedbut also information about the supported features. Not all Bluetoothdevices will support all features that are described in the standard, sodevices have to agree the usage of, e.g., multi-slot packets, encryption, SCOlinks, voice encoding, park/sniff/hold mode, HV2/HV3packets etc.

- Quality of service negotiation: Different parameters control the QoS of aBluetooth device at these lower layers. The poll interval, i.e., the maximumtime between transmissions from a master to a particular slave, controls thelatency and transfer capacity. Depending on the quality of the channel, DMor DH packets may be used (i.e., 2/3 FEC protection or no protection). Thenumber of repetitions for broadcast packets can be controlled. A master canalso limit the number of slots available for slaves' answers to increase itsown bandwidth.
- Power control: A Bluetooth device can measure the received signalstrength. Depending on this signal level the device can direct the sender ofthe measured signal to increase or decrease its transmitting power.

1.11 WIMAX(IEEE802.16)

WiMAX (Worldwide Interoperability for Microwave Access) is a wireless communications standards family designed to provide 30 to 40 megabit-per-second data rates, with the update providing up to 1 Gbit/s for fixed stations.

WiMAX supports several networking usage models:

1. To transfer data across an Internet service provider network, commonly called *backhaul*)
2. A form of fixed wireless broadband Internet access, replacing satellite Internet service
3. A form of mobile Internet access to competes directly with LTE technology

The main aim is to promote and certify compatibility and interoperability of devices based on the 802.16 specification, and to develop such devices for the marketplace. WiMAX is expected to provide about 10 megabits per second of upload and download, at a distance of 10 kilometers from a base station.

WiMax is a standardized wireless version of Ethernet intended primarily as an alternative to wire technologies (such as Cable Modems, DSL and T1/E1 links) to provide broadband access to customer premises.

WiMax transmitters can span a distance of several miles (kilometers) with data rates reaching up to 75 megabits per second (Mbps). A number of wireless signaling options exist for WIMax ranging anywhere from 2 GHz up to 66 GHz bands.

Primarily due to its much higher cost, WiMAX is not a replacement for Wi-Fi home networking or wireless hotspot technologies.

WiMAX would operate similar to WiFi, but at higher speeds over greater distances and for a greater number of users. WiMAX has the ability to provide service even in areas that are difficult for wired infrastructure to reach and the ability to overcome the physical limitations of traditional wired infrastructure.

WiMAX can satisfy a variety of access needs. Potential applications include extending broadband capabilities to bring them closer to subscribers, filling gaps in cable, DSL and T1 services, WiFi, and cellular backhaul, providing last-100 meter access from fibre to the curb and giving service providers another cost-effective option for supporting broadband services. WiMAX can support very high bandwidth solutions where large spectrum deployments (i.e. >10 MHz) are desired using existing infrastructure keeping costs down while delivering the bandwidth needed to support a full range of high-value multimedia services.

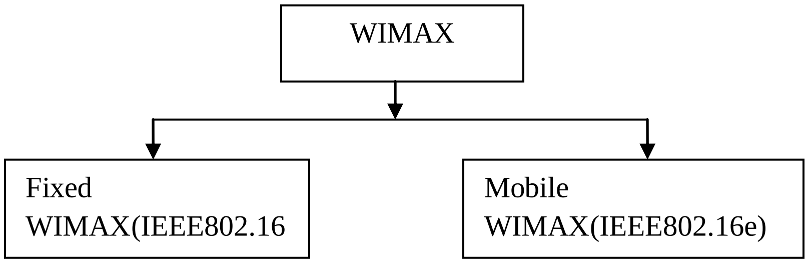
WiMAX can help service providers meet many of the challenges they face due to increasing customer demands without discarding their existing infrastructure investments because it has the ability to seamlessly interoperate across various network types.

WiMAX can provide wide area coverage and quality of service capabilities for applications ranging from real-time delay-sensitive voice-over-IP (VoIP) to real-time streaming video and non-real-time downloads, ensuring that subscribers obtain the performance they expect for all types of communications.

WiMAX, which is an IP-based wireless broadband technology, can be integrated into both wide-area third-generation (3G) mobile and wireless and wire line networks allowing it to become part of a seamless anytime, anywhere broadband access solution.

1.11.1Types of WiMax

WIMAX are of two types



1.11.2 Working Principle

The IEEE 802.16 standard was designed mainly to support point to multipoint topologies in which a base station distributes traffic to many subscriber stations that are mounted on roof tops.

The point to multipoint configuration uses a scheduling mechanism. Wimax doesn’t require stations to listen to one another because they incorporate a large area. This scheduling design suits for Wimax networks because subscriber stations might alleviate traffic from several computers & heavy steady traffic.

Subscriber stations can communicate directly by using a mesh mode supported by 802.16. The mesh mode can help to relax the line of sight requirement & ease the development costs for high frequency bands by allowing subscriber stations to relay the treaffic to one another.

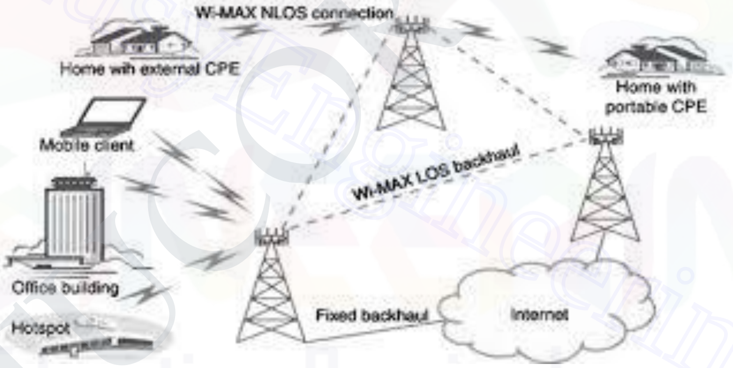


Fig. 1.30 Types of WIMAX Networks

1.11.4 WiMAX Performance

WiMAX is quite powerful, with a speed of up to 70 Mbps, which is a lot. Now what comes after determines the quality of the connection you receive. Some providers try to accommodate too many subscribers on one line (on their servers), which results in poor performances during peak times and for certain applications.

WiMAX has a range of around 50 km in a circle. Terrain, weather and buildings affect this range and this often results in many people not receiving signals good enough for a proper connection. Orientation is also an issue, and some people have to choose to place their WiMAX modems near windows and turned in certain specific directions for good reception.

A WiMAX connection is normally non-line-of-sight, which means that the transmitter and the receiver need not have a clear line between them. But a line-of-sight version exists, where performance and stability is much better, since this does away with problems associated with terrain and buildings.

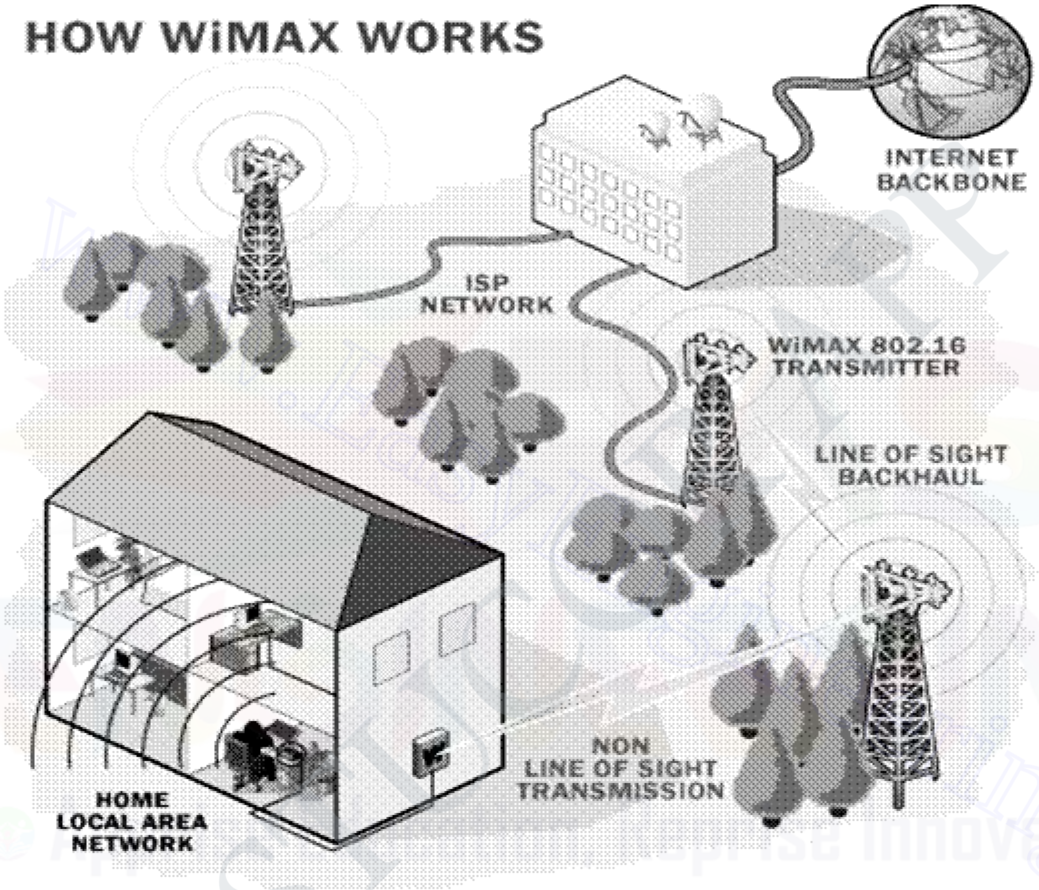


Fig. 1.31 WIMAX Performance

1.11.5 Services offered by Wimax

High Data Rate:

It can support peak downlink data rates of up to 63 Mbps per sector & peak uplink data rates of up to 28 Mbps per sector in a 10 MHz channel.

Quality of Service:

Wimax defines service flows in which can map to different service code points that allows end to end IP based QoS.

Security:

Wimax can work from 1.25 to 20 Mhz.

Mobility:

MobileWimax supports optimized handoff schemes with latency less than 50ms to ensure that real time applications.

1.11.6 WIMAX PHYSICAL LAYER

The 802.16 PHY layer supports TDD and Full and half duplex FDD operations. There are three air interfaces for the 2 – 11 GHz range.

1. Wireless LAN – SC
   It is a single carrier PHY layer for operating in the line of sight conditions with frequencies beyond 11 GHz.
2. Wireless MAN – Sca
   It uses single carrier modulation.
3. Wireless MAN – OFDM
   It uses a 256 carrier OFDM. Different stations are provided with multiple access by the air interface through time division multiple access.

1.11.7 WIMAX MEDIA ACCESS CONTROL

The MAC layer of 802.16 is designed to serve many distributed stations with high data rates. Subscriber stations are not required to listen to one another because this listening might be difficult to achieve in the WImx environment.

The 802.16 MAC protocol is connection oriented & performs link adaptation and ARQ functions to maintain target bit error rate while maximizing the data throughput.

It supports different transport technologies such as IPv4, IPv6, Ethernet& ATM.

Advantages :

1. Long range access.
2. Service can be rendered to variety of users.
3. It works on unlicensed frequency spectrum.
4. Supports data rate of speed 10Mbps at 10 kilometers with line of sight.
5. Symmetrical bandwidth over long range

Disadvantages:

1. It has spectral limitation
2. Poor bandwidth
3. Hogh cost for installation
4. Unreliable service
5. Line – of – sight is required for long distance communication.

UNIT II MOBILE NETWORK LAYER

2.1Introduction

Current versions of the Internet Protocol (IP) assume that the point at which a computer attaches to the Internet or a network is fixed and its IP address identifies the network to which it is attached. Datagrams are sent to a computer based on the location information contained in the IP address.

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

Mobile IP is an enhancement of the Internet Protocol (IP) that adds mechanisms for forwarding Internet traffic to mobile devices (known as mobile nodes) when they are connecting through other than their home network.

If a mobile computer, or mobile node, moves to a new network while keeping its IP address unchanged, its address does not reflect the new point of attachment. Consequently, existing routing protocols cannot route datagrams to the mobile node correctly.

Permanent IP address is one solution. Here emergency communication and quick reachability is possible via the permanent IP address.

Second solution is dynamically adapting the IP address with respect to current location. But the Domain Name System (DNS) has to update the new IP address to the logical name. For millions of nodes frequent updates is not possible.

Another solution is updating the routing table of the router. If the IP address of the receiver is changed, the router will route the data through the new port to which the receiver is now connected. But fast and frequent updating of the router is not possible.

A TCP connection is established using IP addresses of the source and receiver. The change in IP address breaks the existing TCP connection. Next one new TCP connection has to be established.

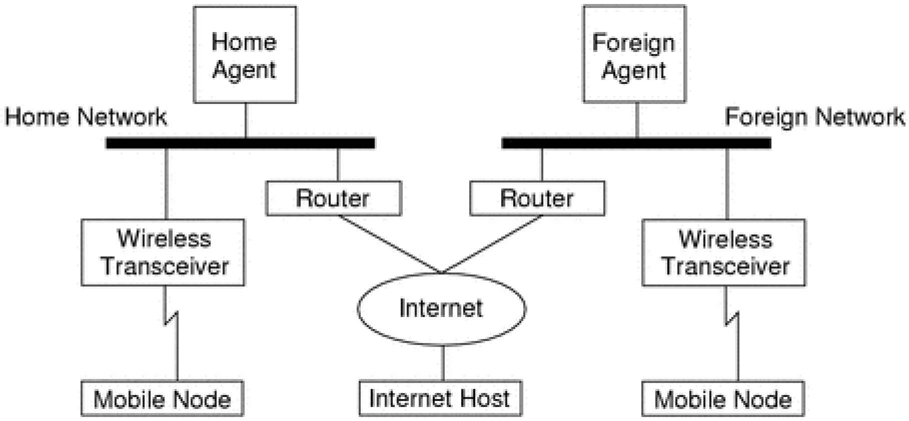


Fig.2.1 Mobile IP Topology

Using the previous illustration's Mobile IP topology, the following scenario shows how a datagram moves from one point to another within the Mobile IP framework.

1. The Internet host sends a datagram to the mobile node using the mobile node's home address (normal IP routing process).
2. If the mobile node is on its home network, the datagram is delivered through the normal IP process to the mobile node. Otherwise, the home agent picks up the datagram.
3. If the mobile node is on a foreign network, the home agent forwards the datagram to the foreign agent.
4. The foreign agent delivers the datagram to the mobile node.
5. Datagrams from the mobile node to the Internet host are sent using normal IP routing procedures. If the mobile node is on a foreign network, the packets are delivered to the foreign agent. The foreign agent forwards the datagram to the Internet host.

2.2 Requirements

The quick solutions are not working properly. The mobile IP is designed as a standard to enable the mobility in the internet.
Requirements of designing mobile IP:

1. Compatibility:

   Mobile IP has to be integrated with the existing operating system, must use the same routers, and network protocols. The mobile IP using device should be able to communicate the devices with normal IP.

2. Transparency:

   The problems with mobility are higher delay and lower bandwidth. The higher layer protocols should be mobility aware.

3. Scalability and efficiency:

   In wireless networks the important consideration is lower bandwidth. For mobility the flooding of the new messages should be restricted. Large numbers of devices are mobile devices. Hence the mobile IP should be scalable over a large number of devices.

4. Security:

   Mobile IP managing messages should be authenticated. The IP layer is responsible for identifying the correct IP address and preventing the fake of IP addresses.

2.3 Entities and terminology of mobile IP:

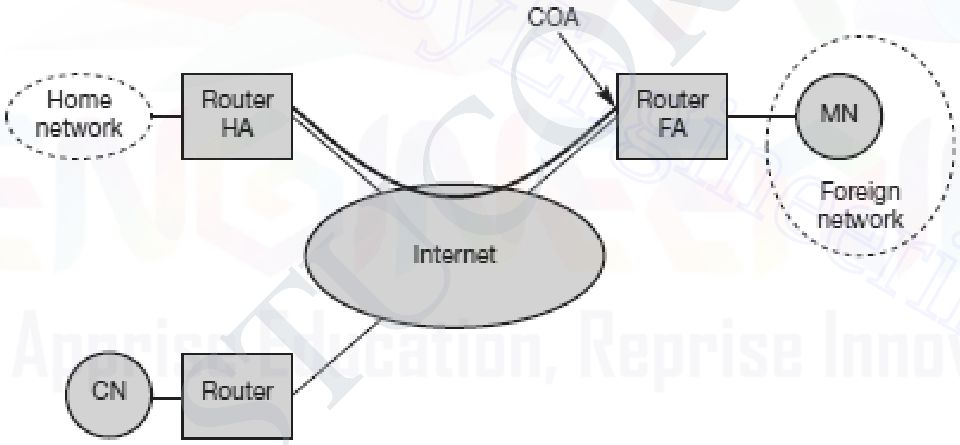


Fig.2.2 Mobile IP example network

1. Mobile Node: It is an end system that can be laptops with antennas, mobile phones or routers.

2. Correspondent node (CN): The CN is either fixed or mobile node acting as partner for communication.

3. Home Network: It is the network to which the mobile node is configured. Within this the mobile IP is not needed.

4. Foreign Network: It is the network at which the MN is currently present.

5. Foreign Agent(FA): It is a default router of the foreign network to the MN.

6. Care-of – address (COA): It defines the current location of the MN. The data is actually addressed to CAO not to the IP address of the MN.

i). foreign agent COA: It is the address of the FA which forwards the data to the MN. In this case, the care-of address is an IP address of the foreign agent. The foreign agent is the endpoint of the tunnel and, on receiving tunneled datagrams, de-encapsulates them and delivers the inner datagram to the mobile node. In this mode, many mobile nodes can share the same care-of address. This sharing reduces demands on the IPv4 address space and can also save bandwidth, because the forwarded packets, from the foreign agent to the mobile node, are not encapsulated. Saving bandwidth is important on wireless links.

ii). Co-located COA: It is the temporarily acquired additional IP address in the MN itself. A mobile node acquires a co-located care-of address as a local IP address through some external means, which the mobile node then associates with one of its own network interfaces. The address might be dynamically acquired as a temporary address by the mobile node, such as through DHCP. The address might also be owned by the mobile node as a long-term address for its use only while visiting some foreign network. When using a co-located care-of address, the mobile node serves as the endpoint of the tunnel and performs de-encapsulation of the datagrams tunneled to it.

7. Home Agent (HA): It is located in the home network. It maintains a location registry for MN. The tunnel of data transmission starts at here. The HA can be implemented on a router. This is best, because all the packets are passing through the router. The HA can also be implemented on an arbitrary node in the subnet.

2.4 IP packet delivery

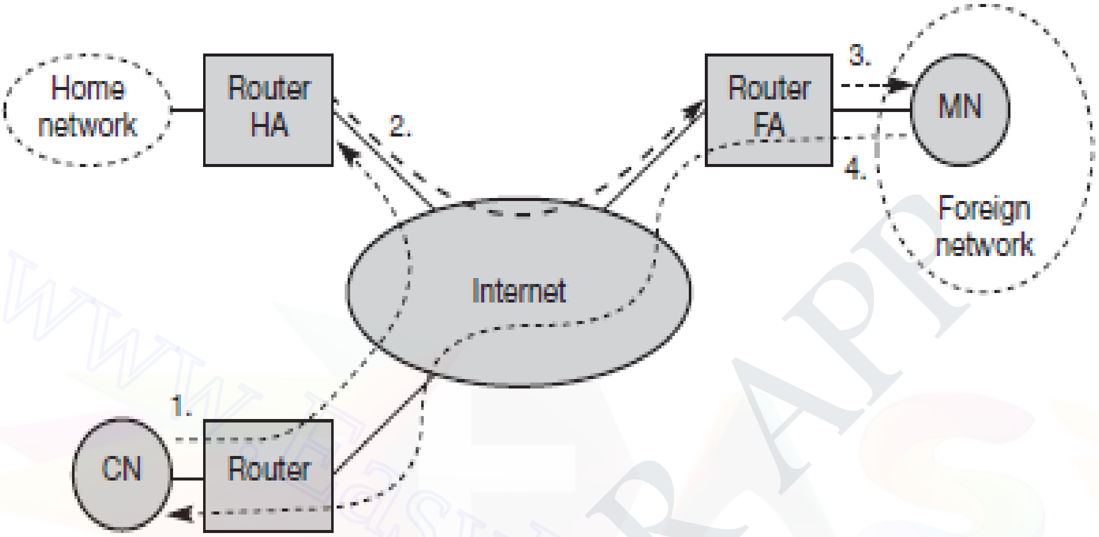


Fig.2. 3 Packet delivery to and from the mobile node

The CN wants to send data to the MN. The sends the data packet in which the source address is the address of the CN and the destination address is the IP address of the MN.

The data packet is forwarded to the HA of the Home network.

The HA knows that the MN is not in the home network. It is in the foreign network. The HA encapsulates the data packet with source address of its own and the destination address of the foreign agent and forwards the packet.

The Foreign agent receives, removes the additional header and forwards the data packet to the MN.

The transmission of data packet from the MN to the CN is very simple. If the CN is fixed one, the MN transmits the packet with its own address as source address and the address of the CN as destination address. It the CN is mobile one, the same procedure is to be followed.

2.5 Agent discovery

The mobile node is moving from one location to another location. During the movement it has to identify the foreign agent. The mobile IP describes two methods to identify the foreign agent.

1. Agent advertisement
2. Agent solicitation.

2.5.1 Agent advertisement

Mobile nodes use agent advertisements to determine their current point of attachment to the Internet or to an organization's network. An agent advertisement is an Internet Control Message Protocol (ICMP) router advertisement that has been extended to also carry a mobility agent advertisement extension.

A foreign agent can be too busy to serve additional mobile nodes. However, a foreign agent must continue to send agent advertisements. This way, mobile nodes that are already registered with it will know that they have not moved out of range of the foreign agent and that the foreign agent has not failed.

Also, a foreign agent that supports reverse tunnels must send it's advertisements with the reverse tunnel flag set on.

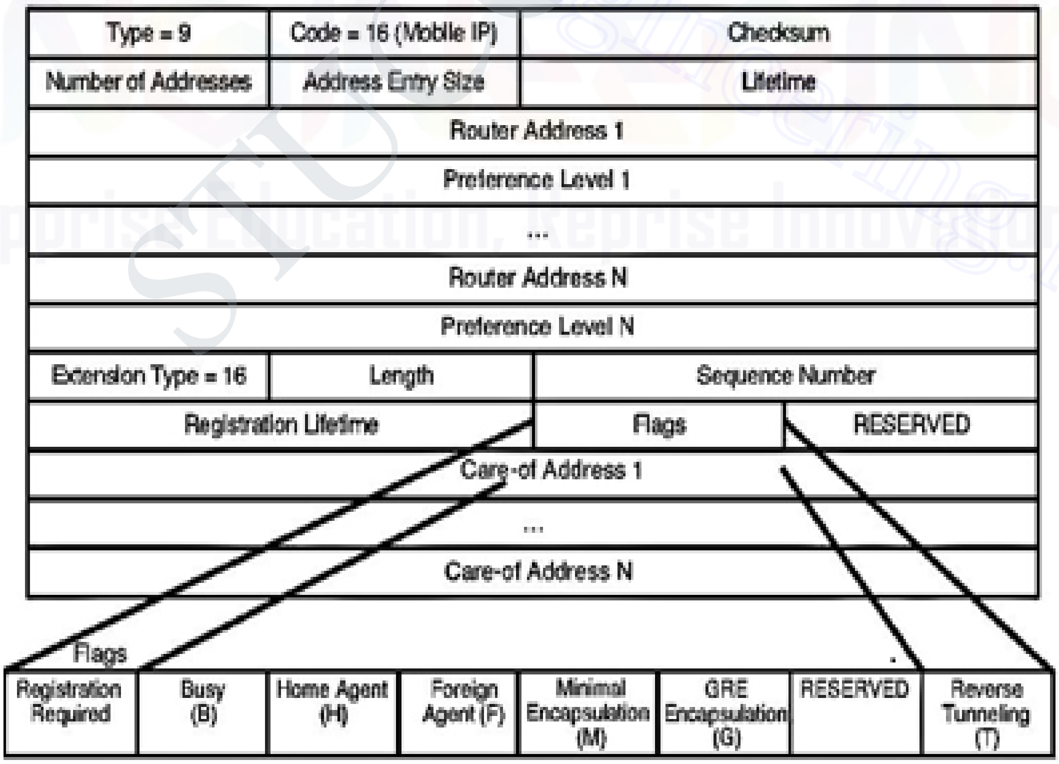


Fig.2.4 The agent advertisement packet format

Foreign agents and home agents are periodically advertising their presence using special agent advertisement messages. Routers also advertising their routing services periodically.

The agent advertisement packet format is shown in the figure.

The upper part represents the ICMP packet. The lower part represents the extension needed for the mobility.

For advertisement the TTL field of the IP packet is set to 1.

The IP destination address is either broadcast address 255.255.255.255, or the multicast address 224.0.0.1.

The fields of the agent advertisement packet are:

Type: It is set to 9.

Code: It is set to 0, when the agent routes traffic from both mobile and non-mobile nodes. It is set to 16, when the agent routes traffic from mobile nodes and not from non-mobile nodes.

Number of Addresses: It shows the number of addresses with this packet.

Lifetime: The length of the time over which this advertisement is valid.

Preference: It defines the preference level of each router. It is used to choose the most preferable one.

The fields of the extension of the packet for mobility:

Type: It is set to 16

Length: It defines the number of COAs provided with the message.

Sequence Number: It gives the total number of advertisements from the beginning.

Registration Lifetime: It specifies the maximum time a MN can request during registration.

Eight bits are used to specify the characteristics of the agent:

R: It specifies that the registration is required with this agent.

B: The agent is busy to accept the new registration.

H: The agent is the Home agent.

F: The agent is the foreign agent.

M: It specifies that the encapsulation method is the Minimal encapsulation method.

G: It specifies that the encapsulation method is the Generic routing encapsulation method.

r : Reserved

T: The FA supports the reverse Tunneling.

2.5.2 Agent Solicitation

When a MN enters a new network, It verifies the advertisement messages. If advertisement messages are not there, it will send agent solicitation message. In high dynamic wireless networks, the MN sends three solicitation messages, one per second. Before getting the agent address the MN will loss many data packets.

When the MN receives the address of the agent, it will use it for data transmission. If, it does not receive the answer, it should decrease the rate of solicitations. The solicitation messages will create collision.

After the advertisements and solicitations, the MN receives the COA for an FA. By using it, the MN can make communication.

2.6 Tunneling and encapsulation

A tunnel is a virtual path between home agent and current COA. Tunneling is a process of sending data packet through the tunnel.

Encapsulation: It is a process of putting one data packet within another packet. The data packet consists of original data and the header. The entire packet is treated as data and one new header is added. The Diagram shows the operation of the encapsulation process.

Decapsulation: It is a process of extracting the data packet from another packet

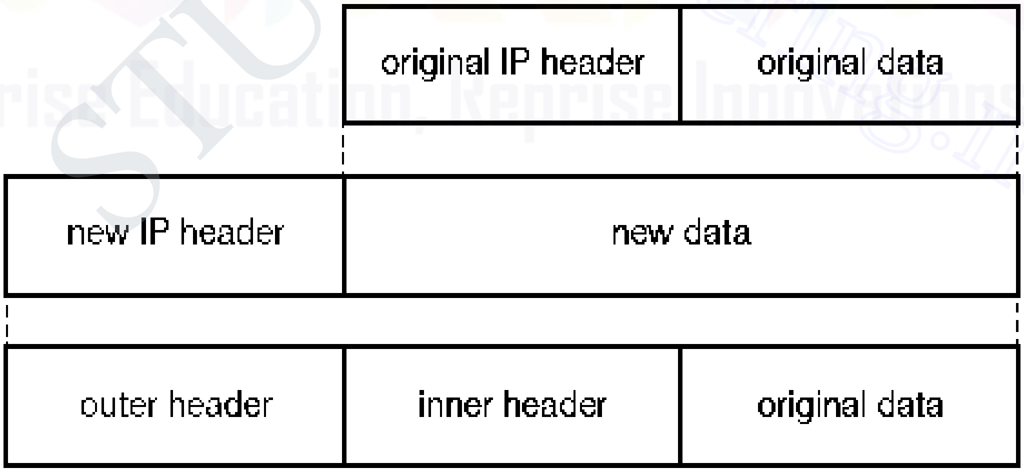


Fig.2.5 IP encapsulation

It consists of two headers . One is inner header which consists of the address of MN and CN.

The second header is the added header which consists of the address of HA and COA.

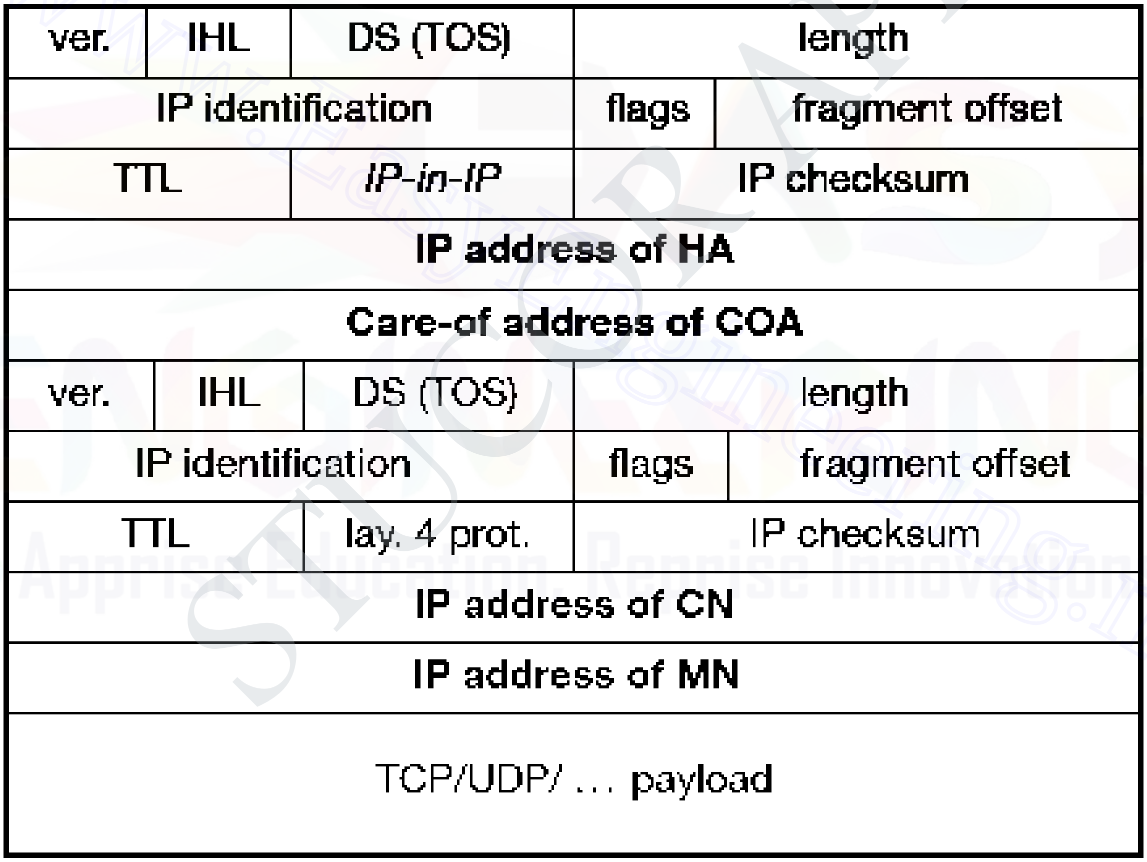
Three categories of encapsulation process

1. IP-in-IP encapsulation
2. Minimal encapsulation
3. Generic routing encapsulation

2.6.1 IP-in-IP encapsulation

Here one IP packet is kept inside of another IP packet.

The figure shows that the data packet contains two IP headers.



<table>
<tr><td>ver.</td><td>IHL</td><td>DS (TOS)</td><td colspan="2">length</td></tr>
<tr><td colspan="3">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="2">TTL</td><td><i>IP-in-IP</i></td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="5">IP address of HA</td></tr>
<tr><td colspan="5">Care-of address of COA</td></tr>
<tr><td>ver.</td><td>IHL</td><td>DS (TOS)</td><td colspan="2">length</td></tr>
<tr><td colspan="3">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="2">TTL</td><td>lay. 4 prot.</td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="5">IP address of CN</td></tr>
<tr><td colspan="5">IP address of MN</td></tr>
<tr><td colspan="5">TCP/UDP/ ... payload</td></tr>
</table>

Fig.2.6. IP-in-IP encapsulation

The fields of the header are

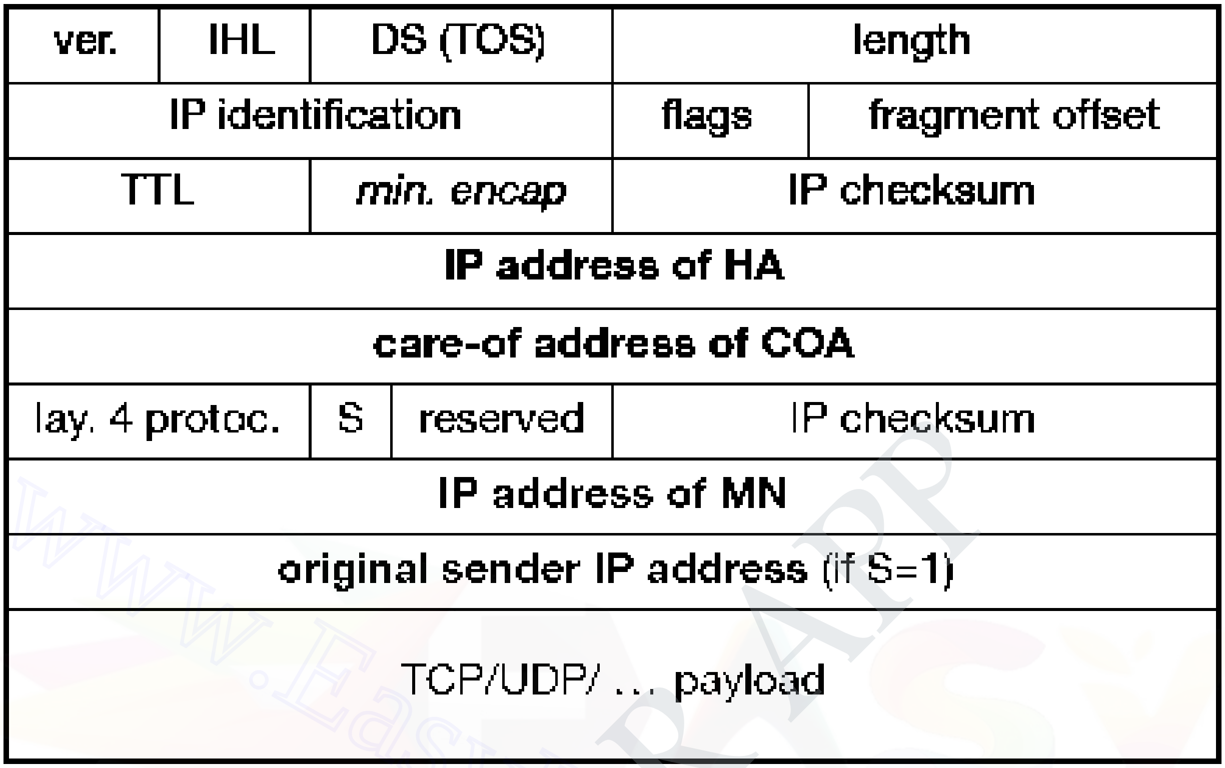
1. Ver: It specifies the current version of IP packet.
2. IHL (Internet Header Length): It defines the length of the outer header.

3. DS(TOS): It specifies the type of service.
4. Length: It covers the length of the entire packet.
5. TTL: Time To Live It specifies the time over which the data packet can travel through the network. It should be high.
6. Type of Protocol: It specifies the type of the protocol which is used in the packet.
7. IP checksum: The checksum is calculated and added in the packet. At receiver the checksum is calculated and compared with the value in the data packet. It is used to identify the error.
8. Source address: In outer header it specifies the address of the Home Agent. In inner header it specifies the address of the CN
9. Destination address: In outer header it specifies the address of the COA. In inner header it specifies the address of the MN.

If any options are there ,those are added at the end of the outer header. If options are not there, the inner header starts after the outer header with the same fields. The TTL value is decremented by 1. That the whole tunnel is considered as on one hop.

2.6.2 Minimal encapsulation

Some fields are redundant in IP-in-IP encapsulation method. Redundant fields are removed from the inner header. If the S bit is set, the original sender address of the CN is included as omitting the source is quite often not an option. No field for fragmentation offset is left in the inner header and minimal encapsulation does not work with already fragmented packets.



ver.	IHL	DS (TOS)		length	
IP identification				flags	fragment offset
TTL		*min. encap*		IP checksum	
IP address of HA					
care-of address of COA					
lay. 4 protoc.		S	reserved	IP checksum	
IP address of MN					
original sender IP address (if S=1)					
TCP/UDP/ … payload					

Fig.2.7 Minimal encapsulation

2.6..3 Generic routing encapsulation

This encapsulation method is applicable for IP and other network layer protocols. It encapsulates the packet of one protocol into the packet of another protocol.

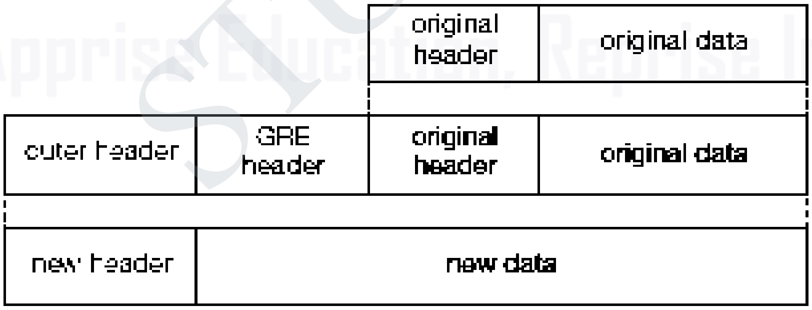
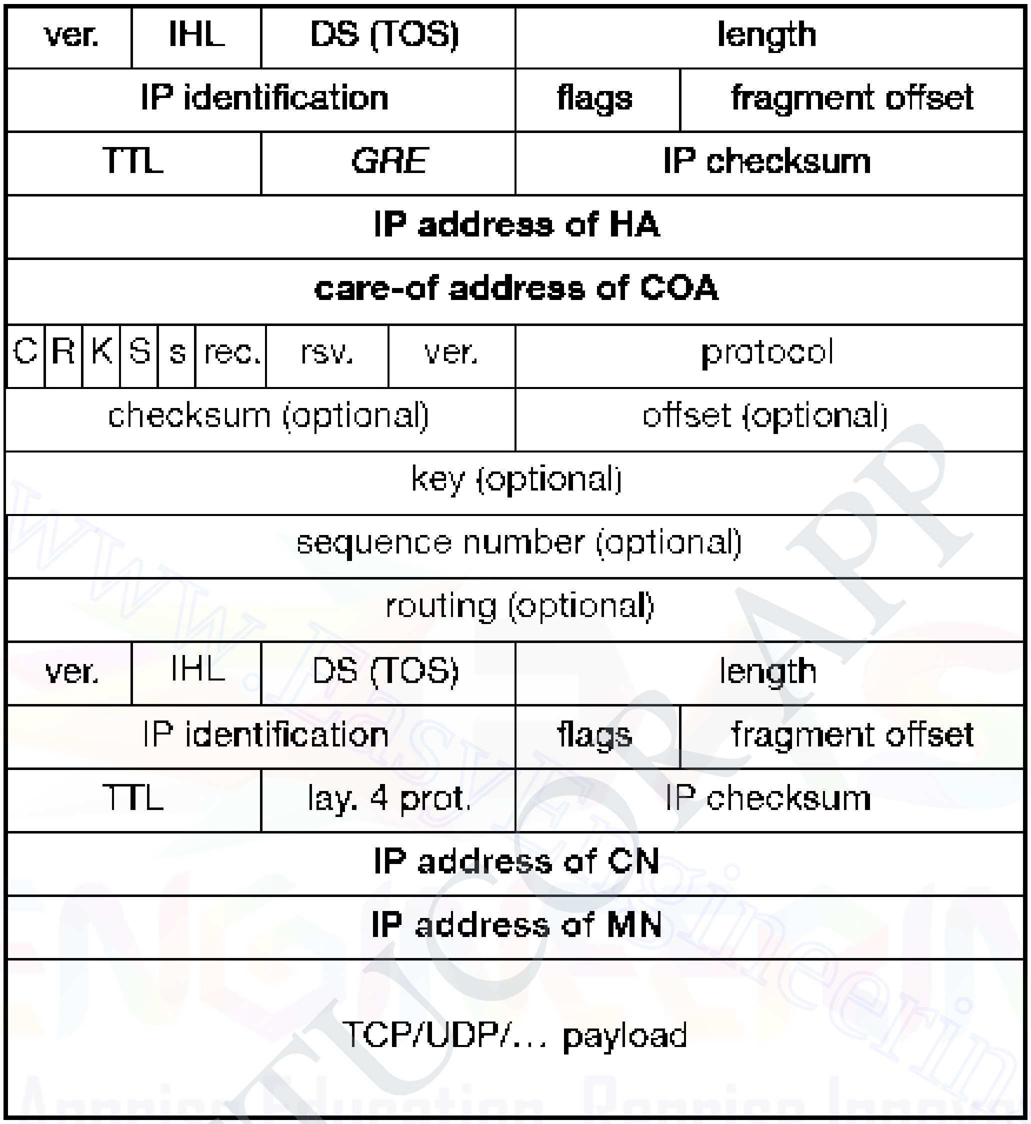


Fig.2.8 Generic routing encapsulation

Here one GRE header is added between inner and outer header.



<table>
<tr><td colspan="4">ver.</td><td colspan="3">IHL</td><td colspan="2">DS (TOS)</td><td colspan="2">length</td></tr>
<tr><td colspan="9">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="7">TTL</td><td colspan="2">GRE</td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="11">IP address of HA</td></tr>
<tr><td colspan="11">care-of address of COA</td></tr>
<tr><td>C</td><td>R</td><td>K</td><td colspan="2">S</td><td>s</td><td>rec.</td><td>rsv.</td><td>ver.</td><td colspan="2">protocol</td></tr>
<tr><td colspan="9">checksum (optional)</td><td colspan="2">offset (optional)</td></tr>
<tr><td colspan="11">key (optional)</td></tr>
<tr><td colspan="11">sequence number (optional)</td></tr>
<tr><td colspan="11">routing (optional)</td></tr>
<tr><td colspan="4">ver.</td><td colspan="3">IHL</td><td colspan="2">DS (TOS)</td><td colspan="2">length</td></tr>
<tr><td colspan="9">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="7">TTL</td><td colspan="2">lay. 4 prot.</td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="11">IP address of CN</td></tr>
<tr><td colspan="11">IP address of MN</td></tr>
<tr><td colspan="11">TCP/UDP/... payload</td></tr>
</table>

Fig.2.8 Protocol fields for GRE

The GRE header is having some flags which are indicating if certain fields are present or not. The flags are

C: If C is set, the checksum field contains a valid IP checksum of the GRE header and the payload.

R: If R is set, the routing fields are present and contain valid information.

K: If K is set, a key field is present and is used for authentication. It does not specify authentication algorithm.

S: If S is set, the sequence number field is present.

s : If s is set, strict source routing is used.

Rec: Recursion Control field is used to represent the count of allowed recursive encapsulations. If this field is zero, additional encapsulation is not allowed. If this field is not zero, additional encapsulation is allowed and this is decremented by one.

Reserved : This field must be zero and are ignored on reception.

Version: It is zero for the GRE version.

Protocol: It contains the protocol of the following packet. For Ethernet the field values are 0 x 6558 and for mobile IP tunnel, the fields contains 0 x 800.

2.7 IPv6

Internet Protocol version 6 (IPv6) is the latest revision of the Internet Protocol (IP) and the first version of the protocol to be widely deployed. IPv6 was developed by the Internet Engineering Task Force (IETF) to deal with the long-anticipated problem of IPv4 address exhaustion. This tutorial will help you in understanding IPv6 and its associated terminologies along with appropriate references and examples.

2.7.1 IPv6 - Features

The successor of IPv4 is not designed to be backward compatible. Trying to keep the basic functionalities of IP addressing, IPv6 is redesigned entirely. It offers the following features:

Larger Address Space

In contrast to IPv4, IPv6 uses 4 times more bits to address a device on the Internet. This much of extra bits can provide approximately 3.4×1038 different combinations of addresses.

Simplified Header

IPv6's header has been simplified by moving all unnecessary information and options (which are present in IPv4 header) to the end of the IPv6 header.

End-to-end Connectivity

After IPv6 is fully implemented, every host can directly reach other hosts on the Internet, with some limitations involved like Firewall, organization policies, etc.
Auto-configuration
IPv6 supports both stateful and stateless auto configuration mode of its host devices.

Faster Forwarding/Routing

Simplified header puts all unnecessary information at the end of the header so it makes routing decision as quickly as possible.

IPSec

Initially it was decided that IPv6 must have IPSec security, making it more secure than IPv4.

No Broadcast

Though Ethernet/Token Ring are considered as broadcast network because they support Broadcasting, IPv6 does not have any broadcast support any more. It uses multicast to communicate with multiple hosts.

Anycast Support

IPv6 has introduced Anycast mode of packet routing. In this mode, multiple interfaces over the Internet are assigned same Anycast IP address. Routers, while routing, send the packet to the nearest destination.

Mobility

IPv6 was designed keeping mobility in mind. This feature enables hosts (such as mobile phone) to roam around in different geographical area and remain connected with the same IP address. The mobility feature of IPv6 takes advantage of auto IP configuration and Extension headers.

Enhanced Priority Support

IPv4 used 6 bits DSCP (Differential Service Code Point) and 2 bits ECN (Explicit Congestion Notification) to provide Quality of Service but it could only be used if the end-to-end devices support it, that is, the source and destination device and underlying network must support it.
In IPv6, Traffic class and Flow label are used to tell the underlying routers how to efficiently process the packet and route it.

Smooth Transition

Large IP address scheme in IPv6 enables to allocate devices with globally unique IP addresses. This mechanism saves IP addresses and NAT is not required. So devices can send/receive data among each other, for example, VoIP and/or any streaming media can be used much efficiently.

Other fact is, the header is less loaded, so routers can take forwarding decisions and forward them as quickly as they arrive.

Extensibility

One of the major advantages of IPv6 header is that it is extensible to add more information in the option part. IPv4 provides only 40-bytes for options, whereas options in IPv6 can be as much as the size of IPv6 packet itself.

2.7.2Transition From IPv4 to IPv6

Complete transition from IPv4 to IPv6 might not be possible because IPv6 is not backward compatible. This results in a situation where either a site is on IPv6 or it is not. It is unlike implementation of other new technologies where the newer one is backward compatible so the older system can still work with the newer version without any additional changes.

To overcome this short-coming, we have a few technologies that can be used to ensure slow and smooth transition from IPv4 to IPv6.

Dual Stack Routers

A router can be installed with both IPv4 and IPv6 addresses configured on its interfaces pointing to the network of relevant IP scheme.

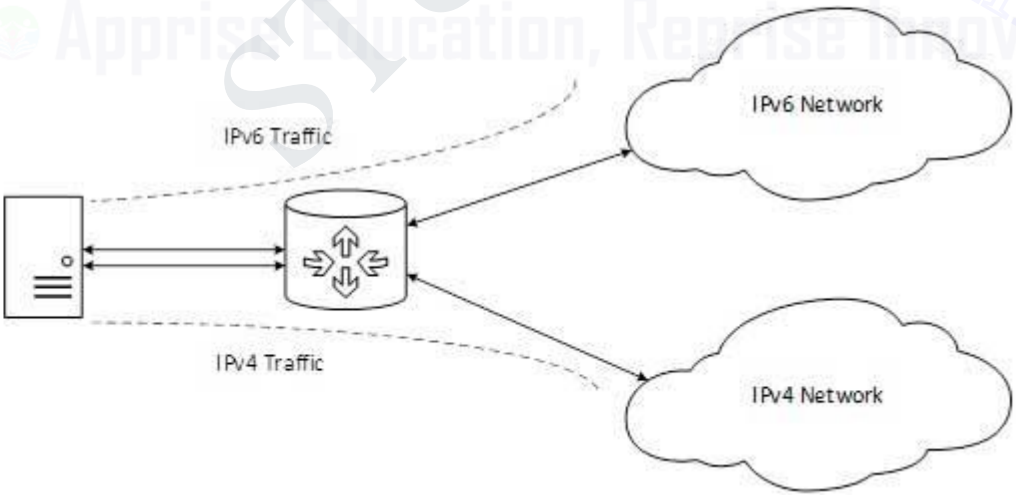


Fig.2.9 Dual Stack Router

In the above diagram, a server having IPv4 as well as IPv6 address configured for it can now speak with all the hosts on both the IPv4 as well as the IPv6 networks with the help of a Dual Stack Router. The Dual Stack Router, can communicate with both the networks. It provides a medium for the hosts to access a server without changing their respective IP versions.

Tunneling

In a scenario where different IP versions exist on intermediate path or transit networks, tunneling provides a better solution where user's data can pass through a non-supported IP version.

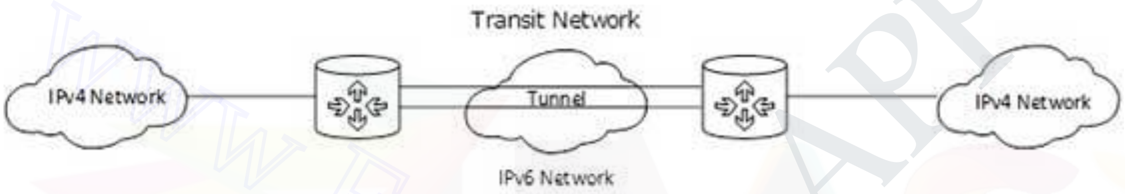


Fig.2.10 Tunneling

The above diagram depicts how two remote IPv4 networks can communicate via a Tunnel, where the transit network was on IPv6. Vice versa is also possible where the transit network is on IPv6 and the remote sites that intend to communicate are on IPv4.

NAT Protocol Translation

This is another important method of transition to IPv6 by means of a NAT-PT (Network Address Translation – Protocol Translation) enabled device. With the help of a NAT-PT device, actual can take place happens between IPv4 and IPv6 packets and vice versa. See the diagram below:

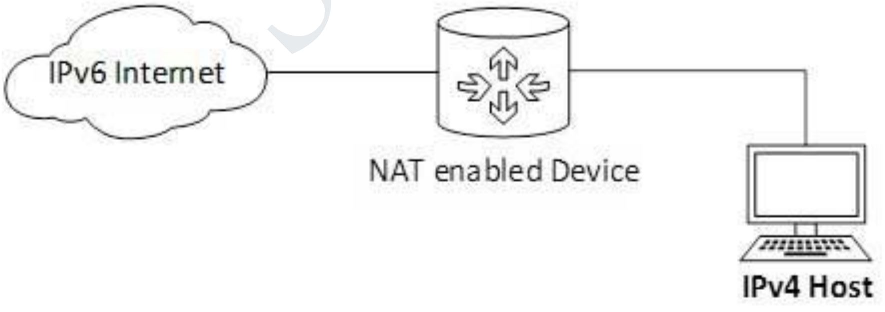


Fig.2.11 NAT - Protocol Translation

A host with IPv4 address sends a request to an IPv6 enabled server on Internet that does not understand IPv4 address. In this scenario, the NAT-PT device can help them communicate. When the IPv4 host sends a request packet to the IPv6 server, the

NAT-PT device/router strips down the IPv4 packet, removes IPv4 header, and adds IPv6 header and passes it through the Internet. When a response from the IPv6 server comes for the IPv4 host, the router does vice versa.

2.7.3Address Structure

An IPv6 address is made of 128 bits divided into eight 16-bits blocks. Each block is then converted into 4-digit Hexadecimal numbers separated by colon symbols.

For example, given below is a 128 bit IPv6 address represented in binary format and divided into eight 16-bits blocks:
0010000000000001 0000000000000000 0011001000111000 1101111111100001 0000000001100011 0000000000000000 0000000000000000 1111111011111011

Each block is then converted into Hexadecimal and separated by ':' symbol:
2001:0000:3238:DFE1:0063:0000:0000:FEFB

Even after converting into Hexadecimal format, IPv6 address remains long. IPv6 provides some rules to shorten the address. The rules are as follows:

Rule.1: Discard leading Zero(es):

In Block 5, 0063, the leading two 0s can be omitted, such as (5th block):
2001:0000:3238:DFE1:63:0000:0000:FEFB

2.7.4 IPv6 - Headers

IPv6 headers have one Fixed Header and zero or more Optional (Extension) Headers. All the necessary information that is essential for a router is kept in the Fixed Header. The Extension Header contains optional information that helps routers to understand how to handle a packet/flow.

Fixed Header

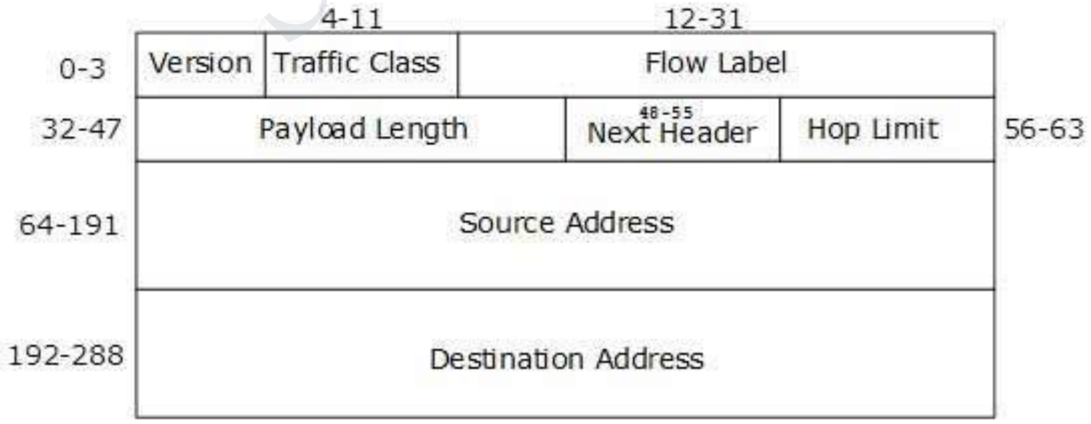


Fig.2.12 IPv6 Fixed Header

IPv6 fixed header is 40 bytes long and contains the following information.

S.N. Field & Description

1 Version (4-bits): It represents the version of Internet Protocol, i.e. 0110.

2 Traffic Class (8-bits): These 8 bits are divided into two parts. The most significant 6 bits are used for Type of Service to let the Router Known what services should be provided to this packet. The least significant 2 bits are used for Explicit Congestion Notification (ECN).

3 Flow Label (20-bits): This label is used to maintain the sequential flow of the packets belonging to a communication. The source labels the sequence to help the router identify that a particular packet belongs to a specific flow of information. This field helps avoid re-ordering of data packets. It is designed for streaming/real-time media.

4 Payload Length (16-bits): This field is used to tell the routers how much information a particular packet contains in its payload. Payload is composed of Extension Headers and Upper Layer data. With 16 bits, up to 65535 bytes can be indicated; but if the Extension Headers contain Hop-by-Hop Extension Header, then the payload may exceed 65535 bytes and this field is set to 0.

5 Next Header (8-bits): This field is used to indicate either the type of Extension Header, or if the Extension Header is not present then it indicates the Upper Layer PDU. The values for the type of Upper Layer PDU are same as IPv4's.

6 Hop Limit (8-bits): This field is used to stop packet to loop in the network infinitely. This is same as TTL in IPv4. The value of Hop Limit field is decremented by 1 as it passes a link (router/hop). When the field reaches 0 the packet is discarded.

7 Source Address (128-bits): This field indicates the address of originator of the packet.

8 Destination Address (128-bits): This field provides the address of intended recipient of the packet.

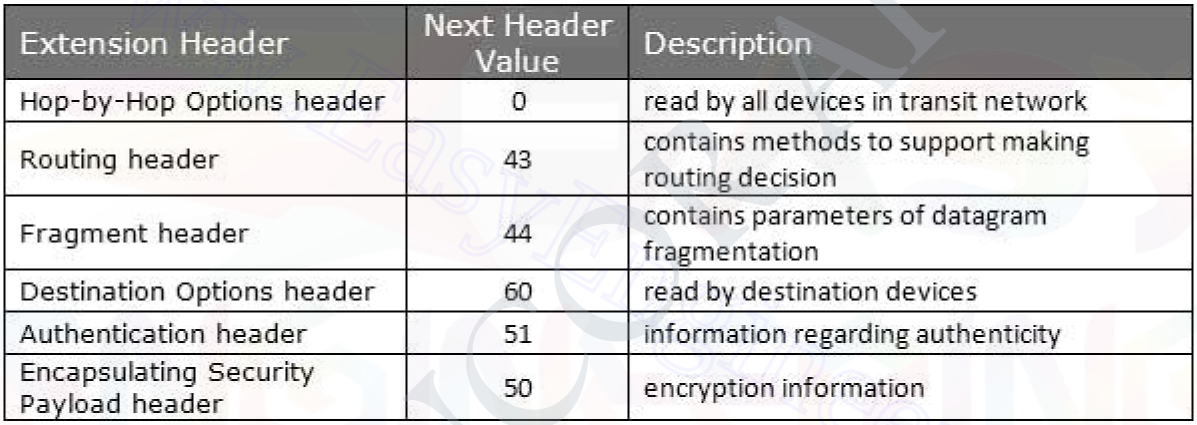
Extension Headers

Rarely used information is put between the Fixed Header and the Upper layer header in the form of Extension Headers. Each Extension Header is identified by a distinct value.

When Extension Headers are used, IPv6 Fixed Header's Next Header field points to the first Extension Header. If there is one more Extension Header, then the first Extension Header's 'Next-Header' field points to the second one, and so on. The last Extension Header's 'Next-Header' field points to the Upper Layer Header. Thus, all the headers points to the next one in a linked list manner.

If the Next Header field contains the value 59, it indicates that there are no headers after this header, not even Upper Layer Header.

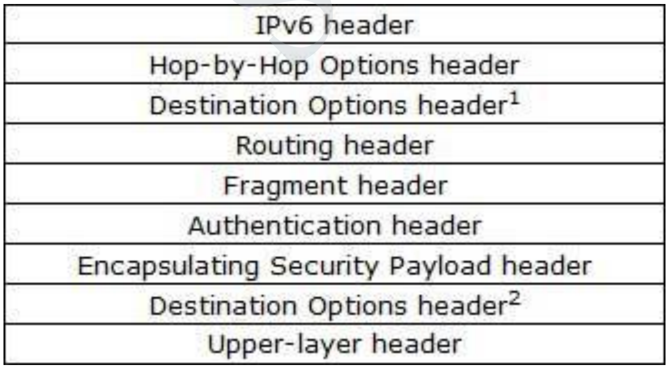
The following Extension Headers must be supported as per RFC 2460:



Extension Header	Next Header Value	Description
Hop-by-Hop Options header	0	read by all devices in transit network
Routing header	43	contains methods to support making routing decision
Fragment header	44	contains parameters of datagram fragmentation
Destination Options header	60	read by destination devices
Authentication header	51	information regarding authenticity
Encapsulating Security Payload header	50	encryption information

Fig.2.

The sequence of Extension Headers should be:



IPv6 header
Hop-by-Hop Options header
Destination Options header[1]
Routing header
Fragment header
Authentication header
Encapsulating Security Payload header
Destination Options header[2]
Upper-layer header

Fig.2.

These headers:

1. should be processed by First and subsequent destinations.

2. should be processed by Final Destination.

Extension Headers are arranged one after another in a linked list manner, as depicted in the following diagram:

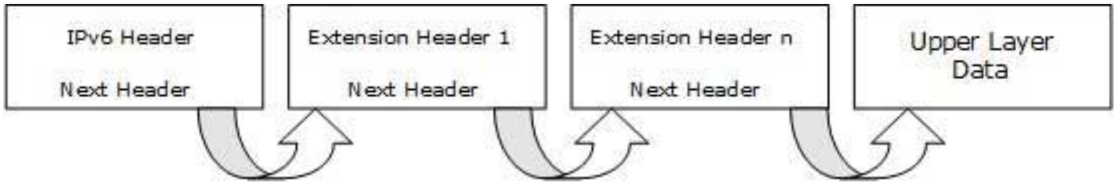


Fig.2.

2.8 SESSION INITIATION PROTOCOL

2.8.1 Introduction

Session Initiation Protocol (SIP) is one of the most common protocols used in VoIP technology. It is an application layer protocol that works in conjunction with other application layer protocols to control multimedia communication sessions over the Internet.

SIP is a signaling protocol used to create, modify, and terminate a multimedia session over the Internet Protocol. A session is nothing but a simple call between two endpoints. An endpoint can be a smartphone, a laptop, or any device that can receive and send multimedia content over the Internet.

SIP takes the help of SDP (Session Description Protocol) which describes a session and RTP (Real Time Transport Protocol) used for delivering voice and video over IP network.

SIP can be used for two-party (unicast) or multiparty (multicast) sessions.

Other SIP applications include file transfer, instant messaging, video conferencing, online games, and steaming multimedia distribution.

Basically SIP is an application layer protocol. It is a simple network signaling protocol for creating and terminating sessions with one or more participants. The SIP protocol is designed to be independent of the underlying transport protocol, so SIP applications can run on TCP, UDP, or other lower-layer networking protocols.

Typically, the SIP protocol is used for internet telephony and multimedia distribution between two or more endpoints. For example, one person can initiate a telephone call to another person using SIP, or someone may create a conference call with many participants.

The SIP protocol was designed to be very simple, with a limited set of commands. It is also text-based, so anyone can read a SIP message passed between the endpoints in a SIP session.

2.8.2 SIP - Network Elements

There are some entities that help SIP in creating its network. In SIP, every network element is identified by a **SIP URI** (Uniform Resource Identifier) which is like an address. Following are the network elements −

- User Agent
- Proxy Server
- Registrar Server
- Redirect Server
- Location Server

User Agent

It is the endpoint and one of the most important network elements of a SIP network. An endpoint can initiate, modify, or terminate a session. User agents are the most intelligent device or network element of a SIP network. It could be a softphone, a mobile, or a laptop.

User agents are logically divided into two parts

- **User Agent Client (UAC)** − The entity that sends a request and receives a response.
- **User Agent Server (UAS)** − The entity that receives a request and sends a response.

SIP is based on client-server architecture where the caller's phone acts as a client which initiates a call and the callee's phone acts as a server which responds the call.

Proxy Server

It is the network element that takes a request from a user agent and forwards it to another user.

- Basically the role of a proxy server is much like a router.
- It has some intelligence to understand a SIP request and send it ahead with the help of URI.
- A proxy server sits in between two user agents.

- There can be a maximum of 70 proxy servers in between a source and a destination.

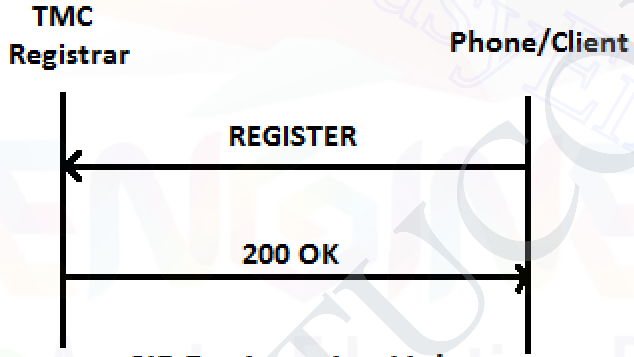
There are two types of proxy servers

- **Stateless Proxy Server** − It simply forwards the message received. This type of server does not store any information of a call or a transaction.
- **Stateful Proxy Server** − This type of proxy server keeps track of every request and response received and can use it in future if required. It can retransmit the request, if there is no response from the other side in time.

Registrar Server

The registrar server accepts registration requests from user agents. It helps users to authenticate themselves within the network. It stores the URI and the location of users in a database to help other SIP servers within the same domain.

Take a look at the following example that shows the process of a SIP Registration.



SIP Registration Link

Here the caller wants to register with the TMC domain. So it sends a REGISTER request to the TMC's Registrar server and the server returns a 200 OK response as it authorized the client.

Redirect Server

The redirect server receives requests and looks up the intended recipient of the request in the location database created by the registrar.

The redirect server uses the database for getting location information and responds with 3xx (Redirect response) to the user. We will discuss response codes later in this tutorial.

Location Server

The location server provides information about a caller's possible locations to the redirect and proxy servers.

Only a proxy server or a redirect server can contact a location server.

The following figure depicts the roles played by each of the network elements in establishing a session.

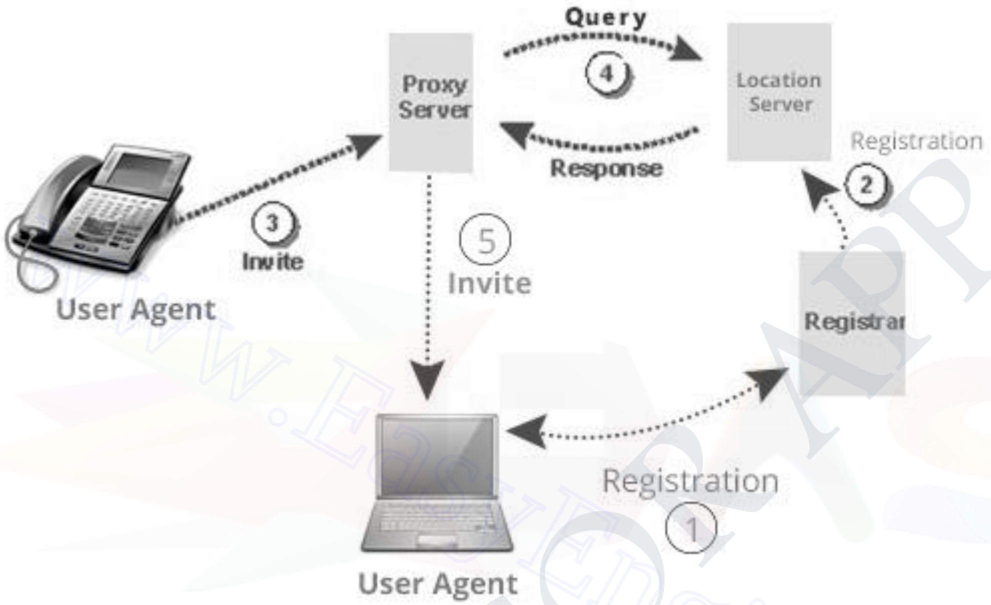


Fig.2.16 Call flow

2.8.3 SIP – System Architecture

SIP is structured as a layered protocol, which means its behavior is described in terms of a set of fairly independent processing stages with only a loose coupling between each stage.

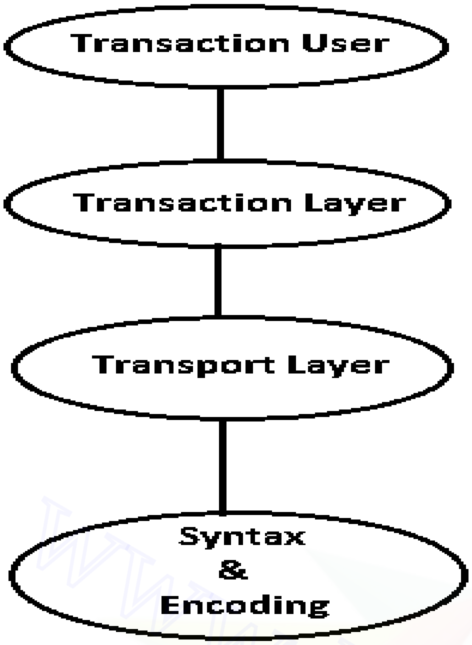


Fig.2.17 SIP – System Architecture

- The lowest layer of SIP is its syntax and encoding. Its encoding is specified using an augmented Backus-Naur Form grammar (BNF).
- At the second level is the transport layer. It defines how a Client sends requests and receives responses and how a Server receives requests and sends responses over the network. All SIP elements contain a transport layer.
- Next comes the transaction layer. A transaction is a request sent by a Client transaction (using the transport layer) to a Server transaction, along with all responses to that request sent from the server transaction back to the client. Any task that a user agent client (UAC) accomplishes takes place using a series of transactions. Stateless proxies do not contain a transaction layer.
- The layer above the transaction layer is called the transaction user. Each of the SIP entities, except the Stateless proxies, is a transaction user.

2.8.4 SIP - Basic Call Flow

The following image shows the basic call flow of a SIP session.

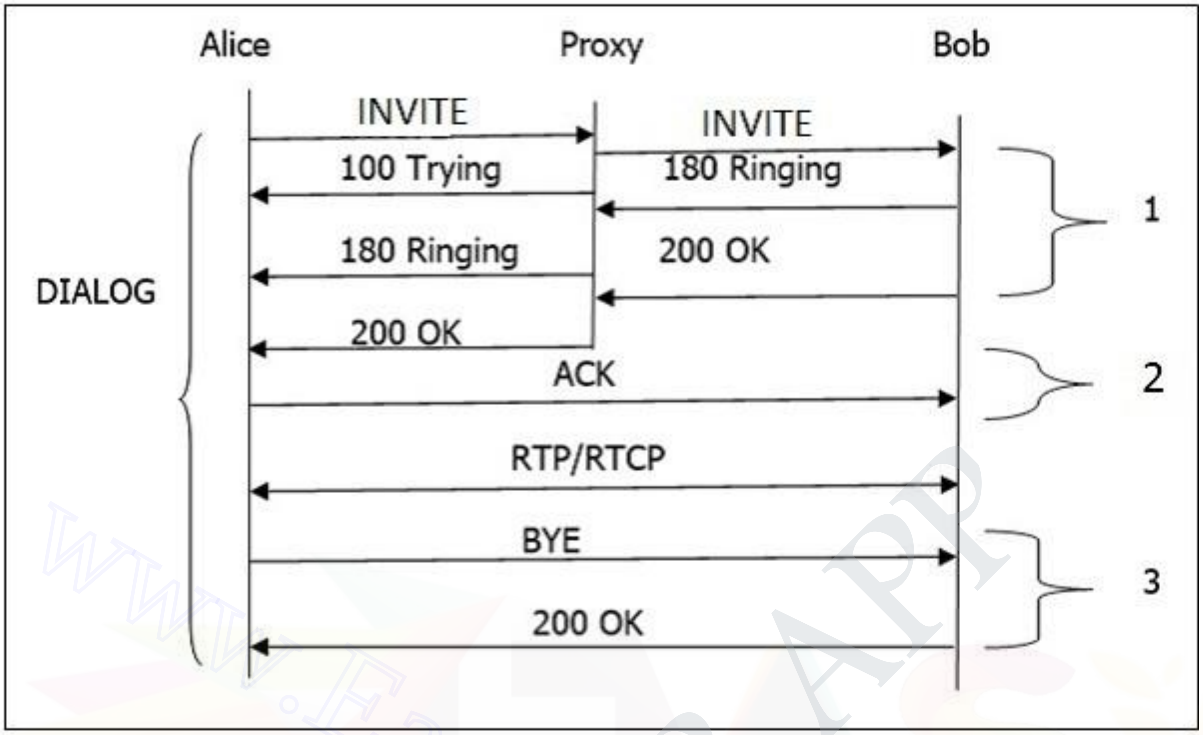


Fig.2.18

Given below is a step-by-step explanation of the above call flow

- An INVITE request that is sent to a proxy server is responsible for initiating a session.
- The proxy server sends a **100 Trying** response immediately to the caller (Alice) to stop the re-transmissions of the INVITE request.
- The proxy server searches the address of Bob in the location server. After getting the address, it forwards the INVITE request further.
- Thereafter, **180 Ringing** (Provisional responses) generated by Bob is returned back to Alice.
- A **200 OK** response is generated soon after Bob picks the phone up.
- Bob receives an **ACK** from the Alice, once it gets **200 OK**.
- At the same time, the session gets established and RTP packets (conversations) start flowing from both ends.
- After the conversation, any participant (Alice or Bob) can send a **BYE** request to terminate the session.
- **BYE** reaches directly from Alice to Bob bypassing the proxy server.
- Finally, Bob sends a **200 OK** response to confirm the BYE and the session is terminated.
- In the above basic call flow, three **transactions** are (marked as 1, 2, 3) available.

The complete call (from INVITE to 200 OK) is known as a **Dialog**.

SIP Trapezoid

How does a proxy help to connect one user with another? Let us find out with the help of the following diagram.

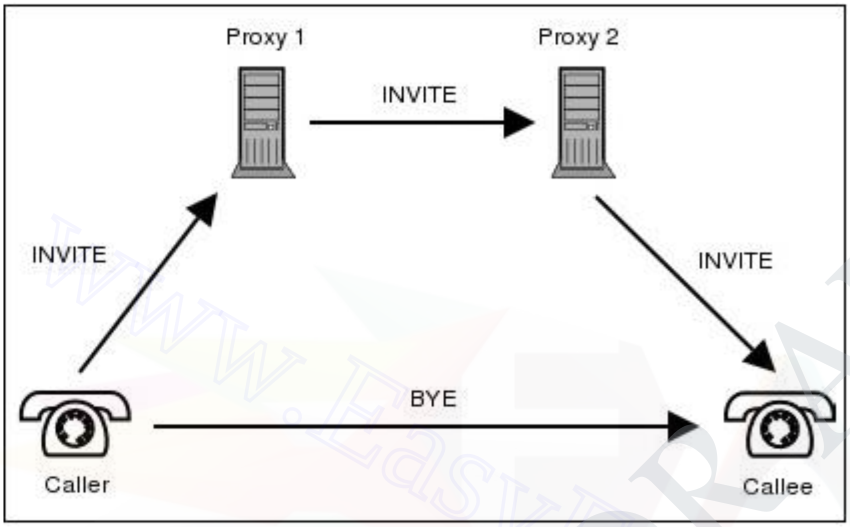


Fig.2.19SIP trapezoid

The topology shown in the diagram is known as a SIP trapezoid. The process takes place as follows −

- When a caller initiates a call, an INVITE message is sent to the proxy server. Upon receiving the INVITE, the proxy server attempts to resolve the address of the callee with the help of the DNS server.
- After getting the next route, caller’s proxy server (Proxy 1, also known as outbound proxy server) forwards the INVITE request to the callee’s proxy server which acts as an inbound proxy server (Proxy 2) for the callee.
- The inbound proxy server contacts the location server to get information about the callee’s address where the user registered.
- After getting information from the location server, it forwards the call to its destination.
- Once the user agents get to know their address, they can bypass the call, i.e., conversations pass directly.

SIP - Messaging

SIP messages are of two types − **requests** and **responses**.

- The opening line of a request contains a method that defines the request, and a Request-URI that defines where the request is to be sent.
- Similarly, the opening line of a response contains a response code.

Request Methods

SIP requests are the codes used to establish a communication. To complement them, there are **SIP responses** that generally indicate whether a request succeeded or failed.

These SIP requests which are known as METHODS make SIP message workable.

- METHODS can be regarded as SIP requests, since they request a specific action to be taken by another user agent or server.
- METHODS are distinguished into two types −
  - Core Methods
  - Extension Methods

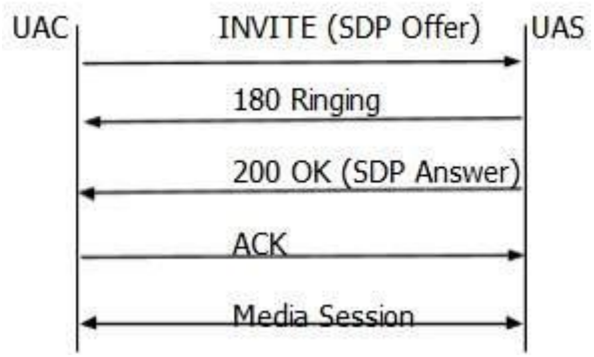
Core Methods

There are six core methods as discussed below.

INVITE

INVITE is used to initiate a session with a user agent. In other words, an INVITE method is used to establish a media session between the user agents.

- INVITE can contain the media information of the caller in the message body.
- A session is considered established if an INVITE has received a success response(2xx) or an ACK has been sent.



- A successful INVITE request establishes a dialog between the two user agents which continues until a BYE is sent to terminate the session.
- An INVITE sent within an established dialog is known as a re-INVITE.
- Re-INVITE is used to change the session characteristics or refresh the state of a dialog.

BYE

BYE is the method used to terminate an established session. This is a SIP request that can be sent by either the caller or the callee to end a session.

It cannot be sent by a proxy server.

BYE request normally routes end to end, bypassing the proxy server.

BYE cannot be sent to a pending an INVITE or an un established session.

REGISTER

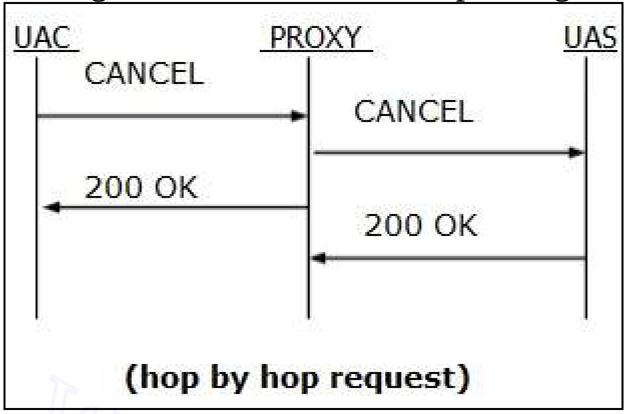
REGISTER request performs the registration of a user agent. This request is sent by a user agent to a registrar server.

- The REGISTER request may be forwarded or proxied until it reaches an authoritative registrar of the specified domain.
- It carries the AOR (Address of Record) in the To header of the user that is being registered.
- REGISTER request contains the time period (3600sec).
- One user agent can send a REGISTER request on behalf of another user agent. This is known as third-party registration. Here, the From tag contains the URI of the party submitting the registration on behalf of the party identified in the To header.

CANCEL

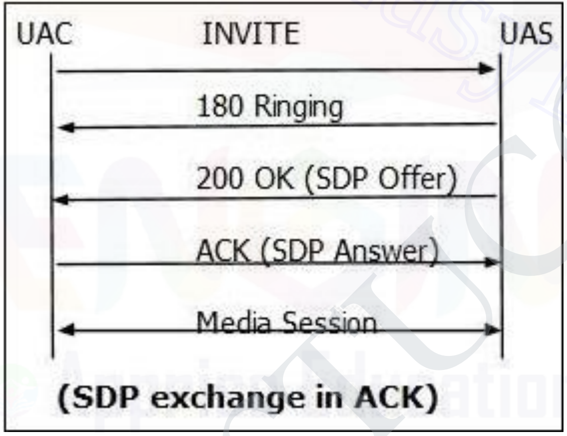
CANCEL is used to terminate a session which is not established. User agents use this request to cancel a pending call attempt initiated earlier.

- It can be sent either by a user agent or a proxy server.
- CANCEL is a hop by hop request, i.e., it goes through the elements between the user agent and receives the response generated by the next stateful element.

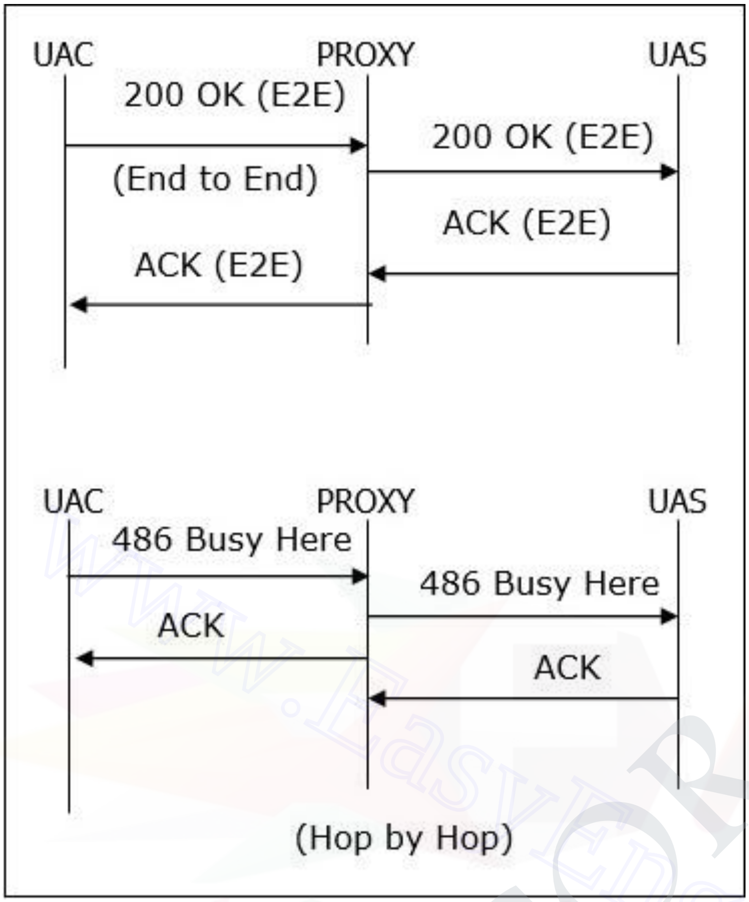


ACK

ACK is used to acknowledge the final responses to an INVITE method. An ACK always goes in the direction of INVITE.ACK may contain SDP body (media characteristics), if it is not available in INVITE.



- ACK may not be used to modify the media description that has already been sent in the initial INVITE.



- A stateful proxy receiving an ACK must determine whether or not the ACK should be forwarded downstream to another proxy or user agent.
- For 2xx responses, ACK is end to end, but for all other final responses, it works on hop by hop basis when stateful proxies are involved.

2.8.5 SIP - Headers

A header is a component of a SIP message that conveys information about the message. It is structured as a sequence of header fields.

SIP header fields in most cases follow the same rules as HTTP header fields. Header fields are defined as Header: field, where Header is used to represent the header field name, and field is the set of tokens that contains the information. Each field consists of a fieldname followed by a colon (":") and the field-value (i.e., field-name: field-value).

SIP Headers - Compact Form

The following image shows the structure of a typical SIP header.

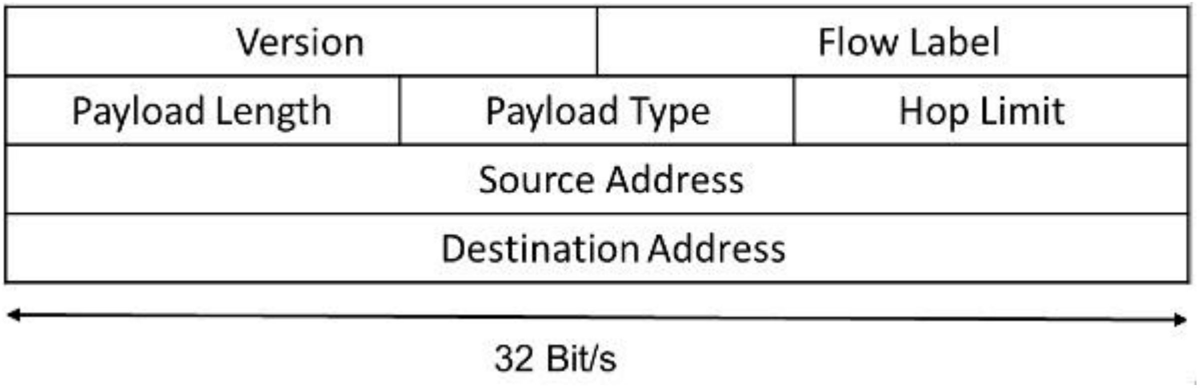


Fig.2.20 SIP Header

Headers are categorized as follows depending on their usage in SIP −

2.8.6 SIP - Mobility

Personal mobility is the ability to have a constant identifier across a number of devices. SIP supports basic personal mobility using the REGISTER method, which allows a mobile device to change its IP address and point of connection to the Internet and still be able to receive incoming calls.

SIP can also support service mobility – the ability of a user to keep the same services when mobile

SIP Mobility During Handover(Pre-call)

A device binds its Contact URI with the address of record by a simple sip registration. According to the device IP address, registration authorizes this information automatically update in sip network.

During handover, the User agent routes between different operators, where it has to register again with a Contact as an AOR with another service provider.

For example, let's take the example of the following call flow. UA which has temporarily received a new SIP URI with a new service provider. The UA then performs a double registration −

The first registration is with the new service operator, which binds the Contact URI of the device with the new service provider's AOR URI.

The second REGISTER request is routed back to the original service provider and provides the new service provider's AOR as the Contact URI.

As shown later in the call flow, when a request comes in to the original service provider's network, the INVITE is redirected to the new service provider who then routes the call to the user.

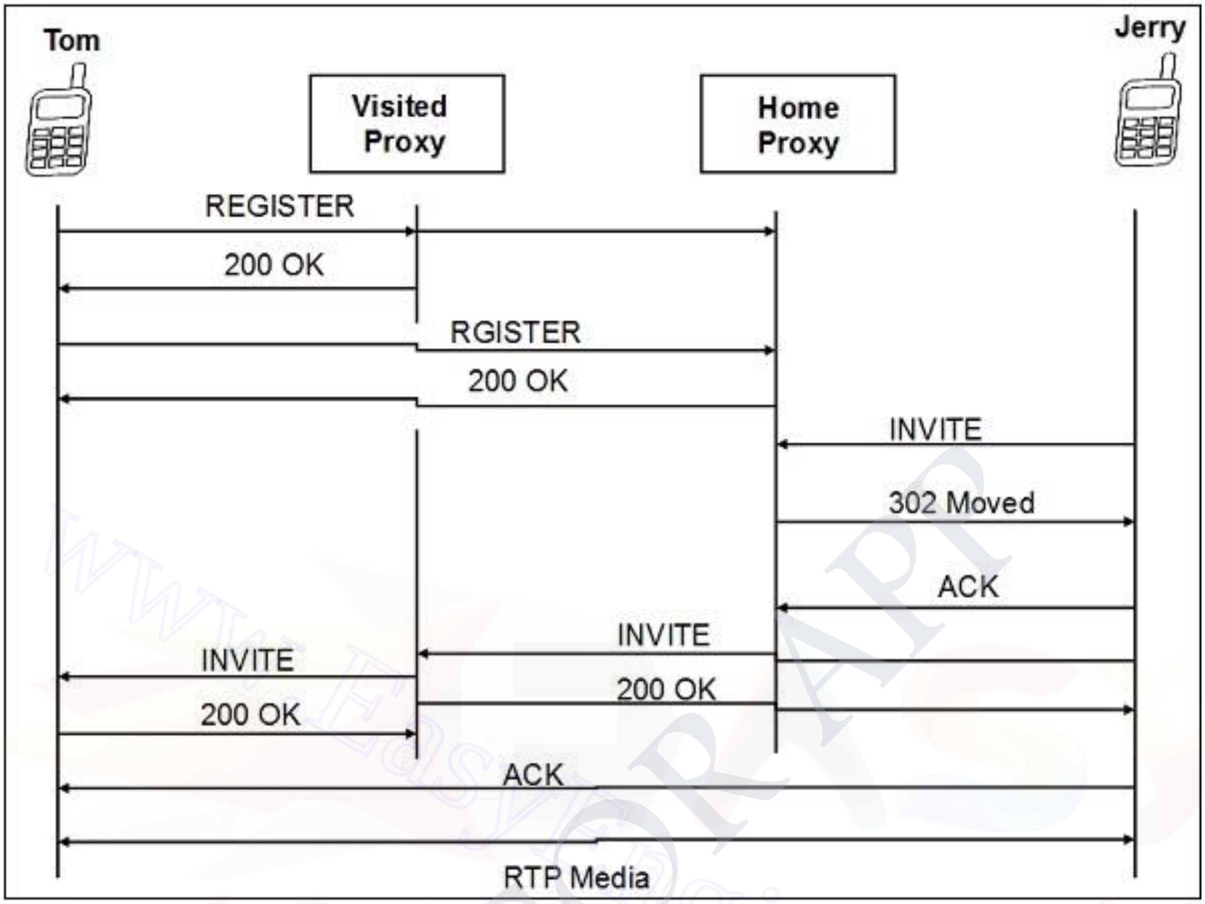


Fig.2.21 SIP Mobility During Handover

The first INVITE that is represents in the above figure would be sent to sip:registrar2.in; the second INVITE would be sent to sip: sip:Tom@registrar2.in, which would be forwarded to sip:Tom@172.22.1.102. It reaches Tom and allows the session to be established. Periodically both registrations would need to be refreshed.

Mobility During a Call(re-Invite)

User Agent may change its IP address during the session as it swaps from one network to another. Basic SIP supports this scenario, as a re-INVITE in a dialog can be used to update the Contact URI and change the media information in the SDP.

Take a look at the call flow mentioned in the figure below.

Here, Tom detects a new network,

Uses DHCP to acquire a new IP address, and

Performs a re-INVITE to allow the signaling and media flow to the new IP address.

If the UA can receive media from both networks, the interruption is negligible. If this is not the case, a few media packets may be lost, resulting in a slight interruption to the call.

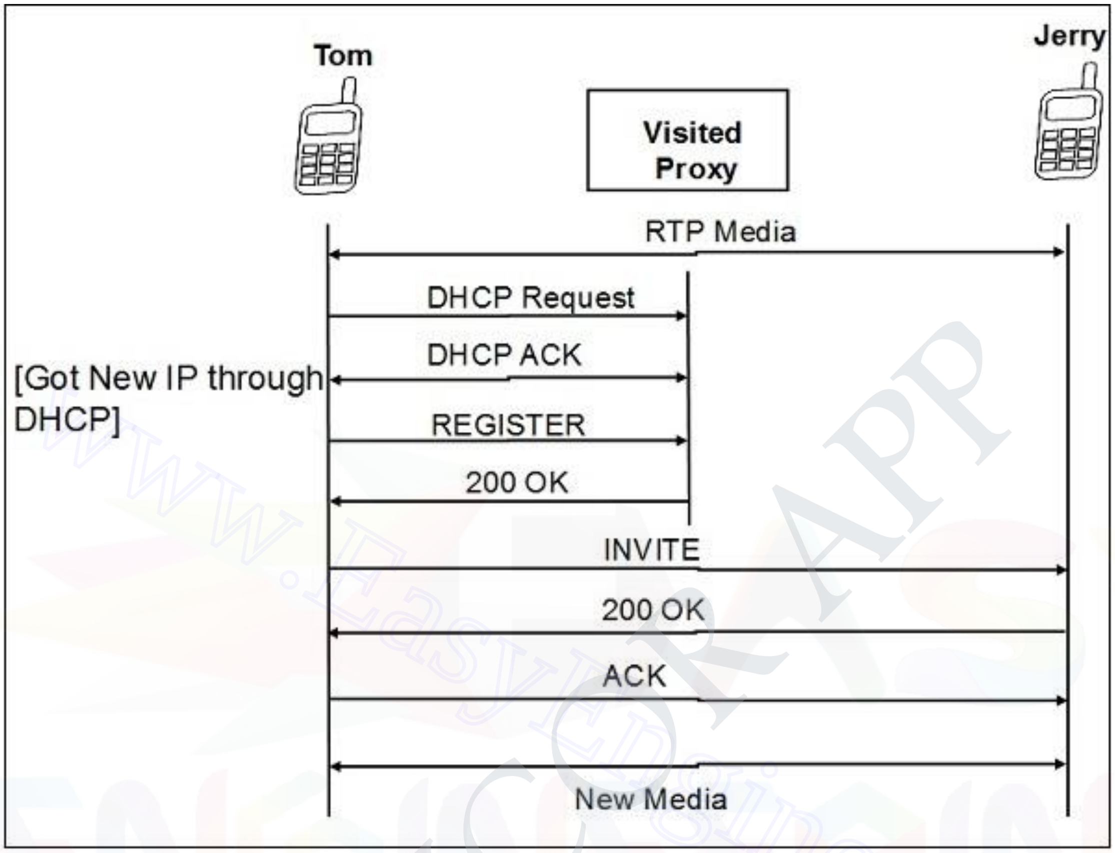


Fig.2.22 SIP Mobility During a Call(re-Invite)

The re-INVITE would appear as follows −

The re-INVITE contains Bowditch's new IP address in the Via and Contact header fields and SDP media information.

Mobility in Mid call (With replace Header)

In mid call mobility, the actual route set (set of SIP proxies that the SIP messages must traverse) must change. We cannot use a re-INVITE in mid call mobility

For example, if a proxy is necessary for NAT traversal, then Contact URI must be changed — a new dialog must be created. Hence, it has to send a new INVITE with a Replaces header, which identifies the existing session.

Note − Suppose A & B both are in a call and if A gets another INVITE (let's say from C) with a replace header (should match existing dialog), then A must accept the INVITE and terminate the session with B and transfer all resource to newly formed dialog.

The call flow is shown in the following Figure. It is similar to the previous call flow using re-INVITE except that a BYE is automatically generated to terminate the existing dialog when the INVITE with the Replaces is accepted.

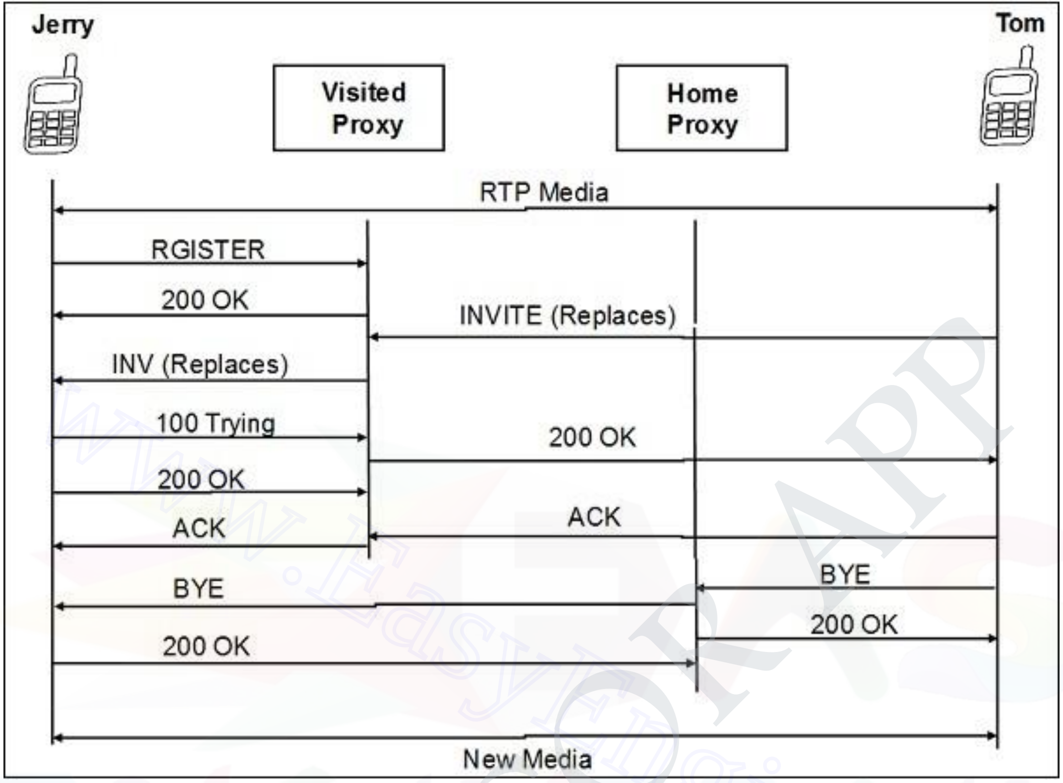


Fig 2.23. SIP Mobility in Midcall (With replace Header)

Given below are the points to note in this scenario

- The existing dialog between Tom and Jerry includes the old visited proxy server.
- The new dialog using the new wireless network requires the inclusion of the new visited proxy server.
- As a result, an INVITE with Replaces is sent by Tom, which creates a new dialog that includes the new visited proxy server but not the old visited proxy server.
- When Jerry accepts the INVITE, a BYE is automatically sent to terminate the old dialog that routes through the old visited proxy server that is now no longer involved in the session.
- The resulting media session is established using Tom's new IP address from the SDP in the INVITE.

Service Mobility

Services in SIP can be provided in either proxies or in UAs. Providing service mobility along with personal mobility can be challenging unless the user's devices are identically configured with the same services.

SIP can easily support service mobility over the Internet. When connected to Internet, a UA configured to use a set of proxies in India can still use those proxies when roaming in Europe. It does not have any impact on the quality of the media session as the media always flows directly between the two UAs and does not traverse the SIP proxy servers.

Endpoint resident services are available only when the endpoint is connected to the Internet. A terminating service such as a call forwarding service implemented in an endpoint will fail if the endpoint has temporarily lost its Internet connection. Hence some services are implemented in the network using SIP proxy servers.

2.9 Mobile ad-hoc networks

Mobile ad-hoc networks are the only choice in the situations where users of a network cannot rely on an infrastructure. Ad-hoc networks are mobile, wireless, multi-hop ad-hoc networks.

Instant infrastructure: Planning and administration of infrastructure is difficult. In those situations ad-hoc connectivity is used.

Disaster relief:

In disaster areas where Hurricanes cut phone and power lines, floods destroy base stations, fires burn servers. Emergency teams must set up an infrastructure extremely fast and reliable. Here mobile ad-hoc connectivity is used.

Remote areas:

For remote areas satellite infrastructures or ad-hoc networks are used.

Effectiveness:

For some applications where existing infrastructure is too expensive, a ad-hoc packet oriented network might be a better solution.

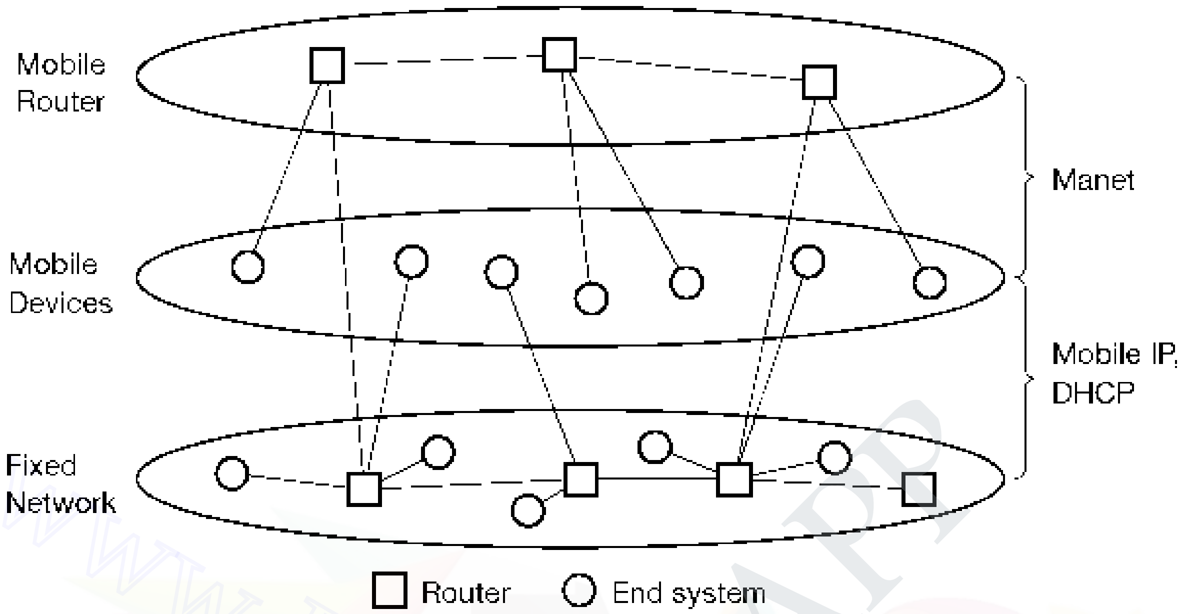


Fig.2.24 MANETs and mobile IP

In ad-hoc networks the mobile node comprises of routing and end system functionality. The above figure shows that Mobile devices can be connected either directly with an infrastructure using Mobile IP for mobility support and DHCP as a source of many parameters, such as an IP address.

2.9.1 Routing

In wireless networks with infrastructure a base station is used and it covers all mobile nodes. But in ad-hoc networks each node must be able to find a path between source and destination to forward the data.

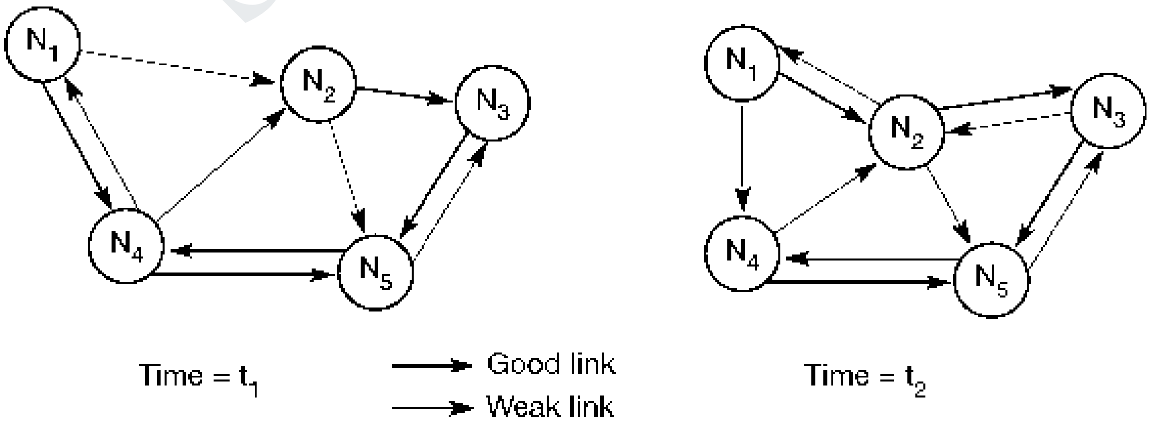


Fig.2.25 Example ad-hoc network

The differences between wired networks and ad-hoc networks are:

Asymmetric links:

A strong link in one direction and weak link in another direction. Routing information of one direction is differentfrom another direction.

Redundant links:

In wired networks, redundant links are used to manage link failure. In ad-hoc networks no administrator is to control redundancy, and computation overhead is high.

Interference:

In wired networks different wires are used as links. So no interference. In ad-hoc networks, one transmission might interfere with another, and nodesmight overhear the transmissions of other nodes. Interference may destroy the data.

Dynamic topology:

The nodes are moving, hence it changes the network topology. Routing algorithms and routing tables must be updated for changing topology. The updation is possible in wired networks and fails in ad-hoc networks.

The source node sends data through one path. But the receiver send acknowledgement through another path. When the topology changes the data transmission path and acknowledgement path are different. In ad-hoc networks, The optimal knowledge for every node would be a description of the current connectivity between all nodes, the expected traffic flows, capacities of all links, delay of each link, and the computing and battery power of each node. But knowing all these factors is difficult.

Periodic updates in ad-hoc network waste the battery power and bandwidth. This is the important problem, since the battery power and bandwidth are important resources.

Considering all the additional difficulties in comparison to wired networks, the following observations concerning routing can be made for ad-hoc networks with moving nodes.

1. Traditional routing algorithms designed for wired networks are designed without considering highly dynamic topology, asymmetric links or interference.

2. The routing algorithms are using the connectivity and interference information from lower layers to find a good path.
3. It will take more time to collect all information, within that the topology may be changed.
4. TO route data at least one router has to be within the range of each node and should have sufficient power.
5. In case of changing environment, nodes have to decide the routing node to forward the data to destination.
6. Flooding:- Forwarding data to all nodes. It will create loops. To avoid loops, a hop counter is used to define the upper bound.
7. Group of nodes form one cluster. For each cluster one head is used to route data between clusters. It makes the routing process as simple and less dynamic.

The routing protocol is subdivided into three categories

1. Flat routing protocol
2. Hierarchical routing protocol
3. Geographic position assisted routing protocol

2.10 Destination sequence distance vector

Distance vector routing is used as routing information protocol in wired networks.

In proactive routing protocol, every node maintains routing information to every other node in the network. The routing information is usually kept in a number of different tables. These tables are periodically updated. The difference between these protocols exists in the way the routing information is updated, detected and the type of information kept at each routing table.

Proactive protocols are not suitable for large networks as they need to maintain node entries for each and every node in the routing table of every node. These protocols maintain different number of routing tables varying from protocol to protocol. There are various well known proactive routing protocols, example: DSDV, OLSR, WRP, etc.

Routing protocols in packet-switched networks traditionally use either distance vector or link-state routing algorithm. Both algorithms allow a host to find the next hop to reach the destination through shortest path. The metric of the shortest path may be the number of hops, time delay in milliseconds, total number of packets queued along the path, etc. Such shortest path protocols have been used in dynamic packet switched networks successfully. The main drawback of both link – state and distance vector protocol are that they take too long to converge and have a high message complexity. Because of the limited bandwidth of wireless links in ad hoc network, message complexity must be kept low and because of the rapidly changing topology, new routing protocols have to be developed to fulfill the basic philosophy.

DSDV uses two things for routing the data.

1. Sequence Number: Sequence number is added in the advertisements. It is used to avoid loops. The advertisement with same sequence number should be discarded.
2. Damping: Advertisements containing changes in the topology currently stored are therefore not disseminated further. A node waits with dissemination if these changes are probably unstable. Waiting time depends on the time between the first and the best
   announcement of a path to a certain destination.

DSDV belongs to the Proactive type of routing protocols. In this protocol, each mobile node in the network keeps a routing table listing all other nodes it has known either directly or through some neighbors. Every node has a single entry in the routing table.

The entry will have information about the node’s IP address, last known sequence number and the hop count to reach that node. Along with these details, the table also keeps track of the next hop neighbor to reach the destination node.
Using the newly added sequence number, the mobile nodes can distinguish state route information from the new and thus prevent the formation of routing loops. The main contribution of the algorithm was to solve the routing loop problem.

Packet Routing and Routing Table Management

In DSDV, using such routing table stored in each mobile node, the packets are transmitted between the nodes of an ad hoc network.

Each node of the ad hoc network updates the routing table with advertisement periodically or when significant new information is available to maintain the consistency of the routing table with the dynamically changing topology of the ad hoc network. Periodically or immediately when network topology changes are detected, each mobile node advertises routing information using broadcasting or multicasting a routing table update packet. The update packet starts out with a metric of one to direct connected nodes. This indicates that each receiving neighbor is one metric (hop) away from the node.

After receiving the update packet, the neighbors update their routing table with incrementing the metric by one and retransmit the update packet to the corresponding neighbors of each of them. The process will be repeated until all the nodes in the ad hoc network have received a copy of the update packet with a corresponding metric.

If a node receives multiple update packets for a same destination during the waiting time period, the routes with more recent sequence numbers.

If the update packets have the same sequence number with the same node, the update packet with the smallest metric will be used and the existing route will be discarded or stored as a less preferable route. In this case, the update packet will be propagated with the sequence number to all mobile nodes in the ad hoc network.

The advertisements of routes that are about to change may be delayed until the best routes have been found. Delaying the advertisement of possibly unstable route can damp the fluctuations of the routing table and reduce the number of rebroadcasts of possible route entries that arrive with the same sequence number. The elements in the routing table of each mobile node change dynamically to keep consistency with dynamically changing topology of an ad hoc network.

To reach this consistency, the routing information advertisement must be frequent or quick enough to ensure that each mobile node can almost always locate all the other mobile nodes in the dynamic ad hoc network. Upon the updated routing information, each node has to relay data packet to other nodes upon request in the dynamically created ad hoc network.

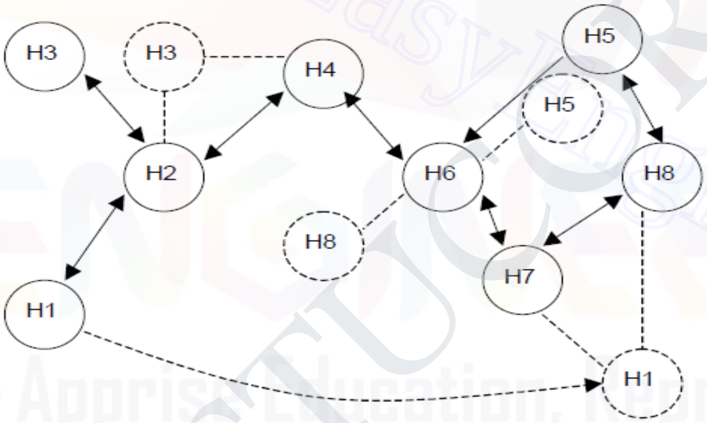


Fig.2.27 An example of the ad hoc networks

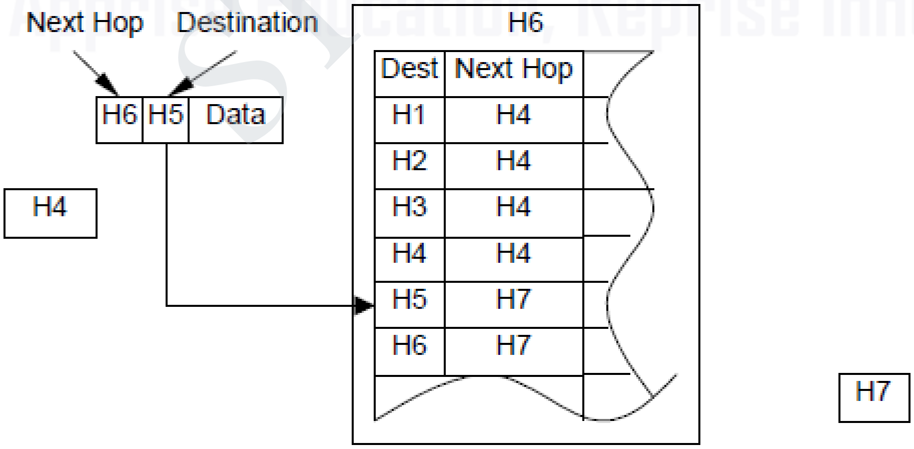
Table 2.1 The routing table of node H6 at one instant

Dest	Next Hop	Metric	Seq.No.	Install
H1	H4	3	S406_H1	T001_H6
H2	H4	2	S128_H2	T001_H6
H3	H4	3	S564_H3	T001_H6
H4	H4	1	S710_H4	T002_H6
H5	H7	3	S392_H5	T001_H6
H6	H6	0	S076_H6	T001_H6
H7	H7	1	S128_H7	T002_H6
H8	H7	2	S050_H8	T002_H6

The table 2.1 is the routing table of the node H6 at the moment before the movement of the nodes. The Install time field in the routing table helps to determine when to delete stale routes.

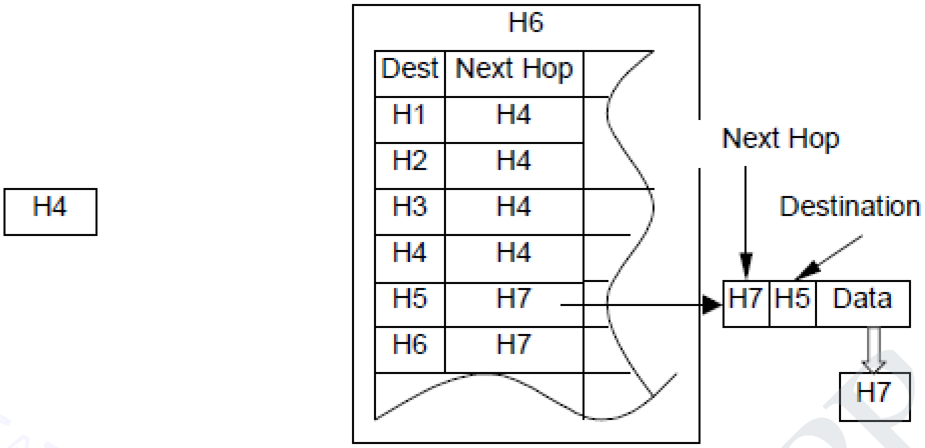
DSDV packet routing

The following figure shows an example of packet routing procedure in DSDV. Node H4 wants to send a packet to the nodeH5 as shown in Figure. The node H4 checks its routing table and locates that the next hop for routing the packet is node H6. Then H4 sends the packet to H6 as shown in Figure. If the sequence number of one node in the newly received routing information update packet is
same as the corresponding sequence number in the routing table, then the metric will be compared and the route with the smallest metric will be used.



Dest	Next Hop
H1	H4
H2	H4
H3	H4
H4	H4
H5	H7
H6	H7

Fig.2.28 Node H6 looks up the destination and route for forwarding the packet in its routing table



Dest	Next Hop
H1	H4
H2	H4
H3	H4
H4	H4
H5	H7
H6	H7

Fig.2.29 Node H6 forwards the packet to the next hop

Node H6 looks up the next hop for the destination node H5 in its routing table when it receives the packet. Node H6 then forwards the packet to the next hop H7 as specified in the routing table. The routing procedure repeated along the path until the packet finally arrives its destination H5.

In the routing information updating process, the original node tags each update packet with a sequence number to distinguish stale updates from the new one. The sequence number is a monotonically increasing number that uniquely identifies each update from a given node. As a result, if a node receives an update from another node, the sequence number must be equal or greater than the sequence number of the corresponding node already in the routing table, or else the newly received routing information in the update packet is stale and should be discarded. If the sequence number of one node in the newly received routing information update packet is same as the corresponding sequence number in the routing table, then the metric will be compared and the route with the smallest metric will be used.

In addition to the sequence number and the metric for each entry of the update packet, the update route information contains also both the address of the final destination and the address of the next hop.

There are two types of update packets, one is called full dump, which carries all of the available routing information.

The other is called incremental, which carries only the routing information changed since the last full dump.

The node H7 advertises its routing information with broadcasting the update packet to its neighbors. When the node H6 receives the update packet, it will check the routing information of each item contained in both the update packet and the its routing table and update the routing table. The entries with higher sequence numbers are always entered into the routing table regardless of whether each of them have a higher metric or not. If an entry has the same sequence number, the route with smaller metric is entered into the routing. The items with old sequence numbers in the update packet are always ignored

Responding to Topology Changes

Links can be broken when the mobile nodes move from place to place or have been shut down etc. The broken link may be detected by the communication hardware or be inferred if no broadcasts have been received for a while from a former neighbor. The metric of a broken link is assigned infinity.

When a link to next hop has broken, any route through that next hop is immediately assigned infinity metric and an updated sequence number. Because link broken qualifies as a significant route change, the detecting node will immediately broadcast an update packet and disclose the modified routes.

To describe the broken links, any mobile node other than the destination node generates a sequence number, which is greater than the last sequence number received from the destination. This newly generated sequence number and a metric of infinity will be packed in an update message and flushed over the network. To avoid nodes themselves and their neighbors generating conflicting sequence numbers when the network topology changes, nodes only generate even sequence numbers for themselves, and neighbors only generate odd sequence numbers for the nodes responding to the link changes. Destination Next Hop Metric Sequence Number

2.11 DYNAMIC SOURCE ROUTING (DSR) PROTOCOL

The Dynamic Source Routing Protocol is a source-routed on-demand routing protocol. A node maintains route caches containing the source routes that it is aware of. The node updates entries in the route cache as and when it learns about new routes.
The two major phases of the protocol:

Route Discovery and Route Maintenance.

Route Discovery

When the source node wants to send a packet to a destination, it looks up its route cache to determine if it already contains a route to the destination. If it finds that an unexpired route to the destination exists, then it uses this route to send the packet. But if the node does not have such a route, then it initiates the route discovery process by broadcasting a route request packet.

Route Request Mechanism

Source node S floods Route Request (RREQ)

Each RREQ, has sender's address, destination's address, and a unique Request ID determined by the sender

Each node appends own identifier when forwarding RREQ

Each intermediate node checks whether it knows of a route to the destination.

If it does not, it appends its address to the route record of the packet and forwards the packet to its neighbors.

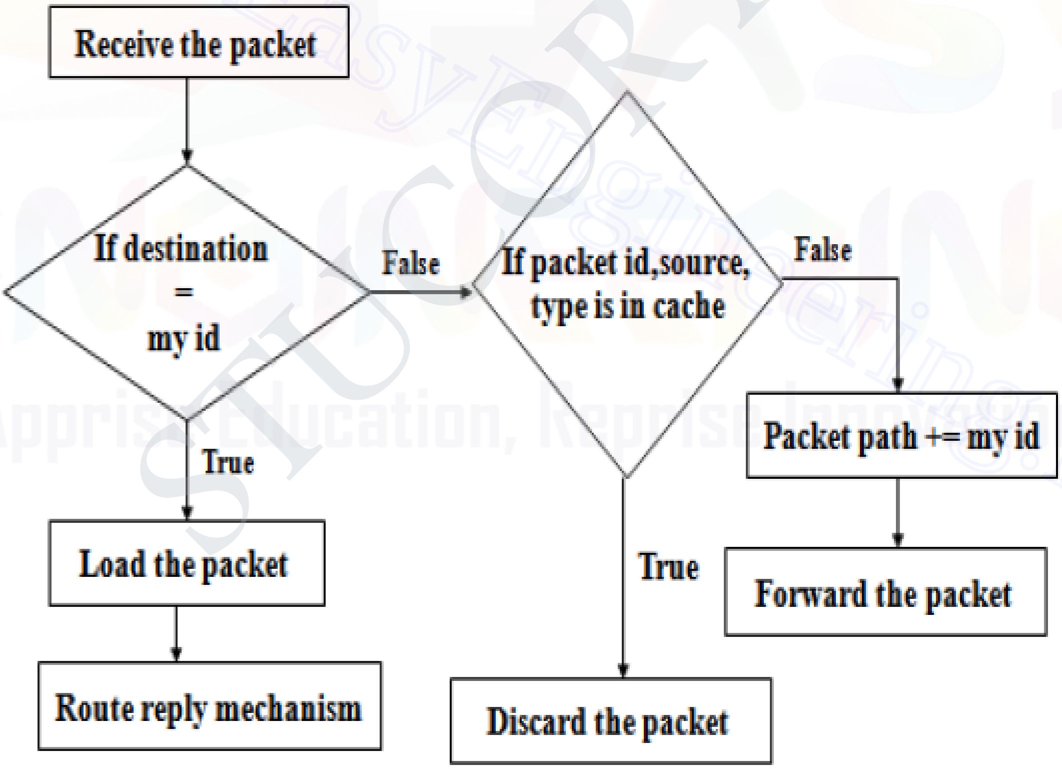


Fig.2.30 Route Request Mechanism

.

If the node has already received the request (which is identified using the unique identifier), it drops the request packet.

If the node recognizes its own address as the destination, the request has reached its target.

Otherwise, the node appends its own address to a list of traversed hops in the packet and broadcasts this updated route request.

To largely eliminate these duplicates, each request should contain a unique request id from the original sender. Each host keeps a cache giving the request id and sender address of recently forward requests, and discards a request rather than propagating it if it has already propagated an earlier copy of the same request id.

Limiting the maximum number of hops over which any route discovery packet can be propagated, can thus further reduce the number of duplicate requests propagated. When processing a received route discovery request rather than forwarding it if it is not the target of the request and if the route recorded in the packet has already reached the maximum length.

During Route Discovery, the sending node saves a copy of the message in the send buffer Send buffer has a copy of every packet that cannot be transmitted by this node due to lack of a route Each packet is time stamped and discarded after a specified time out period, if it cannot be forwarded For packets waiting in the send buffer, the node should occasionally initiate a new route discovery

New Route Discovery rate for the same destination node should be limited if the node is currently unreachable.

Results in wastage of wireless bandwidth due to a large number of RREQs destined for the same destination -> High overhead

To reduce the overhead, the node goes into exponential back-off for the new route discovery of the same target

Packets are buffered that are received during the back-off Nodes on receiving RREP, caches the route included in the RREP

When node S sends a data packet to D, the entire route is included in the packet header hence the name source routing

Intermediate nodes use the source route included in a packet to determine to whom a packet should be forwarded.

N1 broadcasts the request ((N1), id = 42, target = N3), N2 and N4 receive this request.

N2 then broadcasts ((N1, N2), id = 42, target = N3), N4 broadcasts ((N1, N4), id = 42, target = N3). N3 and N5 receive N2's broadcast, N1, N2, and N5 receive N4's broadcast.

N3 recognizes itself as target, N5 broadcasts ((N1, N2, N5), id = 42, target = N3). N3 and N4 receive N5's broadcast. N1, N2, and N5 drop N4's broadcast packet, because they all recognize an already received route request (and N2's broadcast reached N5 before N4's did).
N4 drops N5's broadcast, N3 recognizes (N1, N2, N5) as an alternate, but longer route.

N3 now has to return the path (N1, N2, N3) to N1. This is simple assuming symmetric links working in both directions. N3 can forward the information using the list in reverse order.

Route Reply Mechanism

A route reply is generated when either the destination or an intermediate node with current information about the destination receives the route request packet. A route request packet reaching such a node already contains, in its route record, the sequence of hops taken from the source to this node.

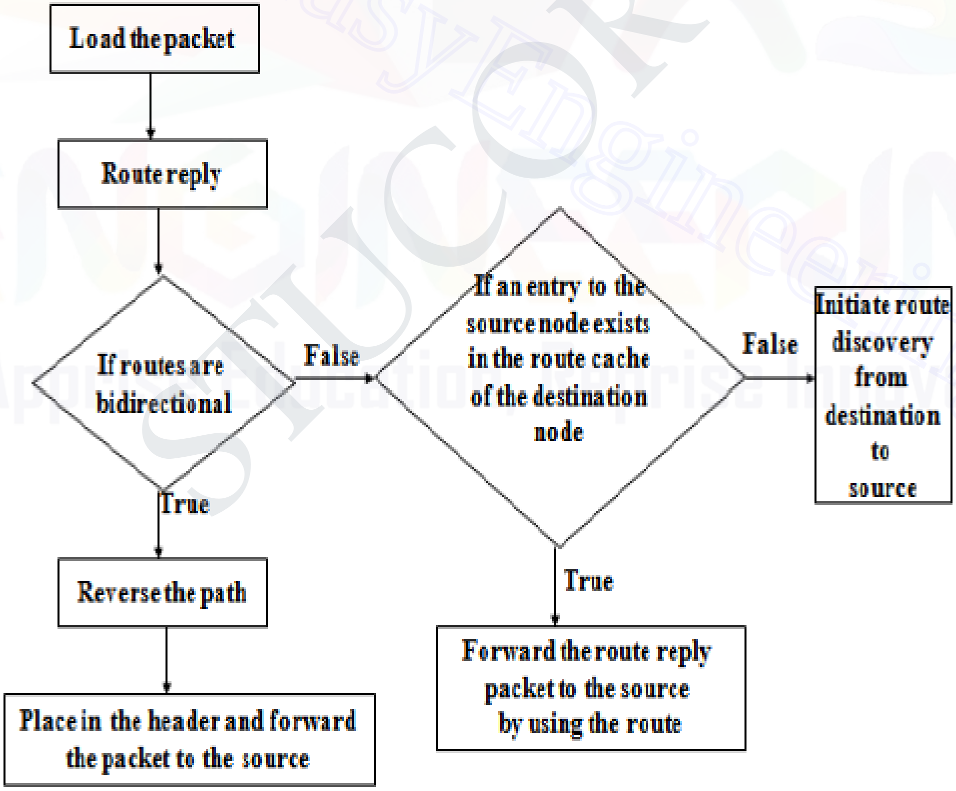

Load the packet
Route reply
If routes are bidirectional
False
True
Reverse the path
Place in the header and forward the packet to the source
If an entry to the source node exists in the route cache of the destination node
False
Initiate route discovery from destination to source
True
Forward the route reply packet to the source by using the route

Fig.2.31 Route Reply Mechanism

In order to return route reply packet to the initiator of the route discovery the target host must have a route to the initiator. If the target has an entry for this destination in its route cache, then it may send the route reply packet using this route in the same way as is used in sending any other packet.

Otherwise the target may reverse the route record from the route request packet, and use this route to send the route reply packet. This however, requires the wireless network communication between each of these pairs of hosts to work equally well in both directions, which may not be true in some environments or with some MAC level protocols.

Each node maintains a Route Cache which records routes it has learned and overheard over time

ROUTE MAINTENANCE

DSR uses two types of packets for route maintenance:

Route Maintenance

Route maintenance performed only while route is in use

Error detection:

Monitors the validity of existing routes by *passively* listening to data packets transmitted at neighboring nodes

Lower level acknowledgements

When problem detected, send *Route Error* packet to original sender to perform new route discovery

Host detects the error and the host it was attempting;

Route Error is sent back to the sender the packet – original src

Route Reply Storms

Using route cache nodes can reply to RREQ, if they have the route. If lots of node replies at the same time, it can cause route reply storm Simultaneous replies from various nodes can cause collision at source (route reply storm)

Also each node may reply with a different route length, e.g. 1 hop (G) , 2 hops (B-G) , and 3 (C-B-G)

Route Request - Hop Limits

Each RREQ message contains a field called hop limit Hop limit controls the propagation of RREQ to the number of hops i.e. how many intermediate nodes are allowed to forward the RREQ

Each receiving node decrements the hop-limit by 1 before forwarding. RREQ is not forwarded & is discarded by node when this limit becomes zero even before reaching the destination. A RREQ with hop-limit zero will determine that the target is the one hop neighbor It also likely that this one hop neighbor has the source route in its cache. If no RREP is received within a timeout period, a new RREQ is sent by the sender with no hop-limit.

Packet Salvaging

When a node discovers that it cannot forward a data packet because the nexthop link is broken, it generates RERR.

It Sends RERR upstream.

Searches its own cache to find an alternate route from itself to destination to forward this packet

If route is found, the node modifies the route as per the route cache and forwards to the next hop node

Otherwise packet is dropped

When a packet is salvaged – its marked as "Salvaged"

A Salvaged packet is salvaged only one time to avoid routing loops when salvaged at multiple locations.

A recommended strategy for salvaging is breakdown the address into two parts – prefix address (hops that are used until now) and suffix address (address from the route cache) this strategy avoids backtracking from the current node to an already traversed node

Route Shortening

Routes may be shortened if one of intermediate nodes becomes unnecessary

Spreading of Route Error Message

When a source node receives an RERR in response to a data packet that it forwarded It piggybacks this RERR on a new RREQ that it forwards to its neighbors.

Neighbors get aware of the RERR and update their route caches.

This helps in reductions in getting the stale routes in RREP sent by the neighbors.

Caching Negative Information

In certain situations, caching of negative information can help DSR. For example,

If A knows that link C-D is broken, it can keep this information in its routing cache for a specified time (using a timer) , e.g. by making the distance to routes through C as infinity

A will not use this path in response to any RREP it receives for subsequent RREQs

After the expiration of timer, the link can be added again in the route cache with correct hop counts, if link is repaired

Advantages

- Routes maintained only between nodes who need to communicate reduces overhead of route maintenance
- Route caching can further reduce route discovery overhead
- A single route discovery may yield many routes to the destination, due to intermediate nodes replying from local caches

Disadvantages

- Packet header size grows with route length due to source routing
- Flood of route requests may potentially reach all nodes in the network
- Care must be taken to avoid collisions between route requests propagated by neighboring nodes
- Insertion of random delays before forwarding RREQ
- Increased contention if too many route replies come back due to nodes replying using their local cache

UNIT – III MOBILE TRANSPORT LAYER

3.0 TCP enhancements for wireless protocols

Transmission Control Protocol (TCP) is the dominant transport protocol in the Internet and supports many of the most popular Internet applications, such as the World Wide Web (WWW), file transfer and e-mail. TCP congestion control algorithms dynamically learn the network bandwidth and delay characteristics of a network and adapt its performance to changes in traffic so as to avoid network collapse

However, networks with wireless links suffer from significant packet losses due to random bit errors and handoffs. Hence TCP performs poorly in networks with wireless links because it treats any packet loss in the network to be a result of network congestion and slows down its transmission rate, or even cause the TCP sender to experience unnecessary timeouts, further reducing its performance.

The development of advance wireless networks, such as WiFi, UMTS and WiMAX, make it necessary to find ways to improve TCP's efficiency and resource utilization, as well as improve the user's experience and reduce latency times.

In order to find effective solutions to this effect, packet losses across wireless links should be distinguished from congestion related packet losses.

Snoop Protocol

Snoop is a TCP-aware link layer protocol. It is designed to improve the performance of TCP in the wired-cum-wireless networks. The Snoop Protocolwas designed to solve the burst/intermittent packet loss due to high bit error rates and short temporary disconnections experienced by TCP in wireless link. The main idea of the protocol is to cache packets at the base station and perform local retransmissions across the wireless link.

The snoop module maintains a cache to temporarily store packets sent from the wired network to mobile hosts that have not been yet acknowledged by the mobile host. For transfer of data from a fixed host to a mobile host, modifications are made only to the routing code at the base station.

These modifications include caching unacknowledged TCP data and performing local retransmissions based on a few policies dealing with acknowledgments (from themobile host) and timeouts. The base station routing code is modified by adding a module, called the snoop that monitors every packet passing through the connection in either direction.

When a new packet sent from a fixed host arrives at the base station, the Snoop module will add it to the cache. Then it forwards the packet to the routing code that routes it to the appropriate mobile host. The snoop module relies on a cache of un-acknowledged packets to improve end to end TCP semantics. Typically, the size of this cache isproportional to the TCP window size.

3.1 TRADITIONAL TCP

TheTransmission Control Protocol(TCP) is one of the core protocols of the Internet protocolsuite, which is simply referred to as TCP/IP. TCP is reliable, delivery of datain order and incorporates congestion control and flow control mechanisms.TCP supports many of the Internet application protocols and resultingapplications, including the World Wide Web,e-mail, File Transfer Protocol.In the Internet protocol suite, TCP is theintermediate layer between the Internet layerand application layers.

The major responsibilities of TCP in are,

- It provides reliable transport of datawhich does not allow losses of data.
- It controls congestions in the networks, will notallow degradation of the networkperformance,
- It controls a packet flow between thetransmitter and the receiver not to exceedthe receiver's capacity.
- It supports full duplex transmission.

Connection –Oriented:

Connection oriented providesanestablishment of virtual connection before any user data is transferred.

The user program is notified once the connection is not established or even interrupted.

Reliable

Reliable means that every transmission of data is acknowledged by the receiver.

If the sender does not receive acknowledgement within a specified amount of time, the sender will retransmits the data for further acknowledgement.

Byte Stream

Stream means that the connection is treated as a stream of bytes. The user application does not need to put data as packages in individual datagrams (as with UDP).

Buffering

TCP is responsible for buffering data and determining the time to send a datagram.
It is possible for an application to tell TCP to send the data it has buffered without waiting for a buffer to fill up.

Full Duplex

TCP provides data transfer in both directions.For the application program these appear as 2 unrelated data streams, although TCP

can piggyback control and data communication by providing control information (such as an ACK) along with user data.

TCP Segments
The chunk of data that TCP asks IP to deliver is called aTCP segment.

Each segment contains:
·Data bytes from the byte stream
·Control information that identifies the data byte

TCP employs a number of tools to achieve high performance and avoid 'congestioncollapse', where network performance can fall by several orders of magnitude. Thesemechanisms control the rate of data entering in to the network, by keeping the data flow below a ratethat would initiate collapse.

There are several mechanisms of TCP that influence the efficiencyof TCP in a mobile environment. Acknowledgments for data sent, or lack of acknowledgments,are used by senders to implicitly interpret network conditions between the TCP sender andreceiver.

Table 3.1Comparison of Protocols with different layers

Layer	**Layer Name**	**Protocol**	**Protocol Data Unit**	**Addressing**
5	Application Layer	HTTP,SMTP etc	Messages	n/a
4	Transport Layer	TCP/UDP	Segments/Datagrams	Port

3	Network/Internet Layer	IP	Packets	IP Address
2	Data Link Layer	Ethernet, Wi-Fi	Frames	MAC Address
1	Physical Layer	10 Base T, 802.11	Bits	n/a

3.1.1 CONGESTION CONTROL

A transport layer protocol such as TCP has been designed for fixed networks with fixed end- systems. Congestion may appear from time to time even in carefully designed networks. The packet buffers of a router are filled and the router cannot forward the packets fast enough because the sum of the input rates of packets destined for one output link is higher than the capacity of the output link. The only thing a router can do in this situation is to drop packets.

A dropped packet is lost for the transmission, and the receiver notices a gap in the packet stream. Now the receiver does not directly tell the sender which packet is missing, but continues to acknowledge all in-sequence packets up to the missing one.

The sender notices the missing acknowledgement for the lost packet and assumes a packet loss due to congestion. Retransmitting the missing packet and continuing at full sending rate would now be unwise, as this might only increase the congestion. To mitigate congestion, TCP slows down the transmission rate dramatically.

All other TCP connections experiencing the same congestion do exactly the same so the congestion is soon resolved. Using UDP is not a solution, because the throughput ishigher compared to a TCP connection just at the beginning. After that, congestion is standard anddata transmission quality is unpredictable. Even under heavy load, TCP guaranteesat least sharing of the bandwidth.

When TCP connection transmits data into connection pipe, the data amount is controlled and limited by congestion control of the sender, where the congestion window determines the essential send rate. The window-based congestion control technique employed by TCP will try to adjust the data flow rate by adjusting the size of the window to avoid network congestion and at the same time, providing a fair share of bandwidth of the network over all possible connections

The congestion control increases exponentially with the packets over the connection, where this initial slow start increasing period must be controlled to avoid declining performance of TCP due to the expected overflow of the receiver buffer. The mechanism of congestion control is classified into 4 main Phases, namely, Slow Start, Congestion Avoidance, Fast Retransmit and Fast-Recovery.

3.1.2 SLOW-START PHASE

The initial phase of implementation of TCP is the slow-start phase, where the slow-start algorithm is used by TCP sender to adjust the data flow rate to the receiver where the period of new slow-start begins with every acknowledgement (ACK) received from the TCP receiver.Inother words, the rate of acknowledgements returned by the receiver determines the rate atwhich the sender can transmit data.

The behavior TCP shows after the detection of congestion is calledslowstart.The sender always calculates acongestion windowfor a receiver. The start size of thecongestion window is one segment (TCP packet). The sender sends one packet and waits foracknowledgement. If this acknowledgement arrives, the sender increases the congestionwindow by one, now sending two packets (congestion window = 2). This scheme doubles thecongestion window every time the acknowledgements come back, which takes one round triptime (RTT). This is called the exponential growth of the congestion window in the slow startmechanism.

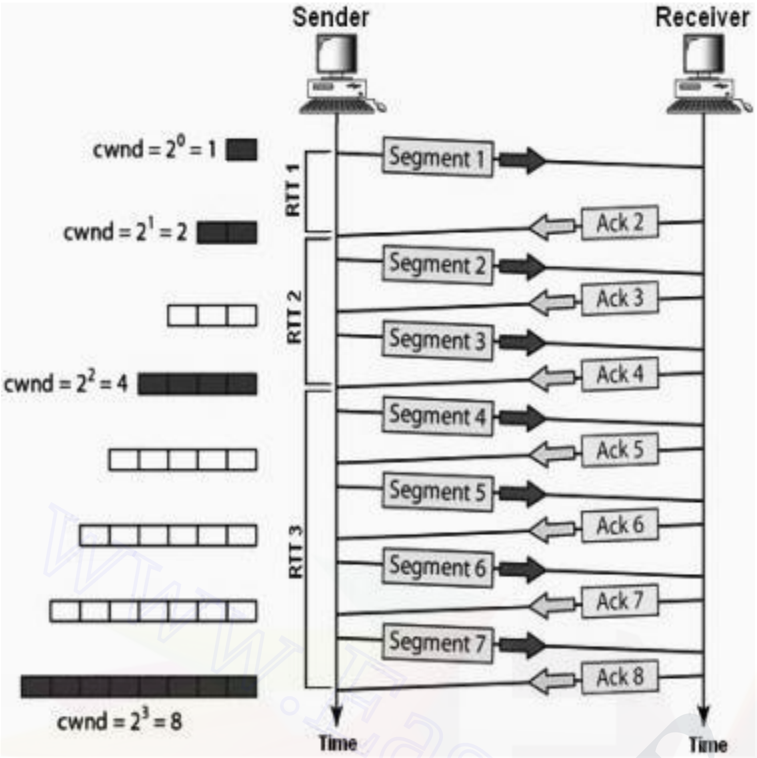


Fig. 3.1 TCP slow-start increases exponentially with RTT.

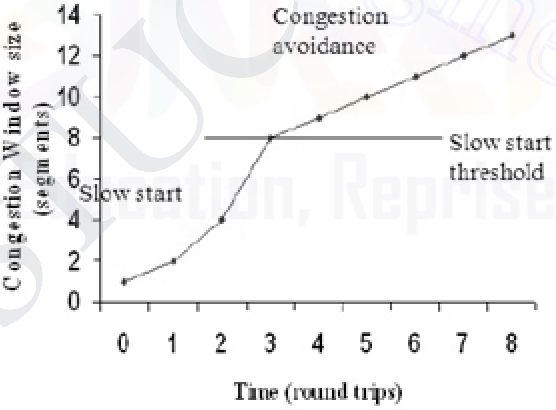


Fig. 3.2 Congestion Window Growth

When a TCP connection first begins, the Slow Start algorithm initializes a congestionwindow to one segment, which is the maximum segment size (MSS) initialized by thereceiver during the connection establishment phase.

When acknowledgements arereturned by the receiver, the congestion window increases by one segment for eachacknowledgement returned. Thus, the sender can transmit the minimum of thecongestion window and the advertised window of the receiver, which is simply called thetransmission window.

Slow Start is actually not very slow when the network is not congested and networkresponse time is good. For example, the first successful transmission andacknowledgement of a TCP segment increases the window to two segments.

Aftersuccessful transmission of these two segments and acknowledgements completes, thewindow is increased to four segments. Then eight segments, then sixteen segments andso on, doubling from there on out up to the maximum window size advertised by thereceiver or until congestion finally does occur.

Doubling the congestion window is too risk. The exponentialgrowth stops at thecongestionthreshold. When the congestionwindow reaches the congestionthreshold, further increase of thetransmission rate is only linear byadding 1 to the congestion windoweach time the acknowledgementscome back.

3.1.3 FAST RETRANSMIT

The congestion threshold can be reduced because of two reasons. First one is if the sender receives continuous acknowledgements for the same packet. It informs the sender that the receiver has got all the packets up to the acknowledged packet in the sequence and also the receiver is receiving something continuously from the sender. The gap in the packet stream is not due to congestion, but a simple packet loss due to a transmission error. The sender can now retransmit the missing packet(s) before the timer expires. This behavior is called Fast retransmit.It is an early enhancement for preventing slow-start to trigger on losses not caused by congestion.

The receipt of acknowledgements shows that there is no congestion to justify a slow start. The sender can continue with the current congestion window.Fast retransmit provides a mechanism to speed-up the retrieval of the connection. This mechanism perceives the losses in segments by duplicate acknowledgements.

.

When a duplicate ACK is received, the sender does not know if it is because a TCPsegment was lost or simply that a segment was delayed and received out of order at thereceiver. If the receiver can re-order segments, it should not be long before the receiversends the latest expected acknowledgement.

Typically no more than one or two duplicateACKs should be received when simple out of order conditions exist. When morethan two duplicate ACKs are received by the sender, it is a strong indication that at leastone segment has been lost. The TCP sender will assume enough time has lapsed for allsegments to be properly re-ordered by the fact that the receiver had enough time to sendthree duplicate ACKs.

When three or more duplicate ACKs are received, the sender does not even wait for aretransmission timer to expire before retransmitting the segment (as indicated by the position of the duplicate ACK in the byte stream). This process is called the FastRetransmit algorithm

Once a loss of segment occurs, the TCP receiver will preserve transmitting ACK segments specifying the next predictable order number. The sequence number will relate to the loss segment. When an individual segment is lost, TCP will preserve creating ACKs for the next segments. This will then induce the reception of duplicate ACKs by the sender. Duplicate ACK represents the lost packet.

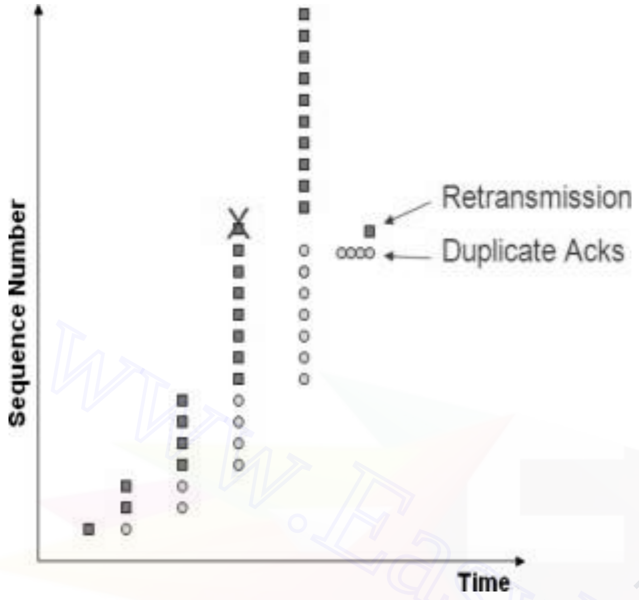


Fig. 3.3 Sequence number with segment loss scenario

In fast retransmit stage, once TCP gets duplicate ACKs, it adopts to resend the segment, where no waiting time is required for the segment timer to expire. This process will speed up the recovery of segment losses. The fast retransmit mechanism based on the concept of resending the unacknowledged segment after three duplicate ACKs are received. When this happens, cwnd size is reset to one packet and the process of slow-start is initiated.Fast retransmit decreases the cwnd to half and at the same time, continues sending segments at this reduced level.

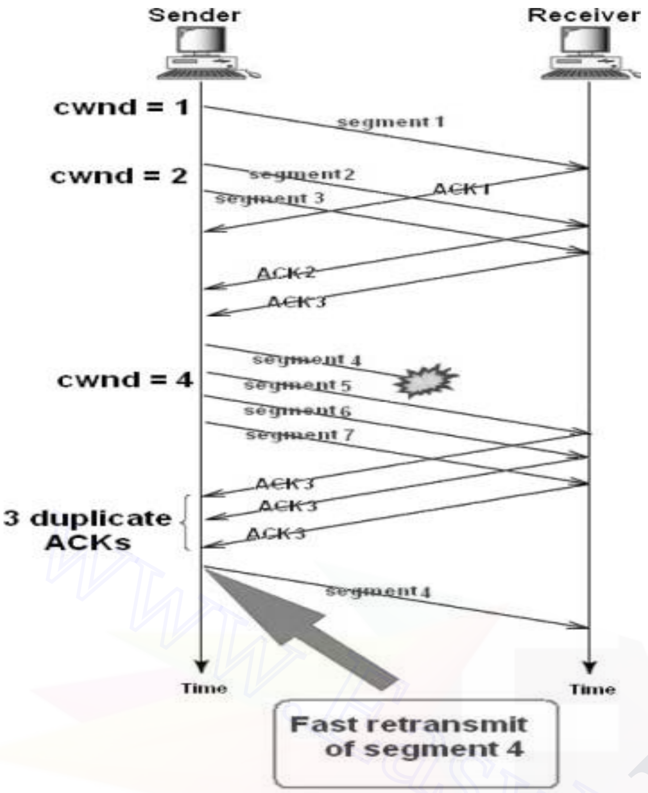


Fig. 3.4 Fast retransmit mechanism

3.1.4 FAST RECOVERY

Fast recovery is a mechanism that replaces slow-start when fast retransmit is used. While duplicate ACKs indicate that a segment has been lost, nevertheless it also indicates that packets are still flowing since the source received a packet with a sequence number higher than the missing packet. In fast recovery, when the segment is lost, the TCP attempts to keep the existing flow rate without returning to slow-start.

The sender can continue with the current congestion window. The sender performs afast recoveryfrom the packet loss. This mechanism can improve the efficiency of TCP. TCP using fast retransmit/fast recovery interprets this congestion in thenetwork and activates the slow start mechanism.

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

The fast recovery is inserted into the function of TCP sender when receiving a primary threshold of duplicate ACK and the value of this threshold is typically fixed to.When the threshold value has been reached, TCP sender decreases cwnd by half and resends single packet.

If any losses happen, the holes of the sequences will be rearranged and the receiver will be ready to accept new segments which are appropriate with its window, but are not the predictable segment.

The receiver will transmit ACKs indicating the sequence hole for each segment over-time, thus duplicated ACKs will be induced. If three or more duplicate ACKs for a segment are transmitted from receiver to sender, the sender adopts immediately that this segment has been missed, and thus, will resend it before RTO expires.

3.1.4.a Fast Recovery Algorithm

The other implementation of fast recovery algorithm is established in TCP NewReno. The fast recovery algorithm is developed in NewReno by means of construing a fractional acknowledgement as a mark of additional lost packet at this sequence number, where fractional ACK is acknowledged after the earlier point, and still in the window of recovery process.

This will further enhance the connection throughput when many packets are lost for data of a single window. Comparatively, TCP Reno is waiting RTO for each packet under this situation, followed by a slow-start, while TCP NewReno deals with this problem by its enhanced recovery.

If segments reaching the destination are out of sequence, the host cannot distribute these segments to the end request as it needs to buffer these segments till the accurate sequence is obtained.

The receiver must transmit an instant duplicate ACK to update the sender of the loss packet (based on the segments that are out of sequence) and the expected number of the sequence. By decreasing cwnd to one single segment after receiving three duplicate ACKs, the network is still able to send segments irrespective of congestion situation.

3.1.4.bAlgorithm Description

With using Fast Retransmit, the congestion window is dropped down to 1 each time network congestion is detected. Thus, it takes an amount of time to reach high link utilization as before. Fast Recovery, however, alleviates this problem by removing the slow-start phase.

The reason for not performing slow-start after receiving 3 dup ACKs is that duplicate ACKs tell the sending side more than a packet has been lost. Since the receiving side can create a duplicate ACK only in the case that it receives an out-of-order packet, the dup ACK shows the sending side that one packet has been left out the network.

Thus, the sending side does not need to drastically decrease cwnd down to 1 and restart slow-start. Moreover, the sender can only decrease cwnd to one-half of the current cwnd and increase cwnd by 1 each time it receives a duplicate ACK.

The Fast-Recovery algorithm is implemented together with the Fast-Retransmit algorithm in the so-called Fast-Retransmit/Fast-Recovery algorithm

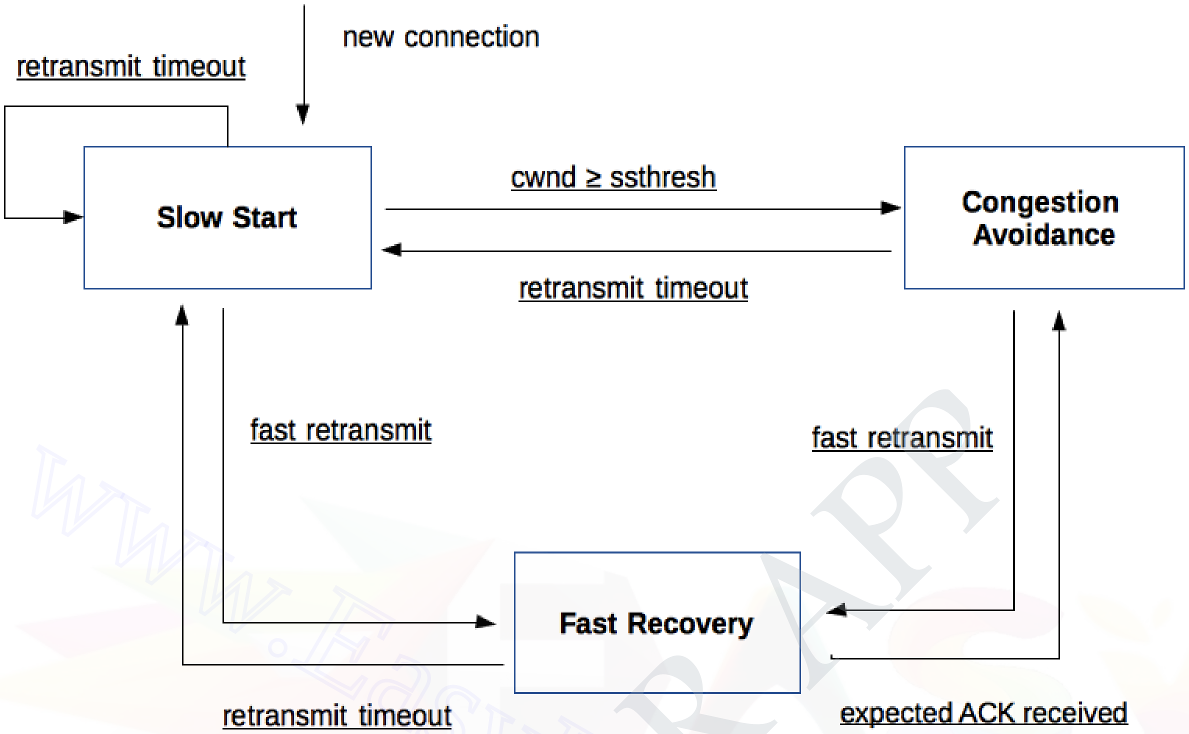


Fig. 3.5 Block Diagram

After receiving three duplicated ACKs in a row:

1. Set ssthresh to half the current send window.
2. Retransmit the missing segment
3. Set cwnd=sshtresh+3.
4. Each time the same duplicated ACK arrives, set cwnd++. Transmit a new packet, if allowed by cwnd.
5. If a non-duplicated ACK arrives, then set cwnd = ssthresh and continue with a linear increase of the cwnd

The following figure shows the fast recovery scenario mechanism and the timing relation with slow-start and congestion avoidance phases.

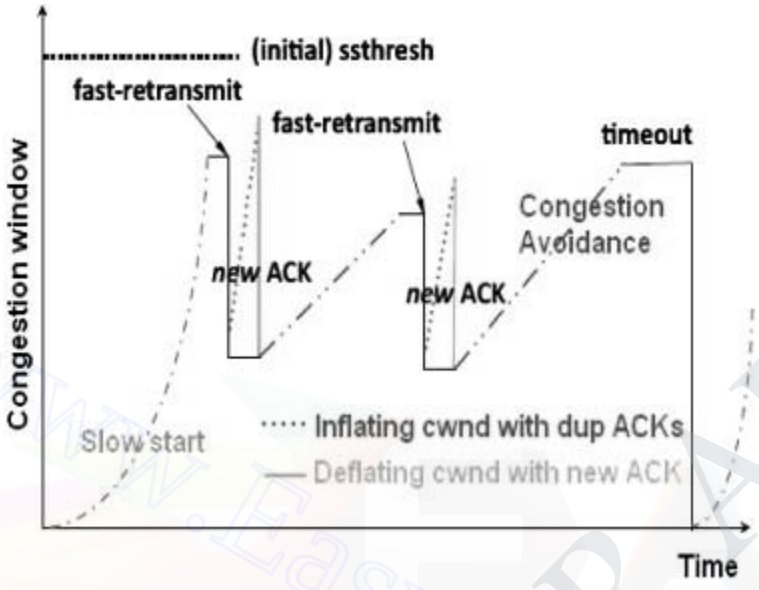


Fig. 3.6 Fast recovery mechanism.

3.1.5 IMPLICATIONS ON MOBILITY

Implications on mobility TCP assumes congestion if packets are dropped typically wrong in wireless networks, here packet loss are due to transmission errors furthermore, mobility itself can cause packet loss, for e.g. a mobile node roams from one access point (e.g. foreign agent in Mobile IP) to another while there are still packets in transit to the wrong access point and forwarding is not possible.

The performance of an unchanged TCP degrades severely however, TCP cannot be changed fundamentally due to the large base of installation in the fixed network, TCP for mobility has to remain compatible the basic TCP mechanisms keep the whole Internet together.

In fixed networksslow start is one of the useful mechanisms; it significantly decreases the efficiency of TCP if used together with mobile receivers or senders. The reason for this is the use of slow start under the wrong assumptions. From a missing acknowledgment, TCP concludes a congestion situation. While this may also happen in networks with mobile and wireless end-systems, it is not the main reason for packet loss.

Error rates on wireless links are orders of magnitude higher compared to fixed fiber or copper links. Packet loss is much more common and cannot always be compensated for by layer 2 retransmission (ARQ) or error correction.

Trying to retransmit on layer 2 could, for example, trigger TCP retransmission if it takes too long. Layer 2 now faces the problem of transmitting the same packet twice over a bad link. Detecting these duplicates on layer 2 is not an option, because more and more connections use end-to-end encryption, making it impossible to look at the packet.

Mobility itself can cause packet loss. There are many situations where a soft handover from one access point to another is not possible for a mobile end system. For example, when using mobile IP, there could still be some packets in transit to the old foreign agent while the mobile node moves to the new foreign agent. The old foreign agent may not be able to forward those packets to the new foreign agent or even buffer the packets if disconnection of the mobile node takes too long. This packet loss has nothing to do with wireless access but is caused by the problems of rerouting traffic.

The TCP mechanism detecting missing acknowledgements via time-outs and concluding packet loss due to congestion cannot distinguish between the different causes. This is a fundamental design problem in TCP: An error control mechanism (i.e. missing acknowledgement due to a transmission error) is misused forcongestion control (missing acknowledgement a due to network overload). In both cases packets are lost (either due to invalid checksums or to dropping in routers). However, the reasons are completely different. TCP cannot distinguish between these two different reasons.

Explicit congestion notification (ECN) mechanisms are currently recommended. Standard TCP reacts with slow start if acknowledgements are missing, which does not help in the case of transmission errors over wireless links and which does not really help during handover. This behavior results in a severe performance degradation of an unchanged TCP if used together with wireless links or mobile nodes.

However, one cannot change TCP completely just to support mobile users or wireless links. The same arguments that were given to keep IP unchanged also apply to TCP. The installed base of computers using TCP is too large to be changed and more important, mechanisms such as slow start keep the Internet operable. Every enhancement to TCP, therefore, has to remain compatible with the standard TCP and must not jeopardize the cautious behavior of TCP in case of congestion.

3.2 CLASSICAL TCP IMPROVEMENTS

Together with the introduction of WLANs in the mid-nineties several researchprojects were started with the goal to increase TCP's performance in wireless andmobile environments.

3.2.1 INDIRECT TCP

.

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

Standard TCP is used between the fixed computer and the access point. Any changes in the internet will not be intimated to the computer. The access point will terminates the standard TCP connection acting as a proxy instead of the mobile host. This means that the access point is used as the mobile host for the fixed host and as the fixed host for the mobile host. A special TCP is used between the access point and the mobile host.

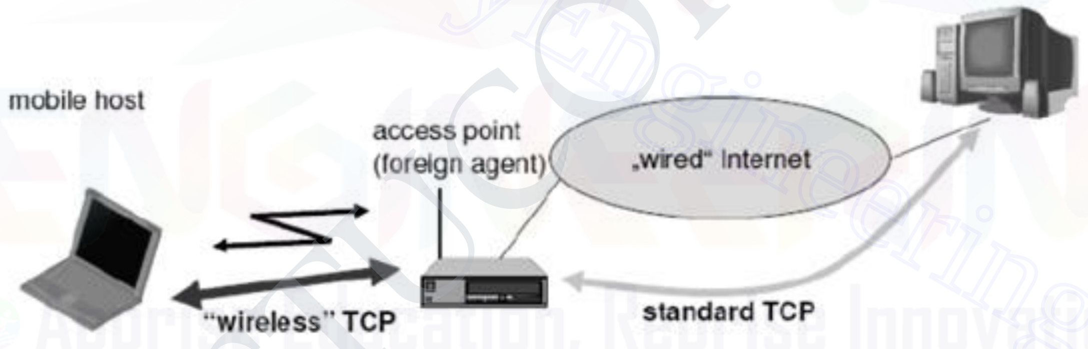


Fig. 3.7Mobile Host Connected Via a Wireless Link

Between the access pointand the mobile host, a special TCP, adapted to wireless links, is used. However,changing TCP for the wireless link is not a requirement. Even an unchangedTCP can benefit from the much shorter round trip time, starting retransmissionmuch faster. A good place for segmenting the connection between mobile hostandcorrespondent host is at the foreign agent of mobile IP.

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

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

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

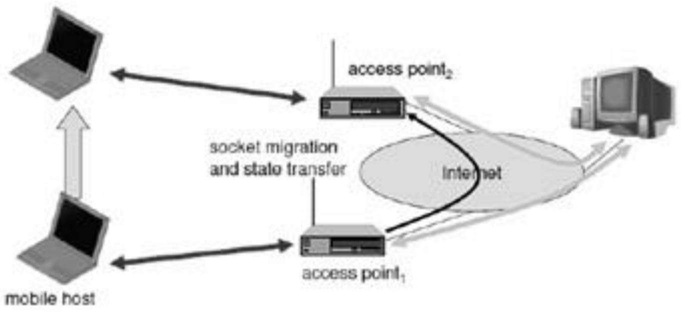
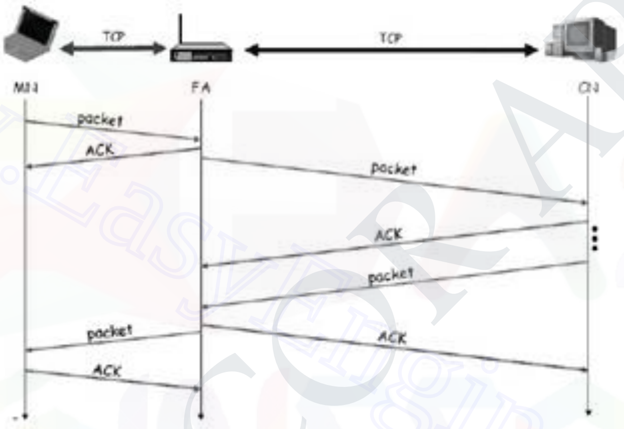


Fig. 3.8 **Packet delivery in I – TCP**



Advantages of I-TCP

- ❖ No changes in the fixed network necessary, no changes for the hosts necessary, all current optimizations to TCP still work.
- ❖ Simple to control, mobile TCP is used only for one hop between a foreign agent and a mobile host.
- ❖ Easy to use different protocols for wired and wireless networks.
- ❖ Transmission errors on the wireless link do not propagate in to the fixed network.

Disadvantages of TCP

- ❖ High latency is possible
- ❖ Loss of end to end semantics
- ❖ Security issues.

3.2.2 SNOOPING TCP

The main drawback of I – TCP is the segmentation of single TCP connection in to two TCP connections, which loses the original end – end TCP semantic. The main function is to buffer the data which is close to the mobile host to perform fast local transmission in case of packet loss.

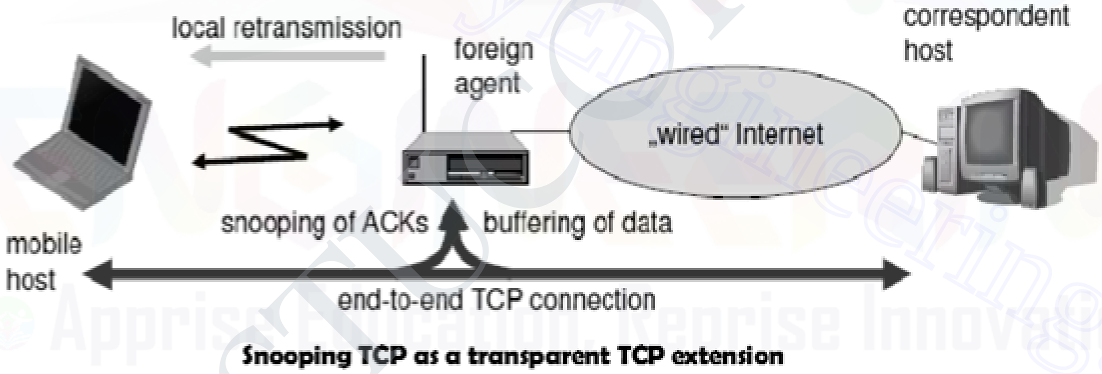


Fig. 3.9 Snooping TCP

In this snooping TCP, the foreign agent buffers all packets withdestination mobile hostandadditionally 'snoops' the packet flow in both directions to recognize acknowledgements. The reason for buffering packets towards the mobile node is to enable the foreign agent to perform a local retransmission in case of packet loss on the wireless link. Theforeign agent buffers every packet until it receives an acknowledgement from the mobile host.If the FA does not receive an acknowledgement from the mobile host within a certain amountof time, either the packet or the acknowledgement has been lost. Alternatively, the foreignagent could receive a duplicate ACK which also shows the loss of a packet. Now, the FAretransmits the packet directly from the buffer thus performing a faster retransmissioncompared to the CH. For transparency, the FA does not acknowledge data to the CH, whichwould violate end-to-end semantic in case of a FA failure.

The foreign agent can filter theduplicate acknowledgements to avoid unnecessary retransmissions of data from thecorrespondent host. If the foreign agent now crashes, the time-out of the correspondent hoststill works and triggers a retransmission.

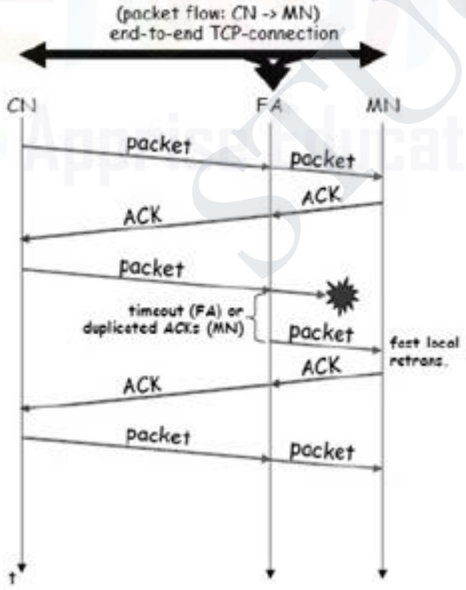


Fig. 3.10 Packet Flow CN – MN

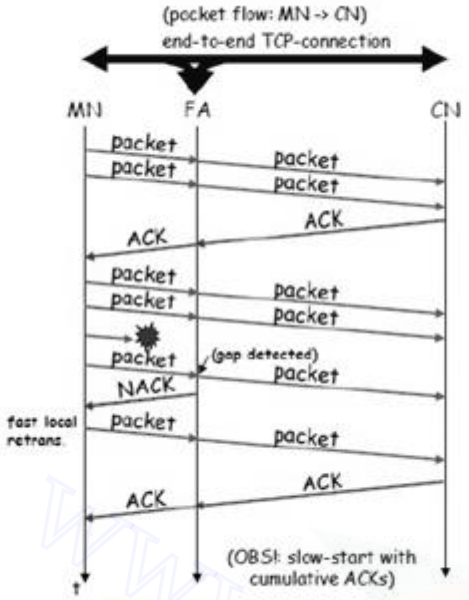


Fig. 3.11 Packet Flow MN – CN

The foreign agent may discard duplicates of packetsalready retransmitted locally and acknowledged by the mobile host. This avoids unnecessarytraffic on the wireless link.For data transfer from the mobile host withdestination correspondent host, the FAsnoops into the packet stream to detect gaps in the sequence numbers of TCP.

As soon as theforeign agent detects a missing packet, it returns a negative acknowledgement (NACK) to themobile host. The mobile host can now retransmit the missing packet immediately. Reorderingof packets is done automatically at the correspondent host by TCP.

Advantages of Snooping TCP:

- It maintains end – end semantics.
- No change is required to the CN.
- It does not need a handover of state
- No problem arises if the new foreign agent uses the enhancement or not.

Dis - Advantages of Snooping TCP:

- Snooping TCP does not isolate the behavior of the wireless link as well as I-TCP.
- It has to change to MH to handle NACKs.
- Snooping may become uselessif end to end encryption schemesare applied between CN and MH.

3.2.3 MOBILE TCP

The M – TCP has the same goals as that of I – TCP and snooping TCP to prevent the sender window from shrinking, if bit errors or disconnection cause current problems. M – TCP will improve the overall throughput by lowering the delay to maintain end – to – end semantics of TCP also to provide a more efficient handover.Additionally, M-TCP isespecially adapted to the problems arising from lengthy or frequent disconnections. M-TCP splits the TCP connection into two parts as I-TCP does.

.

An unmodifiedTCP is used on the standard host-supervisory host (SH) connection, whilean optimized TCP is used on the SH-MH connection. The supervisory host isresponsible for exchanging data between both parts similar to the proxy in ITCP.

The M-TCP approach assumes a relatively low bit error rateon the wireless link. Therefore, it does not perform caching/retransmission ofdata via the SH. If a packet is lost on the wireless link, it has to be retransmittedby the original sender. This maintains the TCP end-to-end semantics.

The SH monitors all packets sent to the MH and ACKs returned from theMH. If the SH does not receive an ACK for some time, it assumes that the MH isdisconnected. It then chokes the sender by setting the sender's window size to 0.

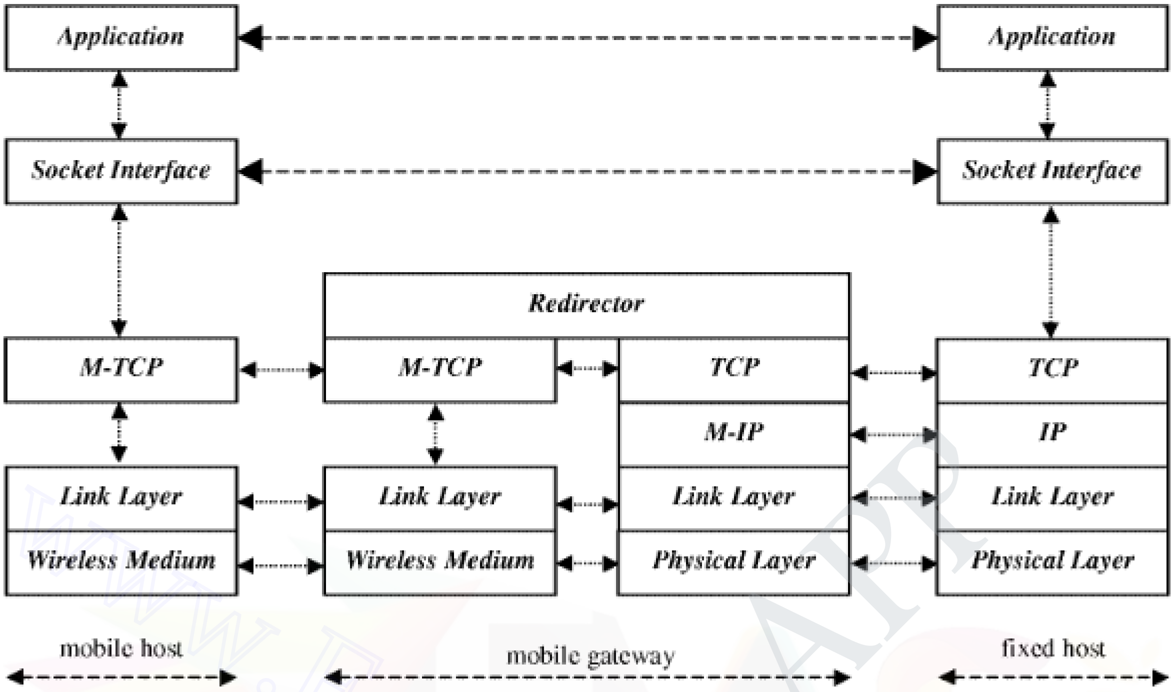
Setting the window size to 0 forces the sender to go into persistent mode, i.e.,the state of the sender will not change no matter how long the receiver is disconnected.

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 thewindow of the sender to the old value. The sender can continue sending at fullspeed. 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 abandwidth manager to implement fair sharing over the wireless link.

The design of MTCP resembles that of a link-layer protocol. Furthermore, since the protocol executes over a single hop, all packets arrive in order and all losses on the wireless segment are non-congestion related.

As a consequence, the design of MTCP can be greatly simplified and optimized, resulting in a significant reduction in processing and communication load over the wireless link. In addition, MTCP is tuned for quick recovery from losses on the wireless channel in order to improve the end-to-end TCP performance.



Mobile-TCP protocol stack

M – TCP Features

- **Elimination of IP processing**
- **Eliminate all congestion control mechanisms**
- **Loss-less header compression**
- **Optimized loss recovery techniques**

M- TCP advantages

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

● If the MH is disconnected, it avoids useless retransmissions, slow starts orbreaking 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 toforward buffers to a new SH. Lost packets will be automatically retransmittedto the new SH.

M- TCP Dis-advantages

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

●A modified TCP on the wireless link not only requires modifications tothe MH protocol software but also new network elements like thebandwidth manager.

3.2.4 TIME OUT – FREEZING

Time out freezing is used where the mobile node (MN) faces long durations of disconnection. While the approaches presented so far can handle short interruptions of the connection,either due to handover or transmission errors on the wireless link, somewere designed for longer interruptions of transmission.

Mobile hosts can be disconnected for a longer time because there is no packet exchange is possible, for e.g., in a tunnel, disconnection occursdue to overloaded cells or mux. With higher priority traffic, TCP disconnects after timeout completely

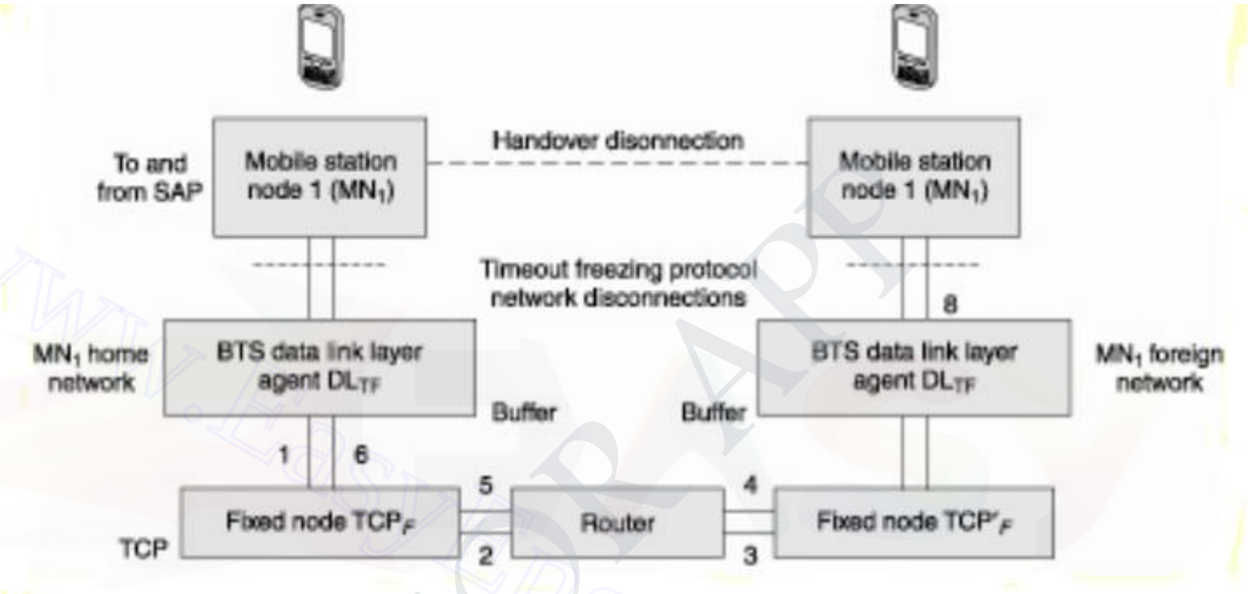

To and from SAP
Mobile station node 1 (MN_1)
Handover disconnection
Mobile station node 1 (MN_1)
Timeout freezing protocol network disconnections
8
MN_1 home network
BTS data link layer agent DL_{TF}
BTS data link layer agent DL_{TF}
MN_1 foreign network
Buffer
Buffer
1
6
5
4
Fixed node TCP_F
Router
Fixed node TCP'_F
TCP
2
3

Fig. 3.12 Time Out Freezing

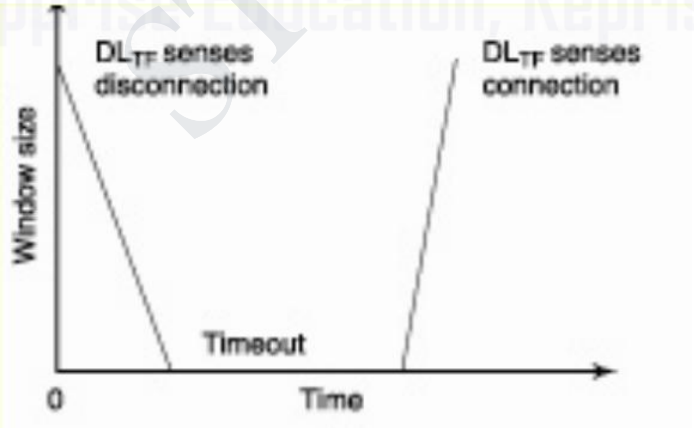


Fig. 3.13 Time out period of Packet Transmission

1. Agent DL_{TF}senses at data-link layer (MAC) the disconnection a little earlier than the TCP layer at the MN

• The TCP layer stops sending packets when disconnection is sensed by the DL_{TF}

• TCP layer presumes some congestion and, therefore, after a timeout the TCP layer freezes completely.

During the timeout period the MN may get some data sequences. After timeout, the TCP transmission freezes.

2. The TCP SYN and ACK data streams can still be received and transmitted through the lower layer with a suitable encoding by the header.

3. When the DL_{TF}agent senses the establishment of connection, it activates the TCP transmission.

TCP freezing

- MAC layer is often able to detect interruption in advance
- MAC can inform TCP layer of upcoming loss of connection
- TCP stops sending, but does now not assume a congested link
- MAC layer signals again if reconnected.

MAC layer will notice the connection problems even before the connection is actually interrupted from a TCP point of view and also identifies the reason for the interruption. The MAC layer can inform the TCP layer of an upcoming loss of connection or that the current interruption is not caused by congestion.

TCP can now stop sending and freezes the current state of its congestion window and further timers. Both the mobile and the correspondent host can be informed about the upcoming interruption which is early noticed by the MAC layer. The reason for the interruption is informed to the correspondent host by means of fast interruption in the wireless links and additional mechanisms.

The correspondent host can goes in to slow start by assuming the congestion and finally breaks the congestion.

Advantages:

1. It offers a way to resume TCP connectionseven after longer interruptions of the connection.
2. It can be used together with encrypted data as it is independent of any other TCP mechanism, such as acknowledgements or sequence numbers.

Dis – Advantages:

1. Freezing the state of TCP does not help in case of some encryption schemes that use time-dependent random numbers.
2. These schemesneed resynchronization after interruption.
3. TCP on mobile host has to be changed, mechanism depends on MAC layer.

3.2.5 SELECTIVE RETRANSMISSION

The standard TCP uses a cumulative acknowledgment scheme, it often does not provide the sender with sufficient information to recover quickly from multiple packet losses within a single transmission window. The TCP can be enhanced with selective retransmission so that it can perform better than standard TCP.

Selective retransmission is the very useful extension of TCP. TCP acknowledgements are often cumulative

- ACK n acknowledges correct and in-sequence receipt of packets up to n packets.
- If single packets are missing quite often a whole packet sequence beginning at the gap has to be retransmitted (go-back-n), thus wasting bandwidth.

Selective retransmission is the solution for using RFC2018 allows for acknowledgements of single packets, not only acknowledgements of in-sequence packet streams without gaps. The receiver can acknowledge single packets, not only trainsof in-sequence packets. The sender can now determine precisely which packet isneeded and can retransmit it.

Advantages:

1. The sender retransmits only the lost packets.
2. Bandwidth requirement is low.
3. Higher efficiency
4. Selective retransmission can be beneficial to all other networks.

Dis – Advantages:

1. More complex software in a receiver, more buffers needed at the receiver.
2. Memory sizes andCPU performance permanently increase the bandwidth of the air interface.

3.2.6 TRANSACTION ORIENTED TCP

T/TCP (Transactional Transmission Control Protocol) is a variant of the Transmission Control Protocol (TCP). It is an experimental TCP extension for efficient transaction-oriented (request/response) service. It was developed to fill the gap between TCP and UDP.

T/TCP is an experimental extension for the TCP protocol. It was designed to address the need for a transaction-based transport protocol in the TCP/IP stack.T/TCP lies between these two protocols, making it an alternative for certain applications.

TCP for Transactions (T/TCP) is a possible successor to both TCP and UDP. It is a transaction-oriented protocol based on a minimum transfer of segments, so it does not have the speed problems associated with TCP. By building on TCP, it does not have the unreliability problems associated with UDP. T/TCP can be considered a superset of the TCP protocol. The reason for this is that T/TCP is designed to work with current TCP machines seamlessly.

The absolute minimum number of packets required in a transaction is two: one request followed by one response. T/TCP has the reliability of TCP and comes very close to realizing the 2-packet exchange (three in fact). T/TCP uses the TCP state model for its timing and retransmission of data, but introduces a new mechanism to allow the reduction in packets.

Three-way handshake

- T/TCP extension avoids 3-way handshakes
- Request/reply data sent with connection messages
- Server caches a connection count (CC) per-client to detect duplicate requests and avoid replaying transaction
- TIME_WAIT is shortened by setting it to 8*RTO
- Latency = RTT + server processing time (SPT)

T/TCP has the reliability of TCP and comes very close to realizing the 2-packet exchange (three in fact). T/TCP uses the TCP state model for its timing and retransmission of data, but introduces a new mechanism to allow the reduction in packets.

Even though three packets are sent using T/TCP, the data is carried on the first two, thus allowing the applications to see the data with the same speed as UDP. The third packet is the acknowledgment to the server by the client that it has received the data, which is how the TCP reliability is incorporated.

Using TCP now requires several packets over the wireless link. First, TCPuses a three-way handshake to establish the connection. At least one additionalpacket is usually needed for transmission of the request, and requires three morepackets to close the connection via a three-way handshake.

Assuming connectionswith a lot of traffic or with a long duration, this overhead is minimal. Butin an example of only one data packet, TCP may need seven packets altogether.Web services are based on HTTP which requires a reliabletransport system. In the internet, TCP is used for this purpose.

Before a HTTP request can be transmitted the TCP connection has to be established. Thisalready requires three messages. If GPRS is used as wide area transport system,one-way delays of 500 ms and more are quite common. The setup of a TCP connectionalready takes far more than a second.

T/TCP can combine packets for connection establishmentand connection release with user data packets. This can reduce the number ofpackets down to two instead of seven.

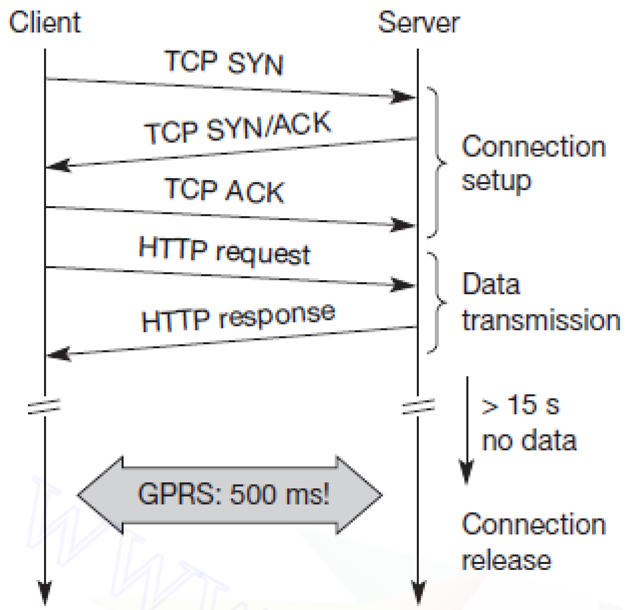


Fig. 3.14 Transaction Oriented TCP

Table 3.2 Comparison of different approaches for "mobile" TCP

Approach	Mechanism	Advantages	Disadvantages
Indirect TCP	splits TCP connection into two connections	isolation of wireless link, simple	loss of TCP semantics, higher latency at handover
Snooping TCP	"snoops" data and acknowledgements, local retransmission	transparent for end-to-end connection, MAC integration possible	problematic with encryption, bad isolation of wireless link
M-TCP	splits TCP connection, chokes sender via window size	Maintains end-to-end semantics, handles long term and frequent disconnections	Bad isolation of wireless link, processing overhead due to bandwidth management
Fast retransmit/ fast recovery	avoids slow-start after roaming	simple and efficient	mixed layers, not transparent
Transmission/ time-out freezing	freezes TCP state at disconnect, resumes after reconnection	independent of content or encryption, works for longer interrupts	changes in TCP required, MAC dependant
Selective retransmission	retransmit only lost data	very efficient	slightly more complex receiver software, more buffer needed
Transaction oriented TCP	combine connection setup/release and data transmission	Efficient for certain applications	changes in TCP required, not transparent

3.3 TCP OVER 3G WIRELESS NETWORKS

Transmission control protocol (TCP) is one of the most widely used transport layer protocols in Internet. Much development and deployment activity has centered on GPRS, UMTS and IMT-2000, also referred to 2.5G/3G wireless networks. However, TCP has not been designed bearing in mind wireless networks. Especially, flow control features can perform less than optimally over wireless interfaces. A number of TCP optimization techniques have been studied to enhance the TCP performance for various wireless environments.

1. **Large window size**

   The traditional TCP specification limits the window size to 64 KB. If the end-to-end capacity is expected to be larger than 64 KB,the window scale option can overcome that limitation. TCP over2.5G/3G should support appropriate window sizes based on theBandwidth Delay Product (BDP) of the end-to-end path. If theestimated path BDP is larger than 64 KB, the window scale option maybe used.

2. **Initial Window Size:**

   The initial TCP window size specifies how many segments that the initiator of a TCP connection sends before waiting for the acknowledgements. Original TCP uses an initial window size of one segment. Using a large initial window size reduces idle times in the beginning of the transmissions speeding up the transmission of small amount of data such as email and web page transmission. Because of larger window size leads to packet loss and poor performance in the congested networks.

3. **Data rates:** Data rates of 2.5G systems are 10–20 kbit/s uplink and 20–50 kbit/s downlink, 3G and future 2.5G systems will initially offer data rates around 64 kbit/s uplink and 115–384 kbit/s downlink.Typically, data rates are asymmetric as it is expected that users will download more data compared to uploading.

   Uploading is limited by the limited battery power. In cellular networks, asymmetry does not exceed 3–6 times, however, considering broadcast systems as additional distribution media (digital radio, satellite systems), asymmetry may reach a factor of 1,000. Serious problems that may reduce throughput dramatically are bandwidth oscillations due to dynamic resource sharing.

4. **Latency**

   The latency of 2.5G/3G links is high mostly due to the extensive processing required at the physical layer of those networks, e.g.,Forward error correction(FEC) and interleaving, and due to transmission delays in the radio access network (including link-level retransmissions).

.

A typical RTT varies between a few hundred milliseconds and one second.The associated radio channels suffer from difficult propagationenvironments. Hence, powerful but complex physical layer techniques need to be applied to provide high capacity in a wide coverage areain a resource efficient way. The rapid improvements in allareas of wireless networks ranging from radio layer techniques oversignal processing to system architecture will ultimately also lead toreduced delays in 3G wireless systems.

5. **Jitter:** Wireless systems suffer from large delay variations or 'delay spikes'. The Reasons for sudden increase in the latency are: link outages due to temporal loss of radio coverage, blocking due to high-priority traffic, or handovers. Handovers are quite often only with outages reaching from some 10 ms (handover in GSM systems) to several seconds (intersystem handover, e.g., from a WLAN to a cellular system using Mobile IP without using additional mechanisms such as multicasting data to multiple access points).

6. **Packet Loss Due to Corruption**

Packets might be lost during handovers or due to corruption.However, recovery at the link layer appears as jitterto the higher layers. Link layer ARQ and FECcan provide a packet service with a negligibly small probability ofundetected errors (failures of the link CRC), and a low level of loss(non-delivery) for the upper layer traffic, e.g., IP. The loss rateof IP packets is low due to the ARQ, but the recovery at the linklayer appears as delay jitter to the higher layers lengthening thecomputed RTO value.

7. **Asymmetry:** 2.5G/3G systems may run asymmetric uplink and downlink data rates. The uplink data rate is limited by battery power consumption and complexity limitations of mobile terminals. However, the asymmetry does not exceed 3-6 times, and can be tolerated by TCP without the need for techniques like ACK congestion control or ACK filtering.

8. **Selective Acknowledgements:**

Original TCP acknowledgements contain the number of the next segments expected to arrive. In case one or more segment is lost, the receiver will include the number of the first lost segment in the acknowledgement sent upon reception of new packets. If the segment received after the lost segments fit in the recipient’s window they will be accepted. However the sender will not know which packets arrived after the lost packet and will either have to retransmit just one segment and then wait for the acknowledgement to find out which segment to send next or retransmit several segments that actually might have reached the recipients already.

UNIT IV-Wide-Area Wireless Networks (WANs)

4.1 INTRODUCTION

The Third Generation (3G) wireless systems offer services and thereby reduce the distinction between the range of services ofwire line and wireless. It is an advanced technology and it enhances the features of second generation and adds its own advanced features. Updating cellular telecommunications network around the world are using 3G technologies.

The main reason for the evolution of 3G was due to the limited capacity of the 2G networks.

2G networks were built for voice calls and slow data transmission. But these services were unable to satisfy the requirements of present wireless revolution. International Telecommunication Union (ITU) has defined the demand for 3G in the International Mobile Telecommunication (IMT)-2000 standards to facilitate growth, increase bandwidth, support diverse applications.

The development like 2.5G or GPRS (General Packet Radio Service) and 2.75G or EDGE (Enhanced Data rates for GSM Evolution) technologies resulted in the transition to 3G. These technologies act like bridge between 2G and 3G.

4.1.1Features of 3G

It provides cost efficient high quality, wireless multimedia applications and enhanced wireless communications.

It supports greater voice and data capacity and high data transmission at low cost. 3G mobiles can operate on 2G and 3G technologies.

It offers greater security features than 2G. It supports network access security, network domain security, user domain security, application security.

It supports video calls and video conferences. It provides support from localized service like accessing traffic and high end services like weather updates. We can listen to music, watch videos online and can download huge files with in less time.

4.1.2 Advantages of 3G

All the functions in a normal 2G mobile devices can be performed in 3G at a higher speed.

It provides faster connectivity, faster internet access and music with improved quality.

4.1.3 Applications of 3G

- The 3G mobile can be used as a modem for computer which can access internet and can download games and songs at high speed.
- It provides high quality voice calls and video calls.
- It provides weather updates, news headlines and TV broadcasting in mobile phone.
- It provides high speed internet facility for many applications. It can provide data transmission speed upto 2Mbits /sec.
- It provides multimedia services such as sharing of digital photos and movies. It provides location based services and real time multi player gaming.
- It supports virtual banking and online selling.
- It supports teleconferencing.

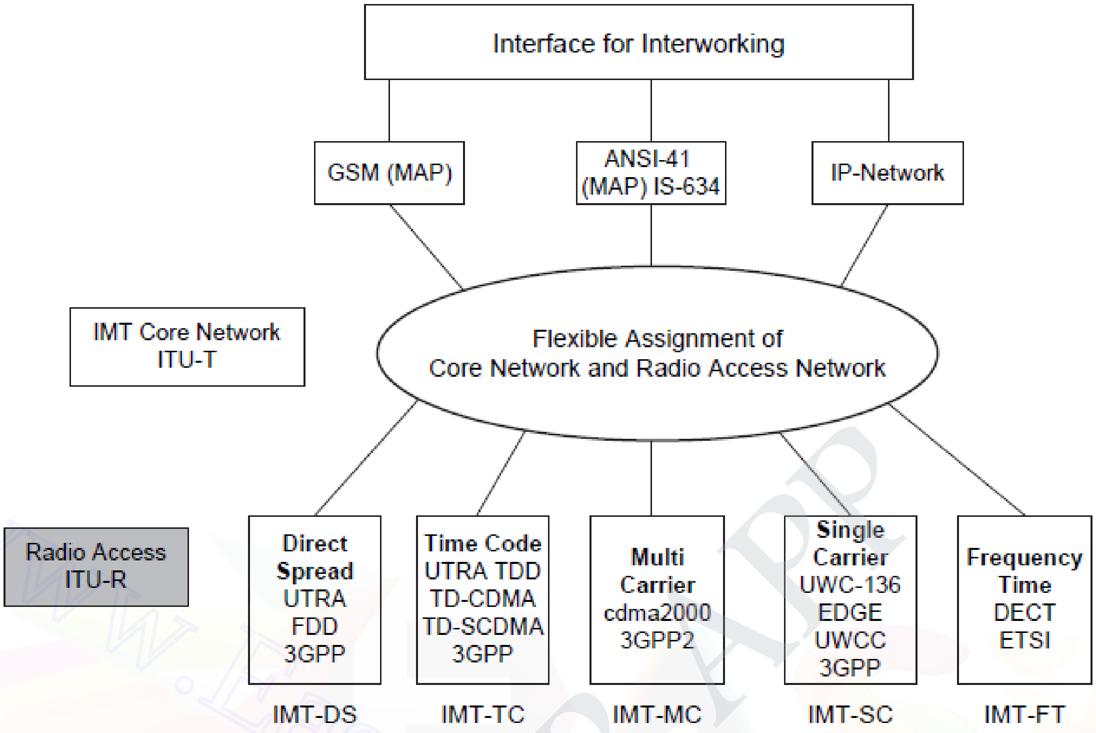
4.1.4 Drawbacks

There are few drawbacks:

- Upgrading the base station and cellular infrastructure to 3G incurs very high costs.
- Service provider has to pay high amount for 3G licensing and agreements.
- Problem with the availability of handsets and few regions and their costs.
- High power consumption.

4.2 IMT Family

The International Telecommunication Union (ITU) identified the long-term spectrumrequirements for the future third-generation (3G) mobile wireless telecommunicationssystems. In 1992, the ITU identified 230 MHz of spectrum in the 2 GHzband to implement the IMT (International Mobile Telecommunications)-2000 system on a worldwide basis for satellite and terrestrial components. The aim of IMT-2000 is to provide universal coverageenabling terminals to have seamless roaming across multiple networks. The ITUaccepted the overall standardization responsibility of IMT-2000 to define radiointerfaces that are applicable in different radio environments including indoor,outdoor, terrestrial, and satellite.


Interface for Interworking
GSM (MAP)
ANSI-41 (MAP) IS-634
IP-Network
IMT Core Network ITU-T
Flexible Assignment of Core Network and Radio Access Network
Radio Access ITU-R
Direct Spread UTRA FDD 3GPP
Time Code UTRA TDD TD-CDMA TD-SCDMA 3GPP
Multi Carrier cdma2000 3GPP2
Single Carrier UWC-136 EDGE UWCC 3GPP
Frequency Time DECT ETSI
IMT-DS
IMT-TC
IMT-MC
IMT-SC
IMT-FT

ITU: International Telecommunication Union
MAP: Mobile Application Part
FDD: Frequency Division Duplex
3GPP: Third-Generation Partnership Project

Fig. 4.1 IMT Family

The above figure provides an overview of the IMT family. IMT-DS is the directspread (DS) technology and includes WCDMA systems. This technology isintended for UMTS terrestrial radio access (UTRA)-FDD and is used in Europeand Japan.

IMT-TC family members are the UTRA-TDD system that uses timedivision (TD) CDMA, and the Chinese TD-synchronous CDMA (TD-SCDMA).Both standards are combined and the third-generation partnership project (3GPP)is responsible for the development of the technology. IMT-MC includes multiplecarrier (MC) cdma2000 technology, an evolution of the cdmaone family.

3GPP2is responsible for standardization. IMT-SC is the enhancement of the US TDMAsystems. UWC-136 is a single carrier (SC) technology. This technology appliesEDGE to enhance the 2 G IS-136 standards. It is now integrated into the 3GPPefforts. IMT-FT is a frequency time (FT) technology. An enhanced version of thecordless telephone standard digital European cordless technology (DECT) has been selected for low mobility applications. The ETSI has the responsibility forstandardization of DECT.

In Europe, 3G systems are intended to support a substantially wider andenhanced range of services compared to the 2G (GSM) system. These enhancementsinclude multimedia services, access to the Internet, high rate data, and soon. The enhanced services impose additional requirements on the fixed networkfunctions to support mobility. These requirements are achieved through an evolutionpath to capitalize on the investments for the 2G system in Europe, Japan, andNorth America.

In North America, the 3G wireless telecommunication system, cdma2000was proposed to ITU to meet most of the IMT requirements in the indoor office,indoor to outdoor pedestrian, and vehicular environment. In addition, thecdma2000 satisfies the requirements for 3G evolution of 2G TIA/EIA 95 familyof standards (cdmaOne).

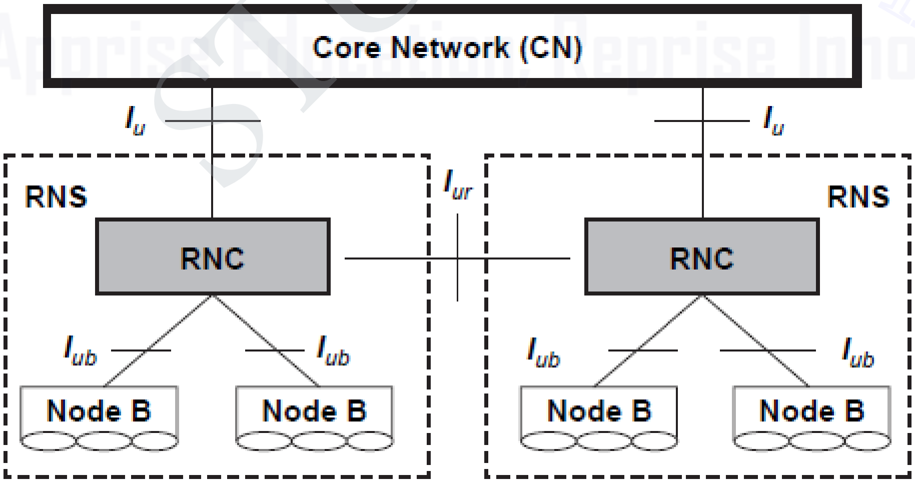
In Japan, evolution of the GSM platform is planned for the IMT (3G) corenetwork due to its flexibility and widespread use around the world. Smooth migrationfrom GSM to IMT-2000 is possible. The service area of the 3G system overlayswith the existing 2G (PDC) system. The 3G system connects and interworkswith 2G systems through an interworking function (IWF). An IMT-2000-PDC dual mode terminal as well as the IMT-2000 single mode terminal is deployed.

UMTS as discussed today and introduced in many countries is based on theinitial release of UMTS standards referred to as release 99 or R99. This (release)is aimed at a cost-effective migration from GSM to UMTS. After R99 the releaseof 2000 or R00 followed. 3GPP decided to split R00 into two standards and callthem release 4 (Rel-4) and release 5 (Rel-5). The version of all standards finalizedfor R99 is now referred to as Rel-3 by 3GPP. Rel-4 introduces QoS in thefixed network plus several execution environments (e.g., MExE, mobile executionenvironment) and new service architectures. Rel-4 was suspended in March 2001.

Rel-5 specifies a new core network. The GSM/GPRS-based core network will bereplaced by an almost all-IP core network. The content of Rel-5 was suspended

4.3 UMTS TERRESTRIAL RADIO ACCESS NETWORK OVERVIEW

The UTRAN (Universal Mobile Telecommunications System**)** consists of a set of radio network subsystems (RNSs). There are two logical elements in RNS. One is node B and another is RNC.



RNC: Radio Network Controller
RNS: Radio Network Subsystem

.

Fig.4.2 UTRAN Logical Architecture

Each cell consists of one group of nodes and one RNC (Radio Network Controller). The RNC is responsible for the use and allocation of all the radio resources of the RNS.

4.3.1The responsibilities of RNC

This element of the UTRAN / radio network subsystem controls the Node Bs whichis connected to it, i.e. the radio resources of the domain. The RNC is responsible for the radio resource management and some of the mobility management functions. It is responsible for data encryption / decryption.

a. Intra UTRAN handover
b. Macro diversity combining/splitting of I_{ub} data streams
c. Frame synchronization
d. Radio resource management
e. Outer loop power control
f. I_u interface user plane setup
g. Serving RNS (SRNS) relocation
h. Radio resource allocation (allocation of codes, etc.)
i. Frame selection/distribution function necessary for soft handover
j. UMTS radio link control (RLC) sub layers function execution.
k. Termination of MAC, RLC, and RRC protocols for transport channels,i.e., DCH, DSCH, RACH, FACHIub's user plane protocols termination.

4.3.2The Node B architecture and responsibilities:

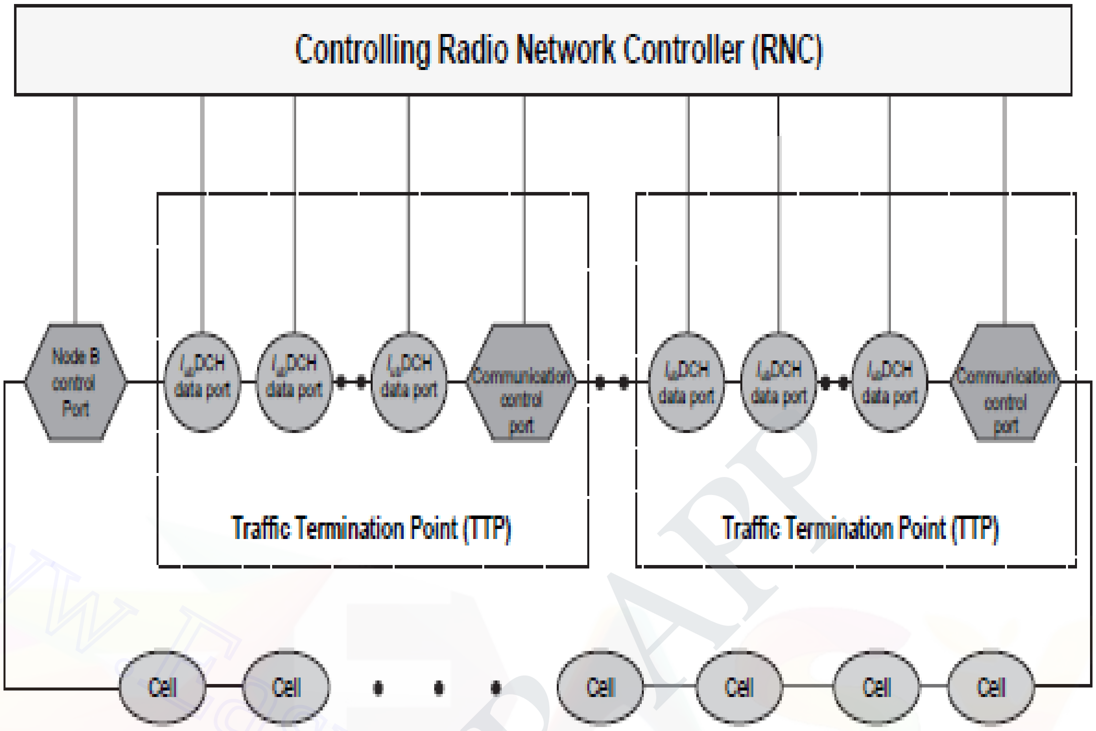


Fig.4.3 Node B logical Architecture

A Node B is responsible for radio transmission and reception in one or more cells to/from the user equipment (UE).

Node B denotes the base station transceiver within UMTS. It contains the transmitter and receiver to communicate with the UEs within the cell. It participates with the RNC in the resource management. NodeB is the 3GPP term for base station, and often the terms are used interchangeably.

The following are the responsibilities of the Node B:

Termination of I_{ub} interface from RNC

Termination of MAC protocol for transport channels RACH, FACH

Termination of MAC, RLC, and RRC protocols for transport channels:BCH, PCH

Radio environment survey (BER estimate, receiving signal strength, etc.)

.

Inner loop power control

Open loop power control

Radio channel coding/decoding

Macro diversity combining/splitting of data streams from its cells (sectors)

Termination of U_u interface from UE

Error detection on transport channels and indication to higher layers

FEC encoding/decoding and interleaving/deinterleaving of transport channels

Multiplexing of transport channels and demultiplexing of coded composite transport channels

Power weighting and combining of physical channels

Modulation and spreading/demodulation and despreading of physical channels

Frequency and time (chip, bit, slot, frame) synchronizationRF processing.

4.3.3 UTRAN Logical Interfaces

The UTRAN protocol structure contains two main layers

The radio network layer(RNL)

The transport network layer (TNL)

Control Plane: It is used for all UMTS- specific signaling. It includes the application protocol (i.e., radio access network application part (RANAP) in I_u, radio network subsystem application part (RNSAP) in I_ur and node B applicationpart (NBAP) in I_{ub}).

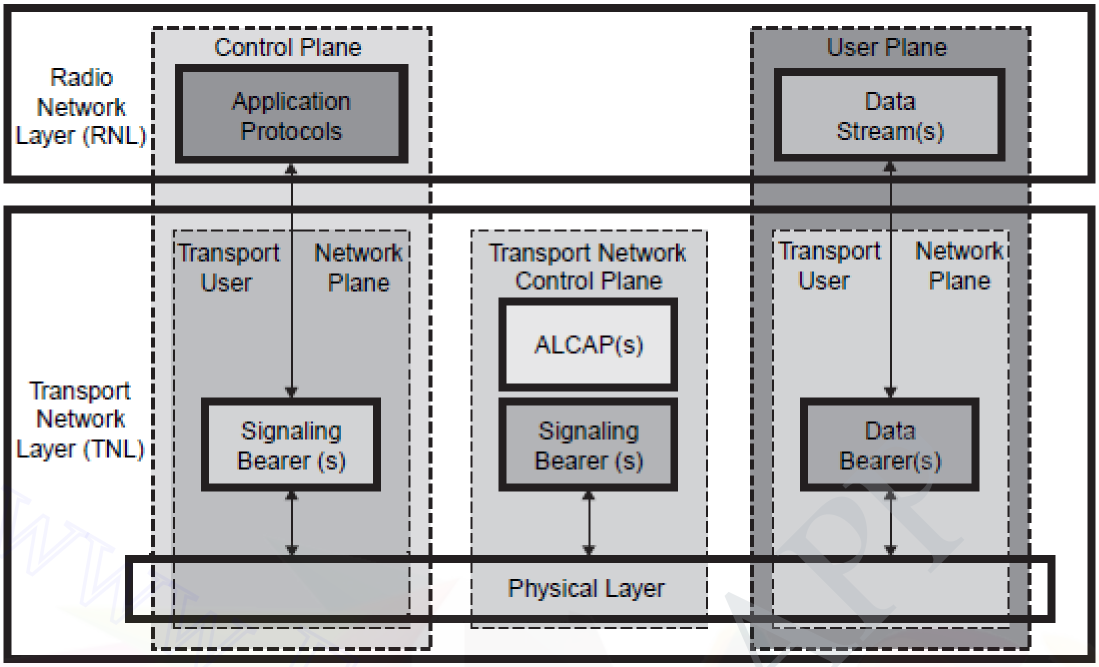

Control Plane
User Plane
Radio Network Layer (RNL)
Application Protocols
Data Stream(s)
Transport Network Layer (TNL)
Transport User Network Plane
Transport Network Control Plane
Transport User Network Plane
ALCAP(s)
Signaling Bearer (s)
Signaling Bearer (s)
Data Bearer(s)
Physical Layer

Fig.4.4 General protocol model for UTRAN interfaces

User Plane:

The user plane carries the user information.It includes data streams and data bearers for data streams.

Transport network control plane:

It carries all control signaling. It contains access link control application part (ALCAP) required to set up the transport bearers (data bearers) for the user plane. It also includes the signaling bearer needed for the ALCAP. The transport plane lies between the control plane and the user plane. The addition of the transport plane in UTRAN allows the application protocol in the radio network control plane to be totally independent of the technology selected for the data bearer in the user plane.

4.3.3.1 lu Interface

The UMTS Iuinterface connects the UTRAN to the UMTS core network (UCN). It consists of three planes.

1. Radio network control plane:

It carries information for the general control of UTRAN radio network operations.

It carries information for control of UTRAN in the context of each specific call.

It carries user call control (CC) and mobility management (MM) signaling messages.

2. The transport network control plane (TNCP):

It carries information for the control of transport network used within UCN.

3. User plane (UP):

It carries user voice and packet data information.

AAL2 is used for the following services: narrowband speech (e.g., EFR, AMR); unrestricted digital information service (up to 64 kbps, i.e., ISDN B channel); any low to average bit rate CS service (e.g., modem service to/from PSTN/ISDN). AAL5 is used for the following services: non-real-time PS data service (i.e., best effort packet access) and real-time PS data.

4.3.3.2 I_{ur} Interface

The I_{ur} interface allows communication between different RNCs within the UTRAN. The open I_{ur} interface enables capabilities like soft handover to occur as well as helping to stimulate competition between equipment manufacturers.

Two different protocol planes are defined for it:

Radio network control plane (RNCP)

Transport network control plane (TNCP)

User plane (UP)

The I_{ur} interface is used to carry:

Information to control the radio resources in the context of specific service request of one mobile on RNCP

Information to control the transport network used within UTRAN on TNCP

User voice and packet data information on UP

The protocols used on this interface are:

Radio access network application part (RANAP)

RANAP signalling protocol resides in the control plane of Radio network layer of Iu interface in the UMTS (Universal Mobile Telecommunication System) protocol stack. Iu interface is the interface between RNC (Radio Network Controller) and CN (Core Network).

DCH frame protocol (DCHFP)

The data transfer takes place using a frame protocol. The procedures belonging to this set include establishment, modification and release of dedicated channel in the DRNC due to hard and soft handover,set-up/release of dedicated transport connections over Iur interface and data transfer for dedicated channels.

RACH frame protocol (RACHFP)

A random-access channel (RACH) is a shared channel used by wireless terminals to access the mobile network (TDMA/FDMA, and CDMA based network) for call set-up and burst data transmission. Whenever mobile wants to make a MO call it schedules the RACH. RACH is transport-layer channel.

FACH frame protocol (FACHFP)

Forward Access Channel

Access link control application part (ALCAP)

Control plane protocol for the transport layer. It is used for multiplexing of different users onto one AAL2 transmission path using channel IDs (CIDs).

Signaling connection control part (SCCP)

Anetwork layer protocol that provides extended routing, flow control, segmentation, connection-orientation, and error correction facilities in Signaling System & telecommunications networks.

Message transfer part 3-B (MTP3-B)

Signaling ATM adaptation layer for network-to-network interface (SAALNNI) (SAAL-NNI is further divided into service specific coordinationfunction for network to network interface (SSCF-NNI), service specific connection oriented protocol (SSCOP), and ATM adaptation layer 5 (AAL5))

Basic inter-RNC mobility support

Support of SRNC relocation

Support of inter-RNC cell and UTRAN registration area update

Support of inter-RNC packet paging

Reporting of protocol errors

Dedicated channel traffic support

Establishment, modification, and release of a dedicated channel in the DRNC due to hard and soft handoff in the dedicated channel state

Setup and release of dedicated transport connections across the Iur interface

Transfer of DCH transport blocks between SRNC and DRNC

Management of radio links in the DRNS via dedicated measurement report procedures and power setting procedures

Common channel traffic support

.

Setup and release of the transport connection across the Iur for common channel data streams

Splitting of the MAC layer between the SRNC (MAC-d) and DRNC (MAC-c and MAC-sh); the scheduling for downlink data transmission is performed in the DRNC

Flow control between the MAC-d and MAC-c/MAC-sh

Global resource management support

Transfer of cell measurements between two RNCs

Transfer of Node B timing between two RNCs

4.3.3.3 I_{ub} Interface

The I_{ub} connects the NodeB and the RNC within the UTRAN. Although when it was launched, a standardization of the interface between the controller and base station in the UTRAN was revolutionary, the aim was to stimulate competition between suppliers, allowing opportunities like some manufacturers who might concentrate just on base stations rather than the controller and other network entities.

Three different protocol planes are defined for it.

Radio network control plane (RNCP)

Transport network control plane (TNCP)

User plane (UP)

The I_{ub} interface is used to carry the information for the general control of Node B for radio network operation on RNCPInformation for the control of radio resources in the context of specific service request of one mobile on

.

RNCPInformation for the control of a transport network used withinUTRANon TCNPUser CC and MM signaling message on RNCP.

UTRA uplink & downlink

At the radio air interface and its associated properties, it is necessary to define the directions in which the transmissions are occurring. Being a full duplex system, i.e. transmitting simultaneously in both directions, it is necessary to be able to define which direction is which.

- Uplink; This may also sometimes be known as the reverse link, and it is the link from the User Equipment (UE) to the Node B or base station.
- Downlink; This may also sometimes be known as the forward link, and it is the link from the Node B or base station to the User Equipment (UE).

UTRA FDD & TDD

In view of the fact that transmissions have to be made in both directions, i.e. in both uplink and downlink. It is necessary to organize the way these transmissions are made. Two techniques are used to ensure concurrent or near concurrent transmissions in both directions: frequency division duplex and time division duplex.

UTRA-FDD: The frequency division duplex version of UTRA uses a scheme whereby transmissions in the uplink and downlink occur on different frequencies. Although this requires double the bandwidth to accommodate the two transmissions, and filters to prevent the transmitted signal from interfering with the receiver. Even though there is a defined split between uplink and downlink, effective filters are required.

.

UTRA-TDD: The time division version of the UTRA uses uplink and downlink transmissions that use the same frequency but are timed to occur at different intervals.

Distribution of UTRAN Functions

Located in the RNC

Radio resource control (L3 Function)

Radio link control (RLC)

Macro diversity combining

Active cell set modification

Assign transport format combination set (centralized data base function)

Multiplexing/demultiplexing of higher layer PDUs into/from transportblock delivered to/from the physical layer on shared dedicated transportchannels (used for soft handover)

L1 function: macro diversity distribution/combining (centralized multipointtermination)

Selection of the appropriate transport format for each transport channel depending upon the instantaneous source rate — collocate with RRCPriority handling between data flows of one user.

Located in Node B

Scheduling of broadcast, paging, and notification messages; location inNode B — to reduce data repetition over I_{ub} and reduce RNC CPU load and memory space

Collision resolution on RACH (in Node B — to reduce nonconstructiveTrafficover I_{ub} interface and reduce round trip delay)

Multiplexing/demultiplexing of higher layer PDUs to/from transport blocksdelivered to/from the physical layer on common transport channels

4.4 UMTS CORE NETWORK ARCHITECTURE

The UMTS network architecture can be divided into three main elements:

1. **User Equipment (UE):** The User Equipment or UE is the name given to what was previous termed the mobile, or cellphone. The new name was chosen because the considerably greater functionality that the UE could have. It could also be anything between a mobile phone used for talking to a data terminal attached to a computer with no voice capability.
2. **Radio Network Subsystem (RNS):** The RNS also known as the UMTS Radio Access Network, UTRAN, is the equivalent of the previous Base Station Subsystem or BSS in GSM. It provides and manages the air interface forthe overall network.
3. **Core Network:** The core network provides all the central processing and management for the system. It is the equivalent of the GSM Network Switching Subsystem or NSS.

The core network is then the overall entity that interfaces to external networks including the public phone network and other cellular telecommunications networks.

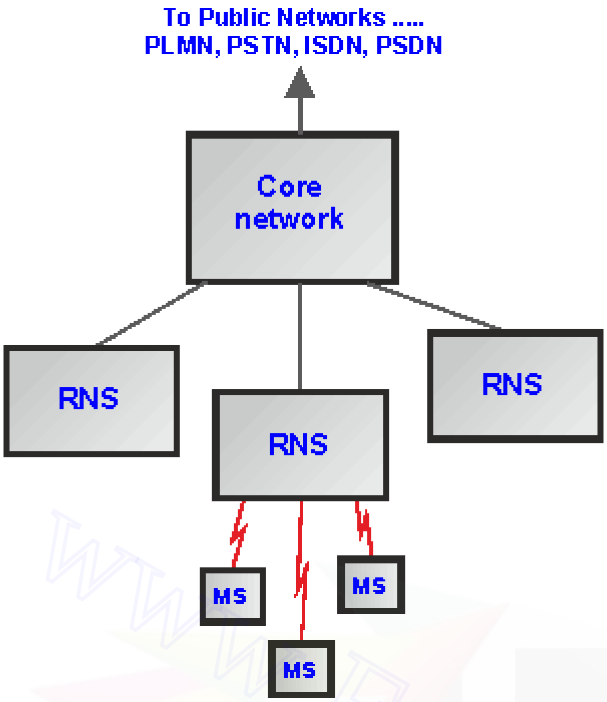


Fig.4.5 UMTS Network Architecture Overview

4.4.1 User Equipment, UE

The USER Equipment or UE is a major element of the overall 3G UMTS network architecture. It forms the final interface with the user. In view of the far greater number of applications and facilities that it can perform, the decision was made to call it user equipment rather than a mobile. However it is essentially the handset (in the broadest terminology), although having access to much higher speed data communications, it can be much more versatile, containing many more applications. It consists of a variety of different elements including RF circuitry, processing, antenna, battery, etc.

There are a number of elements within the UE that can be described separately:

- UE RF circuitry: The RF areas handle all elements of the signal, both for the receiver and for the transmitter. One of the major challenges for the RF power amplifier was to reduce the power consumption. The form of modulation used for W-CDMA requires the use of a linear amplifier. These inherently take more current than nonlinear amplifiers which can be used for the form of modulation used on GSM. Accordingly to maintain battery life, measures were introduced into many of the designs to ensure the optimum efficiency.

- Baseband processing: The base-band signal processing consists mainly of digital circuitry. This is considerably more complicated than that used in phones for previous generations. Again this has been optimized to reduce the current consumption as far as possible.
- Battery: While current consumption has been minimized as far as possible within the circuitry of the phone, there has been an increase in current drain on the battery. With users expecting the same lifetime between charging batteries as experienced on the previous generation phones, this has necessitated the use of new and improved battery technology. Now Lithium Ion (Li-ion) batteries are used. These phones to remain small and relatively light while still retaining or even improving the overall life between charges.
- Universal Subscriber Identity Module, USIM: The UE also contains a SIM card, although in the case of UMTS it is termed a USIM (Universal Subscriber Identity Module). This is a more advanced version of the SIM card used in GSM and other systems, but embodies the same types of information. It contains the International Mobile Subscriber Identity number (IMSI) as well as the Mobile Station International ISDN Number (MSISDN). Other information that the USIM holds includes the preferred language to enable the correct language information to be displayed, especially when roaming, and a list of preferred and prohibited Public Land Mobile Networks (PLMN).

4.4.2 3G UMTS Radio Network Subsystem

.

This is the section of the 3G UMTS / WCDMA network that interfaces to both the UE and the core network. The overall radio access network, i.e. collectively all the Radio Network Subsystem is known as the UTRAN UMTS Radio Access Network.

The radio network subsystem is also known as the UMTS Radio Access Network or UTRAN.

4.4.3 3G UMTS Core Network

The 3G UMTS core network architecture is a migration of that used for GSM with further elements overlaid to enable the additional functionality demanded by UMTS.

In view of the different ways in which data may be carried, the UMTS core network may be split into two different areas:

- **Circuit switched elements:** These elements are primarily based on the GSM network entities and carry data in a circuit switched manner, i.e. a permanent channel for the duration of the call.
    - It is used to provide voice and CS data services.
    - It contains Mobile Switching Center (MSC) and Gateway MSC(GMSC) as functional entities.
- **Packet switched elements:** These network entities are designed to carry packet data. This enables much higher network usage as the capacity can be shared and data is carried as packets which are routed according to their destination.
- It is used to provide packet based services.

It contains

Serving GPRS support node (SGSN),

.

Gateway GPRS support node (GGSN),

Domain Name Server (DNS),

Dynamic Host Configuration Protocol (DHCP) server,

packet charging gateway,

and firewalls.

The core network can be split into the following different functional areas:

Functional entities needed to support PS services (e.g.3G-SGSN, 3G-GGSN)

Functional entities needed to support CS services (e.g. 3G-MSC/VLR)

Functional entities common to both types of services (e.g. 3G-HLR)

Other areas that can be considered part of the core network include:

Network management systems (billing and provisioning, service management,element management, etc.)

IN system (service control point (SCP), service signaling point (SSP), etc.)

ATM/SDH/IP switch/transport infrastructure.

Some network elements, particularly those that are associated with registration are shared by both domains and operate in the same way that they did with GSM.

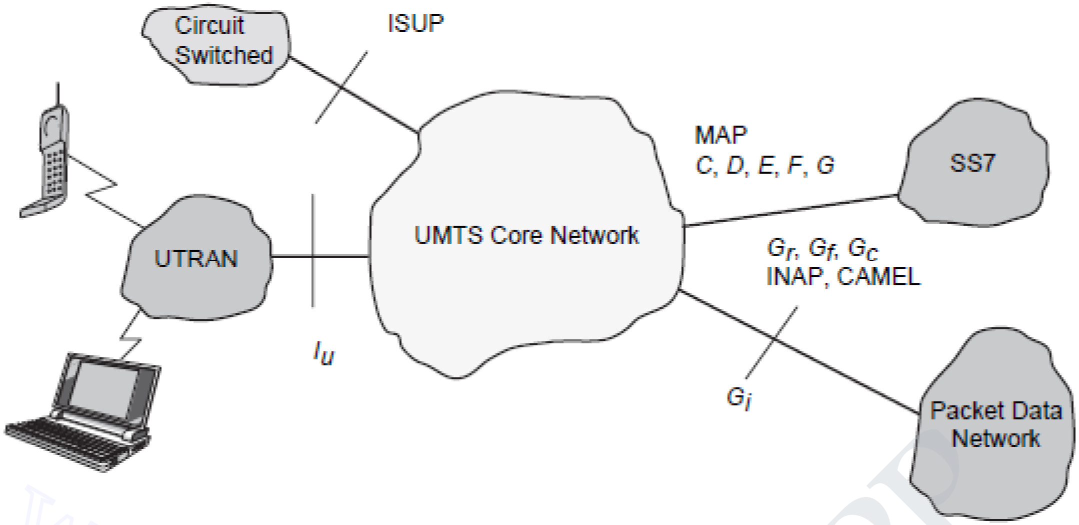
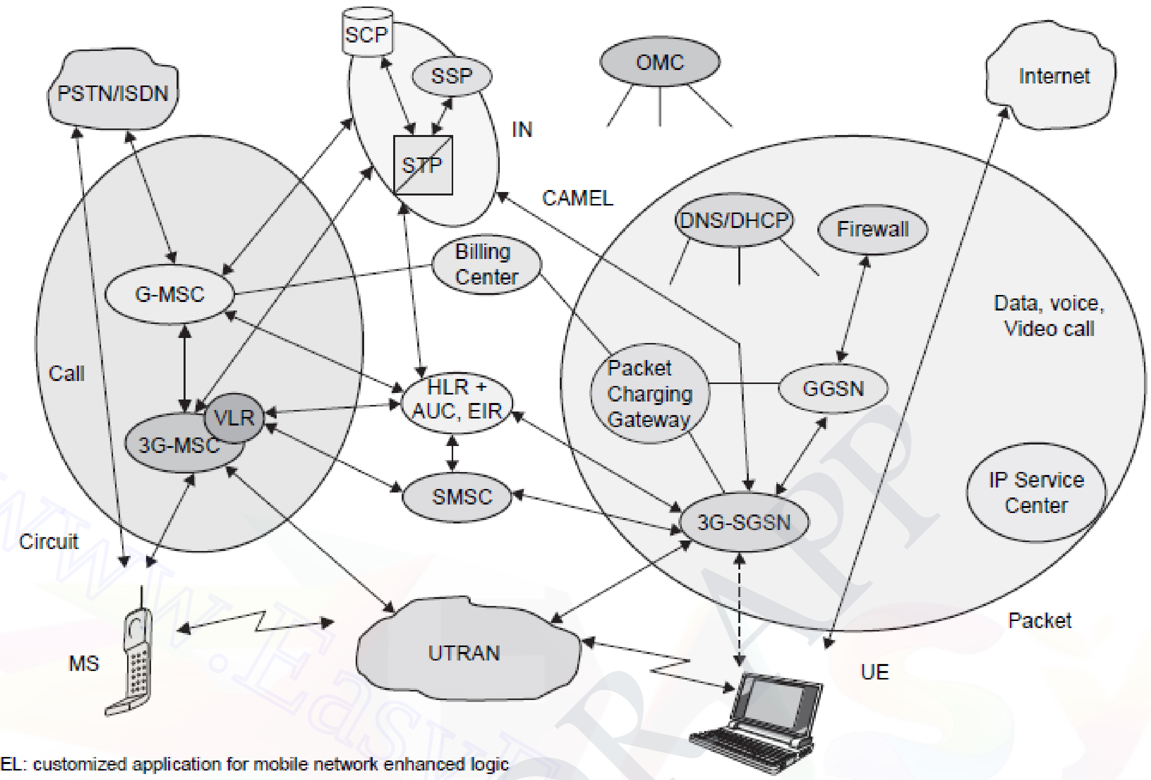


Fig.4.6 UMTS Core network architecture

The above figure shows all the entities that connect to the core network — UTRAN, PSTN, the Internet and the logical connections between terminal equipment (MS,UE), and the PSTN/Internet.



CAMEL: customized application for mobile network enhanced logic
SMSC: short message service center
DNS: domain name server
DHCP: dynamic host configuration protocol

Fig.4.7 Logical architecture of the UMTS core network.

Circuit switched elements

The circuit switched elements of the UMTS core network architecture include the following network entities:

Mobile switching Centre (MSC): This is essentially the same as that within GSM, and it manages the circuit switched calls under way.

Gateway MSC (GMSC): This is effectively the interface to the external networks.

Packet switched elements

The packet switched elements of the 3G UMTS core network architecture include the following network entities:

Serving GPRS Support Node (SGSN):

Gateway GPRS Support Node (GGSN):

4.4.4 3G-MSC

The MSC is the control Centre for the cellular system, coordinating the actions of the BSCs, providing overall control, and acting as the switch and connection into the public telephone network. As such it has a variety of communication links into it which will include fiber optic links as well as some microwave links and some copper wire cables. These enable it to communicate with the BSCs, routing calls to them and controlling them as required. It also contains the Home and Visitor Location Registers, the databases detailing the last known locations of the mobiles. It also contains the facilities for the Authentication Centre, allowing mobiles onto the network. In addition to this it will also contain the facilities to generate the billing information for the individual accounts.

In view of the importance of the MSC, it contains many backup and duplicate circuits to ensure that it does not fail. Obviously backup power systems are an essential element of this to guard against the possibility of a major power failure, because if the MSC became inoperative then the whole network would collapse.

While the cellular network is not seen by the outside world and its operation is a mystery to many, the cellular network is at the very center of the overall cellular system and the success of the whole end to end system is dependent largely on its performance.

This is essentially the same as that within GSM, and it manages the circuit switched calls under way.

It is the main CN element.

It provides CS services.

It provides the necessary control and corresponding signaling interfaces includingSS7, MAP, ISUP (ISDN user part), etc.

It is used to provide the interconnectionto external networks like PSTN and ISDN.

The following functionality is provided by the 3G-MSC.

Mobility management:

Handles attach, authentication, updates to the HLR,SRNS relocation, and intersystems handover.

Call management:

Handles call set-up messages from/to the UE.

Supplementary services:

Handles call-related supplementary services suchas call waiting, etc.

CS data services:

The IWF provides rate adaptation and message translationfor circuit mode data services, such as fax.

Vocoding

SS7, MAP and RANAP interfaces:

The 3G-MSC is able to complete originatingor terminating calls in the network in interaction with other entities ofa mobile network, e.g., HLR, AUC

(Authentication center). It also controls/communicates with RNC using RANAP which may use the services of SS7.

ATM/AAL2

Connection to UTRAN for transportation of user plane trafficacross the Iu interface. Higher rate CS data rates may be supported using adifferent adaptation layer.

Short message services (SMS):

This functionality allows the user to sendand receive SMS data to and from the SMS-GMSC/SMS-IWMSC (Interworking MSC).

VLR functionality:

The VLR is a database that may be located within the3G-MSC and can serve as intermediate storage for subscriber data in orderto support subscriber mobility.

INand CAMEL.

OAM

(operation, administration, and maintenance) agent functionality.

4.4.5 3G-SGSN-Serving GPRS Support Node

The 3G-SGSN is the main CN element for PS services. The 3G-SGSN provides

the necessary control functionality both toward the UE and the 3G-GGSN. It alsoprovides the appropriate signaling and data interfaces including connection toan IP-based network toward the 3G-GGSN, SS7 toward the HLR/EIR/AUC andTCP/IP or SS7 toward the UTRAN.

The 3G-SGSN provides the following functions:

Session management:

Handles session set-up messages from/to the UE andthe GGSN and operates Admission Control and QoS mechanisms.

I_u and G_n MAP interface:

The 3G-SGSN is able to complete originating orterminating sessions in the network by interaction with other entities of amobile network, e.g., GGSN, HLR, AUC. It also controls/communicateswith UTRAN using RANAP.

ATM/AAL5

Physical connection to the UTRAN for transportation of userdata plane traffic across the I_u interface using GPRS tunneling protocol(GTP).

Connection across the G_n interface toward the GGSN for transportation ofuser plane traffic using GTP. Note that no physical transport layer is definedfor this interface.

SMS:

This functionality allows the user to send and receive SMS data to and from the SMS-GMSC /SMS-IWMSC.

Mobility management:

Handles attach, authentication, updates to the HLRand SRNS relocation, and intersystem handover.

Subscriber database functionality:

This database (similar to the VLR) islocated within the 3G-SGSN and serves as intermediate storage for subscriberdata to support subscriber mobility.

Charging:

The SGSN collects charging information related to radio networkusage by the user.

4.4.6 3G-GGSN

The GGSN provides interworking with the external PS network. It is connectedwith SGSN via an IP-based network. The GGSN may optionally support an SS7interface with the HLR to handle mobile terminated packet sessions.

The 3G-GGSN provides the following functions:

It Maintain information locations at SGSN level (macro-mobility)

Gateway between UMTS packet network and external data networks(e.g. IP, X.25)

Gateway-specific access methods to intranet (e.g. PPP termination)

Initiate mobile terminate Route Mobile Terminated packets

User data screening/security can include subscription based, user controlled,or network controlled screening.

User level address allocation: The GGSN may have to allocate (dependingon subscription) a dynamic address to the UE upon PDP context activation.

This functionality may be carried out by use of the DHCP function.

Charging: The GGSN collects charging information related to external data network usage by the user.

4.4.7 SMS-GMSC/SMS-IWMSC

The overall requirement for these two nodes is to handle the SMS from point topoint.

The functionality required can be split into two parts.

The SMS-GMSC isan MSC capable of receiving a terminated short message from a service center,interrogating an HLR for routing information and SMS information, and deliveringthe short message to the SGSN of the recipient UE.

The SMS-GMSC providesthe following functions:

Reception of short message packet data unit (PDU)

Interrogation of HLR for routing information

Forwarding of the short message PDU to the MSC or SGSN using therouting information

The SMS-IWMSC is an MSC capable of receiving an originating shortmessage from within the PLMN and submitting it to the recipient service center.

The SMS-IWMSC provides the following functions:

Reception of the short message PDU from either the 3G-SGSN or3G-MSC

Establishing a link with the addressed service center

Transferring the short message PDU to the service center

Note: The service center is a function that is responsible for relaying, storing,and forwarding a short message. The service center is not part of UCN, althoughthe MSC and the service center may be integrated.

4.4.8 Firewall

A firewall is a network security system, either hardware- or software-based, that controls incoming and outgoing network traffic based on a set of rules.

This entity is used to protect the service providers' backbone data networks fromattack from external packet data networks. The security of the backbone datanetwork can be ensured by applying packet filtering mechanisms based on accesscontrol lists or any other methods deemed suitable.

Introduction

Firewalls are computer security systems that protect your office/home PCs or your network from intruders, hackers & malicious code. Firewalls protect you from offensive software that may come to reside on your systems or from prying hackers. In a day and age when online security concerns are the top priority of the computer users, Firewalls provide you with the necessary safety and protection.

Firewalls are software programs or hardware devices that filter the traffic that flows into you PC or your network through a internet connection. They sift through the data flow & block that which they deem (based on how & for what you have tuned the firewall) harmful to your network or computer system.

When connected to the internet, even a standalone PC or a network of interconnected computers make easy targets for malicious software & unscrupulous hackers. A firewall can offer the security that makes you less vulnerable and also protect your data from being compromised or your computers being taken hostage.

.

Firewalls are setup at every connection to the Internet, therefore subjecting all data flow to careful monitoring. Firewalls can also be tuned to follow "rules". These Rules are simply security rules that can be set up by the network administrators to allow traffic to their web servers, FTP servers, Telnet servers, thereby giving the computer owners/administrators immense control over the traffic that flows in & out of their systems or networks.

Rules will decide who can connect to the internet, what kind of connections can be made, which or what kind of files can be transmitted in out. Basically all traffic in & out can be watched and controlled thus giving the firewall installer a high level of security & protection.

Firewall logic

Firewalls use 3 types of filtering mechanisms:

Packet filtering or packet purity

Data flow consists of packets of information and firewalls analyze these packets to sniff out offensive or unwanted packets depending on what you have defined as unwanted packets.

Proxy

Firewall in this case assumes the role of a recipient & in turn sends it to the node that has requested the information & vice versa.

Inspection

In this case Firewalls instead of sifting through all of the information in the packets, mark key features in all outgoing requests & check for the same matching characteristics in the inflow to decide if it relevant information that is coming through.

.

Firewall Rules

Firewalls rules can be customized as per our needs, requirements & security threat levels.

We can create or disable firewall filter rules based on such conditions as:

IP Addresses

Blocking off a certain IP address or a range of IP addresses, which you think are predatory.

Domain names

Only certain specific domain names are allowed to access our systems/servers or allow access to only some specified types of domain names or domain name extension like .edu or .mil.

Protocols

A firewall can decide which of the systems can allow or have access to common protocols like IP, SMTP, FTP, UDP,ICMP,Telnet or SNMP.

Ports

Blocking or disabling ports of servers that are connected to the internet will help maintain the kind of data flow you want to see it used for & also close down possible entry points for hackers or malignant software.

Keywords

Firewalls also can sift through the data flow for a match of the keywords or phrases to block out offensive or unwanted data from flowing in.

Types of Firewall

Software firewalls

New generation Operating systems come with built in firewalls or you can buy a firewall software for the computer that accesses the internet or acts as the gateway to your home network.

Hardware firewalls

Hardware firewalls are usually routers with a built in Ethernet card and hub. Your computer or computers on your network connect to this router & access the web.

Packet firewalls

The earliest firewalls functioned as packet filters, inspecting the packets that are transferred between computers on the Internet. When a packet passes through a packet-filter firewall, its source and destination address, protocol, and destination port number are checked against the firewall's rule set. Any packets that aren't specifically allowed onto the network are dropped (i.e., not forwarded to their destination). For example, if a firewall is configured with a rule to block Telnet access, then the firewall will drop packets destined for TCP port number 23, the port where a Telnet server application would be listening.

Packet-filter firewalls work mainly on the first three layers of the OSI reference model (physical, data-link and network), although the transport layer is used to obtain the source and destination port numbers. While generally fast and efficient, they have no ability to tell whether a packet is part of an existing stream of traffic. Because they treat each packet in isolation, this makes them vulnerable to spoofing attacks and also limits their ability to make more complex decisions based on what stage communications between hosts are at.

Stateful firewalls

In order to recognize a packet's connection state, a firewall needs to record all connections passing through it to ensure it has enough information to assess whether a packet is the tart of a new connection, a part of an existing connection, or not part of any connection. This is what's called "stateful packet inspection." Stateful inspection was first introduced in 1994 by Check Point Software in its FireWall-1 software firewall, and by the late 1990s, it was a common firewall product feature.

This additional information can be used to grant or reject access based on the packet's history in the state table, and to speed up packet processing; that way, packets that are part of an existing connection based on the firewall's state table can be allowed through without further analysis. If a packet does not match an existing connection, it's evaluated according to the rule set for new connections.

Application-layer firewalls

As attacks against Web servers became more common, so too did the need for a firewall that could protect servers and the applications running on them, not merely the network resources behind them. Application-layer firewall technology first emerged in 1999, enabling firewalls to inspect and filter packets on any OSI layer up to the application layer.

The key benefit of application-layer filtering is the ability to block specific content, such as known malware or certain websites, and recognize when certain applications and protocols -- such as HTTP, FTP and DNS -- are being misused.

Firewall technology is now incorporated into a variety of devices; many routers that pass data between networks contain firewall components and most home computer operating systems include software-based firewalls. Many hardware-based firewalls also provide additional functionality like basic routing to the internal network they protect.

Proxy firewalls

Firewall proxy servers also operate at the firewall's application layer, acting as an intermediary for requests from one network to another for a specific network application. A proxy firewall prevents direct connections between either sides of the firewall; both sides are forced to conduct the session through the proxy, which can block or allow traffic based on its rule set. A proxy service must be run for each type of Internet application the firewall will support, such as an HTTP proxy for Web services.

Firewalls in the perimeterless age

The role of a firewall is to prevent malicious traffic reaching the resources that it is protecting. Some security experts feel this is an outdated approach to keeping information and the resources it resides on safe. They argue that while firewalls still have a role to play, modern networks have so many entry points and different types of users that stronger access control and security at the host is a better technological approach to network security.

Some of the firewall products that you may want to check out are:

McAfee Internet Security

Microsoft Windows Firewall

Norton Personal Firewall

Trend Micro PC-cillin

ZoneAlarm Security Suit

4.4.9 DNS/DHCP

The DNS server is used, as in any IP network, to translate host names intoIP addresses, i.e., logical names are handled instead of raw IP addresses. Also,the DNS server is used to translate the access point name (APN) into the GGSNIP address. It may optionally be used to allow the UE to use logical namesinstead of physical IP addresses.

A dynamic host configuration protocol server is used to manage the allocationof IP configuration information by automatically assigning IP addresses tosystems configured to use DHCP.

DHCP (Dynamic Host Configuration Protocol)

DHCP is a network protocol that is used to assign various network parameters to a device. This greatly simplifies administration, since there is no need to assign static network parameters for each device separately. DHCP is a client-server protocol. A client is a device that is configured to use DHCP to request network parameters from a DHCP server. DHCP server maintains a pool of available IP addresses and assigns one of them to the host. A DHCP server can also provide some other parameters, such as:

- Subnet mask
- Default gateway
- Domain name
- DNS server

DHCP process explained:

DHCP client goes through the four step process:

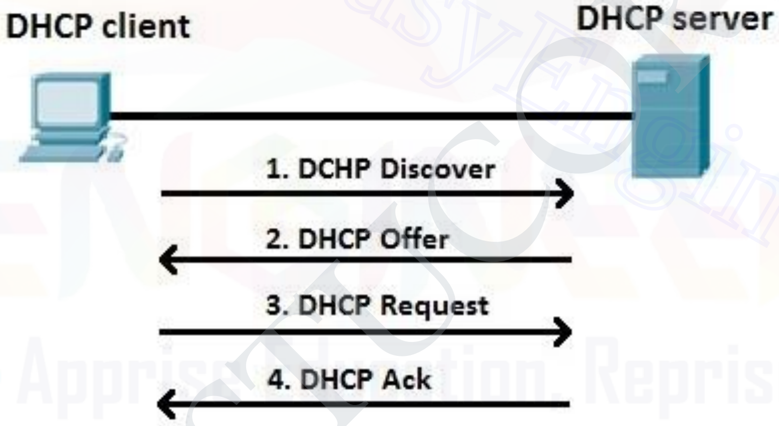


Fig.4.8

1: A DHCP client sends a broadcast packet (DHCP Discover) to discover DHCP servers on the LAN segment.

2: The DHCP servers receive the DHCP Discover packet and respond with DHCP Offer packets, offering IP addressing information.

3: If the client receives the DHCP Offer packets from multiple DHCP servers, the first DHCP Offer packet is accepted. The client responds by broadcasting a DHCP Request packet, requesting network parameters from a single server.

4: The DHCP server approves the lease with a DHCP Acknowledgement packet. The packet includes the lease duration and other configuration information.

DNS (Domain Name System)

DNS is a network protocol used to translate hostnames into IP addresses. DNS is not required to establish a network connection, but it is much more user friendly for human users than the numeric addressing scheme. Consider this example. You can access the Google homepage by typing 74.125.227.99, but it's much easier just to type www.google.com!

To use DNS, you must have a DNS server configured to handle the resolution process. A DNS server has a special-purpose application installed. The application maintains a table of dynamic or static hostname-to-IP address mappings. When a user request some network resource using a hostname, (for example by typing www.google.com in a browser), a DNS request is sent to the DNS server asking for the IP address of the hostname. The DNS server then replies with the IP address.

The figure below explains the concept:

.

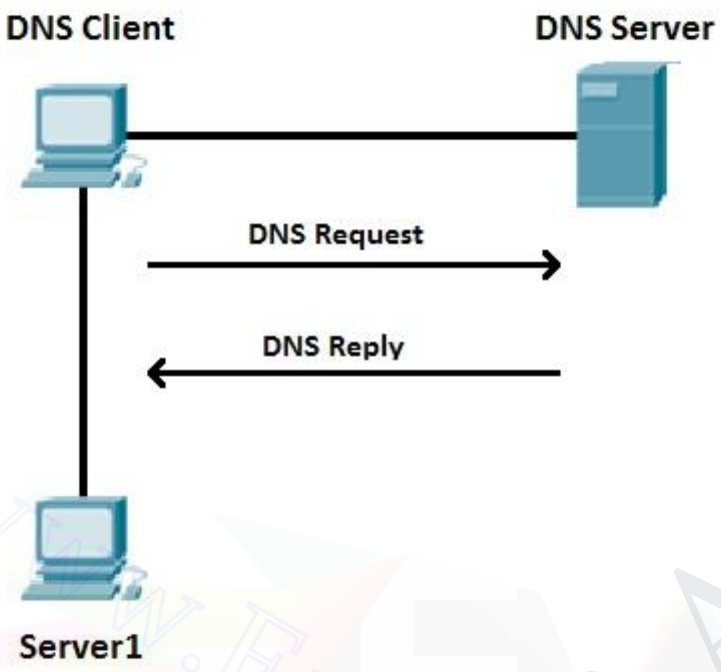


Fig. 4.9 DNS

Suppose that the DNS Client wants to communicate with the server named Server1. Since the DNC Client doesn't know the IP address of Server1, it sends a DNS Request to the DNS Server, asking for Server1's IP address. The DNS Server replies with the IP address of Server1 (DNS Reply).

4.5 HIGH-SPEED DOWNLINK PACKET ACCESS (HSDPA)

In third-generation partnership project (3GPP) standards, Release 4 specificationsprovide efficient IP support enabling provision of services through an all-IP corenetwork. Release 5 specifications focus on HSDPAto provide data rates up to approximately 8–10 Mbps to support packet-basedmultimedia services. Multi input and multi output (MIMO) systems are the workitem in Release 6 specifications, which will support even higher data transmissionrates of up to 20

Mbps. HSDPA is evolved from and backward compatible withRelease 99 WCDMA systems.

HSDPA is based on the same set of technologies as high data rate (HDR)to improve spectral efficiency for data services — such as shared downlink packetdata channel and high peak data rates — using high-order modulation and adaptivemodulation and coding, hybrid ARQ (HARQ) retransmission schemes, fastscheduling and shorter frame sizes.

HSDPA marks a similar boost for WCDMA that EDGE does for GSM. Itprovides a two-fold increase in air interface capacity and a five-fold increase indata speeds in the downlink direction. HSDPA also shortens the round-trip timebetween the network and terminals and reduces variance in downlink transmissiondelay.

The improvements in performance are achieved by:

Bringing some key functions, such as scheduling of data packet transmissionand processing of retransmissions (in case of transmission errors) intothe base station — that is, closer to the air interface.

Using a short frame length to further accelerate packet scheduling fortransmission.

Employing incremental redundancy for minimizing the air-interface load caused by retransmissions.

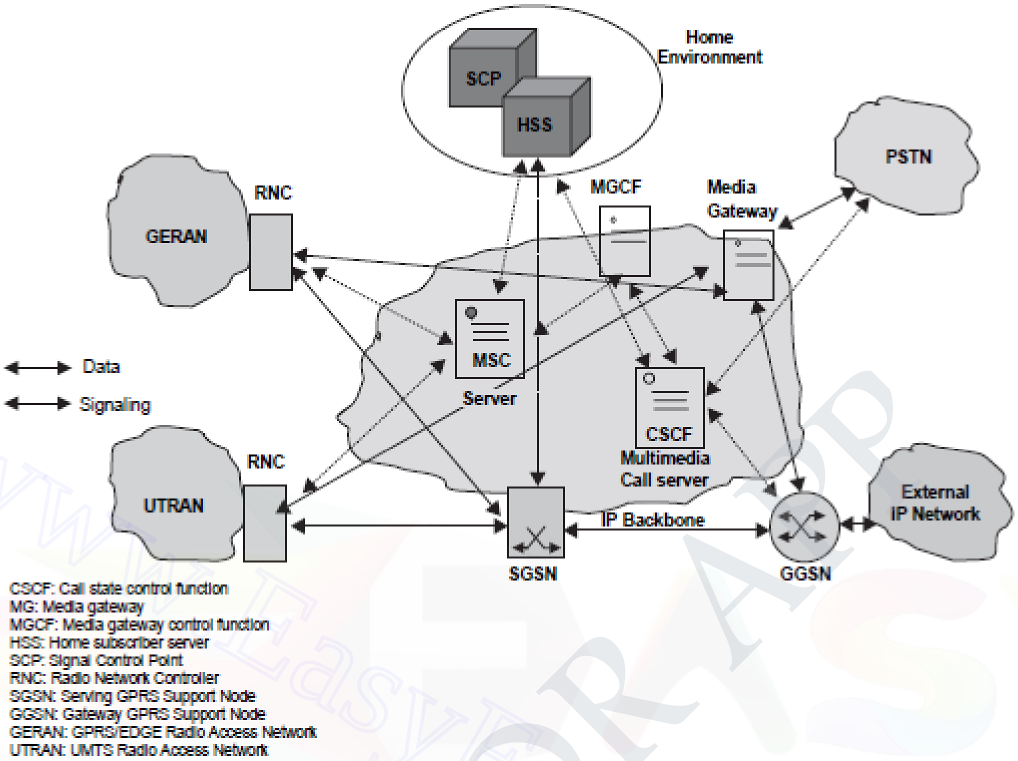

Home
Environment
SCP
HSS
PSTN
RNC
GERAN
MGCF
Media
Gateway
MSC
Server
Data
Signaling
CSCF
Multimedia
Call server
RNC
UTRAN
External
IP Network
IP Backbone
SGSN
GGSN
CSCF: Call state control function
MG: Media gateway
MGCF: Media gateway control function
HSS: Home subscriber server
SCP: Signal Control Point
RNC: Radio Network Controller
SGSN: Serving GPRS Support Node
GGSN: Gateway GPRS Support Node
GERAN: GPRS/EDGE Radio Access Network
UTRAN: UMTS Radio Access Network

Fig.4.10A simplified all-IP UMTS architecture

Adopting a new transport channel type, known as high-speed downlinkshared channel (HS-DSCH) to facilitate air interface channel sharingbetween several users.

Adapting the modulation and coding scheme according to the quality of theradio linkThe primary objective behind HSDPA is to provide a cost-effective, highbandwidth,low-delay, packet-oriented service within UMTS. Backward compatibilityis critical, so the HSDPA architecture adheres to an evolutionary philosophy.From an architectural perspective, HSDPA is a straightforward enhancementof the UMTS Release '99 (R99) architecture, with the addition of a repetition/scheduling entity within the Node B that resides below the R99 media-access control(MAC) layer. From a cellular-network perspective, all R99 techniques can besupported in a network supporting HSDPA, since HSDPA mobile terminals (UEs)are designed to coexist with R99 UEs. HSDPA is particularly suited to extremelyasymmetrical data services, which require significantly higher data rates for thetransmission from the network to the UE, than they do for the transmission fromthe UE to the network.

HSDPA introduces enablers for the high-speed transmission at the physicallayer like the use of a shorter transmission time interval (TTI) (2 ms), and the useof adaptive modulation and coding. HS-DPCCH is used to carry the acknowledgmentsignals to Node B for each block. It is also used to indicate channel quality(CQI) used for adaptive modulation and coding. HS-DSCH uses 2 ms TTI toreduce trip time, to increase the granularity in the scheduling process, and to trackthe time varying radio channel better.

The basic operational principles behind HSDPA are relatively simple. TheRNC routes data packets destined for a particular UE to the appropriate Node B.

Node B takes the data packets and schedules their transmission to the mobileterminal over the air interface by matching the user's priority and estimated channeloperating environment with an appropriately chosen coding and modulationscheme (that is, 16-QAM vs. QPSK).

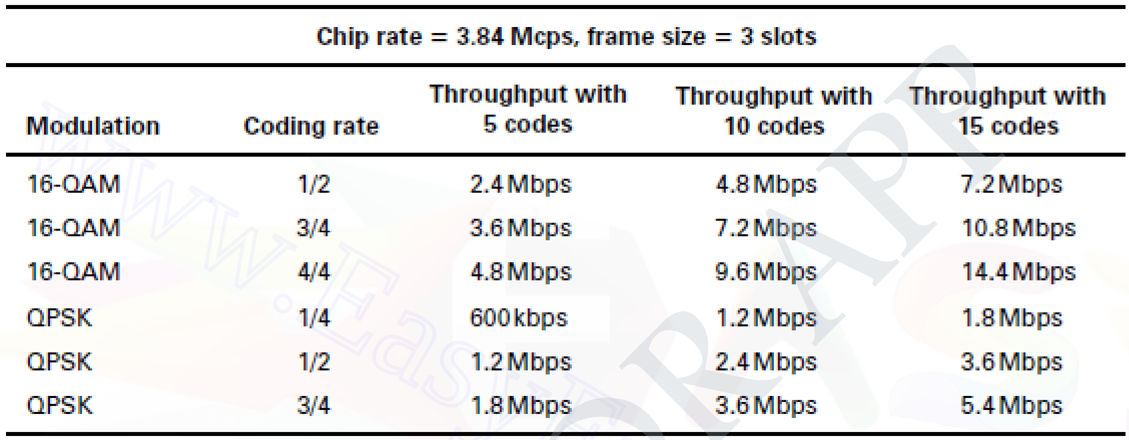
The UE is responsible for acknowledging receipt of the data packet and providing

Node B with information regarding channel condition, power control, andso on. Once it sends the data packet to the UE, Node B waits for an acknowledgment.If it does not receive one within a prescribed time, it assumes that the datapacket was lost and retransmits it.

HSDPA continuously strives, with some modest constraints, to give themaximal bandwidth to the user with the best channel conditions. The data ratesachievable with HSDPA are more than adequate for supporting multimediastreamingservices.

Although conceptually simple, HSDPA's implementation within the contextof a Node B does raise some architectural issues for the designer. In a typical network deployment, the Node B radio cabinet sits in proximity to the radiotower and the power cabinet. For indoor deployments the radio cabinet may bea simple rack, while in outdoor deployments it may be an environmental-controlunit. The guts of the radio cabinet are an antenna interface section (filters, poweramplifiers, and the like), core processing chassis (RF transceivers, combiner, highperformancechannel cards, network interface and system controller card, timingcard, back-plane, and so on), plus mechatronics (power supply, fans, cables, etc.)

Table 15.10 HSDPA data rates.



Chip rate = 3.84 Mcps, frame size = 3 slots				
Modulation	Coding rate	Throughput with 5 codes	Throughput with 10 codes	Throughput with 15 codes
16-QAM	1/2	2.4 Mbps	4.8 Mbps	7.2 Mbps
16-QAM	3/4	3.6 Mbps	7.2 Mbps	10.8 Mbps
16-QAM	4/4	4.8 Mbps	9.6 Mbps	14.4 Mbps
QPSK	1/4	600 kbps	1.2 Mbps	1.8 Mbps
QPSK	1/2	1.2 Mbps	2.4 Mbps	3.6 Mbps
QPSK	3/4	1.8 Mbps	3.6 Mbps	5.4 Mbps

and other miscellaneous elements. The core processing chassis is the cornerstoneof Node B and bears most of the cost. It contains the RF transceiver, combiner,network interface and system controller, timing card, channel card and backplane.

.

Of the core processing chassis elements, only the channel card needs to bemodified to support HSDPA.

The typical UMTS channel card comprises a general-purpose processor thathandles the miscellaneous control tasks, a pool of digital signal processor (DSP)resources to handle symbol-rate processing and chip-rate assist functions, and apool of specialized ASIC (application specific integrated circuit) devices to handleintensive chip-rate operations such as spreading, scrambling, modulation, rakereceiving, and preamble detection.

To support HSDPA, two changes must be made to the channel card. First,the downlink chip-rate ASIC must be modified to support the new 16-QAM modulationschemes and new downlink slot formats associated with HSDPA. In addition,the downlink symbol-rate processing section must be modified to supportHSDPA extensions.

The next change requires a new processing section, called the MAC-hs,which must be added to the channel card to support the scheduling, buffering,transmission, and retransmission of data blocks that are received from theRNC. This is the most intrusive augmentation to the channel card because it requires the introduction of a programmable processing entity together with aretransmission buffer.

Since the channel card already contains both a general-purpose processor anda DSP, one can make convincing arguments that the MAC-hs could be effectivelyrealized using either of the two types of devices. Nonetheless, many designersare finding that, because of the close ties between the MAC-hs function and thelower-layer symbol and chip-rate functions, the DSP is the more practical choice.

Simulations have shown that a retransmission buffer of approximately 2.5 Mbits in size is adequate to handle the buffering requirement of a standard cell with 75or so users.

The new channels introduced in HSDPA are high-speed downlink sharedchannel (HS-DSCH), high-speed shared control channel (HS-SCCH), and highspeeddedicated physical control channel (HS-DPCCH). The HS-DSCH is theprimary radio bearer. Its resources can be shared among all users in a particularsector. The primary channel multiplexing occurs in a time domain, where eachTTI consists of three time slots (each 2 ms). TTI is also referred to as a sub-frame.Within each 2 ms TTI, a constant spreading factor (SF) of 16 is used for codemultiplexing, with a maximum of 15 parallel codes allocated to HS-DSCH. Codesmay all be assigned to one user, or may be split across several users. The numberof codes allocated to each user depends on cell loading, QoS requirements, andUE code capabilities (5, 10, or 15 codes).

The HS-SCCH (a fixed rate 960 kbps, SF _ 128) is used to carry downlinksignaling between Node B and UE before the beginning of each scheduled TTI.It includes UE identity, HARQ-related information and the parameters of theHS-DSCH transport format selected by the link-adaptation mechanism. MultipleHS-SCCHs can be configured in each sector to support parallel HS-DSCH transmissions.

A UE can be allocated a set of up to four HS-SCCHs, which need to bemonitored continuously.

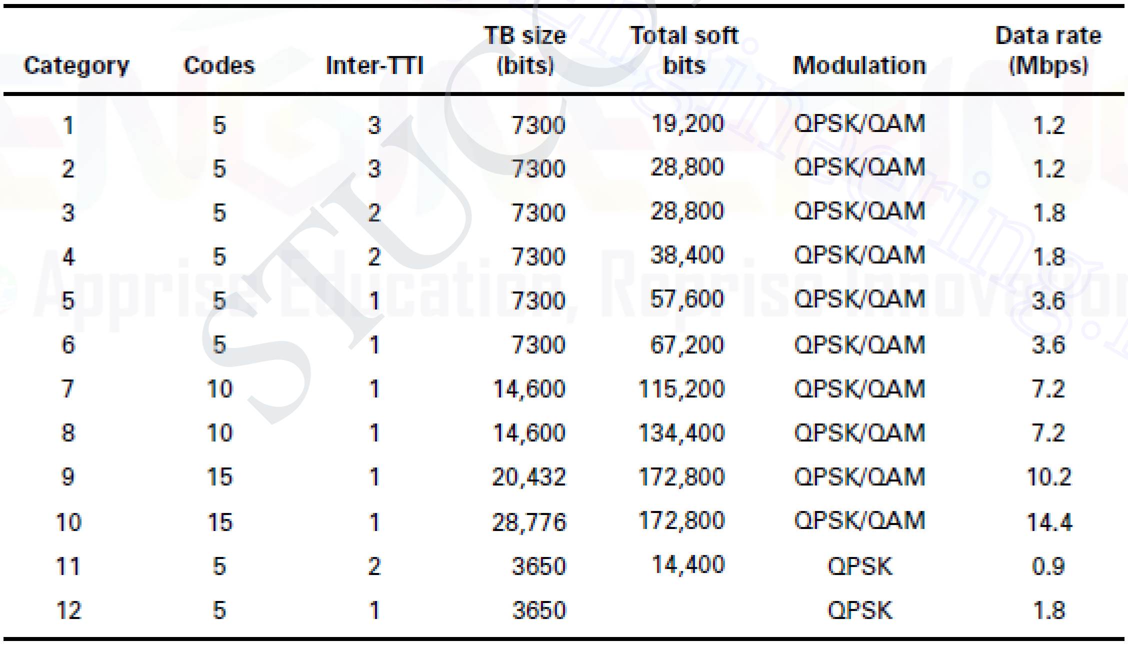
The HS-DPCCH (SF _ 256) carries ACK/NACK signaling to indicate whether

the corresponding downlink transmission was successfully decoded, as well as achannel quality indicator (CQI) to be used for the purpose of link adaptation.

The CQI is based on a common pilot channel (CPICH) and is used to estimate thetransport block size, modulation type, and number of channelization codes thatcan be supported at a given reliability level in downlink transmission. The feedbackcycle of CQI can be set as a network parameter in predefined steps of 2 ms.

UE capabilities include the maximum number of HS-DSCHs supportedsimultaneously (5, 10, or 15), minimum TTI time (minimum time between thebeginning of two consecutive transmissions to the UE), the maximum numberof HS-DSCH transport block (TB) bits received within an HS-DSCH TTI, themaximum number of soft channel bits over all HARQ and supported modulations(QPSK only or both QPSK and 16-QAM). Table 15.11 gives UE categories.

Table 15.11 HSDPA UE categories.



Category	Codes	Inter-TTI	TB size (bits)	Total soft bits	Modulation	Data rate (Mbps)
1	5	3	7300	19,200	QPSK/QAM	1.2
2	5	3	7300	28,800	QPSK/QAM	1.2
3	5	2	7300	28,800	QPSK/QAM	1.8
4	5	2	7300	38,400	QPSK/QAM	1.8
5	5	1	7300	57,600	QPSK/QAM	3.6
6	5	1	7300	67,200	QPSK/QAM	3.6
7	10	1	14,600	115,200	QPSK/QAM	7.2
8	10	1	14,600	134,400	QPSK/QAM	7.2
9	15	1	20,432	172,800	QPSK/QAM	10.2
10	15	1	28,776	172,800	QPSK/QAM	14.4
11	5	2	3650	14,400	QPSK	0.9
12	5	1	3650		QPSK	1.8

4.6LTE

LTE stands for Long Term Evolution and it was started as a project in 2004 by telecommunication body known as the Third Generation Partnership Project (3GPP). SAE (System Architecture Evolution) is the corresponding evolution of the GPRS/3G packet core network evolution. The term LTE is typically used to represent both LTE and SAE.

LTE evolved from an earlier 3GPP system known as the Universal Mobile Telecommunication System (UMTS), which in turn evolved from the Global System for Mobile Communications (GSM). Even related specifications were formally known as the evolved UMTS terrestrial radio access (E-UTRA) and evolved UMTS terrestrial radio access network (E-UTRAN).

A rapid increase of mobile data usage and emergence of new applications such as MMOG (Multimedia Online Gaming), mobile TV, Web 2.0, streaming contents have motivated the 3rd Generation Partnership Project (3GPP) to work on the Long-Term Evolution (LTE) on the way towards fourth-generation mobile.

The main goal of LTE is to provide a high data rate, low latency and packet optimized radio access technology supporting flexible bandwidth deployments. Same time its network architecture has been designed with the goal to support packet-switched traffic with seamless mobility and great quality of service.

Facts about LTE

- LTE is the successor technology not only of UMTS but also of CDMA 2000.

.

- LTE bring up to 50 times performance improvement and much better spectral efficiency to cellular networks.

- LTE introduced to get higher data rates, 300Mbps peak downlink and 75 Mbps peak uplink. In a 20MHz carrier, data rates beyond 300Mbps can be achieved under very good signal conditions.

- LTE is an ideal technology to support high date rates for the services such as voice over IP (VOIP), streaming multimedia, videoconferencing or even a high-speed cellular modem.

- LTE uses both Time Division Duplex (TDD) and Frequency Division Duplex (FDD) mode. In FDD uplink and downlink transmission used different frequency, while in TDD both uplink and downlink use the same carrier and are separated in Time.

- LTE supports flexible carrier bandwidths, from 1.4 MHz up to 20 MHz as well as both FDD and TDD. LTE designed with a scalable carrier bandwidth from 1.4 MHz up to 20 MHz which bandwidth is used depends on the frequency band and the amount of spectrum available with a network operator.

- All LTE devices have to support (MIMO) Multiple Input Multiple Output transmissions, which allow the base station to transmit several data streams over the same carrier simultaneously.

- All interfaces between network nodes in LTE are now IP based, including the backhaul connection to the radio base stations. This is great simplification compared to earlier technologies that were initially based on

E1/T1, ATM and frame relay links, with most of them being narrowband and expensive.

- Quality of Service (QoS) mechanism have been standardized on all interfaces to ensure that the requirement of voice calls for a constant delay and bandwidth, can still be met when capacity limits are reached.
- Works with GSM/EDGE/UMTS systems utilizing existing 2G and 3G spectrum and new spectrum. Supports hand-over and roaming to existing mobile networks.

Advantages of LTE

- **High throughput:** High data rates can be achieved in both downlink as well as uplink. This causes high throughput.
- **Low latency:** Time required to connect to the network is in range of a few hundred milliseconds and power saving states can now be entered and exited very quickly.
- **FDD and TDD in the same platform**: Frequency Division Duplex (FDD) and Time Division Duplex (FDD), both schemes can be used on same platform.
- **Superior end-user experience:** Optimized signaling for connection establishment and other air interface and mobility management procedures have further improved the user experience. Reduced latency (to 10 ms) for better user experience.
- **Seamless Connection**: LTE will also support seamless connection to existing networks such as GSM, CDMA and WCDMA.

- **Plug and play:** The user does not have to manually install drivers for the device. Instead system automatically recognizes the device, loads new drivers for the hardware if needed, and begins to work with the newly connected device.

- **Simple architecture:** Because of Simple architecture low operating expenditure (OPEX).

4.6.1 LTE Network Architecture

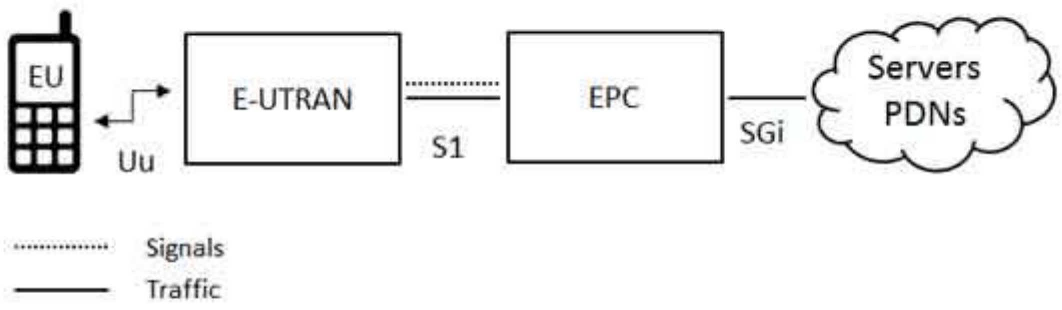
The high-level network architecture of LTE is comprised of following three main components:

- The User Equipment (UE).

- The Evolved UMTS Terrestrial Radio Access Network (E-UTRAN).

- The Evolved Packet Core (EPC).

The evolved packet core communicates with packet data networks in the outside world such as the internet, private corporate networks or the IPmultimedia subsystem. The interfaces between the different parts of the system are denoted Uu, S1 and SGi as shown below:



.

Fig.4.11The LTE network Architecture

The User Equipment (UE):

The internal architecture of the user equipment for LTE is identical to the one used by UMTS and GSM which is actually a Mobile Equipment (ME). The mobile equipment comprised of the following important modules:

- **Mobile Termination (MT)** :This handles all the communication functions.
- **Terminal Equipment (TE)** : This terminates the data streams.
- **Universal Integrated Circuit Card (UICC)** :This is also known as the SIM card for LTE equipments. It runs an application known as the Universal Subscriber Identity Module (USIM).

A **USIM** stores user-specific data very similar to 3G SIM card. This keeps information about the user's phone number, home network identity and security keys etc.

The E-UTRAN (The access network)

The architecture of evolved UMTS Terrestrial Radio Access Network (E-UTRAN) has been illustrated below.

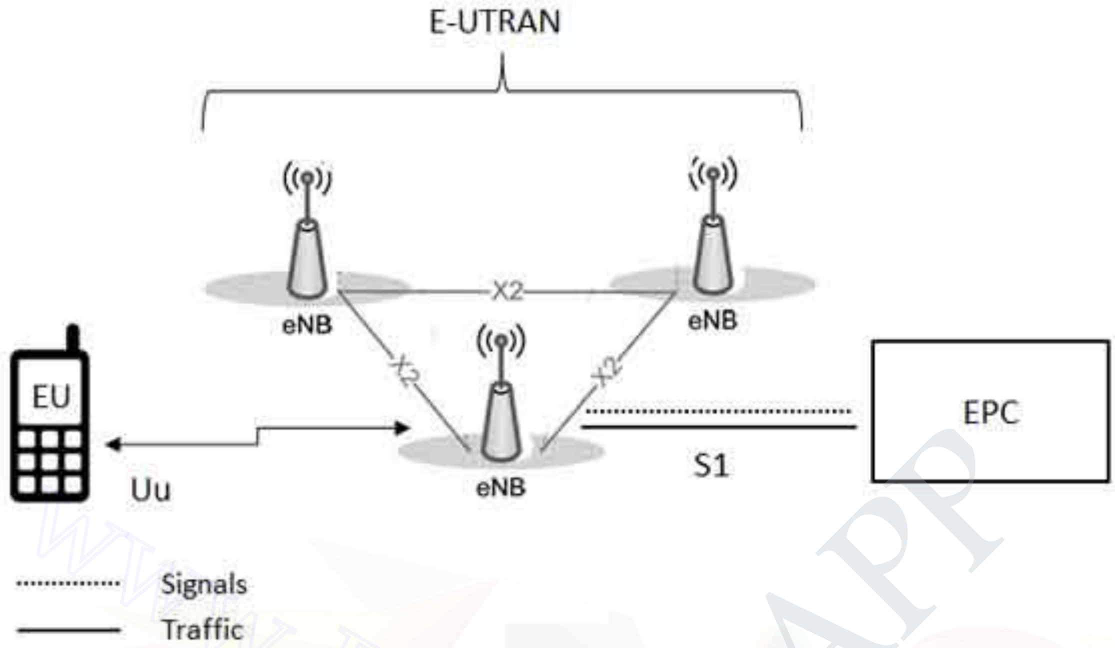


Fig.4.12 The architecture of E-UTRAN

The E-UTRAN handles the radio communications between the mobile and the evolved packet core and just has one component, the evolved base stations, called eNodeB or eNB. Each eNB is a base station that controls the mobiles in one or more cells. The base station that is communicating with a mobile is known as its serving eNB.

LTE Mobile communicates with just one base station and one cell at a time and there are following two main functions supported by eNB:

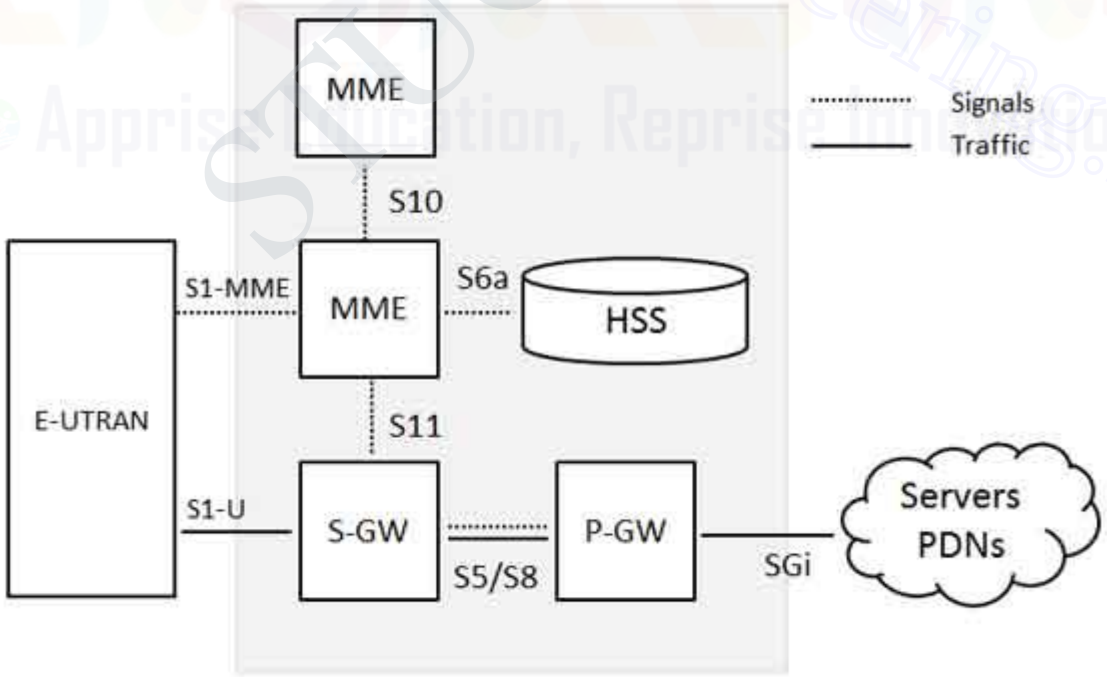
- The eBN sends and receives radio transmissions to all the mobiles using the analogue and digital signal processing functions of the LTE air interface.
- The eNB controls the low-level operation of all its mobiles, by sending them signalling messages such as handover commands.

Each eBN connects with the EPC by means of the S1 interface and it can also be connected to nearby base stations by the X2 interface, which is mainly used for signalling and packet forwarding during handover.

A home eNB (HeNB) is a base station that has been purchased by a user to provide femtocell coverage within the home. A home eNB belongs to a closed subscriber group (CSG) and can only be accessed by mobiles with a USIM that also belongs to the closed subscriber group.

The Evolved Packet Core (EPC) (The core network):

The architecture of Evolved Packet Core (EPC) has been illustrated below. There are few more components which have not been shown in the diagram to keep it simple. These components are like the Earthquake and Tsunami Warning System (ETWS), the Equipment Identity Register (EIR) and Policy Control and Charging Rules Function (PCRF).



.

Fig.4.13 The architecture of Evolved Packet Core

Below is a brief description of each of the components shown in the above architecture:

- The Home Subscriber Server (HSS) component has been carried forward from UMTS and GSM and is a central database that contains information about all the network operator's subscribers.
- The Packet Data Network (PDN) Gateway (P-GW) communicates with the outside world ie. packet data networks PDN, using SGi interface. Each packet data network is identified by an access point name (APN). The PDN gateway has the same role as the GPRS support node (GGSN) and the serving GPRS support node (SGSN) with UMTS and GSM.
- The serving gateway (S-GW) acts as a router, and forwards data between the base station and the PDN gateway.
- The mobility management entity (MME) controls the high-level operation of the mobile by means of signaling messages and Home Subscriber Server (HSS).
- The Policy Control and Charging Rules Function (PCRF) is a component which is not shown in the above diagram but it is responsible for policy control decision-making, as well as for controlling the flow-based charging functionalities in the Policy Control Enforcement Function (PCEF), which resides in the P-GW.

The interface between the serving and PDN gateways is known as S5/S8. This has two slightly different implementations, namely S5 if the two devices are in the same network, and S8 if they are in different networks.

2G/3G Versus LTE

Following table compares various important Network Elements & Signaling protocols used in 2G/3G abd LTE.

2G/3G	LTE
GERAN and UTRAN	E-UTRAN
SGSN/PDSN-FA	S-GW
GGSN/PDSN-HA	PDN-GW
HLR/AAA	HSS
VLR	MME
SS7-MAP/ANSI-41/RADIUS	Diameter
DiameterGTPc-v0 and v1	GTPc-v2
MIP	PMIP

4.6.2 LTE Protocol Stack Layers

The below diagram shows the layers available in E-UTRAN Protocol Stack.

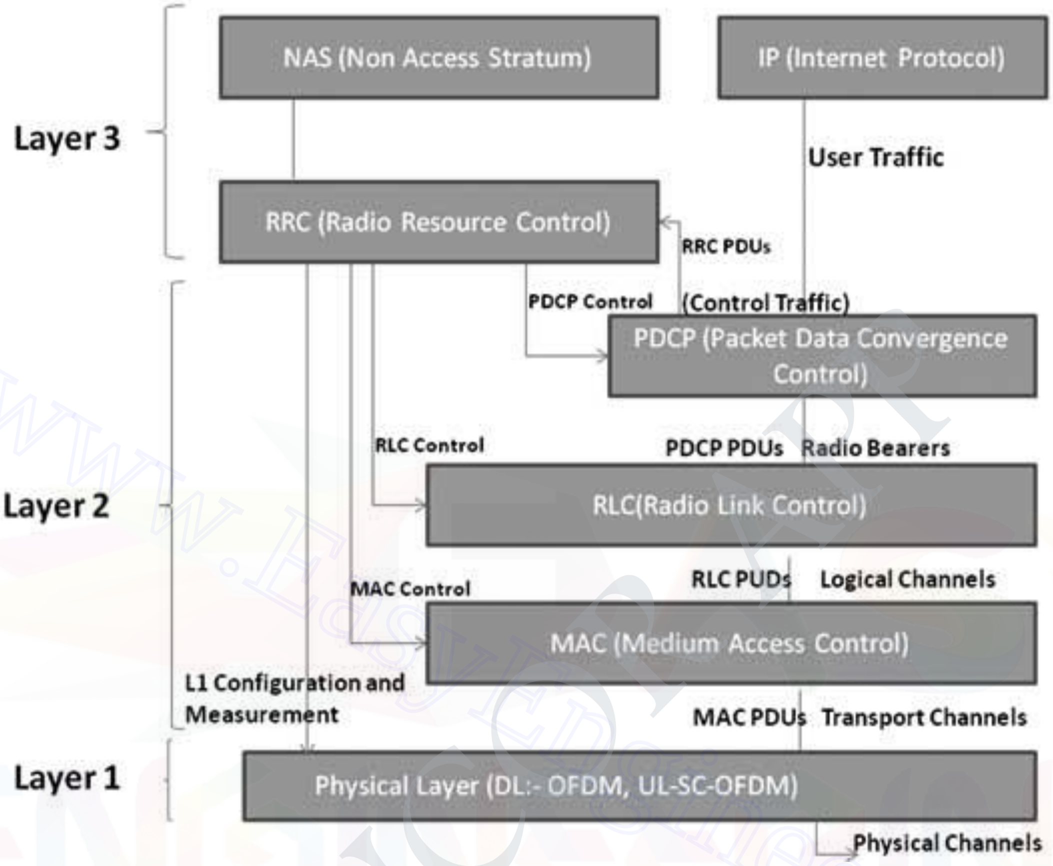


Fig. 4.14 E-UTRAN Protocol Stack

Physical Layer (Layer 1)

Physical Layer carries all information from the MAC transport channels over the air interface. Takes care of the link adaptation (AMC), power control, cell search (for initial synchronization and handover purposes) and other measurements (inside the LTE system and between systems) for the RRC layer.

Medium Access Layer (MAC)

MAC layer is responsible for Mapping between logical channels and transport channels, Multiplexing of MAC SDUs from one or different logical channels onto transport blocks (TB) to be delivered to the physical layer on transport channels, demultiplexing of MAC SDUs from one or different logical channels from transport blocks (TB) delivered from the physical layer on transport channels, Scheduling information reporting, Error correction through HARQ, Priorityhandling between UEs by means of dynamic scheduling, Priority handling between logical channels of one UE, Logical Channel prioritization.

Radio Link Control (RLC)

RLC operates in 3 modes of operation: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM).

RLC Layer is responsible for transfer of upper layer PDUs, error correction through ARQ (Only for AM data transfer), Concatenation, segmentation and reassembly of RLC SDUs (Only for UM and AM data transfer).

RLC is also responsible for re-segmentation of RLC data PDUs (Only for AM data transfer), reordering of RLC data PDUs (Only for UM and AM data transfer), duplicate detection (Only for UM and AM data transfer), RLC SDU discard (Only for UM and AM data transfer), RLC re-establishment, and protocol error detection (Only for AM data transfer).

Radio Resource Control (RRC)

The main services and functions of the RRC sublayer include broadcast of System Information related to the non-access stratum (NAS), broadcast of System Information related to the access stratum (AS), Paging, establishment, maintenance and release of an RRC connection between the UE and E-UTRAN, Security functions including key management, establishment, configuration, maintenance and release of point to point Radio Bearers.

Packet Data Convergence Control (PDCP)

PDCP Layer is responsible for Header compression and decompression of IP data, Transfer of data (user plane or control plane), Maintenance of PDCP Sequence Numbers (SNs), In-sequence delivery of upper layer PDUs at re-establishment of lower layers, Duplicate elimination of lower layer SDUs at re-establishment of lower layers for radio bearers mapped on RLC AM, Ciphering and deciphering of user plane data and control plane data, Integrity protection and integrity verification of control plane data, Timer based discard, duplicate discarding, PDCP is used for SRBs and DRBs mapped on DCCH and DTCH type of logical channels.

Non Access Stratum (NAS) Protocols

The non-access stratum (NAS) protocols form the highest stratum of the control plane between the user equipment (UE) and MME.

NAS protocols support the mobility of the UE and the session management procedures to establish and maintain IP connectivity between the UE and a PDN GW.

UNIT IV – 4G NETWORKS

INTRODUCTION ABOUT 4G

4G stands for Forth Generation of Cellular Communications and is the next step in the evolution of mobile data. 4G provides high mobility with high speed data rates and also supports high capacity IP-based services and applications while it also maintains full backward compatibility.

It is also based on wireless communication that is IP based and is slated on Advanced MIMO technology. 4G technologies follow Multiple Input Multiple Output Technology that uses signal multiplexing between multiple transmitting antennas (space multiplex) and time or frequency.

Fourth generation (4G) technology will offer many advancements to the wireless market, including downlink data rates well over 100 megabits per second (Mbps), low latency, very efficient spectrum use and low-cost implementations.

4G enhancements promise to bring the wireless experience to an entirely new level with impressive user applications, such as sophisticated graphical user interfaces, high-end gaming, high-definition video and high-performance imaging.

4.1 4G VISIONS

The 4G systems are designed to provide a wide variety of new services, from high-quality voice to highdefinition video to high-data-rate wireless channels. The term 4G is used broadly to includeseveral types of BWA communication systems, not only cellular systems. 4G is described as MAGIC — Mobile multimedia, anytime anywhere, Global mobility support, integrated wireless solution, and customized personal service.

The 4G systems willnot only support the next generation mobile services, but also will support the fixed wirelessnetworks. The 4G systems are about seamlessly integrating terminals, networks, andapplications to satisfy increasing user demands.

Accessing information anywhere, anytime, with a seamless connection to a wide range ofinformation and services, and receiving a large volume of information, data, pictures, video,and so on, are the keys of the 4G infrastructure.

The future 4G systems will consist of a set ofvarious networks using IP as a common protocol. 4G systems will have broader bandwidth,higher data rate, and smoother and quicker handoff and will focus on ensuring seamlessservice across a multiple of wireless systems and networks.

The key is to integrate the 4Gcapabilities with all the existing mobile technologies through the advanced techniques ofdigital communications and networking.

Application adaptability and being highly dynamic are the main features of 4G servicesof interest to users.These features mean services can be delivered and be available to the personal preference of different users and support the users' traffic, airinterfaces, radio environment, and quality of service. Connection with the network applications can be transferred into various forms and levels correctly and efficiently.

The following figure illustrates elements and techniques to support the adaptability of the 4G domain. The fourth generation will encompass all systems from various networks, public to private; operator-driven broadband networks to personal areas; and ad hoc networks. The 4G systems will interoperate with 2G and 3G systems, as well as with digital (broadband) broadcasting systems.

In addition, 4G systems will be fully IP-based wireless Internet. This all-encompassing integrated perspective shows the broadrange of systems that the fourth generation intends to integrate, from satellite broadband to high altitude platform to cellular 3G and 3G systems to WLL (wireless local loop) and FWA (fixed wireless access) to WLAN (wireless local area network) and PAN (personal area network),all with IP as the integrating mechanism.

With 4G, a range of new services and models will be available. These services and models need to be further examined for their interface with the design of 4G systems.

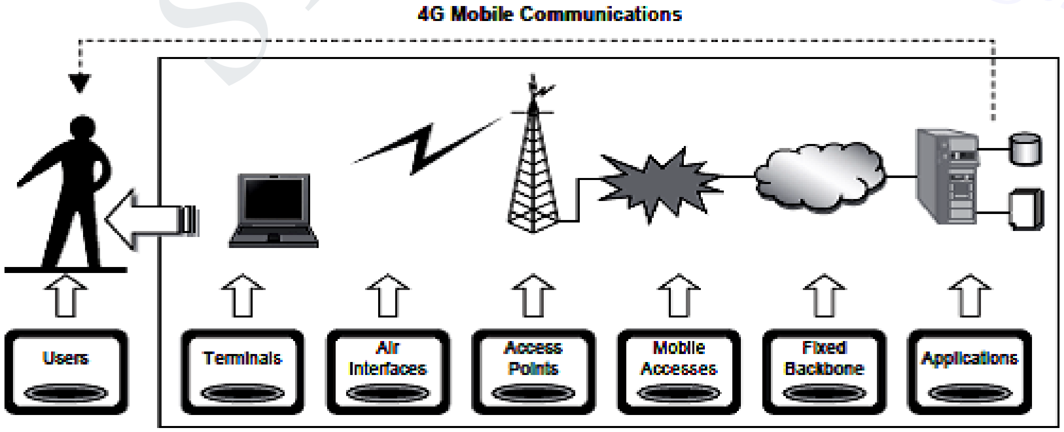


Fig.4.1 4G Visions

Table 4.1.1 Comparison of key parameters of 4G with 3G.

Details	3G including 2.5G (EDGE)	4G
Major requirement driving Architecture	Predominantly voice driven, data was always add on	Converge data and voice over IP
Network architecture	Wide area cell-based	Hybrid-integration of WLAN (WiFi, Bluetooth) and wireless wide-area networks
Speeds	384 kbps to 2 Mbps	20 to 100 Mbps in mobile mode
Frequency band	Dependent on country or continent (1.8 to 2.4 GHz)	Higher frequency bands (2 to 8 GHz)
Bandwidth	5 to 20 MHz	100 MHz or more
Switching design basis	Circuit and packet	All digital with packetized
Access technologies	WCDMA, CDMA2000	OFDM and multicarrier
Forward error correction	Convolutional codes rate 1/2, 1/3	Concatenated coding
Component design	Optimized antenna design, multiband adapters	Smart antenna, software defined multiband and wideband radios
Internet protocol(IP)	Number of air link protocol including IPv5.0	All IP (IPv6.0)
Mobile top speed	200 km/h	200 km/h

4G will need to be highly dynamic in terms of support for:

- The users' traffic
- Air interfaces and terminal types
- Radio environments
- Quality-of-service types
- Mobility patterns.

4.2 4G FEATURES

Some key features of 4G mobile networks are as follows.

- High usability: anytime, anywhere, and with any technology
- Support for multimedia services at low transmission cost
- Personalization
- Integrated services
- Support for interactive multimedia, voice, streaming video, Internet, and other broadband services
- IP based mobile system
- High speed, high capacity, and low cost per bit
- Global access, service portability, and scalable mobile services
- Seamless switching, and a variety of Quality of Service driven services
- Better scheduling and call admission control techniques
- Ad hoc and multi hop networks
- Better spectral efficiency
- Seamless network of multiple protocols and air interfaces

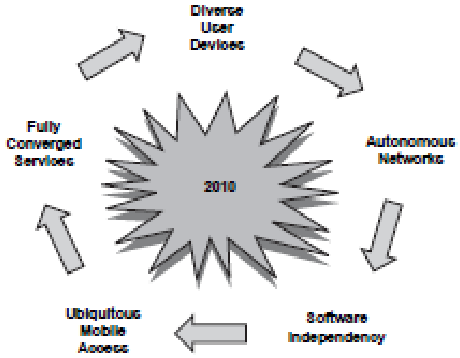


Fig 4.2 4G Features

4G networks will be all-IP-based heterogeneous networks that will allow users to use any system at anytime and anywhere. Users carrying an integrated terminal can use a wide range of applications provided by multiple wireless networks. 4G systems will provide not only telecommunications services, but also data and multimedia services. To support multimedia services, high-data-rate services with system reliability will be provided. At the same time, a low per-bit transmission cost will be maintained by an improved spectral efficiency of the system.

Personalized service will be provided by 4G networks. It is expected that when 4G services are launched, users in widely different locations, occupations, and economic classes will use the services. In order to meet the demands of these diverse users, service providers will design personal and customized service for them. 4G systems will also provide facilities for integrated services. Users can use multiple services from any service provider at the same time.

4G technologies are significant because users joining the network add mobile routers to the network infrastructure. Because users carry much of the network with them, network capacity and coverage is dynamically shifted to accommodate changing user patterns. Users will automatically hop away from congested routes to less congested routes. This permits the network to dynamically and automatically self-balance capacity, and increase network utilization.

In a cellular infrastructure, user's network contribution is nil. They are just consumers competing for resources. But in wireless ad hoc peer-to-peer networks, users cooperate - rather than compete - for network resources.

4.3 4G CHALLENGES

While migrating from 3G to 4G, certain challenges have to be faced.

- **Multimode user terminal:** Multimode user terminal is a device operating in different modes supporting a large type of 4G services and wireless networks by reconfiguring themselves to adapt to different wireless networks. They encounter several design problems like limitations in the device cost, size, backward compatibility to systems and power consumption.
- **Wireless network discovery:** Availing 4G services need the multimode user terminal to find and select the required wireless network. Discovery of 4G systems will be much more challenging than 3G due to the heterogeneity of the networks and their access protocols.
- **Wireless network selection:** 4G will offer the users a option to select a wireless network providing optimized performance and high QoS for a specific place, time and desired service (communication, multimedia). But the parameters that define high QoS and optimized performance at

specific instant must be clearly defined to form the network selection procedure efficient and transparent to the end user.

- **Terminal mobility:** Terminal mobility is an important characteristic to satisfy the "Anytime Anywhere" promise of 4G. It permits the mobile users to move across the geographic boundaries of wireless networks. Two important problems in terminal mobility are location and hand off management. Location management includes tracking the location of the mobile users and maintaining data like the authentication information, QoS capabilities, and the original and the current cell location. Handoff management is maintaining the continuing communication when the terminal roams.
- **Network infrastructure and QoS support:** Unlike previous generation networks such as 2G and 3G, 4G is an integration of IP and non-IP based system. Before 4G, QoS designs were made with a specific wireless system in mind. But in 4G systems QoS designs should consider the integration of various wireless networks to ensure QoS for the end-to-end services.
- **Security:** Most of the security schemes and the encryption/decryption protocols of the present generation networks were designed only for specific services. They appear to be very inflexible to be used across the heterogeneous architecture of 4G that desires dynamically adaptive, reconfigurable and light-weight security mechanism.
- **Fault tolerance:** Wireless networks resemble a tree-like topology. Any failure in one of the levels will affect all the network elements at the levels below. This problem may become more complex because of the multiple tree topologies. Adequate research work is needed to devise a method for fault tolerance in wireless networks.

4.4 APPLICATIONS OF 4G

●**Virtual presence:** 4G will provide user services at all times, even if the user is off-site.

●**Virtual navigation:** 4G will provide users with virtual navigation through which a user can access a database of streets, buildings, etc., of a large city. This requires High-speed transmission.

●**Tele-medicine:** 4G will support the remote health monitoring of patients via video conference assistance for a doctor anytime and anywhere.

●**Tele-geo-processing applications:** 4G will combine geographical information systems (GIS) and global positioning systems (GPS) in which a user will get location querying.

●**Education:** 4G will provide a good opportunity to people anywhere in the world to continue their education on-line in a cost-effective manner.

●**Multimode Software Application**

Multimode software is software that allows the user device to adapt itself to various wireless interfaces networks in order to provide constant net access with high data (packet based) rate.

All the networks will be compatible once the switch is completed, eliminating roaming and areas where only one type of phone is supported.

Once the voice and data networks are superposed there will suddenly be millions of new devices on the network cloud. This will require either reconstruction of the address space for the entire Internet or using different address spaces for the existing wireless networks. The multimode device architecture may improve call completion and expand effective coverage area.

●**Support for Multiple and Efficient Applications and Services**- 4G provides support for unicast, multicast and broadcast services and the applications that rely on them. Prompt enforcement of Service Level Agreements (SLA) along with privacy and other security features.

●**Quality of Service** -Consistent application of admission control and scheduling algorithms regardless of underlying infrastructure and operator diversity leads to an increased quality of service (Qos) to the users.

●**Network Detection Selection**: A mobile terminal that features multiple radio technologies or possibly uses software defined radios if economical, allows participation in multiple networks simultaneously, thereby connecting to the best network with the most appropriate service parameters (cost, QoS and capacity among others) for the application. This requires establishing a uniform process for defining eligibility of a terminal to attach to a network and to determine the validity of link layer configuration.

●**Handover and Service Continuity:** A base station|| that features intra- and inter-technology handovers, assuring service continuity with zero or minimal interruption, without a noticeable loss in service quality. Support for this function requires continuous transparent maintenance of active service instances and inclusion of various access technologies, from Wi-Fi to OFDMA.

●**Crisis Management Application**: In the event of natural disasters where the entire communications infrastructure is in disarray, restoring communications quickly is essential. With wideband wireless mobile communications, limited and even total communication capability (including Internet and video services) could be set up within hours instead of days or even weeks required at present for restoration of wire line communications.

ADVANTAGES OF 4G:-

1. Support for interactive multimedia services like teleconferencing and wireless Internet.

2. Wider bandwidths and higher bitrates.

3. Global mobility and service portability.

4. Scalability of mobile network.

5. Entirely Packet-Switched networks.

6. Digital network elements.

7. Higher band widths to provide multimedia services at lower cost(up to 100 Mbps).

8. Tight network security

4.5 4G TECHNOLOGIES

4.5.1 MULTICARRIER MODULATION

Multi-carrier modulation (MCM) is a method of transmitting data by splitting it into several components, and sending each of these components over separate carrier signals. The individual carriers have narrow bandwidth , but the composite signal can have broad bandwidth.

The advantages of MCM include relative immunity to fading caused by transmission over more than one path at a time (multipath fading), less susceptibility than single-carrier systems to interference caused by impulse noise, and enhanced immunity to inter-symbol interference. Limitations include difficulty in synchronizing the carriers under marginal conditions, and a relatively strict requirement that amplification be linear.

Multicarrier modulation (MCM) is a derivative of frequency-division multiplexing. Forms of multicarrier systems are currently used in DSL modems and digital audio/video broadcast (DAB/DVB). MCM is a baseband process that uses parallel equal bandwidth sub channels to transmit information and is normally implemented with fast Fourier transform (FFT) techniques. MCM's advantages are better performance in the inter symbol-interference environment and avoidance of single-frequency interferers.

However, MCM increases the peak-to-average ratio of the signal, and to overcome inter symbol interference or guard band must be added to the data. The difference, D , of the peak-to-average ratio between MCM and a single carrier system is a function of the number of subcarriers, N , as:

.

$$D(dB) = 10 \log N$$

Any increase in the peak-to-average ratio of a signal requires an increase in linearity of the system to reduce distortion. Linearization techniques can be used, but they increase the cost of the system.

If L_b is the original length of block and the channel's response is of length L_c , the cyclically extended symbol has a new length $L_b + L\ c - 1$. The new symbol of length $L_b + L_c - 1$sampling periods has no inter symbol interference. The cost is an increase in energy and uncoded bits are added to the data. At the MCM receiver, only L_b samples are processed and $L_c - 1$ samples are discarded, resulting in a loss in signal-to-noise ratio (SNR) as:

$$(SNR)_{LOSS} = \frac{10 \log L_b + L_c - 1}{L_b} (dB)$$

Two different types of MCM multicarrier code division multiple access (MC-CDMA) and orthogonal frequency-division multiplexing (OFDM) using time-division multiple access (TDMA). MC-CDMA is actually OFDM with a CDMA overlay.

Similar to single-carrier CDMA systems, the users are multiplexed with orthogonal codes to distinguish users in MC-CDMA. However, in MC-CDMA, each user can be allocated several codes, where the data is spread in time or frequency. Either way, multiple users simultaneously access the system.

In OFDM with TDMA, the users are assigned time slots to transmit and receive data. Typically MC-CDMA uses quadrature phase shift keying (QPSK) for modulation, while OFDM with TDMA could use more high-level modulations, such as multilevel quadrature amplitude modulation (M-QAM. In OFDM the subcarrier pulse shape is a square wave. The task of pulse forming and modulation is performed by a simple inverse FFT (IFFT) which can be implemented very efficiently. To decode the transmission, a receiver needs only to implement FFT.

The OFDM divides a broadband channel into many parallel sub channels. The OFDM receiver senses the channel and corrects distortion on each sub channel before the transmitted data can be extracted. In OFDM, each of the frequencies is an integer multiple of a fundamental frequency. This ensures that even though sub channels overlap, they do not interfere with each other.

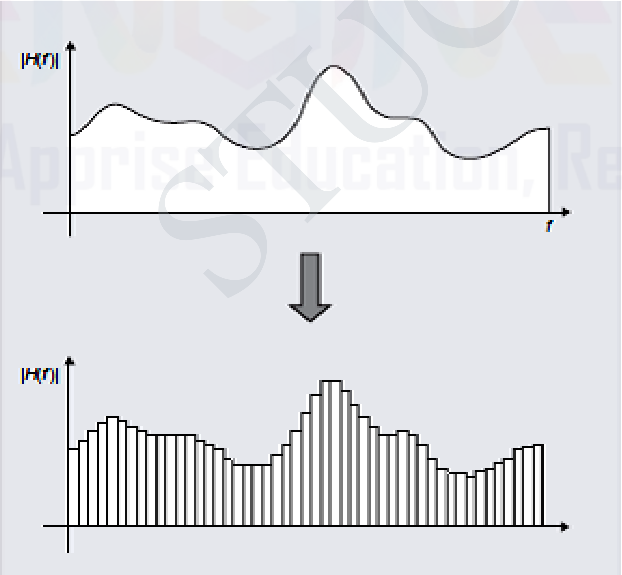


Fig. 4.3 A broadband channel divided into many parallel narrowband channels.

4.5.1.aTHE 4G TRANSCEIVER:

The structure of a 4G transceiver is similar to any other wideband wireless transceiver. A multicarrier modulated signal appears to the RF/IF section of the transceiver as a broadband high PAVR signal. Base stations and mobiles are distinguished in that base stations transmit and receive/ decode more than one mobile, while a mobile is for a single user. A mobile may be a cell phone, a computer, or other personal communication device. The line between RF and baseband will be closer for a 4G system. Data will be converted from analog to digital or vice versa at high data rates to increase the flexibility of the system. Also, typical RF components such as power amplifiers and antennas will require sophisticated signal processing techniques to create the capabilities needed for broadband high data rate signals.

The following figure shows a typical RF/IF section for a transceiver. In the transmit path in phase and quadrature (I&Q) signals are up converted to an IF, and then converted to RF and amplified for transmission. In the receive path the data is taken from the antenna at RF, filtered, amplified, and down converted for baseband processing. The transceiver provides power control, timing and synchronization, and frequency information. When multicarrier modulation is used, frequency information is crucial. If the data is not synchronized properly the transceiver will not be able to decode it.

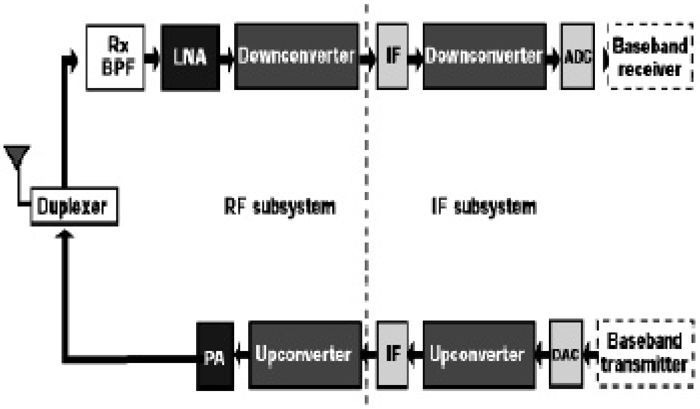


Fig. 4.4 RF/IF Block Diagram

4.5.1.bRECEIVER SECTION:

4G will require an improved receiver section, compared to 3G, to achieve the desired performance in data rates and reliability of communication. The minimum required SNR for reliable communication:

SNR=$2^{C/BW}$

Where C is the channel capacity (which is the data rate), and BW is the bandwidth for 3G, using the 2-Mbps data rate in a 5-MHz bandwidth, the SNR is only 1.2 dB. In 4G, approximately 12-dB SNR is required for a 20-Mbps data rate in a 5-MHz bandwidth. This shows that for the increased data rates of 4G, the transceiver system must perform significantly better than 3G. The receiver front end provides a signal path from the antenna to the baseband processor. It consists of a band pass filter, a low-noise amplifier (LNA), and a down converter.

De-pending on the type of receiver there could be two down conversions (as in a super-heterodyne receiver), where one down conversion converts the signal to an IF. The signal is then filtered and then down converted to or near baseband to be sampled. The other configuration has one down conversion, as in a homodyne (zero IF or ZIF) receiver, where the data is converted directly to base band. The challenge in the receiver design is to achieve the required sensitivity, intermodulation, and spurious rejection, while operating at low power.

APPLICATIONS OF MCM

- In analog military communications
- Digital audio and video broadcast services
- Digital television ,
- Obtaining high data speeds in asymmetric digital subscriber line (ADSL) systems.
- MCM is also used in wireless local area networks (WLAN s).
- Fixed wireless broadband services;
- Mobile wireless broadband communications
- Multiband OFDM for ultra wideband (UWB) communications;

4.5.2 SMART ANTENNA TECHNIQUES

Smart antenna techniques, such as multiple-input multiple-output (MIMO) systems, can extend the capabilities of the 3G and 4G systems to provide customers with increased data throughput for mobile high-speed data applications. MIMO systems use multiple antennas at both the transmitter and the receiver to increase the capacity of the wireless. With MIMO systems, it may be possible to provide in excess of 1 Mbps for 2.5G wireless TDMA EDGE and as high as 20 Mbps for 4G systems.

With MIMO, different signals are transmitted out of each antenna simultaneously in the same bandwidth and then separated at the receiver. With four antennas at the transmitter and receiver, this has the potential to provide four times the data rate of a single antenna system without an increase in transmit power or bandwidth.

MIMO techniques can support multiple independent channels in the same bandwidth, provided the multipath environment is rich enough. The number of transmitting antennas is M , and the number of receiving antennas is N , whereN ≥M.

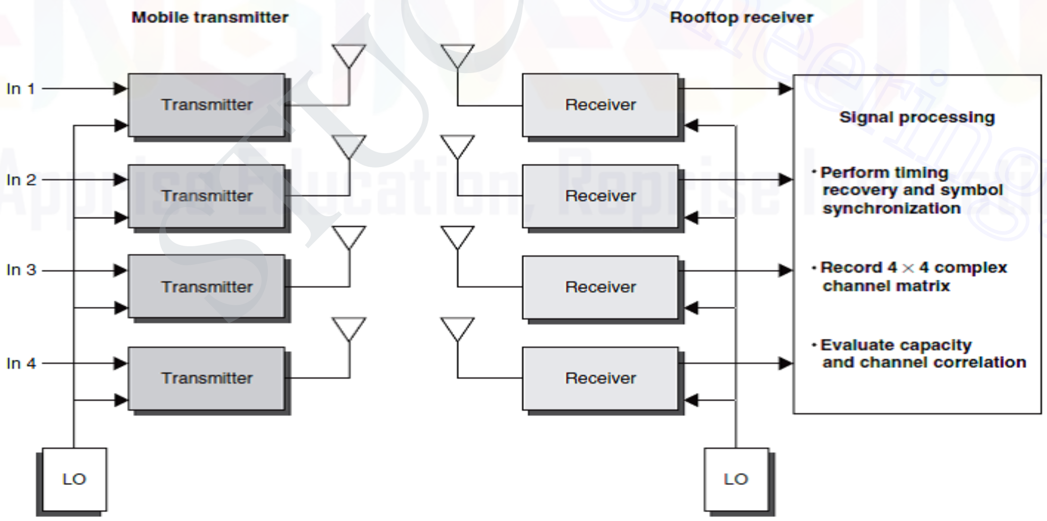


Fig. 4.5 Smart Antenna

There are four cases:

- Single-input, single-output (SISO);
- Single-input, multiple-output (SIMO);
- Multiple-input, single-output (MISO);
- Multiple-input, multiple-output (MIMO).

MIMO – SISO(Single Input Single Output)

The simplest form of radio link can be defined in MIMO terms as SISO – Single Input Single Output. This is effectively a standard radio channel – this transmitter operates with one antenna as does the receiver. There is no diversity and no additional processing required.

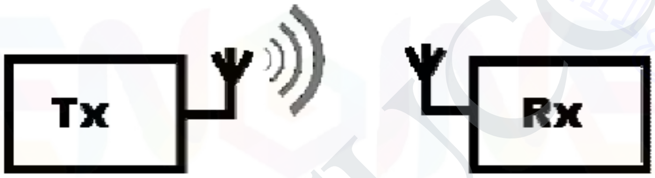


Fig. 4.6 SISO - Single Input Single Output

The advantage of a SISO system is its simplicity. SISO requires no processing in terms of the various forms of diversity that may be used. However the SISO channel is limited in its performance as interference and fading will impact the system more than a MIMO system using some form of diversity. The throughput depends upon the channel bandwidth and the signal to noise ratio.

.

If the channel bandwidth is B, the transmitter power is P_t, the signal at the receiver has an average SNR of SNR_0 , then the Shannon limit on channel capacity C is

$C \approx B \log2 (1 + SNR_0)$

MIMO – SIMO (Single Input Multiple Output)

The SIMO or Single Input Multiple Output version of MIMO occurs where the transmitter has a single antenna and the receiver has multiple antennas. This is also known as receiving diversity. It is often used to enable a receiver system that receives signals from a number of independent sources to struggle the effects of fading.

SIMO has the advantage that it is relatively easy to implement. The use of SIMO may be quite acceptable in many applications, but where the receiver is located in a mobile device such as a cellphone handset, the levels of processing may be limited by size, cost and battery drain.

There are two forms of SIMO that can be used:

Switched diversity SIMO:

This form of SIMO looks for the strongest signal and switches to that antenna.

Maximum ratio combining SIMO:

This form of SIMO takes both signals and sums them to give the combination. In this way, the signals from both antennas contribute to the overall signal.

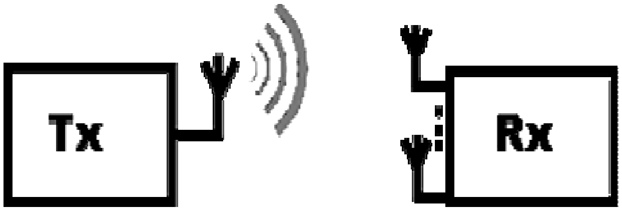


Fig. 4.7 SIMO - Single Input Multiple Output

There are N antennas at the receiver. If the signals received on the antennas have on average the same amplitude, then they can be added coherently to produce an N 2 increase in signal power. There are N sets of noise sources that are added coherently and result in an N –fold increase in noise power. Hence, the overall increase in SNR will be:

$$SNR \approx \frac{N^2 * signal\ power}{N * (noise)} = N * SNR_0$$

The capacity for this channel is approximately equal to

$C \approx B \log2 (1 + N \times SNR_0)$

MIMO – MISO(Multiple Input Single Output)

Multiple Input Single Output (MISO) is also termed transmit diversity. In this case, the same data is transmitted redundantly from the two transmitter antennas. The receiver is then able to receive the optimum signal which it can then use to receive extract the required data.

The advantage of using MISO is that the multiple antennas and the redundancy coding / processing is moved from the receiver to the transmitter. This has a positive impact on size, cost and battery life as the lower level of processing requires less battery consumption.

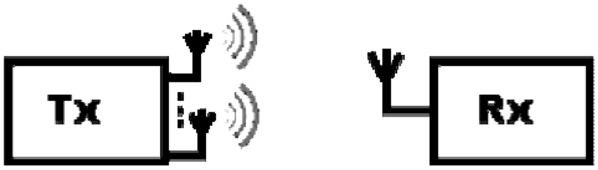


Fig. 4.8 MISO - Multiple Input Single Output

It has M transmitting antennas. The total power is divided into M transmitter branches. If the signals add coherently at the receiving antenna, we get an M - fold increase in SNR as compared to SISO. Because there is only one receiving antenna, the noise level is the same as SISO. The overall increase in SNR is approximately.

$$SNR \approx \frac{M^2 * signal\ power/M}{noise} = M * SNR_0$$

MIMO (Multiple Input Multiple Output)

MIMO is effectively a radio antenna technology as it uses multiple antennas atthe transmitter and receiver to enable a variety of signal paths to carry the data,choosing separate paths for each antenna to enable multiple signal paths to beused.

The two main formats for MIMO are given below:

•**Spatial diversity:**

Spatial diversity used in this narrower sense often refers to transmit and receive diversity. These two methodologies are used to provide improvements in the signal to noise ratio and they are characterized by improving the reliability of the system with respect to the various forms of fading.

•**Spatial multiplexing:**

This form of MIMO is used to provide additional data capacity by utilizing the different paths to carry additional traffic, i.e. increasing the data throughput capability.

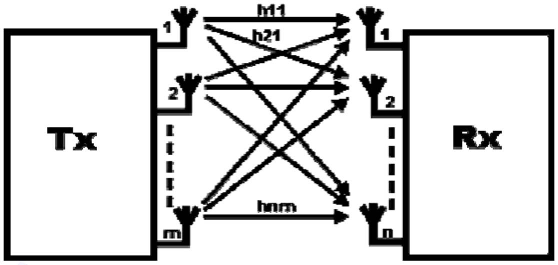


Fig. 4.9 MIMO - Multiple Input Multiple Output

MIMO systems can be viewed as a combination of MISO and SIMO channels. In this case, it is possible to achieve approximately an MN -fold increase in the average SNR 0 giving a channel capacity equal to

$C \approx B \log2 (1 + M \times N \times SNR0)$

4.5.3 OFDM MIMO SYSTEMS

Multiple-input multiple-output (MIMO) wireless technology in combination with orthogonal frequency division multiplexing (MIMOOFDM) is an air-interface solution for next-generation wireless local area networks (WLANs), wireless metropolitan area networks (WMANs), and fourth-generation mobile cellular wireless systems. OFDM and MIMO techniques can be combined to achieve high spectral efficiency and increased throughput.

For high-data rate transmissions, the MIMO channel is frequency selective (multipath). OFDM can transform such channel into a set of parallel frequency-flat channels (reduce Rx complexity).

4.5.3.aMIMO-OFDM Transmitter

The Source bit stream is encoded by the FEC encoder and the coded bitstream mapped to a constellation by digital modulator, and encoded by the MIMO encoder. Each of the parallel output symbol streams are corresponding to a certain Transmitting antenna follows the same Transmitting process:

- Insertion of pilot symbols (synchronization)
- Modulation by inverse FFT
- Attachment of CP and Preamble

Finally, the data frame is transferred to IF/RF stage for Transmitter

4.5.3.bMIMO-OFDM Receiver

The received symbol stream from different Receiving antennas is fist synchronized Preambles and CPs must be extracted from Received symbol stream. The Remaining OFDM symbols demodulated by FFT. Frequency pilots are extracted from the demodulated OFDM symbols, and are used for channel estimation. Estimated channel matrix aids the MIMO decoder and the estimated Transmitted symbols are demodulated and then decoded. All MIMO-OFDM receiversmust perform time synchronization; frequency offset estimation, and correction and parameter estimation. This isgenerally carried out using a preamble consisting of one ormore training sequences. Once the acquisition phase is over, receiver goes into the tracking mode.

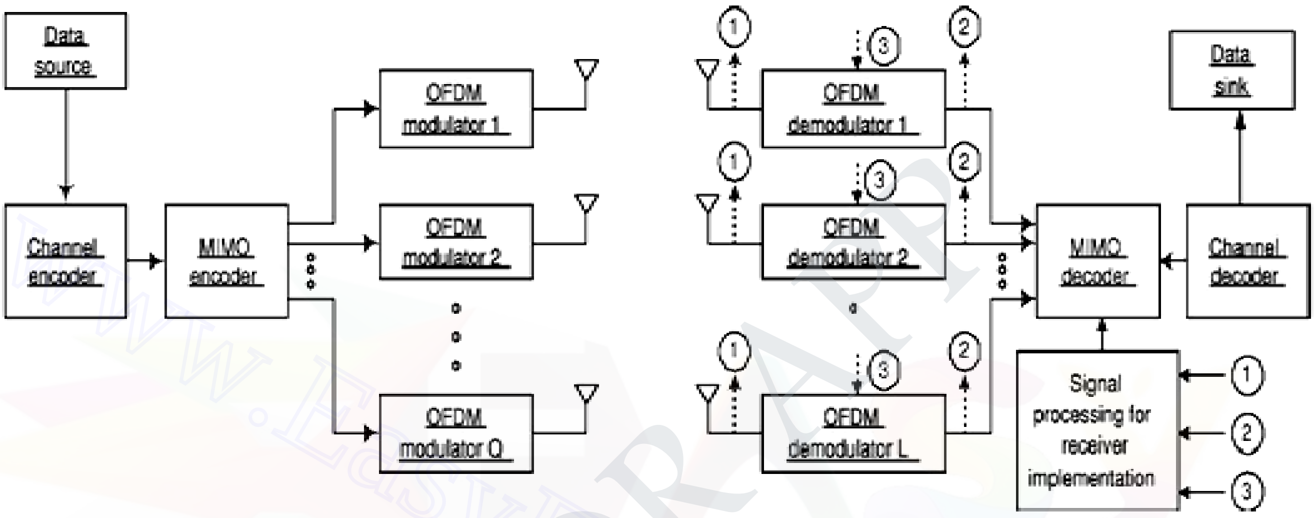


Fig. 4.10 Block Diagram MIMO – OFDM Frame Structure

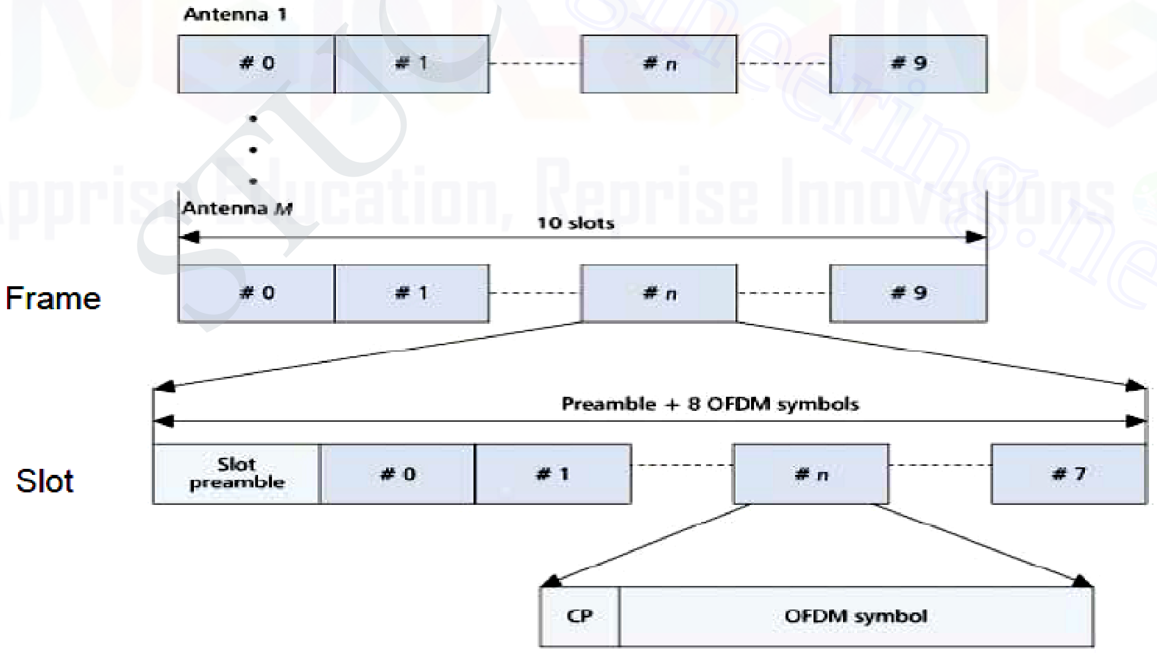


Fig. 4.11 MIMO – OFDM Frame Structure

In the time domain, a frame is a minimum transmission unit that includes 10 slots. Each slot consists of 1 slot preamble and 8 OFDM symbols. The preamble is used for time synchronization. Each OFDM symbol in a slot is attached to a CP that is used to reduce ISI and simplify channel equalizer. The frame is structured such that data and pilot symbols are transmitted over subcarriers (timing phase, timing frequency, and frequency offset estimation)

4.5.3.cSIGNALING IN MIMO-OFDM SYSTEMS

The signaling schemes used in MIMO systems can be grouped into spatial multiplexing which realizes capacity gain, and space-time coding, which improves link reliability through diversity gain. Most multi-antenna signaling schemes, in fact, realize both spatial-multiplexing and diversity gain. A framework for characterizing the trade-off between spatial-multiplexing and diversity gains in flat-fading MIMO channels.

Spatial Multiplexing multiplexes multiple spatial channels to send as many independent data as possible over different antennas.

There are four types of spatial multiplexing schemes:

- Diagonal BLAST
- Horizontal BLAST
- V-BLAST and
- Turbo BLAST

The method to estimate Transmitting signals has three steps:

- Estimate the channel matrix (training sequence)
- Determine optimal detecting order and nulling vectors
- Detect the received signals based on optimal detecting order and successive interference cancellation

4.5.3.dSPATIAL MULTIPLEXING IN MIMO-OFDM SYSTEMS

In an OFDM-based MIMO system, spatial multiplexing is performed by transmitting data streams on a tone-by-tone basis with the total transmit power split uniformly across antennas and tones. Although the use of OFDM eliminates ISI, the computational complexity of MIMO-OFDM spatial-multiplexing receivers can still be high. Computational complexity reductions are

attained by performing channel inversion in the case of a minimum mean-squared error (MMSE) receiver on a subset of tones only and computing the remaining inverses or QR factors, respectively, through interpolation.

4.5.3.eMIMO – OFDM STANDARDS

1. IEEE 802.11n (MIMO) Systems
2. IEEE 802.11a (OFDM) Systems
3. IEEE 802.11a & g (WLAN) Systems
4. IEEE 802.11a & g (WMAN) Systems
5. IEEE 802.16a (WiMAX) Systems

Advantage of OFDM – MIMO systems are:

1. High spectral efficiency& capacity
2. Simple implementation by FFT (fast Fourier transform);
3. Low receiver complexity;
4. Robustability for high-data-rate transmission over multipath fading channel
5. High flexibility in terms of link adaptation
6. Low complexity multiple access schemes such as orthogonal frequency division multiple access.

Applications of OFDM – MIMO systems are:

- Wireless network
- Next generation network(4G)
- Wi-Fi, Wi-MAX, W – MAN
- Digital TV
- Power line control
- Digital audio and video broadcasting
- Discrete Multitone systems

4.5.4 ADAPTIVE MODULATION AND CODING WITH TIME SLOT SCHEDULER

Adaptive modulation or link adaptation is used to improve the spectral efficiency particularly over wireless fading channels. Adaptive modulation offers parameters such as data rate, transmit power, instantaneous BER, symbol rate, and channel code rate to be adjusted relative to the channel fading, by exploiting the channel information that is present at the transmitter. AMC is one of the most important RRM mechanisms that have been used to improve system capacity. AMC adapts the modulation and coding scheme (MCS) according to the channel condition. The channel condition can be reported back by the UE by using Channel Quality Indicator (CQI).

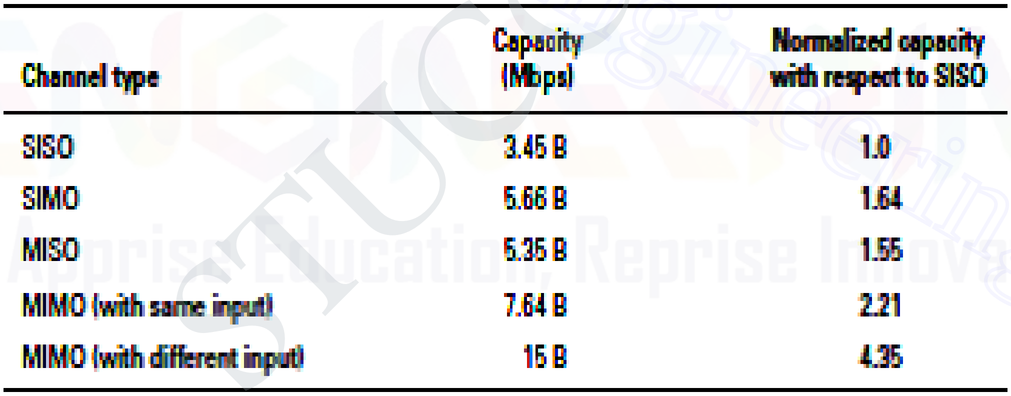
AMC provides the flexibility to match the modulation coding scheme to the average channel conditions for each user. With AMC, the power of the transmitted signal is held constant over a frame interval, and the modulation and coding format is changed to match the current received signal quality or channel conditions.

The implementation of AMC offers several challenges. First, AMC is sensitive to measurement error and delay. In order to select the appropriate modulation, the scheduler must be aware of the channel quality. Errors in the channel estimate will cause the scheduler to select the wrong data rate and either transmits at too high a power, wasting system capacity, or too low a power, raising the block error rate. Delay in reporting channel measurements also reduces the reliability of the channel quality estimate due to the constantly varying mobile channel. Furthermore changes in the interference add to the measurement errors.

A wireless network uses a time-varying channel where packet losses occur due to severe fading. This is misinterpreted by TCP as congestion which leads to inefficient utilization of the available radio link capacity. This results in significant degradation of the wireless system performance.

There is a need for a system with efficient packet data transmission using TCP in 4G. This can be achieved by using a suitable automatic repeat request (ARQ) scheme combined with an adaptive modulation and coding system, and a time-slot scheduler that uses channel predictions. This way, the lower layers are adapted to channel conditions while still providing some robustness through retransmission. The time-slot scheduler shares the spectrum efficiently between users while satisfying the quality-of-service (QoS) requirements.

Table 4.5.3.1 Comparison of channel capacity for different channel types



Channel type	Capacity (Mbps)	Normalized capacity with respect to SISO
SISO	3.45 B	1.0
SIMO	5.66 B	1.64
MISO	5.35 B	1.55
MIMO (with same input)	7.64 B	2.21
MIMO (with different input)	15 B	4.35

4.5.4.aSYSTEM MODEL

Adaptive modulation systems invariably require some Channel State Information (CSI) at the transmitter. This can be achieved by estimating and predicting the channel conditions at the receiver and fed back to the transmitter, so that the transmission scheme can be adapted relative to the channel characteristics.

During each transmission, the modulation scheme is adjusted to maximize the spectral efficiency, under BER and average powerconstraints, based on the instantaneous predictedSINR. The various modulation techniques such asQPSK and M-ary Quadrature AmplitudeModulation (M-QAM) schemes with differentconstellation sizes are provided at the transmitter.

The link adaptation can employ QPSKfor noisychannels, which are more robust and can tolerate higher levels of interference but has lowertransmission bit rate. M-QAMis adapted for clearer channels, and has twice higher bit rate but is moreprone to errors due to interference and noise. To improve the quality of the wireless link the transmitter uses some form of channel coding. The coding can either be in the traditional form ofcoding followed by modulation (each done independent of the other) or joint coding and modulation.

Coding addsredundant bits to the data bits which can correcterrors in the received bits. The degree of coding is determined by its rate, which is the proportion ofdata bits to coded bits. This typically varies from 1/8 to 4/5.

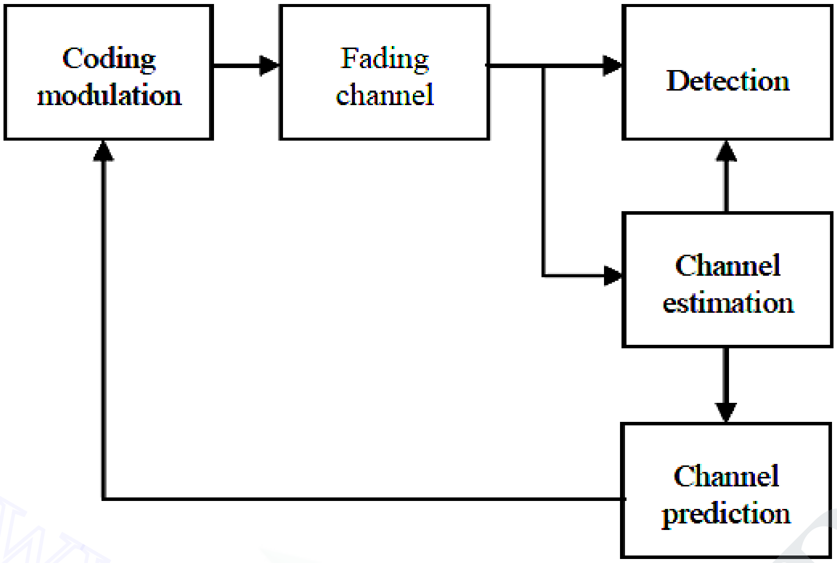


Fig.4.12 Block diagram for Adaptive Modulation & Coding

If the channel quality for each radio link can be predicted for a short duration (say about 10 ms) into the future and accessible by the link layer, then ARQ along with an adaptive modulation and coding system can be selected for each user to satisfy the bit error rate (BER) requirement and provide high throughput. The scheduler uses this information aboutindividual data streams (along with predicted values of different radio links and modulation and coding systems by the link layer) and distributes the time slots among the users. The planning is done so that the desired QoS and associated priority to different users are guaranteed while channel spectrum is efficiently utilized.

4.5.5 COGNITIVE RADIO

Cognitive radio (CR) is a form of wireless communication in which a transceiver can intelligently detect which communication channels are in use and which are not, and instantly move into vacant channels while avoiding occupied ones. This optimizes the use of available radio-frequency (RF) spectrum while minimizing interference to other users. CR is a hybrid technology involving software defined radio (SDR) as applied to spread spectrum communications. Possible functions of cognitive radio include the ability of a transceiver to determine its geographic location, identify and authorize its user, encrypt or decrypt signals, sense neighboring wireless devices in operation, and adjust output power and modulation characteristics.

There are two main types of cognitive radio,

1. Full cognitive radio and
2. Spectrum-sensing cognitive radio.

Full cognitive radio takes into account all parameters that a wireless node or network can be aware of. Spectrum-sensing cognitive radio is used to detect channels in the radio frequency spectrum.

Cognitive radio can able to monitor sense and detect the conditions of their operating environment, and reconfigure their own characteristics dynamically to match those conditions.

The CR can be viewed as an enabling technology that will benefit several types of users by introducing new communications and networking models for the whole wireless world, creating better business opportunities for the operators and new technical dimensions for smaller operators, and helping shape an overall more efficient approach regarding spectrum requirements and usage in the next generation of wireless networks.

.

The primary objectives of the cognitive radio are to provide highly reliable communications whenever and wherever needed and to utilize the radio spectrum efficiently. The key issues in the cognitive radio are awareness, intelligence, learning, adaptability, reliability, and efficiency.

The three major tasks of the cognitive radio include [:

(1) Radio-scene analysis,

(2) Channel identification, and

(3) Dynamic spectrum management and transmit-power control.

The radio-scene analysis includes the detection of spectrum holes by for example sensing the radio frequency spectrum. The channel identification includes estimation of the channel state information which is needed at the receiver for coherent detection.

The transmitter power controland dynamic spectrum management select the transmission power levels and frequency holes for transmission based on the results of radio scene analysis and channel identification. The first two tasks are carried out in the receiver (RX) while the third task is carried out in the transmitter (TX), which requires some form of feedback between RX and TX.

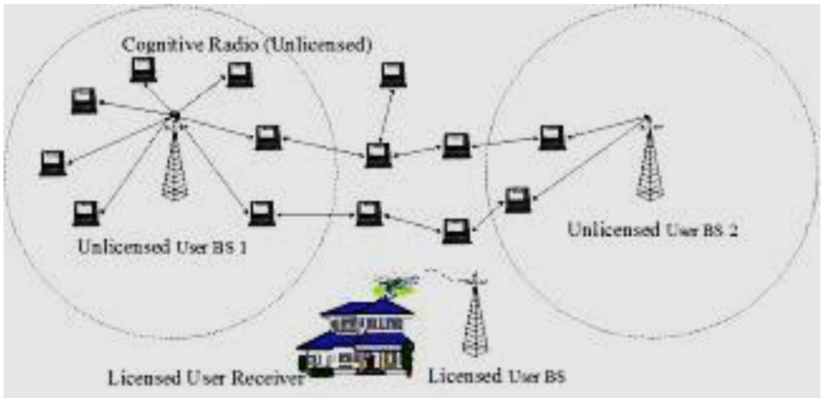


Fig.4.13 Cognitive Radio

CR ARCHITECTURE

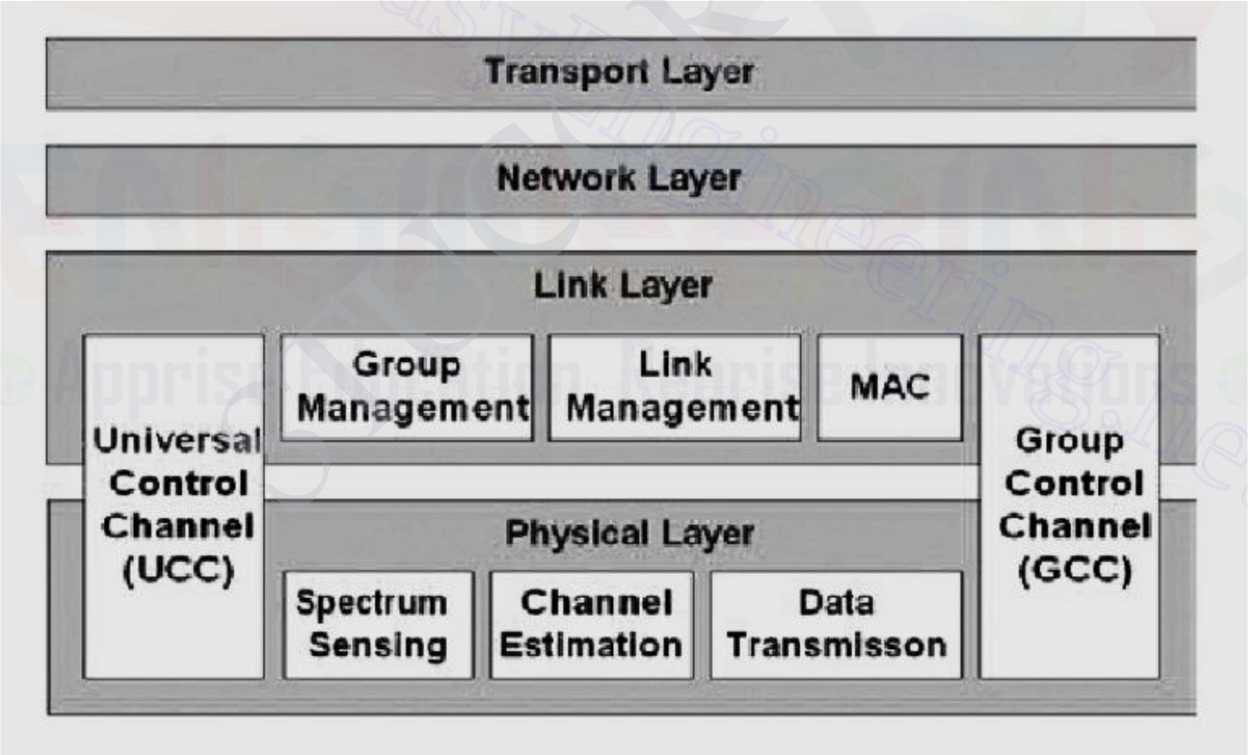


Fig.4.14 Architecture of CR

The main function of the physical layer is to sense the spectrum over all available degrees of freedom (time, frequency and space) in order to identify sub-channels currently available for transmission. In order to set up the link, channel sounding is used to estimate the quality of sub-channels between SUs that want to communicate. The transmission parameters (transmit power, bit rate, coding, etc.) are determined based on the channel sounding results. Link Management covers the setup of a link in order to enable the communication between two SUs and afterwards the maintenance of this SU Link for the duration of the communication.

Link Management covers the setup of a link in order to enable the communication between two SUs and afterwards the maintenance of this SU Link for the duration of the communication. Medium Access Control as long as it can be assured that all Sub-Channels are used exclusively, i.e. all Sub-Channels used by one SU Link cannot be used by any other SU Link this problem comes down to a simple token-passing algorithm ensuring that only one of the two communication peers is using the link.

CR's optimally uses the available spectrum as determined by the spectrum sensing and channel estimation functions. Therefore it should have the ability to operate at variable symbol rates, modulation formats (e.g. low to high order QAM), different channel coding schemes, power levels and be able to use multiple antennas for interference nulling, capacity increase MIMO or range extensions. In group management, it is assumed that any secondary station will belong to a SU Group. A newly arriving user can either join one of the existing groups or create a new one through the Universal Control Channel.

4.5.5.aFUNCTIONS OF CR

The main functions of Cognitive Radios are:

Spectrum Sensing: It detects the unused spectrum and sharing it without harmful interference with other users.

Spectrum sensing techniques can be classified into three categories:
Transmitter detection: Cognitive radios must have the capability to determine if a signal from a primary transmitter is locally present in a certain spectrum, there are several approaches proposed: matched filter detection energy detection

Match filter detection is obtained by correlating a known signal, or template, with an unknown signal to detect the presence of the template in the unknown signal.
Cooperative detection refers to spectrum sensing methods where information from multiple Cognitive radio users is incorporated for primary user detection.

4.5.5.bISSUES IN CR

- Spectrum utilization: presence of white spaces
- Spectral co-existence
- Spectrum sharing
- Spectrum management

4.5.5.cAPPLICATIONS

- **Incumbent Wireless Carriers:** It allows incumbent wireless providers to increase the capacity of their networks by reducing the interference.
- **Rural Broadband Telecommunications:** It can help the local providers to overcome the obstacle of high cost/ scarcity of spectrum to deploying wireless in undeserved rural markets.
- **Defense/Military Applications:** Satisfies the need for a mobile interference – resistant scalable, frequency agile, cost efficient wireless system.
- **Public Safety:** Provides resilience and continuity of operations in wireless impaired environment that is not available from commercial wireless system.
- **Utilities/ Smart Grids**: Provides an efficient means of measuring and reporting usage.
- **Mobile Content Providers**: Gives content providers to control the content of their own delivery systems.
- **Cable Markets**: Allows operators without spectrum to have a competitive entry in to the wireless market.