UNIT - 1

WIRELESS LAN

PART – A

1. **What is infrared (IR) transmission?**
Infrared light transmission is one of the important technologies used in wireless LAN. It is based on the transmission of infrared light at 900 nm wavelength.

2. **What is the principle behind infrared technology?**
Infrared technology uses diffuse light reflected at walls, furniture etc. Or directed light when line of sight (LOS) exists between sender and receiver.

3. **What are the advantages of infrared technology?**
Shielding is easy and no need for license for infrared technology. Electrical devices do not interfere with infrared transmission.

4. **What are the disadvantages of infrared transmission?**
Low bandwidth Cannot penetrate through walls or other obstacles.

5. **Define – Spread Spectrum**
Spread spectrum involves spreading the bandwidth needed to transmit data. The main advantage of using spread spectrum is the resistance to narrow interference.

6. **What are the spread spectrum techniques?**
There are two basic methods for spread spectrum transmissions. Direct Sequence Spread Spectrum (DSSS) Frequency Hopping Spread Spectrum (FHSS)

7. **What is the principle behind FHSS?**
Frequency Hopping Spread Spectrum is evolved in order to avoid jamming. Hence in this method, the transmitter shifts the center frequency of transmitted signal. The shifts in frequency or frequency hops, occur

according to a random pattern is known only to the transmitter and receiver.

8. **What are the major issues in WLAN?**
   Two major issues in WLAN are as follows Hidden station problem Exposed station problem

9. **List out the applications of WLAN.**
   Transfer of medical images Remote access to patient records Remote monitoring of patients Remote diagnosis of patients at home or in an ambulance In telemedicine Surveillance Internet supporting database.

10. **What is IEEE 802.11?**
    The IEEE 802.11 is the first WLAN standard that has secured the market in large extent. The primary goal of the standard was the specification of a simple and robust that offers time bounded and asynchronous services.

11. **Define Short Inter Frame Space (SIFS)**
    Short IFS is the shortest IFS used for the high priority frames like acknowledgement frames, CTS frames, poll response etc.

12. **Define Distributed coordination function Inter Frame Space (DIFS)**
    DCF-IFS is used for transmitting data frames. It is equal to SIFS plus two time slots and is the longest inter frame gap.

13. **What are the functions of MAC layer in IEEE 802.11?**
    The functions of MAC layer are Media Access Control Reliable delivery of data units Management functions Authentication encryption

14. **What is the need for WATM?**
    WATM systems had to be designed for transferring voice, classical data, video, multimedia etc.

15. **What is HIPERLAN?**
    The HIPERLAN stands for High PERformance Radio LAN is an initiation of RES-10 group of the ETSI as a PAN European standard for high speed wireless local networks.

16. Give any two requirements of HIPERLAN.

Data rates of 23.529 Mbps Multi-hop and Ad-hoc networking Support of time bounded services

17. What are the three phases in channel access in HIPERLAN-1?

Prioritization phase Contention phase Transmission phase

18. Give any three differences between HIPERLAN 1 and HIPERLAN 2.

HIPERLAN 1 HIPERLAN 2 Application Wireless LAN Access to ATM fixed networks Range 50 m 50 – 100 m Data rate 23.5 M bits/sec > 20 M bits/s

19. What is meant by BRAN?

The BRAN (Broadband Radio Access Networks (BRAN) is standardized by the European Telecommunications Standards Institute (ETSI). Primary motivation of BRAN is the deregulation and privatization of the telecommunication sector. BRAN technology is independent from the protocols of the fixed network. BRAN can be used for ATM and TCP/IP networks.

20. List the functional requirements of HIPERLANS.

The functional requirements of HIPERLAN are Data rates of 23.529 Mbps Support both synchronous and asynchronous traffic Power saving support Video at 2 Mbps, 100 ns latency and audio at 32 Kbps, 10 ns latency To coverage multihub features Low mobility of 1.4 m/s Support of time bound services Asynchronous file transfer at 13.4 Mbps

21. What is Bluetooth?

Bluetooth is an inexpensive personal area Ad-hoc network operating in unlicensed bands and owned by the user. It is an open specification for short range wireless voice and data communications that was developed for cable replacement in PAN (Personal Area Network).

22. What is the advantage of piconet /scatternet. ?

The advantage of the Piconet / Scatternet scheme is that it allows many devices to share the same physical area and make efficient use of bandwidth.

23. What are the states of Bluetooth?

Bluetooth specifies four states, they are Master-M SlaveS StandbySB ParkedP

24. List the two major states in the operation of Bluetooth.

The major states in the operation of Bluetooth are Standby state Connection state

25. What is Piconet and Scatternet?

Bluetooth specification defines a small cell called as piconet which has upto 8 devices grouped together. Two or more piconets grouped together know as scatternet.

26. What type of modulation used in Bluetooth?

Bluetooth uses Gaussian-shaped Frequency Shift Keying (GFSK) modulation with a nominal modulation index of K = 0.3

27. What is the data rate of Bluetooth?

The maximum data rate is 721Mbps for asymmetric mode.

28. List the logical channels provided by L2CAP.

L2CAP provides three types of logical channels. They are: Connectionless Connection oriented Signaling

29. What is the need for WIMAX?

The main reason for the development of WIMAX(World Interoperability Microwave Access) is the demand of high data rates not only the faster downloading but also for the use of new applications like Voip, Video, streaming multimedia conferencing and interactive gaming.

30. What is WIMAX?

WIMAX is the air interface for the actual radio interface network, where both fixed and mobile users can have access to the network. Its specification is IEEE 802.16.

31. Write the throughput feature in WIMAX.

WIMAX supports throughput up to 63 Mbps on the downlink and 28 Mbps on the uplink, assuming a 10 MHZ bandwidth channel with TDD frames and with 64 QAM 5/6 as modulation scheme.

32. What are the frequency bands of IEEE 802.16?

The 802.16 standard defines a number of air interfaces that can be divided into, 10-66 GHz licensed band Below 11 GHz licensed bands Below 11 GHZ unlicensed bands

PART – B

1. Discuss about the Advantages and Disadvantages of Wireless LAN.

WLANs are typically restricted in their diameter to buildings, a campus, single rooms etc. and are operated by individuals, not by large-scale network providers. The global goal of WLANs is to replace office cabling, to enable tetherless access to the internet and, to introduce a higher flexibility for ad-hoc communication in, e.g., group meetings.

Some **advantages** of WLANs are:

- **Flexibility:** Within radio coverage, nodes can communicate without further restriction. Radio waves can penetrate walls, senders and receivers can be placed anywhere (also non-visible, e.g., within devices, in walls etc.). Sometimes wiring is difficult if firewalls separate buildings (real firewalls made out of, e.g., bricks, not routers set up as a firewall). Penetration of a firewall is only permitted at certain points to prevent fire from spreading too fast.
- **Planning:** Only wireless ad-hoc networks allow for communication without previous planning, any wired network needs wiring plans. As long as devices follow the same standard, they can communicate. For wired networks, additional cabling with the right plugs and probably interworking units (such as switches) have to be provided.
- **Design:** Wireless networks allow for the design of small, independent devices which can for example be put into a pocket. Cables not only restrict users but also designers of small PDAs, notepads etc. Wireless senders and receivers can be hidden in historic buildings, i.e., current networking technology can be introduced without being visible.
- **Robustness:** Wireless networks can survive disasters, e.g., earthquakes or users pulling a plug. If the wireless devices survive, people can still communicate. Networks requiring a wired infrastructure will usually break down completely.

- **Cost:** After providing wireless access to the infrastructure via an access point for the first user, adding additional users to a wireless network will not increase the cost. This is, important for e.g., lecture halls, hotel lobbies or gate areas in airports where the numbers using the network may vary significantly. Using a fixed network, each seat in a lecture hall should have a plug for the network although many of them might not be used permanently. Constant plugging and unplugging will sooner or later destroy the plugs. Wireless connections do not wear out.

But WLANs also have several **disadvantages**:

- **Quality of service:** WLANs typically offer lower quality than their wired counterparts. The main reasons for this are the lower bandwidth due to limitations in radio transmission (e.g., only 1-10 Mbit/s user data rate instead of 100-1,000 Mbit/s), higher error rates due to interference (e.g., 10^{-4} instead of 10^{-12} for fiber optics), and higher delay/delay variation due to extensive error correction and detection mechanisms.

- **Proprietary solutions:** Due to slow standardization procedures, many companies have come up with proprietary solutions offering standardized functionality plus many enhanced features (typically a higher bit rate using a patented coding technology or special inter-access point protocols). However, these additional features only work in a homogeneous environment, i.e., when adapters from the same vendors are used for all wireless nodes. At least most components today adhere to the basic standards IEEE 802.11b or 802.11a .

- **Restrictions:** All wireless products have to comply with national regulations. Several government and non-government institutions worldwide regulate the operation and restrict frequencies to minimize interference. Consequently, it takes a very long time to establish global solutions like, e.g., IMT-2000, which comprises many individual standards.WLANs are limited to low-power senders and certain license-free frequency bands, which are not the same worldwide.

- **Safety and security:** Using radio waves for data transmission might interfere with other high-tech equipment in, e.g., hospitals. Senders and receivers are operated by laymen and, radiation has to be low. Special pre-cautions have to be taken to prevent safety hazards. The open radio interface makes eavesdropping much easier in WLANs than, e.g., in the case of fiber optics. All standards must offer (automatic) encryption, privacy mechanisms, support for anonymity etc.

2. **Discuss the design goals to be taken into account for WLANs to ensure their commercial success.**

- **Global operation:** WLAN products should sell in all countries so, national and international frequency regulations have to be considered.
- **Low power:** Devices communicating via a WLAN are typically also wireless devices running on battery power. The LAN design should take this into account and implement special power-saving modes and power management functions.
- **License-free operation:** The equipment must operate in a license-free band, such as the 2.4 GHz ISM band.
- **Robust transmission technology:** Compared to their wired counterparts, WLANs operate under difficult conditions. If they use radio transmission, many other electrical devices can interfere with them . WLAN transceivers cannot be adjusted for perfect transmission in a standard office or production environment.
- Antennas are typically omnidirectional, not directed. Senders and receivers may move.
- **Simplified spontaneous cooperation:** To be useful in practice, WLANs should not require complicated setup routines but should operate spontaneously after power-up.
- **Easy to use:** In contrast to huge and complex wireless WANs, wireless LANs are made for simple use. They should not require complex management, but rather work on a plug-and-play basis.
- **Protection of investment:** A lot of money has already been invested into wired LANs. The new WLANs should protect this investment by being interoperable with the existing networks. This means that simple bridging between the different LANs should be enough to interoperate, i.e., the wireless LANs should support the same data types and services that standard LANs support.

- **Safety and security:** Wireless LANs should be safe to operate, especially regarding low radiation if used, e.g., in hospitals. Users cannot keep safety distances to antennas. The equipment has to be safe for pacemakers, too. Users should not be able to read personal data during transmission, i.e., encryption mechanisms should be integrated. The networks should also take into account user privacy, i.e., it should not be possible to collect roaming profiles for tracking persons if they do not agree.
- **Transparency for applications:** Existing applications should continue to run over WLANs, the only difference being higher delay and lower bandwidth. The fact of wireless access and mobility should be hidden if it is not relevant, but the network should also support location aware applications, e.g., by providing location information.

3. Explain the principle of Infra red and radio transmission technologies and also mention its advantages and disadvantages.

Today, two different basic transmission technologies can be used to set up WLANs. One technology is based on the transmission of infra red light (e.g., at 900 nm wavelength), the other one, which is much more popular, uses radio transmission in the GHz range (e.g., 2.4 GHz in the license-free ISM band). Both technologies can be used to set up ad-hoc connections for work groups, to connect, e.g., a desktop with a printer without a wire, or to support mobility within a small area.

Infra red technology uses diffuse light reflected at walls, furniture etc. or directed light if a line-of-sight (LOS) exists between sender and receiver. Senders can be simple light emitting diodes (LEDs) or laser diodes. Photodiodes act as receivers.

- The main **advantages** of infra red technology are its simple and extremely cheap senders and receivers which are integrated into nearly all mobile devices available today. PDAs, laptops, notebooks, mobile phones etc. have an infra red data association (IrDA) interface. Version 1.0 of this industry standard implements data rates of up to 115 kbit/s, while IrDA 1.1 defines higher data rates of 1.152 and 4 Mbit/s. No licenses are needed for infra red technology and shielding is very simple. Electrical devices do not interfere with infra red transmission.

- **Disadvantages** of infra red transmission are its low bandwidth compared to other LAN technologies. Typically, IrDA devices are internally connected to a serial port limiting transfer rates to 115 kbit/s. Even 4 Mbit/s is not a particularly high data rate. However, their main disadvantage is that infra red is quite easily shielded. Infra red transmission cannot penetrate walls or other obstacles. Typically, for good transmission quality and high data rates a LOS, i.e., direct connection, is needed.

- **Advantages** of radio transmission include the long-term experiences made with radio transmission for wide area networks (e.g., microwave links) and mobile cellular phones. Radio transmission can cover larger areas and can penetrate (thinner) walls, furniture, plants etc. Additional coverage is gained by reflection. Radio typically does not need a LOS if the frequencies are not too high. Furthermore, current radio-based products offer much higher transmission rates (e.g., 54 Mbit/s) than infra red (directed laser links, which offer data rates well above 100 Mbit/s). ● Again, the main advantage is also a big **disadvantage** of radio transmission. Shielding is not so simple. Radio transmission can interfere with other senders, or electrical devices can destroy data transmitted via radio. Additionally, radio transmission is only permitted in certain frequency bands. Very limited ranges of license-free bands are available worldwide and those that are available are not the same in all countries.

Of the three WLAN technologies Only one (IEEE 802.11) standardized infra red transmission in addition to radio transmission. The other two (HIPERLAN and Bluetooth) rely on radio. The main reason for this are the shielding problems of infra red. WLANs should, e.g., cover a whole floor of a building and not just the one room where LOSs exist. Future mobile devices may have to communicate while still in a pocket or a suitcase so cannot rely on infra red. The big advantage of radio transmission in everyday use is indeed the ability to penetrate certain materials and that a LOS is not required. Many users experience a lot of difficulties adjusting infra red ports of, e.g., mobile phones to the infra red port of their PDA. Using, e.g., Bluetooth is much simpler.

4. Explain the architecture used in IEEE 802.11 Infrastructure and ad-hoc networks in detail.

Many WLANs of today need an **infrastructure** network. Infrastructure networks not only provide access to other networks, but also include forwarding functions, medium access control etc. In these infrastructure-based wireless networks, communication typically takes place only between the wireless nodes and the access point (see Figure 1.1), but not directly between the wireless nodes.

The access point does not just control medium access, but also acts as a bridge to other wireless or wired networks. Figure 1.1 shows three access points with their three wireless networks and a wired network. Several wireless networks may form one logical wireless network, so the access points together with the fixed network in between can connect several wireless networks to form a larger network beyond actual radio coverage.

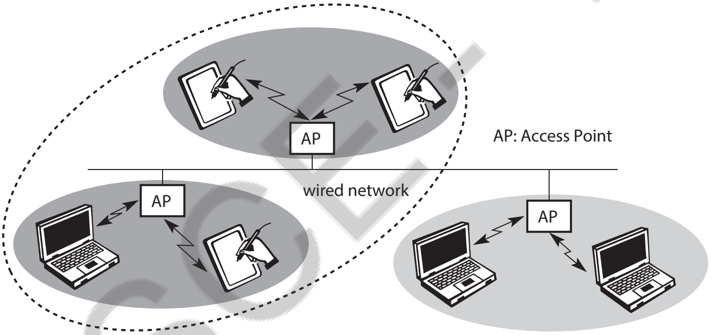


Figure 1.1 Example of three infrastructure-based wireless networks

Typically, the design of infrastructure-based wireless networks is simpler because most of the network functionality lies within the access point, whereas the wireless clients can remain quite simple. This structure is reminiscent of switched Ethernet or other star-based networks, where a central element (e.g., a switch) controls network flow. This type of network can use different access schemes with or without collision. Collisions may occur if medium access of the wireless nodes and the access point is not coordinated. However, if only the access point controls medium access, no collisions are possible. This setting may be useful for quality of service guarantees such as minimum bandwidth for certain nodes. The access point may poll the single wireless nodes to ensure the data rate.

Infrastructure-based networks lose some of the flexibility wireless networks can offer, e.g., they cannot be used for disaster relief in cases where no infrastructure is left. Typical cellular phone networks are infrastructure based networks for a wide area . Also satellite-based cellular phones have an infrastructure the satellites. Infrastructure does not necessarily imply a wired fixed network.

Ad-hoc wireless networks, however, do not need any infrastructure to work. Each node can communicate directly with other nodes, so no access point controlling medium access is necessary. Figure 1.2 shows two ad-hoc networks with three nodes each. Nodes within an ad-hoc network can only communicate if they can reach each other physically, i.e., if they are within each other's radio range or if other nodes can forward the message. Nodes from the two networks shown in Figure 1.2 cannot, therefore, communicate with each other if they are not within the same radio range.

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

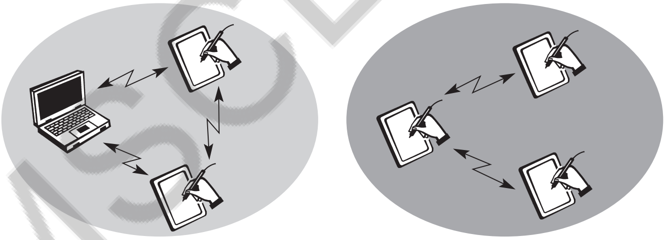


Figure 1.2 Example of two ad-hoc wireless networks

Clearly, the two basic variants of wireless networks , infrastructure-based and ad-hoc, do not always come in their pure form. There are networks that rely on access points and infrastructure for basic services (e.g., authentication of access, control of medium access for data with associated quality of service, management functions), but that also allow for direct communication between the wireless nodes.

However, ad-hoc networks might only have selected nodes with the capabilities of forwarding data. Most of the nodes have to connect to such a special node first to transmit data if the receiver is out of their range.

From the three WLANs presented, IEEE 802.11 and HiperLAN2 are typically infrastructure-based networks, which additionally support ad-hoc networking. However, many implementations only offer the basic infrastructure-based version. The third WLAN, Bluetooth , is a typical wireless ad-hoc network. Bluetooth focuses precisely on spontaneous ad-hoc meetings or on the simple connection of two or more devices without requiring the setup of an infrastructure.

5. Explain in detail about IEEE 802.11 with neat illustrations.

(OR)

Explain the System architecture and Protocol architecture of IEEE 802.11 in detail.

The IEEE standard 802.11 (IEEE, 1999) specifies the most famous family of WLANs in which many products are available. As the standard's number indicates, this standard belongs to the group of 802.x LAN standards, e.g., 802.3 Ethernet or 802.5 Token Ring. This means that the standard specifies the physical and medium access layer adapted to the special requirements of wireless LANs, but offers the same interface as the others to higher layers to maintain interoperability.

The primary goal of the standard was the specification of a simple and robust WLAN which offers time-bounded and asynchronous services. The MAC layer should be able to operate with multiple physical layers, each of which exhibits a different medium sense and transmission characteristic. Candidates for physical layers were infra red and spread spectrum radio transmission techniques.

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.

System architecture

Wireless networks can exhibit two different basic system architectures infrastructure-based or ad-hoc. Figure 1.3 shows the components of an infrastructure and a wireless part as specified for IEEE 802.11. Several nodes, called **stations (STA_i)**, 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 (BSS$_i$)**. The example shows two BSSs BSS$_1$ and BSS$_2$ 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. The distribution system connects the wireless networks via the APs with a **portal**, which forms the interworking unit to other LANs.

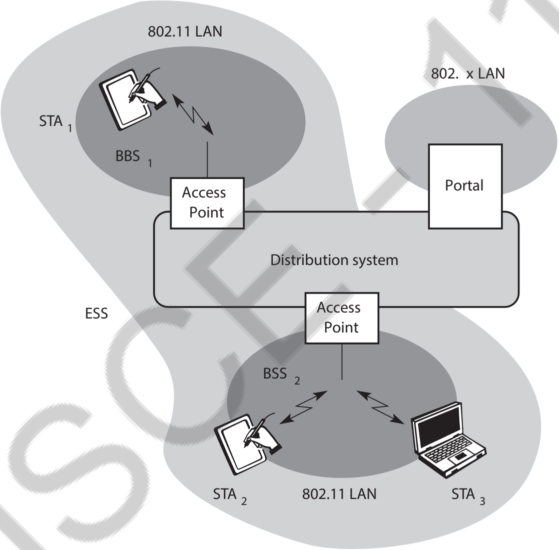


Figure 1.3 Architecture of an infrastructure-based IEEE 802.11

It could consist of bridged IEEE LANs, wireless links, or any other networks. However, **distribution system services** are defined in the standard Stations can select an AP and associate with it. The APs support roaming (i.e., changing access points), the distribution system handles data transfer between the different APs. APs provide synchronization within a BSS, support power management, and can control medium access to support time-bounded service.

In addition to infrastructure-based networks, IEEE 802.11 allows the building of ad-hoc networks between stations, thus forming one or more independent BSSs (IBSS) as shown in Figure 1.4. In this case, an IBSS comprises a group of stations using the same radio frequency. Stations STA_1, STA_2, and STA_3 are in $IBSS_1$, STA_4 and STA_5 in $IBSS_2$. This means for example that STA_3 can communicate directly with STA2 but not with STA5. Several IBSSs can either be formed via the distance between the IBSSs or by using different carrier frequencies (then the IBSSs could overlap physically). IEEE 802.11 does not specify any special nodes that support routing, forwarding of data or exchange of topology information as, e.g., HIPERLAN 1 or Bluetooth.

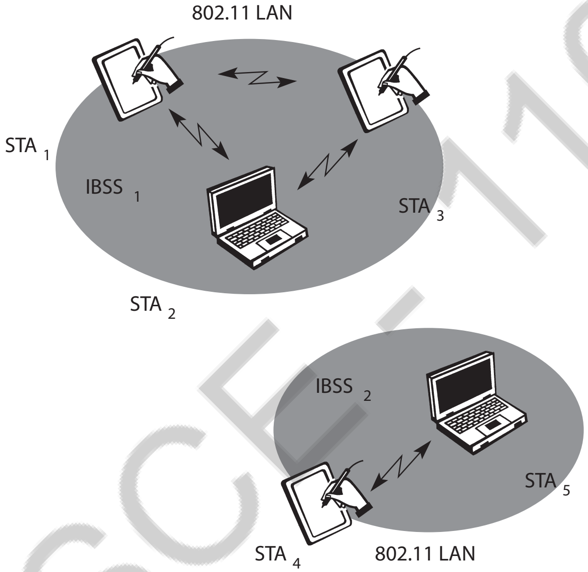


Figure 1.4 Architecture of IEEE 802.11 ad-hoc wireless LANs

Protocol architecture

As indicated by the standard number, IEEE 802.11 fits seamlessly into the other 802.x standards for wired LANs. Figure 1.5 shows the most common scenario: an IEEE 802.11 wireless LAN connected to a switched IEEE 802.3 Ethernet via a bridge. Applications should not notice any difference apart from the lower bandwidth and perhaps higher access time from the wireless LAN. The WLAN behaves like a slow wired LAN. Consequently, the higher layers (application, TCP, IP) look the same for wireless nodes as for wired nodes. 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 the different media. In many of today's networks, no explicit LLC layer is visible.

The IEEE 802.11 standard only covers the physical layer **PHY** and medium access layer **MAC** like the other 802.x LANs do. The physical layer is subdivided into the **physical layer convergence protocol (PLCP)** and the **physical medium dependent** sublayer **PMD** . The basic tasks of the MAC layer comprise medium access, fragmentation of user data, and encryption. The PLCP sublayer provides a carrier sense signal, called clear channel assessment (CCA), and provides a common PHY service access point (SAP) independent of the transmission technology. Finally, the PMD sublayer handles modulation and encoding/decoding of signals. The PHY layer (comprising PMD and PLCP) and the MAC layer will be explained in more detail in the following sections.

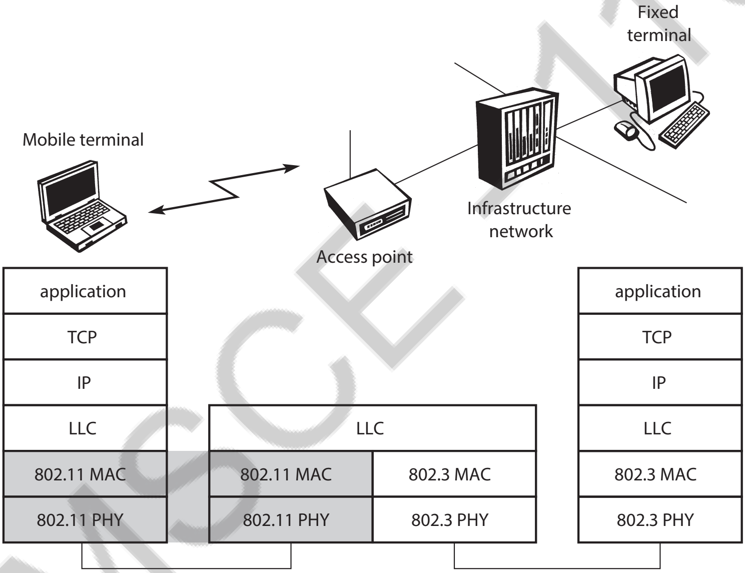


Figure 1.5 IEEE 802.11 protocol architecture and bridging

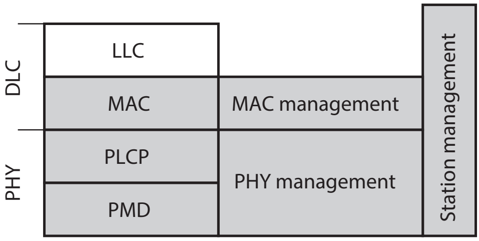


Figure 1.6 Detailed IEEE 802.11 protocol architecture and management

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

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

Physical layer

IEEE 802.11 supports three different physical layers: one layer based on infra red and two layers based on radio transmission (primarily in the ISM band at 2.4 GHz, which is available worldwide). All PHY variants include the provision of the **clear channel assessment** signal **(CCA)**. This is needed for the MAC mechanisms controlling medium access and indicates if the medium is currently idle. The transmission technology determines exactly how this signal is obtained.

The PHY layer offers a service access point (SAP) with 1 or 2 Mbit/s transfer rate to the MAC layer (basic version of the standard). The remainder of this section presents the three versions of a PHY layer defined in the standard.

Frequency hopping spread spectrum

Frequency hopping spread spectrum (FHSS) is a spread spectrum technique which allows for the coexistence of multiple networks in the same area by separating different networks using different hopping sequences.

The original standard defines 79 hopping channels for North America and Europe, and 23 hopping channels for Japan (each with a bandwidth of 1 MHz in the 2.4 GHz ISM band). The selection of a particular channel is achieved by using a pseudo-random hopping pattern. National restrictions also determine further parameters, e.g., maximum transmit power is 1 W in the US, 100 mW EIRP (equivalent isotropic radiated power) in Europe and 10 mW/MHz in Japan.

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 , a 4 level GFSK for 2 Mbit/s (i.e., 2 bits are mapped to one frequency). While sending and receiving at 1 Mbit/s is mandatory for all devices, operation at 2 Mbit/s is optional. This facilitated the production of low-cost devices for the lower rate only and more powerful devices for both transmission rates in the early days of 802.11.

Figure 1.7 shows a frame of the physical layer used with FHSS. The frame consists of two basic parts, the PLCP part (preamble and header) and the payload part. While the PLCP part is always transmitted at 1 Mbit/s, payload, i.e. MAC data, can use 1 or 2 Mbit/s. Additionally, MAC data is scrambled using the polynomial $s(z) = z^7 + z^4 + 1$ for DC blocking and whitening of the spectrum. 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 of potential receivers and signal detection by the CCA.
- **Start frame delimiter (SFD):** The following 16 bits indicate the start of the frame and provide frame synchronization. The SFD pattern is 0000110010111101.
- **PLCP_PDU length word (PLW):** This first field of the PLCP header indicates the length of the payload in bytes including the 32 bit CRC at the end of the payload. PLW can range between 0 and 4,095.
- **PLCP signalling field (PSF):** This 4 bit field indicates the data rate of the payload following. All bits set to zero (0000) indicates the lowest data rate of 1 Mbit/s. The granularity is 500 kbit/s, thus 2 Mbit/s is indicated by 0010 and the maximum is 8.5 Mbit/s (1111). This system obviously does not accommodate today's higher data rates.
- **Header error check (HEC)**: Finally, the PLCP header is protected by a 16 bit checksum with the standard ITU-T generator polynomial $G(x) = x^{16} + x^{12} + x^5 + 1$.

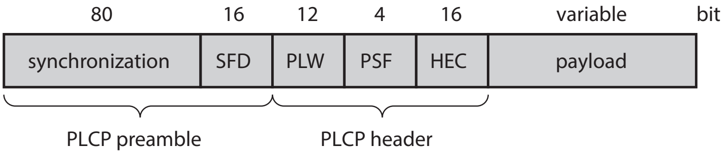


Figure 1.7 Format of an IEEE 802.11 PHY frame using FHSS

Direct sequence spread spectrum

Direct sequence spread spectrum (DSSS) is the alternative spread spectrum method separating by code and not by frequency. In the case of IEEE 802.11 DSSS, 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 robustness against interference and its insensitivity to multipath propagation (time delay spread). 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 and 2 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. Again, the maximum transmit power is 1 W in the US, 100 mW EIRP in Europe and 10 mW/MHz in Japan. The symbol rate is

1 MHz, resulting in a chipping rate of 11 MHz. All bits transmitted by the DSSS PHY are scrambled with the polynomial s(z) = $z^7 + z^4 + 1$ for DC blocking and whitening of the spectrum. Many of today's products offering 11 Mbit/s according to 802.11b are still backward compatible to these lower data rates.

Figure 1.8 shows a frame of the physical layer using DSSS. The frame consists of two basic parts, the PLCP part (preamble and header) and the payload part. 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 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 offset compensation. The synchronization field only consists of scrambled 1 bits.
- **Start frame delimiter (SFD):** This 16 bit field is used for synchronization at the beginning of a frame and consists of the pattern 1111001110100000.

- **Signal:** Originally, only two values have been defined for this field to indicate the data rate of the payload. The value 0x0A indicates 1 Mbit/s (and thus DBPSK), 0x14 indicates 2 Mbit/s (and thus DQPSK). Other values have been reserved for future use, i.e., higher bit rates.
- **Service:** This field is reserved for future use; however, 0x00 indicates an IEEE 802.11 compliant frame.
- **Length:** 16 bits are used in this case for length indication of the payload in microseconds.
- **Header error check (HEC):** Signal, service, and length fields are protected by this checksum using the ITU-T CRC-16 standard polynomial.

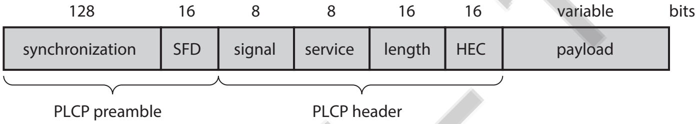


Figure 1.8 Format of an IEEE 802.11 PHY frame using DSSS

Infra red

The PHY layer, which is based on infra red (IR) transmission, uses near visible light at 850-950 nm. Infra red light is not regulated apart from safety restrictions (using lasers instead of LEDs). The standard does not require a line-of-sight between sender and receiver, but should also work with diffuse light. This allows for point-to-multipoint communication. The maximum range is about 10 m if no sunlight or heat sources interfere with the transmission. Typically, such a network will only work in buildings, e.g., classrooms, meeting rooms etc. Frequency reuse is very simple a wall is more than enough to shield one IR based IEEE 802.11 network from another.

6. Explain any two MAC mechanism used in IEEE 802.11 WLAN systems. (May/June 2012)

Medium access control layer

The MAC layer has to fulfill several tasks. First of all, it has to control medium access, but it can also offer support for roaming, authentication, and power conservation. The basic services provided by the MAC layer are the mandatory **asynchronous data service** and an optional **time-bounded service**. While 802.11 only offers the asynchronous service in ad-hoc network mode, both service types can be offered using an infrastructure-based network together with the access point coordinating medium access. The asynchronous service supports broadcast and multi-cast packets, and packet exchange is based on a 'best effort' model, i.e., no delay bounds can be given for transmission.

The following three basic access mechanisms have been defined for IEEE 802.11: the mandatory basic method based on a version of CSMA/CA, an optional method avoiding the hidden terminal problem, and finally a contention-free polling method for time-bounded service. The first two methods are also summarized as **distributed coordination function (DCF)**, the third method is called **point coordination function (PCF)**. DCF only offers asynchronous service, while PCF offers both asynchronous and time-bounded service but needs an access point to control medium access and to avoid contention. The MAC mechanisms are also called **distributed foundation wireless medium access control (DFWMAC)**.

For all access methods, several parameters for controlling the waiting time before medium access are important. Figure 1.9 shows the three different parameters that define the priorities of medium access. The values of the parameters depend on the PHY and are defined in relation to a **slot** time. Slot time is derived from the medium propagation delay, transmitter delay, and other PHY dependent parameters. Slot time is 50 μs for FHSS and 20 μs for DSSS.

The medium, as shown, can be busy or idle (which is detected by the CCA). If the medium is busy this can be due to data frames or other control frames. During a contention phase several nodes try to access the medium.

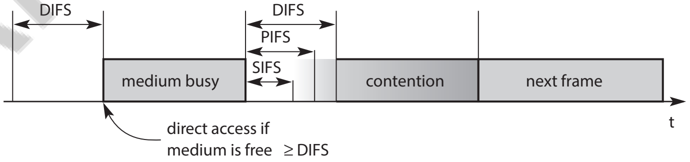


Figure 1.9 Medium access and inter-frame spacing

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

- **PCF inter-frame spacing (PIFS):** A waiting time between DIFS and SIFS (and thus a medium priority) is used for a time-bounded service. An access point polling other nodes only has to wait PIFS for medium accessPIFS is defined as SIFS plus one slot time.
- **DCF inter-frame spacing (DIFS):** This parameter denotes the longest waiting time and has the lowest priority for medium access. This waiting time is used for asynchronous data service within a contention period. DIFS is defined as SIFS plus two slot times.

Basic DFWMAC-DCF using CSMA/CA

The mandatory access mechanism of IEEE 802.11 is based on **carrier sense multiple access with collision avoidance** (CSMA/CA), which is a random access scheme with carrier sense and collision avoidance through random backoff. The basic CSMA/CA mechanism is shown in Figure 1.10. If the medium is idle for at least the duration of DIFS (with the help of the CCA signal of the physical layer), a node can access the medium at once. This allows for short access delay under light load. But as more and more nodes try to access the medium, additional mechanisms are needed.

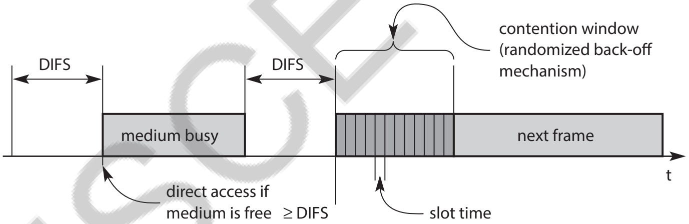


Figure 1.10 Contention window and waiting time

If the medium is busy, nodes have to wait for the duration of DIFS, entering a contention phase afterwards. Each node now chooses a **random backoff time** within a **contention window** and delays medium access for this random amount of time. The node continues to sense the medium. As soon as a node senses the channel is busy, it has lost this cycle and has to wait for the next chance, i.e., until the medium is idle again for at least DIFS. But if the randomized additional waiting time for a node is over and the medium is still idle, the node can access the medium immediately . The additional waiting time is measured in multiples of the above-mentioned slots. This additional randomly distributed delay helps to avoid collisions otherwise all stations would try to transmit data after waiting for the medium becoming idle again plus DIFS.

Independent of the overall time a node has already waited for transmission; each node has the same chances for transmitting data in the next cycle. To provide fairness, IEEE 802.11 adds a **backoff timer.** Again, each node selects a random waiting time within the range of the contention window. If a certain station does not get access to the medium in the first cycle, it stops its backoff timer, waits for the channel to be idle again for DIFS and starts the counter again. As soon as the counter expires, the node accesses the medium. This means that deferred stations do not choose a randomized backoff time again, but continue to count down. Stations that have waited longer have the advantage over stations that have just entered, in that they only have to wait for the remainder of their backoff timer from the previous cycle(s).

Figure 1.11 explains the basic access mechanism of IEEE 802.11 for five stations trying to send a packet at the marked points in time. Station$_3$ has the first request from a higher layer to send a packet (packet arrival at the MAC SAP). The station senses the medium, waits for DIFS and accesses the medium, i.e., sends the packet. Station$_1$, station$_2$, and station$_5$ have to wait at least until the medium is idle for DIFS again after station$_3$ has stopped sending. Now all three stations choose a backoff time within the contention window and start counting down their backoff timers.

Figure 1.11 shows the random backoff time of station$_1$ as sum of bo$_e$ (the elapsed backoff time) and bo$_r$ (the residual backoff time). The same is shown for station$_5$. Station$_2$ has a total backoff time of only bo$_e$ and gets access to the medium first. No residual backoff time for station$_2$ is shown. The backoff timers of station$_1$ and station$_5$ stop, and the stations store their residual backoff times. While a new station has to choose its backoff time from the whole contention window, the two old stations have statistically smaller backoff values. The older values are on average lower than the new ones.

Now station$_4$ wants to send a packet as well, so after DIFS waiting time, three stations try to get access. It can now happen, as shown in the figure, that two stations accidentally have the same backoff time, no matter whether remaining or newly chosen. This results in a collision on the medium as shown, i.e., the trans-

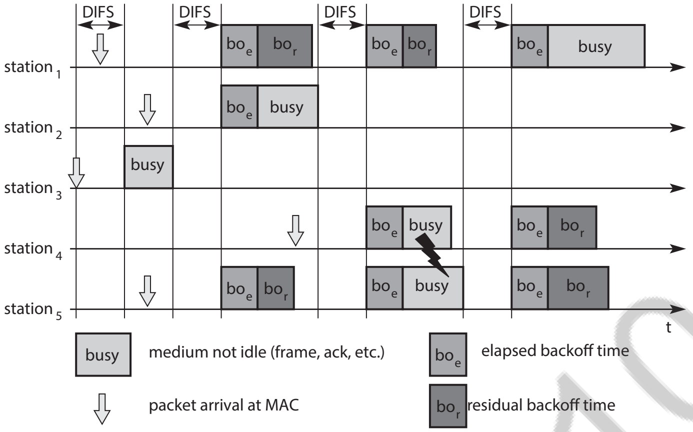


Figure Basic DFWMAC DCF with several competing sensors

mitted frames are destroyed. Station1 stores its residual backoff time again..In the last cycle shown station$_1$ finally gets access to the medium, while station$_4$ and station$_5$ have to wait. A collision triggers a retransmission with a new random selection of the backoff time. Retransmissions are not privileged.

Still, the access scheme has problems under heavy or light load. Depending on the size of the contention window (CW), the random values can either be too close together (causing too many collisions) or the values are too high (causing unnecessary delay). The system tries to adapt to the current number of stations trying to send.

The contention window starts with a size of, e.g., $CW_{min} = 7$. Each time a collision occurs, indicating a higher load on the medium, the contention window doubles up to a maximum of, e.g., $CW_{max} = 255$ (the window can take on the values 7, 15, 31, 63, 127, and 255). The larger the contention window is, the greater is the resolution power of the randomized scheme. It is less likely to choose the same random backoff time using a large CW. However, under a light load, a small CW ensures shorter access delays. This algorithm is also called **exponential backoff** and is already familiar from IEEE 802.3 CSMA/CD in a similar version.

While this process describes the complete access mechanism for broadcast frames, an additional feature is provided by the standard for unicast data transfer. Figure 1.12 shows a sender accessing the medium and sending its data. But now, the receiver answers directly with an

acknowledgement (ACK). The receiver accesses the medium after waiting for a duration of SIFS so no other station can access the medium in the meantime and cause a collision. The other stations have to wait for DIFS plus their backoff time. This acknowledgement ensures the correct reception (correct checksum CRC at the receiver) of a frame on the MAC layer, which is especially important in error-prone environments such as wireless connections. If no ACK is returned, the sender automatically retransmits the frame. But now the sender has to wait again and compete for the access right. There are no special rules for retransmissions. The number of retransmissions is limited, and final failure is reported to the higher layer.

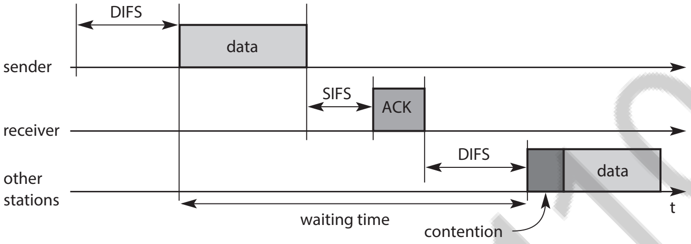


Figure 1.12 IEEE 802.11 unicast data transfer

DFWMAC-DCF with RTS/CTS extension

Hidden terminal problem occurs if one station can receive two others, but those stations cannot receive each other. The two stations may sense the channel is idle, send a frame, and cause a collision at the receiver in the middle. To deal with this problem, the standard defines an additional mechanism using two control packets, RTS and CTS. The use of the mechanism is optional; however, every 802.11 node has to implement the functions to react properly upon reception of RTS/CTS control packets.

Figure 1.13 illustrates the use of RTS and CTS. After waiting for DIFS (plus a random backoff time if the medium was busy), the sender can issue a **request to send (RTS)** control packet. The RTS packet thus is not given any higher priority compared to other data packets. The RTS packet includes the receiver of the data transmission to come and the duration of the whole data transmission. This duration specifies the time interval necessary to transmit the whole data frame and the acknowledgement related to it. Every node receiving this RTS now has to set its **net allocation vector (NAV)** in accordance with the duration field. The NAV then specifies the earliest point at which the station can try to access the medium again.

If the receiver of the data transmission receives the RTS, it answers with a **clear to send (CTS)** message after waiting for SIFS. This CTS packet contains the duration field again and all stations receiving this packet from the receiver of the intended data transmission have to adjust their NAV. The latter set of receivers need not be the same as the first set receiving the RTS packet. Now all nodes within receiving distance around sender and receiver are informed that they have to wait more time before accessing the medium. Basically, this mechanism reserves the medium for one sender exclusively (this is why it is sometimes called a virtual reservation scheme).

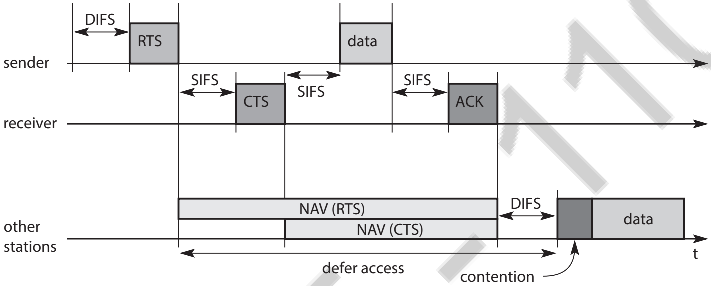


Figure 1.13 IEEE 802.11 hidden node provisions for contention-free access

Finally, the sender can send the data after SIFS. The receiver waits for SIFS after receiving the data packet and then acknowledges whether the transfer was correct. The transmission has now been completed, the NAV in each node marks the medium as free and the standard cycle can start again.

Within this scenario (i.e., using RTS and CTS to avoid the hidden terminal problem), collisions can only occur at the beginning while the RTS is sent. Two or more stations may start sending at the same time (RTS or other data packets). Using RTS/CTS can result in a non-negligible overhead causing a waste of bandwidth and higher delay. An RTS threshold can determine when to use the additional mechanism (basically at larger frame sizes) and when to disable it (short frames).

Wireless LANs have bit error rates in transmission that are typically several orders of magnitude higher than, e.g., fiber optics. The probability of an erroneous frame is much higher for wireless links assuming the same frame length. One way to decrease the error probability of frames is to use shorter frames. In this case, the bit error rate is the same, but now only short frames are destroyed and, the frame error rate decreases.

However, the mechanism of fragmenting a user data packet into several smaller parts should be transparent for a user. The MAC layer should have the possibility of adjusting the transmission frame size to the current error rate on the medium. The IEEE 802.11 standard specifies a fragmentation mode. Again, a sender can send an RTS control packet to reserve the medium after a waiting time of DIFS. This RTS packet now includes the duration for the transmission of the first fragment and the corresponding acknowledgement. A certain set of nodes may receive this RTS and set their NAV according to the duration field. The receiver answers with a CTS, again including the duration of the transmission up to the acknowledgement. A (possibly different) set of receivers gets this CTS message and sets the NAV.

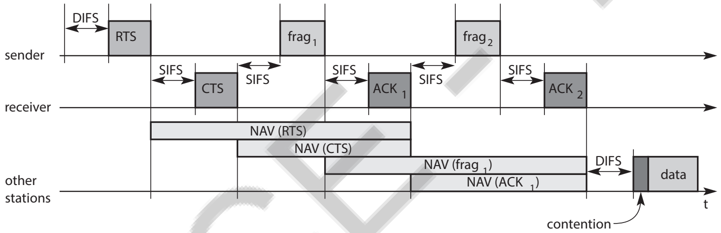


Figure 1.14 IEEE 802.11 fragmentation of user data

As shown in Figure 1.13, the sender can now send the first data frame, frag1, after waiting only for SIFS. The new aspect of this fragmentation mode is that it includes another duration value in the frame frag_1. This duration field reserves the medium for the duration of the transmission following, comprising the second fragment and its acknowledgement. Again, several nodes may receive this reservation and adjust their NAV. If all nodes are static and transmission conditions have not changed, then the set of nodes receiving the duration field in frag_1 should be the same as the set that has received the initial reservation in the RTS control packet. However, due to the mobility of nodes and changes in the environment, this could also be a different set of nodes.

The receiver of frag_1 answers directly after SIFS with the acknowledgement packet ACK_1 including the reservation for the next transmission as shown. Again, a fourth set of nodes may receive this reservation and adjust their NAV (which again could be the same as the second set of nodes that has received the reservation in the CTS frame).

If $frag_2$ was not the last frame of this transmission, it would also include a new duration for the third consecutive transmission. (In the example shown, $frag_2$ is the last fragment of this transmission so the sender does not reserve the medium any longer.) The receiver acknowledges this second fragment, not reserving the medium again. After ACK_2, all nodes can compete for the medium again after having waited for DIFS.

DFWMAC-PCF with polling

The two access mechanisms presented so far cannot guarantee a maximum access delay or minimum transmission bandwidth. To provide a time-bounded service, the standard specifies a **point coordination function (PCF)** on top of the standard DCF mechanisms. Using PCF requires an access point that controls medium access and polls the single nodes. Ad-hoc networks cannot use this function so, provide no QoS but 'best effort' in IEEE 802.11 WLANs.

The **point co-ordinator** in the access point splits the access time into super frame periods as shown in Figure 1.15. A **super frame** comprises a **contentionfree period** and a **contention period**. The contention period can be used for the two access mechanisms presented above. The figure also shows several wireless stations (all on the same line) and the stations' NAV (again on one line).

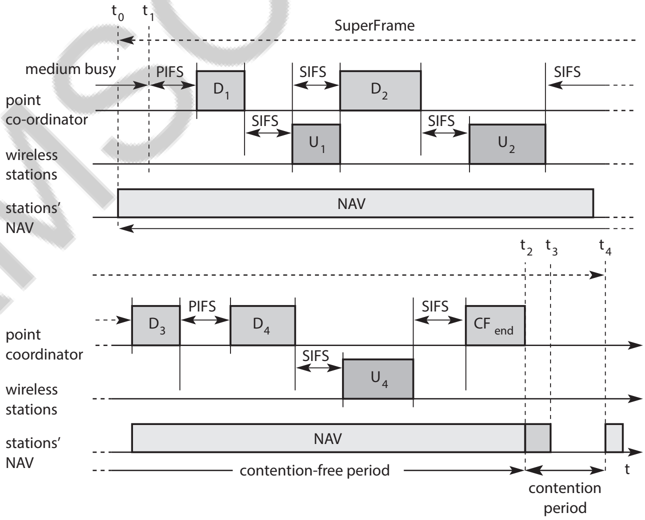


Figure 1.15 Contention-free access using polling mechanisms (PCF)

At time to the contention free period of the super frame should therotically start, but another station is still transmitting data (i.e., the medium is busy). This means that PCF also defers to DCF, and the start of the super frame may be postponed. The only possibility of avoiding variations is not to have any contention period at all. After the medium has been idle until t_1, the point coordinator has to wait for PIFS before accessing the medium. As PIFS is smaller than DIFS, no other station can start sending earlier.

The point coordinator now sends data D_1 downstream to the first wireless station. This station can answer at once after SIFS . After waiting for SIFS again, the point coordinator can poll the second station by sending D_2. This station may answer upstream to the coordinator with data U_2. Polling continues with the third node. This time the node has nothing to answer and the point coordinator will not receive a packet after SIFS.

After waiting for PIFS, the coordinator can resume polling the stations. Finally, the point coordinator can issue an end marker (CF_{end}), indicating that the contention period may start again. Using PCF automatically sets the NAV, preventing other stations from sending. In the example, the contention-free period planned initially would have been from t_0 to t_3. However, the point coordinator finished polling earlier, shifting the end of the contention-free period to t_2. At t_4, the cycle starts again with the next super frame.

The transmission properties of the whole wireless network are now determined by the polling behavior of the access point. If only PCF is used and polling is distributed evenly, the bandwidth is also distributed evenly among all polled nodes. This would resemble a static, centrally controlled time division multiple access (TDMA) system with time division duplex (TDD) transmission. This method comes with an overhead if nodes have nothing to send, but the access point polls them permanently. Anastasi (1998) elaborates the example of voice transmission using 48 byte packets as payload. In this case, PCF introduces an overhead of 75 byte.

7. Explain the basic structure of an IEEE 802.11 MAC data frame in detail.

MAC frames

Figure 1.16 shows the basic structure of an IEEE 802.11 MAC data frame together with the content of the frame control field. The fields in the figure refer to the following:

- **Frame control:** The first 2 bytes serve several purposes. They contain several sub-fields as explained after the MAC frame.

- **Duration/ID:** If the field value is less than 32,768, the duration field contains the value indicating the period of time in which the medium is occupied (in μs). This field is used for setting the NAV for the virtual reservation mechanism using RTS/CTS and during fragmentation. Certain values above 32,768 are reserved for identifiers.

- **Address 1 to 4:** The four address fields contain standard IEEE 802 MAC addresses (48 bit each), as they are known from other 802.x LANs. The meaning of each address depends on the DS bits in the frame control field and is explained in more detail in a separate paragraph.

- **Sequence control:** Due to the acknowledgement mechanism frames may be duplicated. Therefore a sequence number is used to filter duplicates.

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

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

- The frame control field shown in Figure 1.16 contains the following fields:

- **Protocol version:** This 2 bit field indicates the current protocol version and is fixed to 0 by now. If major revisions to the standard make it incompatible with the current version, this value will be increased.

- **Type:** The type field determines the function of a frame: management (=00), control (=01), or data (=10). The value 11 is reserved. Each type has several subtypes as indicated in the following field.

- **Subtype:** Example subtypes for management frames are: 0000 for associ-ation request, 1000 for beacon. RTS is a control frame with subtype 1011, CTS is coded as 1100. User data is transmitted as data frame with subtype 0000.

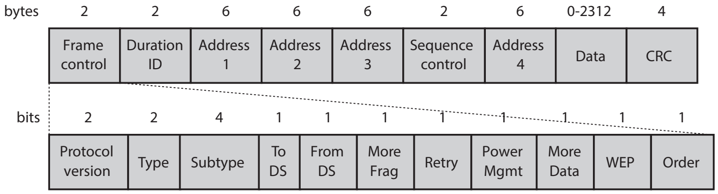


Figure 1.16 IEEE 802.11 MAC packet structure

- **To DS/From DS:** Explained in the following in more detail.
- **More fragments:** This field is set to 1 in all data or management frames that have another fragment of the current MSDU to follow.
- **Retry:** If the current frame is a retransmission of an earlier frame, this bit is set to 1. With the help of this bit it may be simpler for receivers to eliminate duplicate frames.
- **Power management:** This field indicates the mode of a station after successful transmission of a frame. Set to 1 the field indicates that the station goes into power-save mode. If the field is set to 0, the station stays active.
- **More data:** In general, this field is used to indicate a receiver that a sender has more data to send than the current frame. This can be used by an access point to indicate to a station in power-save mode that more packets are buffered. Or it can be used by a station to indicate to an access point after being polled that more polling is necessary as the station has more data ready to transmit. • **Wired equivalent privacy (WEP):** This field indicates that the standard security mechanism of 802.11 is applied. However, due to many weaknesses found in the WEP algorithm higher layer security should be used to secure an 802.11 network.
- **Order:** If this bit is set to 1 the received frames must be processed in strict order.

Table 1.1 gives an overview of the four possible bit values of the DS bits and the associated interpretation of the four address fields.

Table 1.1 Interpretation of the MAC addresses in an 802.11

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

Every station, access point or wireless node, filters on **address 1**. This address identifies the physical receiver(s) of the frame. Based on this address, a station can decide whether the frame is relevant or not. The second address, **address 2**, represents the physical transmitter of a frame. This information is important because this particular sender is also the recipient of the MAC layer acknowledgement. If a packet from a transmitter (address 2) is received by the receiver with address 1, this receiver in turn acknowledges the data packet using address 2 as receiver address as shown in the ACK packet in Figure 1.17. The remaining two addresses, **address 3** and **address 4**, are mainly necessary for the logical assignment of frames (logical sender, 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 constitutes a packet which is exchanged between two wireless nodes without a distribution system. **DA** indicates the **destination address**, **SA** the **source address** of the frame, which are identical to the physical receiver and sender addresses respectively. The third address identifies the **basic service set (BSSID)** (see Figure 1.4), the fourth address is unused.

- **Infrastructure network, from AP:** If only the 'from DS' bit is set, the frame physically originates from an access point. DA is the logical and physical receiver, the second address identifies the BSS, the third address specifies the logical sender, the source address of the MAC frame. This case is an example for a packet sent to the receiver via the access point.

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

✷ **Infrastructure network, within DS:** For packets transmitted between two access points over the distribution system, both bits are set. The first **receiver address (RA)**, represents the MAC address of the receiving access point. Similarly, the second address **transmitter address (TA)**, identifies the sending access point within the distribution system. Now two more addresses are needed to identify the original destination DA of the frame and the original source of the frame SA. Without these additional addresses, some encapsulation mechanism would be necessary to transmit MAC frames over the distribution system transparently.

8. Explain IEEE 802.11 special control packets: ACK, RTS, and CTS in detail.

Figure 1.17 shows three control packets as examples for many special packets defined in the standard. The **acknowledgement packet (ACK)** is used to acknowledge the correct reception of a data frame as shown in Figure 7.12. The receiver address is directly copied from the address 2 field of the immediately previous frame. If no more fragments follow for a certain frame the duration field is set to 0. Otherwise the duration value of the previous frame (minus the time required to transmit the ACK minus SIFS) is stored in the duration field.

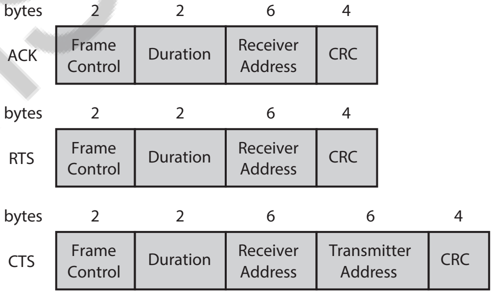


Figure 1.17 IEEE 802.11 special control packets: ACK, RTS, and CTS

For the MACA algorithm the RTS/CTS packets are needed. As Figure 1.13 shows, these packets have to reserve the medium to avoid collisions. Therefore, the **request to send (RTS)** packet contains the receiver address of the intended recipient of the following data transfer and the transmitter address of the station transmitting the RTS packet. The duration (in μs) comprises the time to send the CTS, data, and ACK plus three SIFS. The immediately following **clear to send (CTS)** frame copies the transmitter address from the RTS packet into its receiver address field. Additionally, it reads the duration field, subtracts the time to send the CTS and a SIFS and writes the result into its own duration field.

9. Explain the MAC management, power management and synchronisation in an IEEE 802.11.

MAC management plays a central role in an IEEE 802.11 station as it more or less controls all functions related to system integration, i.e., integration of a wireless station into a BSS, formation of an ESS, synchronization of stations etc. The following functional groups have been identified and will be discussed in more detail in the following sections:

- **Synchronization:** Functions to support finding a wireless LAN, synchronization of internal clocks, generation of beacon signals.
- **Power management:** Functions 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).

Synchronization

Each node of an 802.11 network maintains an internal clock. To synchronize the clocks of all nodes, IEEE 802.11 specifies a **timing synchronization function (TSF)**. As we will see in the following section, synchronized clocks are needed for power management, but also for coordination of the PCF and for synchronization of the hopping sequence in an FHSS system. Using PCF, the local timer of a node can predict the start of a super frame, i.e., the contention free and contention period. FHSS physical layers need the same hopping sequences so that all nodes can communicate within a BSS.

Within a BSS, timing is conveyed by the (quasi)periodic transmissions of a beacon frame. A **beacon** contains a timestamp and other management information used 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 not required to hear every beacon to stay synchronized; however, from time to time internal clocks should be adjusted. The transmission of a beacon frame is not always periodic because the beacon frame is also deferred if the medium is busy.

Within **infrastructure-based** networks, the access point performs synchronization by transmitting the (quasi)periodic beacon signal, whereas all other wireless nodes adjust their local timer to the time stamp. This represents the simple case shown in Figure 1.18. The access point is not always able to send its beacon B periodically if the medium is busy. However, the access point always tries to schedule transmissions according to the expected beacon interval (**target beacon transmission time**), i.e., beacon intervals are not shifted if one beacon is delayed. The timestamp of a beacon always reflects the real transmit time, not the scheduled time.

For ad-hoc networks, the situation is slightly more complicated as they do not have an access point for beacon transmission. In this case, each node maintains its own synchronization timer and starts the transmission of a beacon frame after the beacon interval. Figure 1.19 shows an example where multiple stations try to send their beacon. However, the standard random backoff algorithm is also applied to the beacon frames so only one beacon wins. All other stations now adjust their internal clocks according to the received beacon and

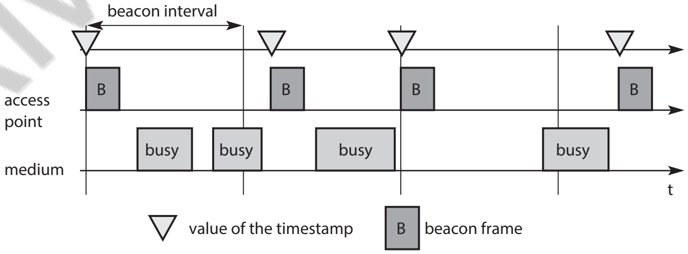


Figure 1.18 Beacon transmission in a busy 802.11 infrastructure network

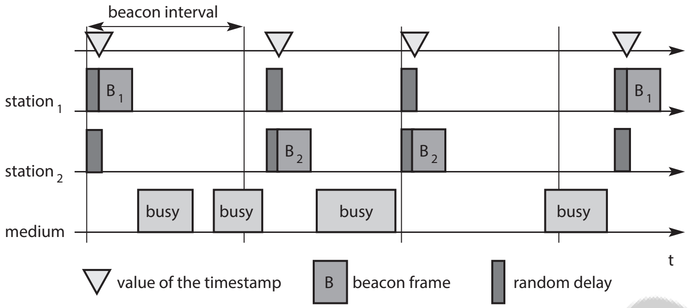


Figure 1.19 Beacon transmission in a busy 802.11 ad hoc network

suppress their beacons for this cycle. If collision occurs, the beacon is lost. In this scenario, the beacon intervals can be shifted slightly because all clocks may vary as may the start of a beacon interval from a node's point of view. However, after successful synchronization all nodes again have the same consistent view.

Power management

Wireless devices are battery powered (unless a solar panel is used). Therefore, power-saving mechanisms are crucial for the commercial success of such devices. Standard LAN protocols assume that stations are always ready to receive data, although receivers are idle most of the time in lightly loaded networks. However, this permanent readiness of the receiving module is critical for battery life as the receiver current may be up to 100 mA .

The basic idea of IEEE 802.11 power management is to switch off the transceiver whenever it is not needed. For the sending device this is simple to achieve as the transfer is triggered by the device itself. However, since the power management of a receiver cannot know in advance when the transceiver has to be active for a specific packet, it has to 'wake up' the transceiver periodically. Switching off the transceiver should be transparent to existing protocols and should be flexible enough to support different applications. However, throughput can be traded-off for battery life. Longer off-periods save battery life but reduce average throughput and vice versa.

The basic idea of power saving includes two states for a station: **sleep** and **awake**, and buffering of data in senders. If a sender intends to communicate with a power-saving station it has to buffer data if the station is asleep. The sleeping station on the other hand has to wake up periodically and stay awake for a certain time. During this time, all senders can announce the destinations of their buffered data frames. If a station detects that it is a destination of a buffered packet it has to stay awake until the transmission takes place. Waking up at the right moment requires the **timing synchronization function (TSF) .**

All stations have to wake up or be awake at the same time.

Power management in **infrastructure**-based networks is much simpler compared to ad-hoc networks. The access point buffers all frames destined for stations operating in power-save mode. With every beacon sent by the access point, a **traffic indication map (TIM)** is transmitted. The TIM contains a list of stations for which unicast data frames are buffered in the access point.

The TSF assures that the sleeping stations will wake up periodically and listen to the beacon and TIM. If the TIM indicates a unicast frame buffered for the station, the station stays awake for transmission. For multi-cast/ broadcast transmission, stations will always stay awake. Another reason for waking up is a frame which has to be transmitted from the station to the access point. A sleeping station still has the TSF timer running.

Figure 1.20 shows an example with an access point and one station. The state of the medium is indicated. Again, the access point transmits a beacon frame each beacon interval. This interval is now the same as the TIM interval. Additionally, the access point maintains a **delivery traffic indication map (DTIM)** interval for sending broadcast/multicast frames. The DTIM interval is always a multiple of the TIM interval.

All stations (in the example, only one is shown) wake up prior to an expected TIM or DTIM. In the first case, the access point has to transmit a broadcast frame and the station stays awake to receive it. After receiving the broadcast frame, the station returns to sleeping mode. The station wakes up again just before the next TIM transmission. This time the TIM is delayed due to a 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 access point indicates that the station is the destination for a buffered frame. The station answers with a **PS** (power saving) **poll** and stays awake to receive data. The access point then transmits the data for the station, the station acknowledges the receipt and may also send some

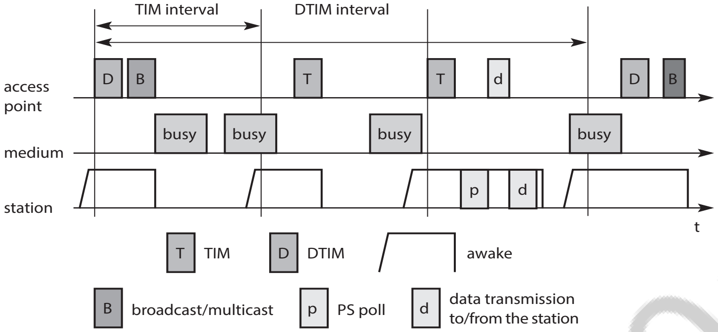


Figure 1.20 Power management in IEEE 802.11 infrastructure networks

data . This is acknowledged by the access point . Afterwards, the station switches to sleep mode again.

Finally, the access point has more broadcast data to send at the next DTIM interval, which is again deferred by a busy medium. Depending on internal thresholds, a station may stay awake if the sleeping period would be too short. This mechanism clearly shows the trade-off between short delays in station access and saving battery power. The shorter the TIM interval, the shorter the delay, but the lower the power-saving effect.

In ad-hoc networks, power management is much more complicated than in infrastructure networks. In this case, there is no access point to buffer data in one location but each station needs the ability to buffer data if it wants to communicate with a power-saving station. All stations now announce a list of buffered frames during a period when they are all awake. Destinations are announced using **ad-hoc traffic indication map (ATIMs)** the announcement period is called the **ATIM window**.

Figure 1.21 shows a simple ad-hoc network with two stations. Again, the beacon interval is determined by a distributed function (different stations may send the beacon). However, due to this synchronization, all stations within the ad-hoc network wake up at the same time. All stations stay awake for the ATIM interval as shown in the first two steps and go to sleep again if no frame is buffered for them. In the third step, station$_1$ has data buffered for station$_2$. This is indicated in an ATIM transmitted by station$_1$. Station$_2$ acknowledges this ATIM and stays awake for the transmission. After the ATIM window, station$_1$ can transmit the data frame, and station$_2$ acknowledges its receipt. In this case, the stations stay awake for the next beacon.

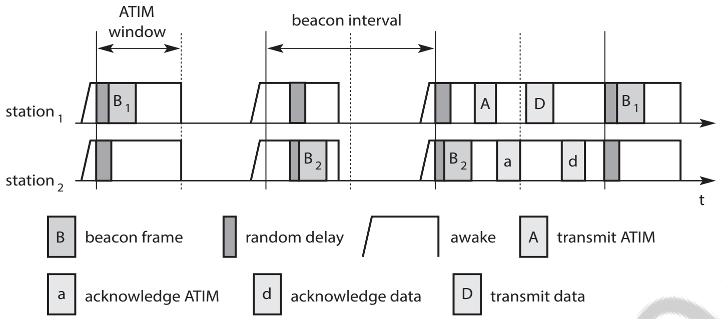


Figure 1.21 Power management in IEEE 802.11 ad hoc networks

One problem with this approach is that of scale. If many stations within an ad-hoc network operate in power-save mode, they may also want to transmit their ATIM within the ATIM window. More ATIM transmissions take place, more collisions happen and more stations are deferred. The access delay of large networks is difficult to predict. QoS guarantees can not be given under heavy load.

10. Explain Roaming in WLAN.

Typically, wireless networks within buildings require more than just one access point to cover all rooms. Depending on the solidity and material of the walls, one access point has a transmission range of 10-20 m if transmission is to be of decent quality. Each storey of a building needs its own access point(s) as quite often walls are thinner than floors. If a user walks around with a wireless station, the station has to move from one access point to another to provide uninterrupted service. Moving between access points is called **roaming**. The term “handover” or “handoff” as used in the context of mobile or cellular phone systems would be more appropriate as it is simply a change of the active cell. However, for WLANs roaming is more common.

The steps for roaming between access points are:

- A station decides that the current link quality to its access point AP_1 is to poor. The station then starts **scanning** for another access point.
- Scanning involves the active search for another BSS and can also be used for setting up a new BSS in case of ad-hoc networks. IEEE 802.11 specifies scanning on single or multiple channels (if

available at the physical layer) and differentiates between passive scanning and active scanning. **Passive scanning** simply means listening into the medium to find other networks, i.e., receiving the beacon of another network issued by the synchronization function within an access point. **Active scanning** comprises sending a

✸ **Probe** on each channel and waiting for a response. Beacon and probe responses contain the information necessary to join the new BSS. ● The station then selects the best access point for roaming based on, e.g., signal strength, and sends an **association request** to the selected access point AP_2.

✸ The new access point AP_2 answers with an **association response**. If the response is successful, the station has roamed to the new access point AP_2. Otherwise, the station has to continue scanning for new access points. ● The access point accepting an association request indicates the new station in its BSS to the distribution system (DS). The DS then updates its database, which contains the current location of the wireless stations. This database is needed for forwarding frames between different BSSs, i.e. between the different access points controlling the BSSs, which combine to form an ESS Additionally, the DS can inform the old access point AP_1 that the station is no longer within its BSS.

The standard **IEEE 802.11f (Inter Access Point Protocol, IAPP)** provide a compatible solution for all vendors. It also includes load-balancing between access points and key generation for security algorithms based on IEEE 802.1x.

11. Explain about IEEE 802.11b in detail.

Soon after the first commercial 802.11 products came on the market some companies offered proprietary solutions with 11 Mbit/s. To avoid market segmentation, a common standard, **IEEE 802.11b** soon followed and was added as supplement to the original standard (Higher-speed physical layer extension in the 2.4 GHz band). This standard describes a new PHY layer and is by far the most successful version of IEEE 802.11 available today.

As the name of the supplement implies, this standard only defines a new PHY layer. Depending on the current interference and the distance between sender and receiver 802.11b systems offer 11, 5.5, 2, or 1 Mbit/s. Maximum user data rate is approx 6 Mbit/s. The lower data rates 1 and 2 Mbit/s use the 11-chip Barker sequence and DBPSK or DQPSK, respectively. The new data rates, 5.5 and 11 Mbit/s, use 8-chip **complementary code keying (CCK)** .

The standard defines several packet formats for the physical layer. The mandatory format interoperates with the original versions of 802.11. The optional versions provide a more efficient data transfer due to shorter headers/different coding schemes and can coexist with other 802.11 versions. However, the standard states that control all frames shall be transmitted at one of the basic rates, so they will be understood by all stations in a BSS.

Figure 1.22 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 1.8. One difference is the rate encoded in the signal field this is encoded in multiples of 100 kbit/s. Thus, 0x0A represents 1 Mbit/s, 0x14 is used for 2 Mbit/s, 0x37 for 5.5 Mbit/s and 0x6E for 11 Mbit/s. Note that the preamble and the header are transmitted at 1 Mbit/s using DBPSK. The optional **short PLCP PPDU** format differs in several ways. The short synchronization field consists of 56 scrambled zeros instead of scrambled ones. The short start frame delimiter SFD consists of a mirrored bit pattern compared to the SFD of the long format: 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 receive the short format will not detect the start of a frame (but will sense the medium

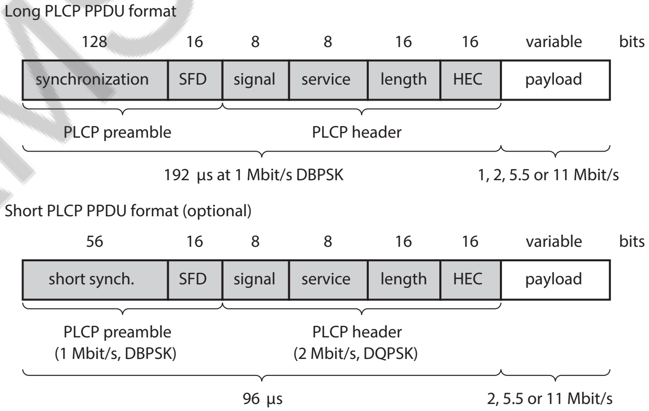


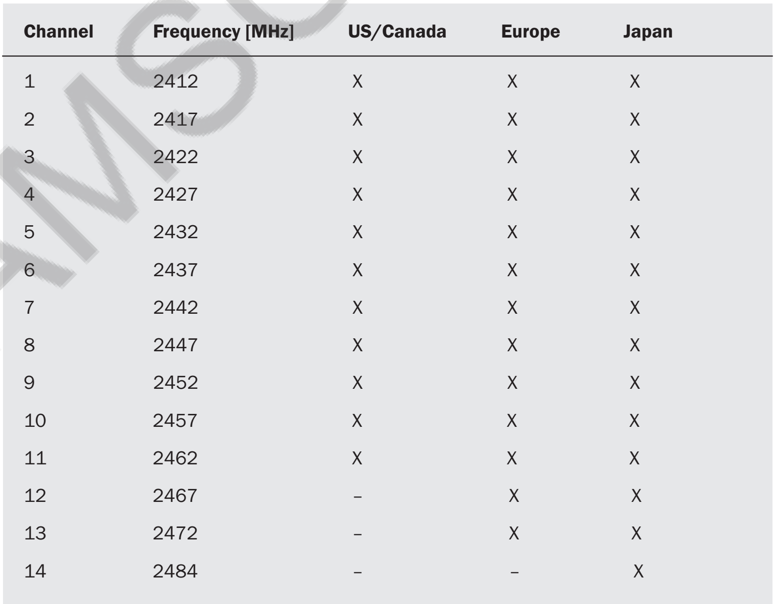
Figure 1.22. IEEE 802.11b PHY packet formats

is busy). Only the preamble is transmitted at 1 Mbit/s, DBPSK. The following header is already transmitted at 2 Mbit/s, DQPSK, which is also the lowest available data rate. As Figure 1.22 shows, the length of the overhead is only half for the short frames (96 μs instead of 192 μs). This is useful for, e.g., short, but timecritical, data transmissions.

The standards operates (like the DSSS version of 802.11) on certain frequencies in the 2.4 GHz ISM band. For each channel the center frequency is given. Depending on national restrictions 11 (US/Canada), 13 (Europe with some exceptions) or 14 channels (Japan) can be used.

Figure 1.23 illustrates the non-overlapping usage of channels for an IEEE 802.11b installation with minimal interference in the US/Canada and Europe. 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 the additional channels (12 and 13) in Europe as their hardware is limited to 11 channels. Some European installations use channel 13 to minimize interference. Users can install overlapping cells for WLANs using the three non-overlapping channels to provide seamless coverage. This is similar to the cell planning for mobile phone systems.

Table 1.2 Channel plan IEEE 802.11b



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

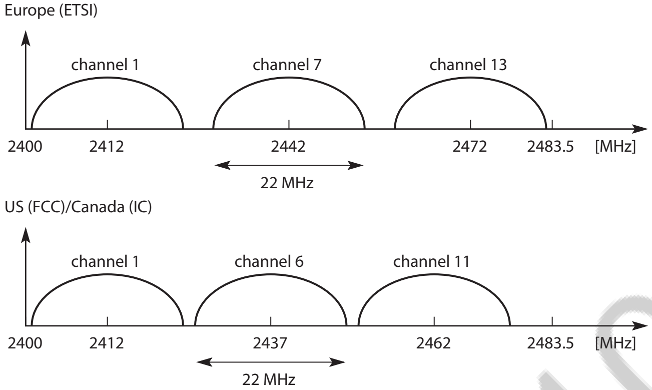


Figure 1.23. IEEE 802. 11 b non overlapping channel selection

12. Explain about IEEE 802.11a in detail.

Initially aimed at the US 5 GHz U-NII (Unlicensed National Information Infrastructure) bands **IEEE 802.11a** offers up to 54 Mbit/s using OFDM (IEEE, 1999). The first products were available in 2001 and can now be used (after some harmonization between IEEE and ETSI) in Europe. The FCC (US) regulations offer three different 100 MHz domains for the use of 802.11a, each with a different legal maximum power output: 5.15-5.25 GHz/50 mW, 5.25-5.35 GHz/250 mW, and 5.725-5.825 GHz/1 W. ETSI (Europe) defines different frequency bands for Europe: 5.15-5.35 GHz and 5.47-5.725 GHz and requires two additional mechanisms for operation: dynamic frequency selection (DFS) and transmit power control (TPC) Maximum transmit power is 200 mW EIRP for the lower frequency band (indoor use) and 1 W EIRP for the higher frequency band (indoor and outdoor use). DFS and TPC are not necessary, if the transmit power stays below 50 mW EIRP and only 5.15-5.25 GHz are used. Japan allows operation in the frequency range

5.15-5.25 GHz and requires carrier sensing every 4 ms to minimize interference. Up to now, only 100 MHz are available ‘worldwide’ at 5.15-5.25 GHz. The physical layer of IEEE 802.11a and the ETSI standard HiperLAN2 has been jointly developed, so both physical layers are almost identical. However, HiperLAN2 differs in the MAC layer, the PHY layer packet formats, and the offered services (quality of service, real time etc.). It should be noted that most of the development for the physical layer for 802.11a was adopted from the HiperLAN2 standardization but 802.11a products were available first and are already in widespread use.

Again, IEEE 802.11a uses the same MAC layer as all 802.11 physical layers do. To be able to offer data rates up to 54 Mbit/s IEEE 802.11a uses many different technologies. The system uses 52 subcarriers (48 data + 4 pilot) that are modulated using BPSK, QPSK, 16-QAM, or 64-QAM. To mitigate transmission errors, FEC is applied using coding rates of 1/2, 2/3, or 3/4. Table 1.3 gives an overview of the standardized combinations of modulation and coding schemes together with the resulting data rates. To offer a data rate of 12 Mbit/s, 96 bits are coded into one OFDM symbol. These 96 bits are distributed over 48 subcarriers and

2 bits are modulated per sub-carrier using QPSK (2 bits per point in the constellation diagram). Using a coding rate of 1/2 only 48 data bits can be transmitted.

Table 1.3 Rate dependent parameters 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

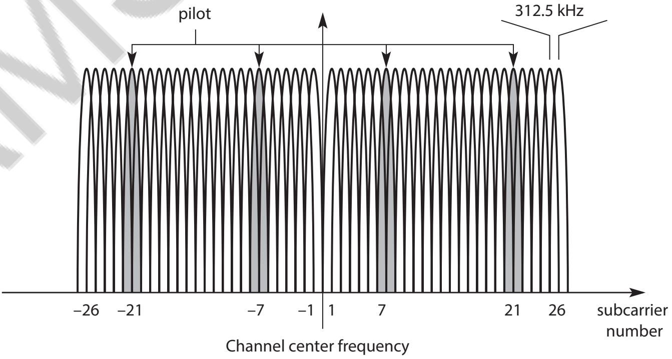


Figure 1.24 Usage of OFDM in IEEE 802.11a

Figure 1.24 shows the usage of OFDM in IEEE 802.11a. The basic idea of OFDM (or MCM in general) was the reduction of the symbol rate by distributing bits over numerous subcarriers. IEEE 802.11a uses a fixed symbol rate of 250,000 symbols per second independent of the data rate (0.8 μs guard interval for ISI mitigation plus 3.2 μs used for data results in a symbol duration of 4 μs). As Figure 1.24 shows, 52 subcarriers are equally spaced around a center frequency. (Center frequencies will be explained later). The spacing between the subcarriers is 312.5 kHz. 26 subcarriers are to the left of the center frequency and 26 are to the right. The center frequency itself is not used as subcarrier. Subcarriers with the numbers -21, -7, 7, and 21 are used for pilot signals to make the signal detection robust against frequency offsets.

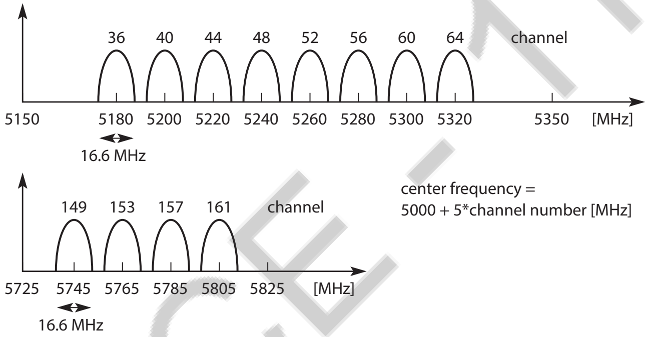


Figure 1.25 Operating channels of IEEE 802.11a in the U-NII bands

Similar to 802.11b several operating channels have been standardized to minimize interference. Figure 1.25 shows the **channel layout** for the US U-NII bands. The center frequency of a channel is 5000 + 5*channel number [MHz]. This definition provides a unique numbering of channels with 5 MHz spacing starting from 5 GHz. Depending on national regulations, different sets of channels may be used. Eight channels have been defined for the lower two bands in the U-NII (36, 40, 44, 48, 52, 56, 60, and 64); four more are available in the high band (149, 153, 157, and 161). Using these channels allows for interference-free operation of overlapping 802.11a cells. Channel spacing is 20 MHz, the occupied bandwidth of 802.11a is 16.6 MHz. How is this related to the spacing of the sub-carriers? 20 MHz/64 equals 312.5 kHz. 802.11a uses 48 carriers for data,

4 for pilot signals, and 12 carriers are sometimes called virtual subcarriers. (Set to zero, they do not contribute to the data transmission

but may be used for an implementation of OFDM with the help of FFT, Multiplying 312.5 kHz by 52 subcarriers and adding the extra space for the center frequency results in approximately 16.6 MHz occupied bandwidth per channel .

Due to the nature of OFDM, the PDU on the physical layer of IEEE 802.11a looks quite different from 802.11b or the original 802.11 physical layers. Figure 1.26 shows the basic structure of an **IEEE 802.11a PPDU**.

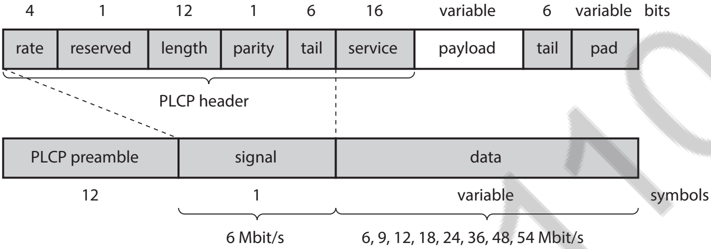


Figure 1.26 IEEE 802.11a physical layer PDU

- The **PLCP preamble** consists of 12 symbols and is used for frequency acquisition, channel estimation, and synchronization. The duration of the preamble is 16 μs.
- The following OFDM symbol, called **signal**, contains the following fields and is BPSK-modulated. The 4 bit **rate** field determines the data rate and the modulation of the rest of the packet (examples are 0x3 for 54 Mbit/s, 0x9 for 24 Mbit/s, or 0xF for 9 Mbit/s). The **length** field indicates the number of bytes in the payload field. The **parity** bit shall be an even parity for the first 16 bits of the signal field (rate, length and the reserved bit). Finally, the six **tail** bits are set to zero.
- The **data** field is sent with the rate determined in the rate field and contains a **service** field which is used to synchronize the descrambler of the receiver (the data stream is scrambled using the polynomial $x^7 + x^4 + 1$) and which contains bits for future use. The **payload** contains the MAC PDU (1-4095 byte). The **tail** bits are used to reset the encoder. Finally, the **pad** field ensures that the number of bits in the PDU maps to an integer number of OFDM symbols.

Compared to IEEE 802.11b working at 2.4 GHz IEEE 802.11a at 5 GHz offers much higher data rates. However, shading at 5 GHz is much more severe compared to 2.4 GHz and depending on the SNR, propagation conditions and the distance between sender and receiver, data rates may drop fast (e.g., 54 Mbit/s may be available only in an LOS or near LOS condition). Additionally, the MAC layer of IEEE 802.11 adds overheads. User data rates are therefore much lower than the data rates listed above. Typical user rates in Mbit/s are (transmission rates in brackets) 5.3 (6), 18 (24), 24 (36), and 32 (54). The following section presents some additional developments in the context of 802.11, which also comprise a standard for higher data rates at 2.4 GHz that can benefit from the better propagation conditions at lower frequencies.

13. Explain about the Newer developments in IEEE Standards.

While many products that follow the IEEE 802.11a and 802.11b standards are available, several new groups have been formed within the IEEE to discuss enhancements of the standard and new applications. The completed standards **IEEE 802.11c** and **802.11d** cover additions for bridging support and updates for physical layer requirements in different regulatory domains

- **802.11e (MAC enhancements):** Currently, the 802.11 standards offer non quality of service in the DCF operation mode. Some QoS guarantees can be given, only via polling using PCF. For applications such as audio, video, or media stream, distribution service classes have to be provided. For this reason, the MAC layer must be enhanced compared to the current standard. ● **802.11f (Inter-Access Point Protocol):** The current standard only describes the basic architecture of 802.11 networks and their components. The implementation of components, such as the distribution system, was deliberately not specified. Specifications of implementations should generally be avoided as they hinder improvements. However, a great flexibility in the implementation combined with a lack of detailed interface definitions and communication protocols, e.g., for management severely limits the interoperability of devices from different vendors. For example, seamless roaming between access points of different vendors is often impossible. 802.11f standardizes the necessary exchange of information between access points to support the functions of a distribution system.
- **802.11g (Data rates above 20 Mbit/s at 2.4 GHz):** Introducing new modulation schemes, forward error correction and OFDM also allows for higher data rates at 2.4 GHz. This approach should be backward compatible to 802.11b and should benefit from the better propagation characteristics at 2.4 GHz compared to 5 GHz.

Currently, chips for 54 Mbit/s are available as well as first products. An alternative (or additional) proposal for 802.11g suggests the socalled packet binary convolutional coding (PBCC) to reach a data rate of 22 Mbit/s (Heegard, 2001). While the 54 Mbit/s OFDM mode is mandatory, the 22 Mbit/s PBCC mode can be used as an option. The decision between 802.11a and 802.11g is not obvious. Many 802.11a products are already available and the 5 GHz band is (currently) not as crowded as the 2.4 GHz band where not only microwave ovens, but also Bluetooth, operate (see section 7.5). Coverage is better at 2.4 GHz and fewer access points are needed, lowering the overall system cost. 802.11g access points can also communicate with 802.11b devices as the current 802.11g products show. Dual mode (or then triple mode) devices will be available covering 802.11a and b (and g). If a high traffic volume per square meter is expected (e.g., hot spots in airport terminals), the smaller cells of 802.11a access points and the higher number of available channels (to avoid interference) at 5 GHz are clear advantages.

- **802.11h (Spectrum managed 802.11a):** The 802.11a standard was primarily designed for usage in the US U-NII bands. The standardization did not consider non-US regulations such as the European requirements for power control and dynamic selection of the transmit frequency. To enable the regulatory acceptance of 5 GHz products, dynamic channel selection (DCS) and transmit power control (TPC) mechanisms have been added. With this extension, 802.11a products can also be operated in Europe. These additional mechanisms try to balance the load in the 5 GHz band.

- **802.11i (Enhanced Security mechanisms):** As the original security mechanisms (WEP) proved to be too weak soon after the deployment of the first products , this working group discusses stronger encryption and authentication mechanisms. IEEE 802.1x will play a major role in this process. Additionally, IEEE 802.11 has several **study groups** for new and upcoming topics. The group ‘Radio Resource Measurements’ investigates the possibilities of 802.11 devices to provide measurements of radio resources. Solutions for even higher throughput are discussed in the ‘High Throughput’ study group. The first study group recently became the IEEE project 802.11k ‘Radio Resource Measurement Enhancements.’

14. Explain in detail about the phases , Quality of service and support in HIPERLAN-1 and compare with its remaining type.

In 1996, the ETSI standardized HIPERLAN 1 as a WLAN allowing for node mobility and supporting ad-hoc and infrastructure-based topologies. (HIPERLAN stands for **high performance local area network**.) **HIPERLAN 1** was originally one out of four HIPERLANs envisaged, as ETSI decided to have different types of networks for different purposes. The key feature of all four networks is their integration of time-sensitive data transfer services. Over time, names have changed and the former HIPERLANs 2, 3, and 4 are now called HiperLAN2, HIPERACCESS, and HIPERLINK. The current focus is on HiperLAN2, a standard that comprises many elements from ETSI's **BRAN** (broadband radio access networks) and **wireless ATM** activities. Neither wireless ATM nor HIPERLAN 1 were a commercial success. However, the standardization efforts had a lot of impact on QoS supporting wireless broadband networks such as **HiperLAN2**.

Historical: HIPERLAN 1

ETSI describes HIPERLAN 1 as a wireless LAN supporting priorities and packet life time for data transfer at 23.5 Mbit/s, including forwarding mechanisms, topology discovery, user data encryption, network identification and power conservation mechanisms. HIPERLAN 1 should operate at 5.1-5.3 GHz with a range of 50 m in buildings at 1 W transmit power.

The service offered by a HIPERLAN 1 is compatible with the standard MAC services known from IEEE 802.x LANs. Addressing is based on standard 48 bit MAC addresses. A special HIPERLAN 1 identification scheme allows the concurrent operation of two or more physically overlapping HIPERLANs without mingling their communication. Confidentiality is ensured by an encryption/decryption algorithm that requires the identical keys and initialization vectors for successful decryption of a data stream encrypted by a sender.

An innovative feature of HIPERLAN 1, which many other wireless networks do not offer, is its ability to forward data packets using several relays. Relays can extend the communication on the MAC layer beyond the radio range. For power conservation, a node may set up a specific wake-up pattern. This pattern determines at what time the node is ready to receive, so that at other times, the node can turn off its receiver and save energy. These nodes are called p-savers and need so-called p-supporters that contain information about the wake-up patterns of all the p-savers they are responsible for. A p-supporter only forwards data to a p-saver at the moment the p-saver is awake. This action also requires buffering mechanisms for packets on p-supporting forwarders.

The following describes only the medium access scheme of HIPERLAN 1, a scheme that provides QoS and a powerful prioritization scheme. However, it turned out that priorities and QoS in general are not that important for standard LAN applications today. IEEE 802.11 in its standard versions does not offer priorities, the optional PCF is typically not implemented in products yet 802.11 is very popular.

Elimination-yield non-preemptive priority multiple access (EY-NPMA) is not only a complex acronym, but also the heart of the channel access providing priorities and different access schemes. EY-NPMA divides the medium access of different 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.

In a case where several nodes compete for the medium, all three phases are necessary (called 'channel access in **synchronized channel condition**'). If the channel is free for at least 2,000 so-called high rate bit-periods plus a dynamic extension, only the third phase, i.e. transmission, is needed (called 'channel

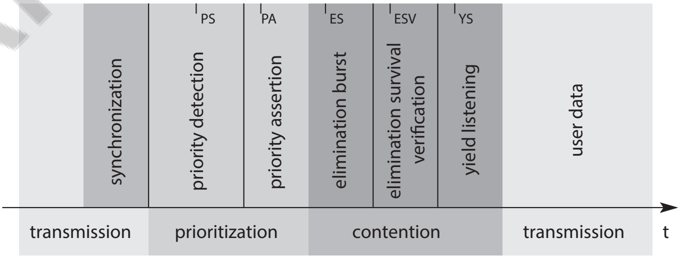


Figure 1.27 Phases of the HIPERLAN 1 EY-NPMA access scheme

access in **channel-free condition**’). The dynamic extension is randomly chosen between 0 and 3 times 200 high rate bit-periods with equal likelihood. This extension further minimizes the probability of collisions accessing a free channel if stations are synchronized on higher layers and try to access the free channel at the same time. HIPERLAN 1 also supports ‘channel access in the **hidden elimination condition**’ to handle the problem of hidden terminals .

The contention phase is further subdivided into an **elimination phase** and a **yield phase**. The purpose of the elimination phase is to eliminate as many contending nodes as possible. The result of the elimination phase is a more or less constant number of remaining nodes, almost independent of the initial number of competing nodes. Finally, the yield phase completes the work of the elimination phase with the goal of only one remaining node.

Figure 1.27 gives an overview of the three main phases and some more details which will be explained in the following sections. For every node ready to send data, the access cycle starts with synchronization to the current sender. The first phase, prioritization, follows. After that, the elimination and yield part of the contention phase follow. Finally, the remaining node can transmit its data. Every phase has a certain duration which is measured in numbers of slots and is determined by the variables I_{PS}, I_{PA}, I_{ES}, I_{ESV}, and I_{YS}.

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 the right to send. The first objective of the prioritization phase is to make sure that no node with a lower priority gains access to the medium while packets with higher priority are waiting at other nodes. This mechanism always grants nodes with higher priority access to the medium, no matter how high the load on lower priorities.

In the first step of the prioritization phase, the priority detection, time is divided into five slots, slot 0 (highest priority) to slot 4 (lowest priority). Each slot has a duration of IPS = 168 high rate bit-periods. If a node has the access priority p, it has to listen into the medium for p slots (priority detection). If the node senses the medium is idle for the whole period of p slots, the node asserts the priority by immediately transmitting a burst for the duration I_{PA} = 168 high rate bit-periods (priority assertion). The burst consists of the following high rate bit sequence, which is repeated as many times as necessary for the duration of the burst:

11111010100010011100000110010110

If the node senses activity in the medium, it stops its attempt to send data in this transmission cycle and waits for the next one. The whole prioritization phase ends as soon as one node asserts the access priority with a burst. This means that the prioritization phase is not limited by a fixed length, but depends on the highest priority.

Let us assume, for example, that there are three nodes with data ready to be sent, the packets of node 1 and node 2 having the priority 2, the packet of node

3 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, the prioritization phase has taken three slots.

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

Elimination phase

Several nodes may now enter the elimination phase. Again, time is divided into slots, using the elimination slot interval I_{ES} = 212 high rate bit periods. The length of an individual elimination burst is 0 to 12 slot intervals long, the probability of bursting within a slot is 0.5. The probability $P_E(n)$ of an elimination burst to be n elimination slot intervals long is given by:

- $P_E(n) = 0.5^{n+1}$ for $0 \leq n < 12$
- $P_E(n) = 0.5^{12}$ for $n = 12$

The elimination phase now resolves contention by means of elimination bursting and elimination survival verification. Each contending node sends an elimination burst with length n as determined via the probabilities and then listens to the channel during the survival verification interval I_{ESV} = 256 high rate bit periods. The burst sent is the same as for the priority assertion. A contending node survives this elimination phase if, and only if, it senses the channel is idle during its survival verification period. Otherwise, the node is eliminated and stops its attempt to send data during this transmission cycle.

The whole elimination phase will last for the duration of the longest elimination burst among the contending nodes plus the survival verification time. One or more nodes will survive this elimination phase, and can then continue with the next phase.

Yield phase

During the yield phase, the remaining nodes only listen into the medium without sending any additional bursts. Again, time is divided into slots, this time called yield slots with a duration of I_{YS} = 168 high rate bit-periods. The length of an individual yield listening period can be 0 to 9 slots with equal likelihood. The probability $P_Y(n)$ 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 channel is idle during the whole period, it has survived the yield listening. Otherwise, it withdraws for the rest of the current transmission cycle. This time, the length of the yield phase is determined by the shortest yield-listening period among all the contending nodes. At least one node will survive this phase and can start to transmit data. This is what the other nodes with longer yield listening period can sense. It is important to note that at this point there can still be more than one surviving node so a collision is still possible.

Transmission phase

A node that has survived the prioritization and contention phase can now send its 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 a unicast transmission, the sender expects to receive an immediate acknowledgement from the destination, called an acknowledgement HCPDU (AK-HCPDU), which is an LBR HCPDU containing only an LBR part.

Quality of service support and other specialties

The speciality of HIPERLAN 1 is its QoS support. The quality of service offered by the MAC layer is based on three parameters (**HMQoS-**

parameters). The user can set a priority for data, priority = 0 denotes a high priority, priority = 1, a low priority. The user can determine the lifetime of an MSDU to specify timebounded delivery. The **MSDU lifetime** specifies the maximum time that can elapse between sending and receiving an MSDU. Beyond this, delivery of the MSDU becomes unnecessary. The MSDU lifetime has a range of 0-16,000 ms. The **residual MSDU lifetime** shows the remaining lifetime of a packet.

Besides data transfer, the MAC layer offers functions for looking up other HIPERLANs within radio range as well as special power conserving functions. **Power conservation** is achieved by setting up certain recurring patterns when a node can receive data instead of constantly being ready to receive. Special group-attendance patterns can be defined to enable multicasting. All nodes participating in a multicast group must be ready to receive at the same time when a sender transmits data.

HIPERLAN 1 MAC also offers user data **encryption** and **decryption** using a simple XOR-scheme together with random numbers. A key is chosen from a set of keys using a key identifier (KID) and is used together with an initialization vector IV to initialize the pseudo random number generator. This random sequence is XORed with the user data (UD) to generate the encrypted data. Decryption of the encrypted UD works the same way, using the same random number sequence. This is not a strong encryption scheme encryption is left to higher layers.

Table 1.4 Mapping of the normalized residual lifetime to the CAC priority

NRL	MSDU priority = 0	MSDU priority = 1
NRL < 10 ms	0	1
10 ms ≤ NRL < 20 ms	1	2
20 ms ≤ NRL < 40 ms	2	3
40 ms ≤ NRL < 80 ms	3	4
80 ms ≤ NRL	4	4

It is interesting to see how the HIPERLAN 1 MAC layer selects the next PDU for transmission if several PDUs are ready and how the waiting time of a PDU before transmission is reflected in its channel access priority. The selection has to reflect the user priority (0 or 1) and the residual lifetime to guarantee a timebounded service. The MAC layer then has to map this information onto a channel access priority used by the CAC, competing with other nodes for the transmit rights.

First of all, the MAC layer determines the **normalized residual HMPDU lifetime (NRL)**. This is the residual lifetime divided by the estimated number of hops the PDU has to travel. The computation reflects both the waiting time of a PDU in the node and the distance, and the additional waiting times in other nodes. Then the MAC layer computes the channel access priority for each PDU following the mapping shown in Table 1.4.

The final selection of the most important HMPDU (HIPERLAN 1 MAC PDU) is performed in the following order:

- HMPDUs with the highest priority are selected;
- from these, all HMPDUs with the shortest NRL are selected;
- from which finally any one without further preferences is selected from the remaining HMPDUs.

Besides transferring data from a sender to a receiver within the same radio coverage, HIPERLAN 1 offers functions to forward traffic via several other wireless nodes a feature which is especially important in wireless ad-hoc networks without an infrastructure. This forwarding mechanism can also be used if a node can only reach an access point via other HIPERLAN 1 nodes.

15. Explain about WATM in detail.

WATM Wireless ATM (WATM; sometimes also called wireless, mobile ATM, wmATM) does not only describe a transmission technology but tries to specify a complete communication system. While many aspects of the IEEE WLANs originate from the data communication community, many

WATM aspects come from the telecommunication industry . This specific situation can be compared to the case of competition and merging with regard to the concepts TCP/IP and ATM (IP-switching, MPLS). Similar to fixed networks where ATM never made it to the desktop, WATM will not make it to mobile terminals. However, many concepts found in WATM can also be found in QoS supporting WLANs such as HiperLAN2.

Motivation for WATM

Several reasons led to the development of WATM:

- The need for seamless integration of wireless terminals into an ATM network. This is a basic requirement for supporting the same integrated services and different types of traffic streams as ATM does in fixed networks.
- ATM networks scale well from LANs to WANs and mobility is needed in local and wide area applications. Strategies were needed to extend ATM for wireless access in local and global environments.
- For ATM to be successful, it must offer a wireless extension. Otherwise it cannot participate in the rapidly growing field of mobile communications.
- WATM could offer QoS for adequate support of multi-media data streams. Many other wireless technologies (e.g., IEEE 802.11) typically only offer best effort services or to some extent, time-bounded services. However, these services do not provide as many QoS parameters as ATM networks do.
- For telecommunication service providers, it appears natural that merging of mobile wireless communication and ATM technology leads to wireless ATM. One goal in this context is the seamless integration of mobility into B-ISDN which already uses ATM as its transfer technology.

It is clear that WATM will be much more complex than most of the other wireless systems. While, for example, IEEE 802.11 only covers local area access methods, Bluetooth only builds up piconets. Mobile IP only works on the network layer, but WATM tries to build up a comprehensive system covering physical layer, media access, routing, integration into the fixed ATM network, service integration into B-ISDN etc.

Wireless ATM working group

To develop this rather complex system, the ATM Forum formed the **Wireless ATM Working Group** in 1996, which aimed to develop a set of specifications that extends the use of ATM technology to wireless networks. These wireless networks should cover many different networking scenarios, such as private and public, local and global, mobility and wireless access .

The main goal of this working group involved ensuring the compatibility of all new proposals with existing ATM Forum standards. It should be possible to upgrade existing ATM networks, i.e., ATM switches and ATM end-systems, with certain functions to support mobility and radio access if required. Two main groups of open issues have been identified in this context: the extensions needed for the ‘fixed’ ATM to support mobility and all protocols and mechanisms related to the radio access.

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

- **Location management:** Similar to other cellular networks, WATM networks must be able to locate a wireless terminal or a mobile user, i.e., to find 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 currently responsible for the wireless terminal. Each time a user moves to a new access point, the system must reroute traffic.
- **Handover signalling:** The network must provide mechanisms which search for new access points, set up new connections between intermediate systems and signal the actual change of the access point.
- **QoS and traffic control:** In contrast to wireless networks offering only best effort traffic, and to cellular networks offering only a few different types of traffic, WATM should be able to offer many QoS parameters. To maintain these parameters, all actions such as rerouting, handover etc. have to be controlled. The network must pay attention to the incoming traffic (and check if it conforms to some traffic contract) in a similar way to today’s ATM (policing).
- **Network management:** All extensions of protocols or other mechanisms also require an extension of the management functions to control the network **Radio resource control:** As for any wireless network, radio frequencies, modulation schemes, antennas, channel coding etc. have to be determined.
- **Wireless media access:** Different media access schemes are possible, each with specific strengths and weaknesses for, e.g., multi-media or voice applications. Different centralized or distributed access schemes working on ATM cells can be imagined.
- ● **Wireless data link control:** The data link control layer might

offer header compression for an ATM cell that carries almost 10 per cent overhead using a 5 byte header in a 53 byte cell. This layer can apply ARQ or FEC schemes to improve reliability.

- ● **Handover issues:** During handover, cells cannot only be lost but can also be out of sequence (depending on the handover mechanisms). Cells must be re-sequenced and lost cells must be retransmitted if required.

WATM services

- **Office environments:** This includes all kinds of extensions for existing fixed networks offering a broad range of Internet/Intranet access, multi-media conferencing, online multi-media database access, and telecommuting. Using WATM technology, the office can be virtually expanded to the actual location of an employee.
- **Universities, schools, training centres:** The main foci in this scenario are distance learning, wireless and mobile access to databases, internet access, or teaching in the area of mobile multi-media computing.
- **Industry:** WATM may offer an extension of the Intranet supporting data base connection, information retrieval, surveillance, but also real-time data transmission and factory management.
- **Hospitals:** Due to the quality of service offered for data transmission, WATM was thought of being the prime candidate for reliable, high-bandwidth mobile and wireless networks. Applications could include the transfer of medical images, remote access to patient records, remote monitoring of patients, remote diagnosis of patients at home or in an ambulance, as well as tele-medicine. The latter needs highly reliable networks with guaranteed quality of service to enable, e.g., remote surgery.
- **Home:** Many electronic devices at home (e.g., TV, radio equipment, CD-player, PC with internet access) could be connected using WATM technology. Here, WATM would permit various wireless connections, e.g., a PDA with TV access.
- **Networked vehicles:** All vehicles used for the transportation of people or goods will have a local network and network access in the future. Currently, vehicles such as trucks, aircraft, buses, or cars

only have very limited communication capabilities (e.g., via GSM, UTMS), WATM could provide them with a high-quality access to the internet, company databases, multimedia conferencing etc. On another level, local networks among the vehicles within a certain area are of increasing importance, e.g., to prevent accidents or increase road capacity by platooning (i.e., forming a train of cars or trucks on the road with very low safety distance between single vehicles).

Mobility within an ATM network is provided by the **ATM mobility extension service (AMES)**. AMES facilitates the use of these ATM networks by different equipment and applications requiring mobility. Wireless equipment should obtain equivalent services from the network as wired terminals from a user's perspective. AMES comprises the extensions needed to support terminal portability for home and business use. Users can rearrange devices without losing access to the ATM network and retain a guaranteed service quality.

WATM should offer a personal cellular system (PCS) **access service**. PCSs like GSM, IS-95, UMTS etc. may use the mobility supporting capabilities of the fixed ATM network to route traffic to the proper base station controller. Public services for users could be a multimedia telephony service, a symmetric service offering speech and low bit rate video with medium mobility, as well as the asymmetrical service of real-time online data transfer, e.g., web browsing, e-mail and downloading of files. Private services could include a multi-media cordless telephone with higher quality compared to the public version. Special private data transfer services, e.g., carrying production data, could be deployed on a campus.

Another field of services is provided by **satellite ATM services (SATM)**. Future satellites will offer a large variety of TV, interactive video, multimedia, Internet, telephony and other services . The main advantage in this context is the ubiquitous wide area coverage in remote, rural, and even urban areas. Satellites can be used directly (direct user access service), e.g., via a mobile phone or a terminal with antenna, which enables the user to access the ATM network directly. A whole network can be connected to a satellite using a mobile switch (fixed access service). For example, all computers in a school in a remote area could be connected to a switch, which connects to a satellite. Even ships can carry ATM networks and can then use the seamless integration of their onboard ATM network to a global

ATM network (mobile platform service).

Generic reference model

Figure 1.28 shows a generic reference model for wireless mobile access to an ATM network. A mobile ATM (MATM) terminal uses a WATM terminal adapter to gain wireless access to a WATM RAS (Radio Access System). MATM terminals could be represented by, e.g., laptops using an ATM adapter for wired access plus software for mobility. The WATM terminal adapter enables wireless access, i.e., it includes the transceiver etc., but it does not support mobility. The RAS with the radio transceivers is connected to a mobility enhanced ATM switch (EMAS-E), which in turn connects to the ATM network with mobility aware switches (EMAS-N) and other standard ATM switches. Finally, a wired, non-mobility aware ATM end system may be the communication partner in this example.

The radio segment spans from the terminal and the terminal adapter to the access point, whereas the fixed network segment spans from the access point to the fixed end system. The fixed mobility support network, comprising all mobility aware switches EMAS-E and EMAS-N, can be distinguished from the standard ATM network with its non-mobility aware switches and end systems.

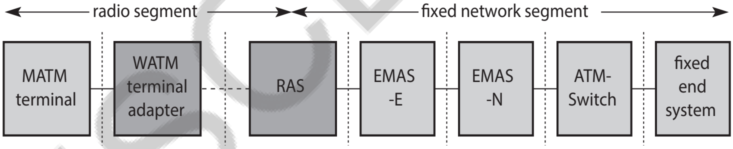


Figure 1.28 Example of a generic WATM reference model

Handover

One of the most important topics in a WATM environment is handover. Connectionless, best-effort protocols supporting handover, such as mobile IP on layer 3 and IEEE 802.11 with IAPP on layer 2, do not have to take too much care about handover quality. These protocols do not guarantee certain traffic parameters as WATM does. The main problem for WATM during the handover is rerouting all connections and maintaining connection quality. While in connectionless, best-effort environments, handover mainly involves rerouting of a packet stream without reliable transport, an end-system in WATM networks could maintain many connections, each with a different quality of service requirements (e.g., limited delay, bounded jitter, minimum bandwidth etc.). Handover not only involves rerouting of connections, it also involves reserving resources in switches, testing of availability of radio bandwidth, tracking of terminals to perform look-ahead reservations etc.

Many different requirements have been set up for handover.

- **Handover of multiple connections:** As ATM is a connection-oriented technology where end-systems can support many connections at the same time, handover in WATM must support more than only one connection. This results in the rerouting of every connection after handover. However, resource availability may not allow rerouting of all connections or forces QoS degradation. The terminal may then decide to accept a lower quality or to drop single connections.
- **Handover of point-to-multi-point connections:** Seamless support of point-to-multi-point connections is one of the major advantages of the ATM technology. WATM handover should also support these types of connection. However, due to the complexity of the scheme, some restrictions might be necessary.
- **QoS support:** Handover should aim to preserve the QoS of all connections during handover. However, due to limited resources, this is not always possible. Functions for QoS re-negotiation and dropping of connections on a priority basis may be required. Candidate access points should advertise their resources to the terminal, and this information could then be used by a handover algorithm to optimize handover and to balance the load between different access points.
- **Data integrity and security:** WATM handover should minimize cell loss and avoid all cell duplication or re-ordering. Security associations between the terminal and the network should not be compromised by handover.
- **Signaling and routing support:** WATM must provide the means to identify mobility-enabled switches in the network, to determine radio adjacent switches by another switch, and to reroute partial connections in the handover domain.
- **Performance and complexity:** The fact that WATM systems are complex by nature is mainly due to their support of connections with QoS. The simplicity of the handover functionality should

be the central goal of the handover design. Modifications to the mobility-enabled switches should be extremely limited, but the functions required could have rather stringent processing time requirements. Due to performance reasons, ATM switches are very much hardware based and it is more difficult to integrate updates and new features. The handover code needed for the terminals should be rather simple due to the fact that increasing code size also requires more processing power, i.e., more battery power, which is typically a serious limitation in the design of mobile terminals.

Location management

As for all networks supporting mobility, special functions are required for looking up the current position of a mobile terminal, for providing the moving terminal with a permanent address, and for ensuring security features such as privacy, authentication, or authorization. These and more functions are grouped under the term **location management**.

Several requirements for location management have been identified :

- **Transparency of mobility:** A user should not notice the location management function under normal operation. Any change of location should be performed without user activity. This puts certain constraints on the permissible time delay of the functions associated with location management. Transparent roaming between different domains (private/private, private/public, public/public) should be possible. This may include roaming between networks based on different technologies using, for example, a dual mode terminal.
- **Security:** To provide a security level high enough to be accepted for mission-critical use (business, emergency etc.), a WATM system requires special features. All location and user information collected for location management and accounting should be protected against unauthorized disclosure. This protection is particularly important for roaming profiles that allow the precise tracking of single terminals. As the air interface is very simple to access, special access restrictions must be implemented to, e.g., keep public users out of private WATM networks. Users should also be able to determine the network their terminal is allowed to access. Essential security features include authentication of users and terminals, but also of access points. Encryption is also necessary, at least between

terminal and access point, but preferably end-to-end.

- **Efficiency and scalability:** Imagine WATM networks with millions of users like today's mobile phone networks. Every function and system involved in location management must be scalable and efficient. This includes distributed servers for location storage, accounting and authentication. The performance of all operations should be practically independent of network size, number of current connections and network load. The clustering of switches and hierarchies of domains should be possible to increase the overall performance of the system by dividing the load. In contrast to many existing cellular networks, WATM should work with a more efficient, integrated signaling scheme. All signaling required for location management should therefore be incorporated into existing signaling mechanisms, e.g., by adding new information elements to existing messages. This allows for the utilization of the existing signaling mechanisms in the fixed ATM network which are efficient.

- **Identification**: Location management must provide the means to identify all entities of the network. Radio cells, WATM networks, terminals, and switches need unique identifiers and mechanisms to exchange identity information. This requirement also includes information for a terminal concerning its current location (home network or foreign network) and its current point of attachment. In addition to the permanent **ATM end system address (AESA)**, a terminal also needs a routable temporary AESA as soon as it is outside its home network. This temporary AESA must be forwarded to the terminal's home location.

- **Inter-working and standards:** All location management functions must cooperate with existing ATM functions from the fixed network, especially routing. Location management in WATM has to be harmonized with other location management schemes, such as location management in GSM and UMTS networks, the internet using Mobile IP, or Intranets with special features. This harmonization could, for instance, lead to a two-level location management if Mobile IP is used on top of WATM. All protocols used in WATM for database updates, registration etc. have to be standardized to permit mobility across provider network boundaries. However, inside an administrative domain, proprietary enhancements and optimizations could be applied.

Mobile quality of service

Quality of service (QoS) guarantees are one of the main advantages envisaged for WATM networks compared to, e.g., mobile IP working over packet radio networks. While the internet protocol IP does not guarantee QoS, ATM networks do (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 same QoS properties as any wired ATM network. Typical traditional QoS parameters are link delay, cell delay variation, bandwidth, cell error rate etc.
- **Wireless QoS:** The QoS properties of the wireless part of a WATM network differ from those of the wired part. Again, link delay and error rate can be specified, but now error rate is typically some order of magnitude that is higher than, e.g., fiber optics. Channel reservation and multiplexing mechanisms at 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 access points, cell loss during handover, or the speed of the whole handover procedure represent critical factors for QoS.

The WATM system has to map the QoS specified by an application onto these sets of QoS parameters at connection setup and has to check whether the QoS requested can be satisfied. However, applications will not specify single parameters in detail, but end-to-end requirements, such as delay or bandwidth. The WATM system must now map, e.g., end-to-end delay onto the cell delays on each segment, wired and wireless. To handle the complexity of such a system, WATM networks will initially only offer a set of different service classes to applications.

Additionally, applications must be adaptive to some degree to survive the effects of mobility, such as higher cell loss, delay variations etc. Applications could, for example, negotiate windows of QoS parameters where they can adapt without breaking the connection.

A crucial point in maintaining QoS over time is QoS support in hand-over protocols. These protocols can support two different types of QoS during handover:

- **Hard handover QoS:** While the QoS with the current RAS may be guaranteed due to the current availability of resources, no QoS guarantees are given after the handover. This is comparable to the traditional approach for, e.g., GSM networks with voice connections. If a terminal can set up a connection, the connection's quality is guaranteed. If there are not enough resources after handover (too many users are already in the target cell), the system cuts off the connection. This is the only possible solution if the applications and terminals cannot adapt to the new situation.

Soft handover QoS: Even for the current wireless segment, only statistical QoS guarantees can be given, and the applications also have to adapt after the handover. This assumes adaptive applications and at least allows for some remaining QoS guarantees during, e.g., periods of congestion or strong interference.

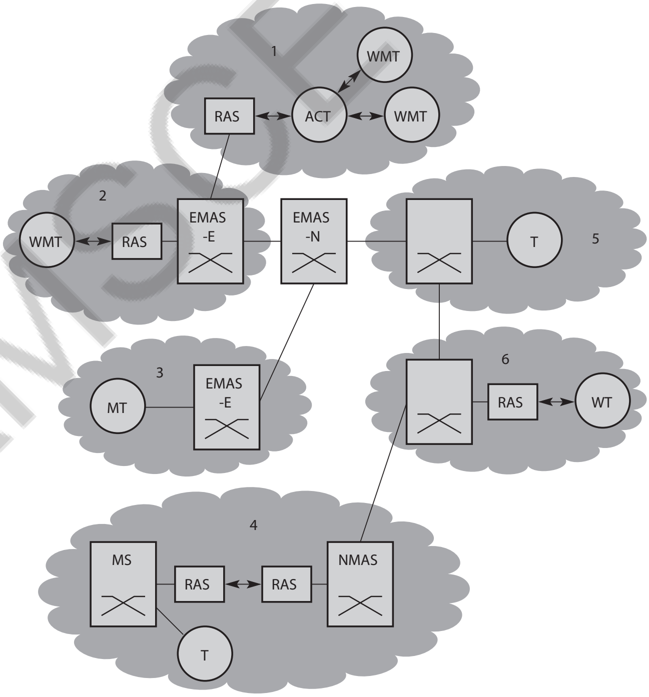


Figure 1.29 WATM reference model with several access

Access scenarios

Figure 1.29 shows possible access scenarios for WATM and illustrates what was planned during the specification of WATM. While this section has focused on the wireless access of mobile ATM terminals, several other configurations are possible (Bhat, 1998). As additional entities, Figure 1.29 shows the following components:

- **T (terminal):** A standard ATM terminal offering ATM services defined for fixed ATM networks.

- **MT (mobile terminal):** A standard ATM terminal with the additional capability of reconnecting after access point change. The terminal can be moved between different access points within a certain domain. ● **WT (wireless terminal):** This terminal is accessed via a wireless link, but the terminal itself is fixed, i.e., the terminal keeps its access point to the network.

- **WMT (wireless mobile terminal):** The combination of a wireless and a mobile terminal results in the WMT. This is exactly the type of terminal presented throughout this WATM section, as it has the ability to change its access point and uses radio access.

- **RAS (radio access system):** Point of access to a network via a radio link as explained in this chapter.

- **EMAS (end-user mobility supporting ATM switch, -E: edge, -N: network):** Switches with the support of end-user mobility.

- **NMAS (network mobility-supporting ATM switch):** A whole network can be mobile not just terminals. Certain additional functions are needed to support this mobility from the fixed network.

- **MS (mobile ATM switch):** ATM switches can also be mobile and can use wireless access to another part of the ATM network.

- **ACT (ad-hoc controller terminal):** For the configuration of ad-hoc networks, special terminal types might be required within the wireless network. These terminals could, for example, control wireless access without an RAS.

Based on these entities, we can define several scenarios which should be supported by WATM if fully specified.

- **Wireless ad-hoc ATM network (scenario 1):** WMTs can communicate with each other without a fixed network. Communication can be set up without any infrastructure. Access control can be accomplished via the ACT. If the ad-hoc network needs a connection to a fixed network, this can be provided by means of an RAS.

- **Wireless mobile ATM terminals (scenario 2):** The configuration discussed throughout this chapter is the wireless and mobile terminal accessing the fixed network via an RAS. In this configuration, a WMT cannot communicate without the support provided by entities within the fixed network, such as an EMAS-E.

- **Mobile ATM terminals (scenario 3):** This configuration supports device portability and allows for simple network reconfiguration. Users can change the access points of their ATM equipment over time without the need for reconfiguration by hand. Again, this scenario needs support through entities in the fixed network (e.g., EMAS-E).

- **Mobile ATM switches (scenario 4):** An even more complex configuration comprises mobile switches using wireless access to other fixed ATM networks. Now entities supporting switch mobility are needed within the fixed network (NMAS). There are many applications for this scenario, e.g., net works in aircraft, trains, or ships. Within the mobile network either fixed, mobile, wireless, or mobile and wireless terminals can be used. This is the most complex configuration ever envisaged within an ATM environment.

- **Fixed ATM terminals (scenario 5):** This configuration is the standard case. Terminals and switches do not include capabilities for mobility or wireless access. This is also the reference configuration for applications which work on top of an ATM network. Convergence layers have to hide the special characteristics of mobility and wireless access because no special applications should be required for the scenarios presented here.

- **Fixed wireless ATM terminals (scenario 6):** To provide simple access to ATM networks without wiring, a fixed wireless link is the ideal solution. Many alternative carriers are using or planning to use this way of accessing customers as they do not own the wired infrastructure. This scenario does not require any changes or enhancements in the fixed network.

The main difference between WATM and other approaches is the integration of a whole system into the specification. WATM specifies radio access, mobility management, handover schemes, mobile QoS, security etc. The main complexity of WATM lies within the functions and protocols needed for handover, due to its desired ability to maintain QoS parameters for connections during handover, and the connection-oriented paradigm of ATM. Consequencently there is a need for resource reservation, checking for available resources at access points, and rerouting of connections.

As WATM was planned as an integrated approach, issues like location management, security, and efficiency of the whole system had to be considered. To minimize overheads, WATM tried to harmonize the functions required with those available in fixed ATM. Overall, the approach was already too ambitious to be realized as a stand-alone network. All configurations should have been able to interact with existing cellular systems and Internet technology.

16. Explain about BRAN in detail.

The broadband radio access networks (BRAN), which have been standardized by the European Telecommunications Standards Institute (ETSI), could have been an RAL for WATM .

The main motivation behind BRAN is the deregulation and privatization of the telecommunication sector in Europe. Many new providers experience problems getting access to customers because the telephone infrastructure belongs to a few big companies.

One possible technology to provide network access for customers is radio. The advantages of radio access are high flexibility and quick installation. Different types of traffic are supported, one can multiplex traffic for higher efficiency, and the connection can be asymmetrical (as, e.g., in the typical www scenario where many customers pull a lot of data from servers but only put very small amounts of data onto them).

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

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

- **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 HIPERACCESS nodes with a high-speed link, HIPERLINK technology can be chosen. HIPERLINK provides a fixed point-to-point connection with up to 155 Mbit/s. Currently, there are no plans regarding this standard.

Common characteristics of HIPERLAN/2, HIPERACCESS, and HIPERLINK include their support of the ATM service classes CBR, VBR-rt, VBR-nrt, UBR, and ABR. It is clear that only HiperLAN2 can be a candidate for the RAL [illegible] is tec[illegible]ls the requirements of ATM QoS support, [illegible] less a[illegible]h bandwidth.

As an access network, [illegible] from the protocols of the fixed network. BRAN can be used for ATM and TCP/IP networks as illustrated in Figure 1.[illegible] physical layers, the DLC layer of BRAN offers a common interface to higher

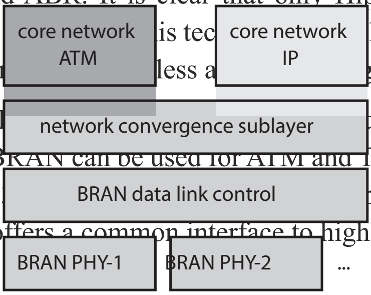


Figure 1.30 Layered model of BRAN wireless access networks

layers. To cover special characteristics of wireless links and to adapt directly to different higher layer network technologies, BRAN provides a network convergence sublayer. 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 ATM network would use services of the BRAN network convergence sublayer.

17. Explain the features of HiperLAN2 .

While HIPERLAN 1 did not succeed HiperLAN2 might have a better chance. (This is also written as HIPERLAN/2, HiperLAN/2, H/2; official name: HIPERLAN Type 2.) Standardized by ETSI (2000a) this wireless network works at 5 GHz (Europe: 5.15-5.35 GHz and 5.47-5.725 GHz license exempt bands; US: license free U-NII bands and offers data rates of up to 54 Mbit/s including QoS support and enhanced security features.

In comparison with basic IEEE 802.11 LANs, HiperLAN2 offers more features in the mandatory parts of the standard.

- **High-throughput transmission:** Using OFDM in the physical layer and a dynamic TDMA/TDD-based MAC protocol, HiperLAN2 not only offers up to 54 Mbit/s at the physical layer but also about 35 Mbit/s at the network layer. The overheads introduced by the layers (medium access, packet headers etc.) remains almost constant over a wide rage of user packet sizes and data rates. HiperLAN2 uses MAC frames with a constant length of 2 ms.

- **Connection-oriented:** Prior to data transmission HiperLAN2 networks establish logical connections between a sender and a receiver (e.g., mobile device and access point). Connection set-up is used to negotiate QoS parameters. All connections are time-division-multiplexed over the air interface (TDMA with TDD for separation of up/downlink). Bidirectional point-topoint as well as unidirectional point-to-multipoint connections are offered. Additionally, a broadcast channel is available to reach all mobile devices in the transmission range of an access point.

- **Quality of service support:** With the help of connections, support of QoS is much simpler. Each connection has its own set of QoS parameters (bandwidth, delay, jitter, bit error rate etc.). A more simplistic scheme using priorities only is available.

- **Dynamic frequency selection:** HiperLAN2 does not require frequency planning of cellular networks or standard IEEE 802.11 networks. All access points have built-in support which automatically selects an appropriate frequency within their coverage area. All APs listen to neighboring APs as well as to other radio sources in the environment. The best frequency is chosen depending on the current interference level and usage of radio channels.
- **Security support:** Authentication as well as encryption are supported by HiperLAN2. Both, mobile terminal and access point can authenticate each other. This ensures authorized access to the network as well as a valid network operator. However, additional functions (directory services, key exchange schemes etc.) are needed to support authentication. All user traffic can be encrypted using DES, Triple-DES, or AES to protect against eavesdropping or man-in-the-middle attacks.
- **Mobility support:** Mobile terminals can move around while transmission always takes place between the terminal and the access point with the best radio signal. Handover between access points is performed automatically. If enough resources are available, all connections including their QoS parameters will be supported by a new access point after handover. However, some data packets may be lost during handover.
- **Application and network independence:** HiperLAN2 was not designed with a certain group of applications or networks in mind. Access points can connect to LANs running ethernet as well as IEEE 1394 (Firewire) systems used to connect home audio/video devices. Interoperation with 3G networks is also supported, so not only best effort data is supported but also the wireless connection of, e.g., a digital camera with a TV set for live streaming of video data.
- **Power save:** Mobile terminals can negotiate certain wake-up patterns to save power. Depending on the sleep periods either short latency require ments or low power requirements can be supported.

18. Explain in detail reference architecture of HiperLAN2 .(May/ June 2012)

Reference model and configurations

Figure 1.31 shows the standard architecture of an infrastructure-based HiperLAN2 network. In the example, two **access points** (AP) are attached to a core network. Core networks might be Ethernet LANs, Firewire connections between audio and video equipment, ATM networks, UMTS 3G cellular phone networks etc. Each AP consists of an **access point controller** (APC) and one or more **access point transceivers** (APT). An APT can comprise one or more sectors (shown as cell here). Finally, four **mobile terminals** (MT) are also shown. MTs can move around in the cell area as shown. The system automatically assigns the APT/AP with the best transmission quality. No frequency planning is necessary as the APs automatically select the appropriate frequency via **dynamic frequency selection**

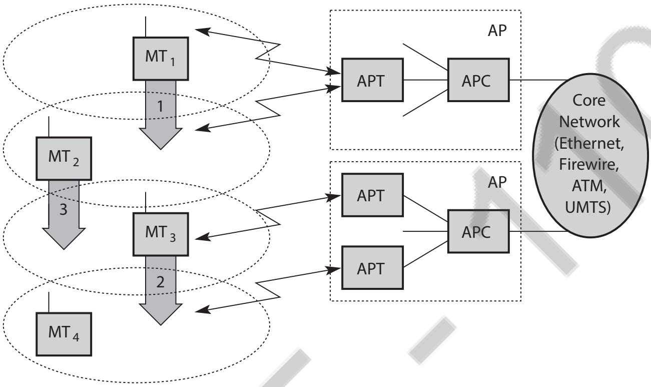


Figure 1.31 HiperLAN2 basic structure and handover scenarios

Three handover situations may occur:

- **Sector handover** (Inter sector): If sector antennas are used for an AP, which is optional in the standard, the AP shall support sector handover. This type of handover is handled inside the DLC layer so is not visible outside the AP (as long as enough resources are available in the new sector).
- **Radio handover** (Inter-APT/Intra-AP): As this handover type, too, is handled within the AP, no external interaction is needed. In the

example of Figure 7.31 the terminal MT_3, moves from one APT to another of the same AP. All context data for the connections are already in the AP (encryption keys, authentication, and connection parameters) and does not have to be renegotiated. ● **Network handover** (Inter-AP/Intra-network): This is the most complex situation: MT_2 moves from one AP to another. In this case, the core network and higher layers are also involved. This handover might be supported by the core network (similar to the IAPP, IEEE 802.11f). Otherwise, the MT must provide the required information similar to the situation during a new association. HiperLAN2 networks can operate in two different modes (which may be used simultaneously in the same network).

✷ **Centralized mode** (CM): This infrastructure-based mode is shown again in a more abstract way in Figure 7.32 (left side). All APs are connected to a core network and MTs are associated with APs. Even if two MTs share the same cell, all data is transferred via the AP. In this mandatory mode the AP takes complete control of everything.

✷ **Direct mode** (DM): The optional ad-hoc mode of HiperLAN2 is illustrated on the right side of Figure 1.32. Data is directly exchanged between MTs if they can receive each other, but the network still has to be controlled. This can be done via an AP that contains a central controller (CC) anyway or via an MT that contains the CC functionality. There is no real difference between an AP and a CC besides the fact that APs are always connected to an infrastructure but here only the CC functionality is needed. This is why the standard coined two different names. IEEE 802.11, too, offers an ad-hoc mode, but not the CC functionality for QoS support.

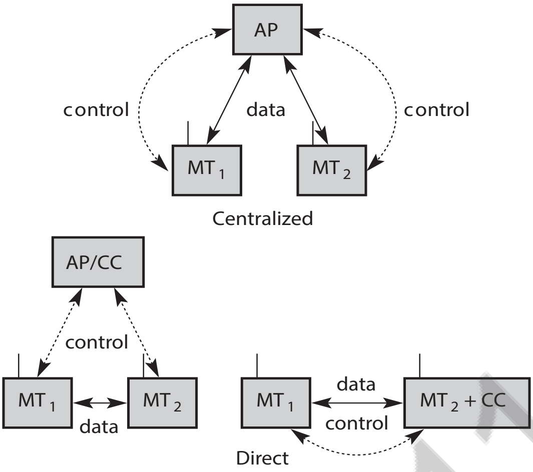


Figure 1.32 HiperLAN2 centralized vs direct mode

Figure 1.33 shows the HiperLAN2 protocol stack as used in access points. Protocol stacks in mobile terminals differ with respect to the number of MAC and RLC instances (only one of each). The lowest layer, the **physical layer**, handles as usual all functions related to modulation, forward error correction, signal detection, synchronization etc. Section 7.4.4.2 describes the physical layer in more detail. The **data link control** (DLC) layer contains the MAC functions, the RLC sublayer and error control functions. If an AP comprises several APTs then each APT requires an own MAC instance. The **MAC** of an AP assigns each MT a certain capacity to guarantee connection quality depending on available resources. Above the MAC DLC is divided into a control and a user part. This separation is common in classical connection-oriented systems such as cellular phones or PSTN. The user part contains **error control** mechanisms. HiperLAN2 offers reliable data transmission using acknowledgements and retransmissions. For broadcast transmissions a repetition mode can be used that provides increased reliability by repeating data packets. Additionally, unacknowledged data transmission is available. The **radio link control** (RLC) sublayer comprises most control functions in the DLC layer (the CC part of an AP). The **association control function** (ACF) controls association and authentication of new MTs as well as synchronization of the radio cell via beacons. The **DLC user connection control** (DCC or DUCC) service controls connection setup, modification, and release.

Finally, the **radio resource control** (RRC) handles handover between APs and within an AP. These functions control the dynamic frequency selection and power save mechanisms of the MTs.

On top of the DLC layer there is the **convergence layer**. This highest layer of HiperLAN2 standardization may comprise segmentation and reassembly functions and adaptations to fixed LANs, 3G networks etc. The following sections give some more insight into the 3 HiperLAN2 layers.

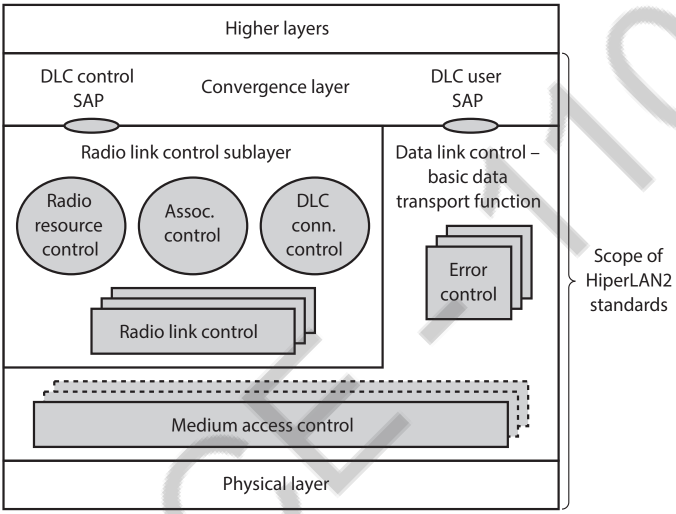


Figure 1.33 HiperLAN2 protocol stack

Physical layer

Many functions and features of HiperLAN2's physical layer (ETSI, 2001a) served as an example for IEEE 802.11a. It is not surprising that both standards offer similar data rates and use identical modulation schemes. Table 1.5 gives an overview of the data rates offered by HiperLAN2 together with other parameters such as coding (compare this with Table 1.3).

Figure 1.34 illustrates the reference configuration of the transmission chain of a HiperLAN2 device. After selecting one of the above transmission modes, the DLC layer passes a PSDU to the physical layer (PSDUs are called DLC PDU trains

Table 1.5 rate dependent parameters for HiperLAN 2

Data rate [Mbit/s]	Modulation	Coding rate	Coded bits per sub-carrier	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
27	16-QAM	9/16	4	192	108
36	16-QAM	3/4	4	192	144
54	64-QAM	3/4	6	288	216

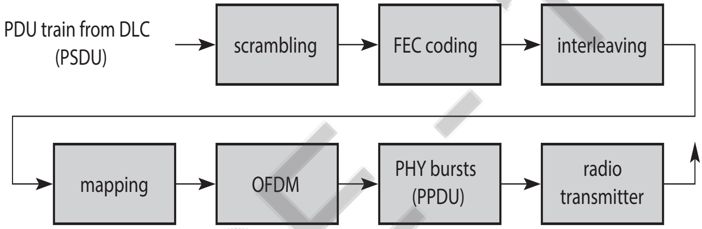


Figure 1.34 HiperLAN2 physical layer reference configuration

in the HiperLAN2 context). The first step then is **scrambling** of all data bits with the generator polynomial $x^7 + x^4 + 1$ for DC blocking and whitening of the spectrum. The result of this first step are **scrambled bits**. The next step applies **FEC coding** for error protection. Coding depends on the type of data (broadcast, uplink, downlink etc.) and the usage of sector or omni-directional antennas. The result of this step is an **encoded bit**. For mitigation of frequency selective fading **interleaving** is applied in the third step. Interleaving ensures that adjacent encoded bits are mapped onto non-adjacent subcarriers (48 subcarriers are used for data transmission). Adjacent bits are mapped alternately onto less and more significant bits of the constellation. The result is an **interleaved bit**.

The following **mapping** process first divides the bit sequence in groups of 1, 2, 4, or 6 bits depending on the modulation scheme (BPSK, QPSK, 16-QAM, or 64QAM). These groups are mapped onto the appropriate modulation symbol according to the constellation diagrams standardized in (ETSI, 2001a). The results of this mapping are **subcarrier modulation symbols**. The **OFDM** modulation step converts these symbols into a baseband signal with the help of the inverse FFT. The symbol interval is 4 μs with 3.2 μs useful part and 0.8 μs guard time. Pilot sub-carriers (sub-carriers -21, -7, 7, 21) are added. The last step before radio transmission is the creation of **PHY bursts** (PPDUs in ISO/OSI terminology). Each burst consists of a preamble and a payload. Five different PHY bursts have been defined: broadcast, downlink, uplink with short preamble, uplink with long preamble, and direct link (optional). The bursts differ in their preambles.

The final **radio transmission** shifts the baseband signal to a carrier frequency depending on the channel number and the formula already used for 802.11a: carrier_number = (carrier_frequency 5000 MHz)/5 MHz. All nominal carrier frequencies are spaced 20 MHz apart, resulting in a frequency allocation table for Europe as illustrated in Figure 1.35.

Maximum transmit power is 200 mW EIRP for the lower frequency band (indoor use) and 1 W EIRP for the higher frequency band (indoor and outdoor use). DFS and TPC are not necessary, if the transmit power stays below 50 mW EIRP and only 5.15-5.25 GHz are used (be aware that national differences exist even within Europe and regulation may change over time).

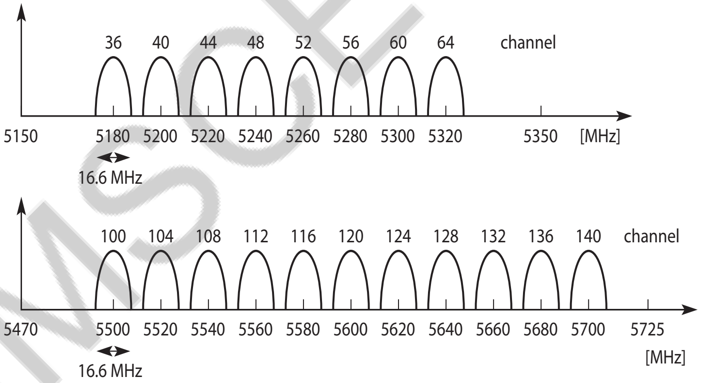


Figure 1.35 Operating channels of HiperLAN2 in Europe

As described above, the DLC layer is divided into MAC, control and data part (which would fit into the LLC sublayer according to ISO/OSI). ETSI (2001b) standardizes the basic data transport functions, i.e., user part with error control and MAC, while ETSI (2002a) defines RLC functionality.

The medium access control creates frames of 2 ms duration as shown in Figure 7.36. With a constant symbol length of four μs this results in 500 OFDM symbols. Each MAC frame is further sub-divided into four phases with variable 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).

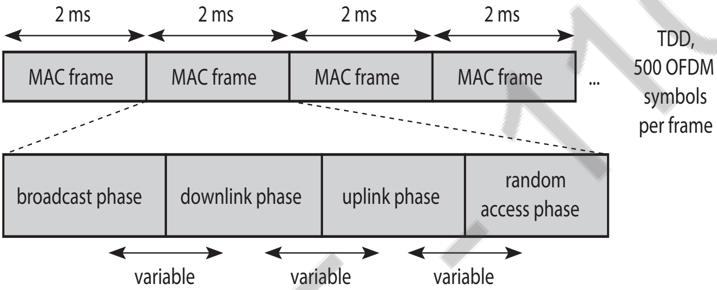


Figure 1.36 Basic structure of HiperLAN2 MAC frames

An optional **direct link phase** can be inserted between the downlink and the uplink phase. The access to the common physical medium is always controlled by the CC (typically in an AP).

HiperLAN2 defines six different so-called transport channels for data transfer in the above listed phases. These transport channels describe the basic message format within a MAC frame.

- **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). This also comprises the PHY mode used. 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. As the access during the RCHs is based on slotted Aloha,

collision at the AP may occur. The ACH signals back which slot was successfully transmitted. 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. Access is via slotted Aloha so, collisions may occur. Collision resolution is performed with the help of an exponential back-off scheme (ETSI, 2001b). The length is 9 bytes. A maximum number of 31 RCHs is currently supported.

BCH, FCH and ACH are used in the broadcast phase only and use BPSK with code rate 1/2. LCH and SCH can be used in the downlink, uplink or (optional) direct link phase. RCH is used in the uplink only for random access (BPSK, code rate 1/2). HiperLAN2 defines further how many of the channels are used within a MAC frame. This configuration may change from MAC frame to MAC frame depending on the connection QoS, resource requests, number of MTs etc. Figure

1.37 shows valid combinations of channels/transfer phases within MAC frames. It is required that the transport channels BCH, FCH and ACH are present plus at least one RCH. While the duration of the BCH is fixed (15 byte), the duration of the others may vary (either due to a variable size of the channel or due to the multiple use of channels). However, the order BCH-FCH-ACH-DL phase-UL phase-RCH must be kept from an MT's point of view (centralized mode). For the direct mode the DiL phase is inserted between the DL and UL phases.

Data between entities of the DLC layer are transferred over so-called **logical channels** (just another name for any distinct data path). The type of a logical

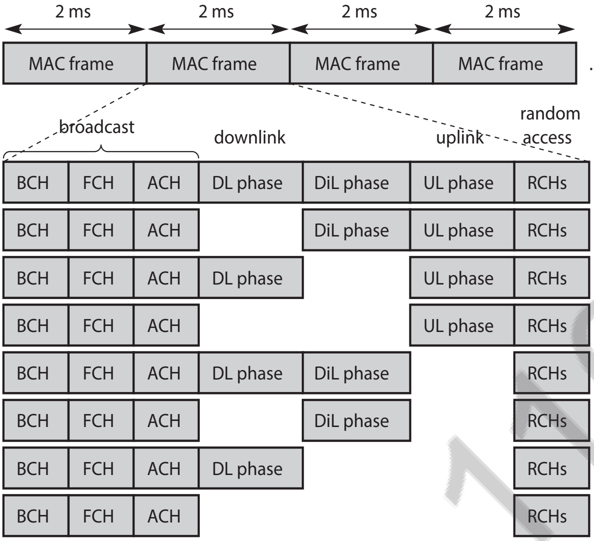


Figure 1.37 Valid configurations of MAC frames

channel is defined by the type of information it carries and the interpretation of the values in the corresponding messages. This is a well-known concept from, e.g., cellular phone systems like GSM. The following logical channels are defined in HiperLAN2 (logical channels use 4 letter acronyms):

- **Broadcast control channel (BCCH):** This channel on the downlink conveys a constant amount of broadcast information concerning the whole radio cell. Examples are the seed for the scrambler, network/ access point/sector identifiers, AP transmission power, expected AP reception power, pointers to the FCH/RCH, number of RCHs (1 to 31), load indicator, number of sectors etc.

- **Frame control channel (FCCH):** The FCCH describes the structure of the remaining parts of the MAC frame. This comprises resource grants for SCHs and LCHs belonging to certain MTs. Resource grants contain the MAC address the grant belongs to, the number of LCHs and SCHs, their PHY modes etc. This scheme allows for a precise reservation of the medium with associated QoS properties.

- **Random access feedback channel (RFCH):** This channel informs MTs that have used an RCH in the previous frame about the success of their access attempt.

- **RLC broadcast channel (RBCH):** This channel transfers information regarding RLC control information, MAC IDs during an association phase, information from the convergence layer, or seeds for the encryption function only if necessary.

- **Dedicated control channel (DCCH):** This channel carries RLC messages related to a certain MT and is established during the association of an MT.
- **User broadcast channel (UBCH):** A UBCH transfers broadcast messages from the convergence layer. Transmission is performed in the unacknowledged or repetition mode.
- **User multi-cast channel (UMCH):** This channel performs unacknowledged transmission of data to a group of MTs.
- **User data channel (UDCH):** Point-to-point data between an AP and an MT (CM) or between two MTs (DM) use this channel. Error protection via an ARQ scheme is possible.
- **Link control channel (LCCH):** This bi-directional channel conveys ARQ feedback and discards messages between the error control functions of an AP and an MT (CM) or between two MTs (DM). A LCCH is typically assigned to a UDCH.
- **Association control channel (ASCH):** This channel is only used in the uplink and for currently non-associated MTs (related to a certain AP). This is the case for a new association request (new MT in the network) or a handover request on behalf of the RLC.

The reader may have noticed that some transport channels transfer exactly the information of one logical channel as their descriptions were identical. This is indeed the case for some channels (BCCH-BCH, FCCH-FCH, RFCH-ACH) as the scheme of mapping logical and transport channels shows (see Figure 7.38). This figure also shows in which mode which channel can be used (uplink and downlink in the centralized mode, direct link in the direct mode).

Figure 1.39 gives an example for mapping the logical channel UDCH to the transport channel LCH. The payload of the LCH is used for a sequence number plus the payload of the UDCH.

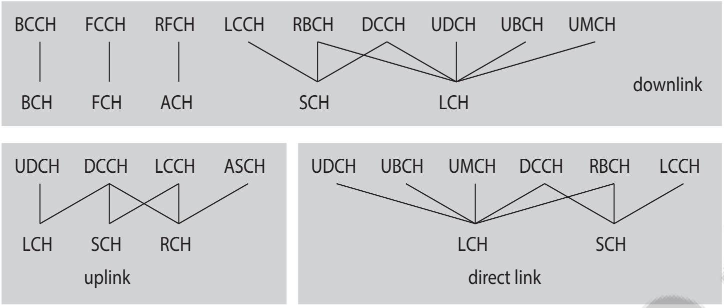


Figure 1.38 Mapping of logical and transport channels downlink

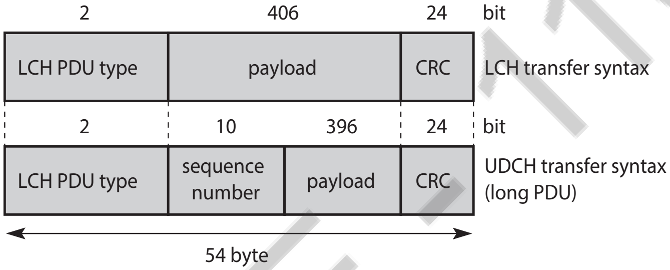


Figure 1.39 HiperLAN2 LCH and UDCH transfer syntax

The **radio link control** sublayer in connection oriented systems offering QoS like HiperLAN2 is quite complex and comprises many protocols, functions, and messages. ETSI (2002a) defines three main services for the RLC sublayer:

- **Association control function (ACF):** ACF contains all procedures for association, authentication, and encryption. An MT starts the association process. The first step is the synchronization with a beacon signal transmitted in each BCCH of a MAC frame. The network ID may be obtained via the RBCH. The next step is the MAC ID assignment. This unique ID is used to address the MT. From this point on, all RLC control messages are transmitted via a DCCH. During the following link capability negotiation, lists of supported convergence layers, authentication and encryption procedures are exchanged. Depending on these parameters the following steps may take place: encryption start-up, authentication, obtaining the ID of the MT. If all necessary steps are successful the MT is associated with the AP. Disassociation may take place at any time, either explicitly (MT or AP initiated) or implicitly (loss of the radio connection). The AP may send MT-alive messages to check if an MT is still available.

- **Radio resource control (RRC):** An important function of the RRC is handover support as already shown in Figure 7.31. Each associated MT continuously measures the link quality. To find handover candidates the MT additionally checks other frequencies. If only one transceiver is available the MT announces to the AP that it is temporarily unavailable (MT absence). Based on radio quality measurements, an AP can change the carrier frequency dynamically (DFS). The RLC offers procedures to inform all MTs. To minimize interference with other radio sources operating at the same frequency (HiperLAN2s or other WLANs) transmission power control (TPC) must be applied by the RRC. An MT can save power by negotiating with an AP a sleeping period of *n* MAC frames. After these *n* frames the MT may wake up because data is ready to be sent, or the AP signals data to be received. If the MT misses the wakeup message from the AP it starts the MT alive procedure. If no data has to be transmitted the MT can again fall asleep for *n* frames.
- **DLC user connection control (DCC or DUCC):** This service is used for set-
- ting up, releasing, or modifying unicast connections. Multi-cast and broadcast connections are implicitly set-up by a group/broadcast join during the association procedure.

Convergence layer

As the physical layer and the data link layer are independent of specific core network protocols, a special **convergence layer (CL)** is needed to adapt to the specific features of these network protocols. HiperLAN2 supports two different types of CLs: cell-based and packet-based. The **cell-based** CL expects data packets of fixed size (cells, e.g., ATM cells), while the **packet-based** CL (ETSI, 2000d) handles packets that are variable in size (e.g., Ethernet or Firewire frames). For the packet-based CL additional functionality is necessary for segmentation and reassembling of packets that do not fit into the DLC payload of HiperLAN2 (49.5 byte). Three examples of convergence layers follow:

- **Ethernet:** This sublayer supports the transparent transport of

Ethernet frames over a HiperLAN2 wireless network (ETSI, 2001d). This includes the mapping of Ethernet multicast and broadcast messages onto HiperLAN2 multicast and broadcast messages. A collision domain can also be emulated. This sublayer also supports priorities according to IEEE 802.1p. The standard supports the traffic classes best effort, background, excellent effort, controlled load, video, voice, and network control. The sublayer does not transmit the Ethernet preamble, start of frame delimiter, and frame check sequence. These fields of an Ethernet frame are not necessary during transmission and will be appended in the receiver's Ethernet sublayer.

- **IEEE 1394 (Firewire):** As a high-speed real-time bus for connecting, e.g., audio and video devices, timing and synchronization is of special importance for IEEE 1394. ETSI (2001e) supports synchronization of timers via the air and treats isochronous data streams with special regard to jitter

- **ATM:** The cell-based CL is used for this type of network (ETSI, 2000c). As the payload of an ATM cell is only 48 byte, which fits into the 49.5 byte of a DLC-PDU, segmentation and reassembly is not necessary. In this case, the sublayer only has to control connection identifiers and MAC IDs.

19. Explain about architecture and protocols of Bluetooth in detail.

Compared to the WLAN technologies ,Bluetooth technology aims at so-called **ad-hoc piconets**, which are local area networks with a very limited coverage and without the need for an infrastructure. This is a different type of network is needed to connect different small devices in close proximity (about 10 m) without expensive wiring or the need for a wireless infrastructure. The envisaged gross data rate is 1 Mbit/s, asynchronous (data) and synchronous (voice) services should be available. Many of today's devices offer an infra red data association (IrDA) interface with transmission rates of, e.g., 115 kbit/s or 4 Mbit/s. There are various problems with IrDA: its very limited range (typically 2 m for built-in interfaces), the need for a line-of-sight between the interfaces, and, it is usually limited to two participants, i.e., only point-to-point connections are supported. IrDA has no internet working functions, has no media access, or any other enhanced communication mechanisms. The big advantage of IrDA is its low cost, and it can be found in almost any mobile device (laptops, PDAs, mobile phones).

It took a thousand years before the Swedish IT-company Ericsson initiated some studies in 1994 around a so-called multi-communicator link (Haartsen, 1998). The project was renamed (because a friend of the designers liked the Vikings) and Bluetooth was born. In spring 1998 five companies (Ericsson, Intel, IBM, Nokia, Toshiba) founded the Bluetooth consortium with the goal of developing a single-chip, low-cost, radio-based wireless network technology. Many other companies and research institutions joined the special interest group around Bluetooth (2002), whose goal was the development of mobile phones, laptops, notebooks, headsets etc. including Bluetooth technology

In 2001, the first products hit the mass market, and many mobile phones, laptops, PDAs, video cameras etc. are equipped with Bluetooth technology today. At the same time the Bluetooth development started, a study group within IEEE 802.11 discussed **wireless personal area networks (WPAN)** under the following five criteria:

- **Market potential:** How many applications, devices, vendors, customers are available for a certain technology?
- **Compatibility:** Compatibility with IEEE 802.
- **Distinct identity:** Originally, the study group did not want to establish a second 802.11 standard. However, topics such as, low cost, low power, or small form factor are not addressed in the 802.11 standard.
- **Technical feasibility:** Prototypes are necessary for further discussion, so the study group would not rely on paper work.
- **Economic feasibility:** Everything developed within this group should be cheaper than other solutions and allow for high-volume production.

Obviously, Bluetooth fulfills these criteria so the WPAN group cooperated with the Bluetooth consortium. IEEE founded its own group for WPANs, IEEE 802.15, in March 1999. This group should develop standards for wireless communications within a **personal operating space.** A POS has been defined as a radius of 10 m around a person in which the person or devices of this person communicate with other devices.

User

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

- **Connection of peripheral devices:** Today, most devices are connected to a desktop computer via wires (e.g., keyboard, mouse, joystick, headset, speakers). This type of connection has several disadvantages: each device has its own type of cable, different plugs are needed, wires block office space. In a wireless network, no wires are needed for data transmission. However, batteries now have to replace the power supply, as the wires not only transfer data but also supply the peripheral devices with power.
- **Support of ad-hoc networking:** Imagine several people coming together, discussing issues, exchanging data. For instance, students might join a lecture, with the teacher distributing data to their personal digital assistants (PDAs). Wireless networks can support this type of interaction; small devices might not have WLAN adapters following the IEEE 802.11 standard, but cheaper Bluetooth chips built in.

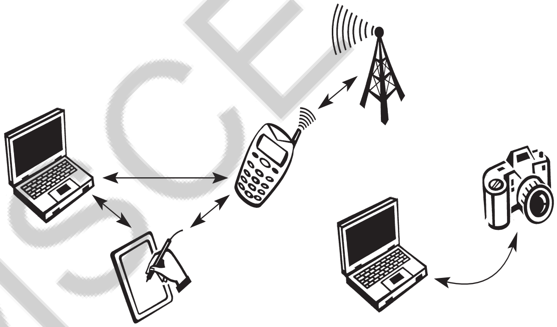


Figure 1.40 Example configurations with a Bluetooth-based piconet

- **Bridging of networks:** Using wireless piconets, a mobile phone can be connected to a PDA or laptop in a simple way. Mobile phones will not have full WLAN adapters built in, but could have a Bluetooth chip. The mobile phone can then act as a bridge between the local piconet and, e.g., the global GSM network . For instance, on arrival at an airport, a person’s mobile phone could receive e-mail via GSM and forward it to the laptop which is still in a suitcase. Via a piconet, a fileserver could update local information stored on a laptop or PDA while the person is walking into the office. When comparing Bluetooth with other WLAN technology we have to keep in mind that one of its goals was to provide local wireless access at very low cost. From a technical point of view, WLAN technologies like those above could also be used, however, WLAN adapters, e.g., for IEEE 802.11, have been designed for higher bandwidth and larger range and are more expensive and consume a lot more power.

Architecture

Like IEEE 802.11b, Bluetooth operates in the 2.4 GHz ISM band. However, MAC, physical layer and the offered services are completely different. After presenting the overall architecture of Bluetooth and its specialty, the piconets, the following sections explain all protocol layers and components in more detail.

Networking

To understand the networking of Bluetooth devices a quick introduction to its key features is necessary. Bluetooth operates on 79 channels in the 2.4 GHz band with 1 MHz carrier spacing. Each device performs frequency hopping with 1,600 hops/s in a pseudo random fashion. Bluetooth applies FHSS for interference mitigation (and FH-CDMA for separation of networks).

A very important term in the context of Bluetooth is a **piconet**. A piconet is a collection of Bluetooth devices which are synchronized to the same hopping sequence. Figure 1.41 shows a collection of devices with different roles. One device in the piconet can act as **master** (M), all other devices connected to the master must act as **slaves** (S). The master determines the hopping

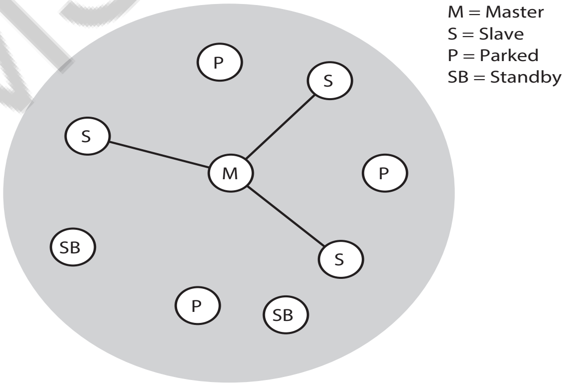


Figure 1.41 Simple Bluetooth piconet

pattern in the piconet and the slaves have to synchronize to this pattern. Each piconet has a unique hopping pattern. If a device wants to participate it has to synchronize to this. Two additional types of devices are shown: parked devices (P) can not actively participate in the piconet (i.e., they do not have a connection), but are known and can be reactivated within some milliseconds .Devices in stand-by (SB) do not participate in the piconet. Each piconet has exactly one master and up to seven simultaneous slaves. More than 200 devices can be parked. The reason for the upper limit of eight active devices, is the 3-bit address used in Bluetooth. If a parked device wants to communicate and there are already seven active slaves, one slave has to switch to park mode to allow the parked device to switch to active mode.

Figure 1.42 gives an overview of the formation of a piconet. As all active devices have to use the same hopping sequence they must be synchronized. The first step involves a master sending its clock and device ID. All Bluetooth devices have the same networking capabilities, i.e., they can be master or slave. There is no distinction between terminals and base stations, any two or more devices can form a piconet. The unit establishing the piconet automatically becomes the master, all other devices will be slaves. The hopping pattern is determined by the device ID, a 48-bit worldwide unique identifier. The phase in the hopping pattern is determined by the master's clock. After adjusting the internal clock according to the master a device may participate in the piconet. All active devices are assigned a 3-bit **active member address** (AMA). All parked devices use an 8-bit **parked member address** (PMA). Devices in stand-by do not need an address.

All users within one piconet have the same hopping sequence and share the same 1 MHz channel. As more users join the piconet, the throughput per user drops quickly (a single piconet offers less than 1 Mbit/s gross data rate). This

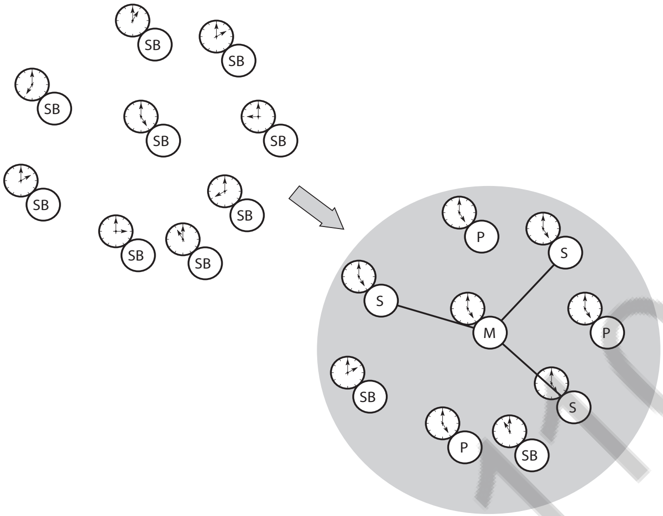


Figure 1.42 Forming a Bluetooth piconet

led to the idea of forming groups of piconets called **scatternet.** Only those units that really must exchange data share the same piconet, so that many piconets with overlapping coverage can exist simultaneously.

In the example, the scatternet consists of two piconets, in which one device participates in two different piconets. Both piconets use a different hopping sequence, always determined by the master of the piconet. Bluetooth applies **FH-CDMA** for separation of piconets. In an average sense, all piconets can share the total of 80 MHz bandwidth available. Adding more piconets leads to a graceful performance degradation of a single piconet because more and more collisions may occur. A collision occurs if two or more piconets use the same carrier frequency at the same time. This will probably happen as the hopping sequences are not coordinated.

If a device wants to participate in more than one piconet, it has to synchronize to the hopping sequence of the piconet it wants to take part in. If a device acts as slave in one piconet, it simply starts to synchronize with the hopping sequence of the piconet it wants to join. After synchronization, it acts as a slave in this piconet and no longer participates in its former piconet. To enable synchronization, a slave has to know the identity of the master that determines the hopping sequence of a piconet. Before leaving one piconet, a slave informs the current master that it will be unavailable for a certain amount of time. The remaining devices in the piconet continue to communicate as usual.

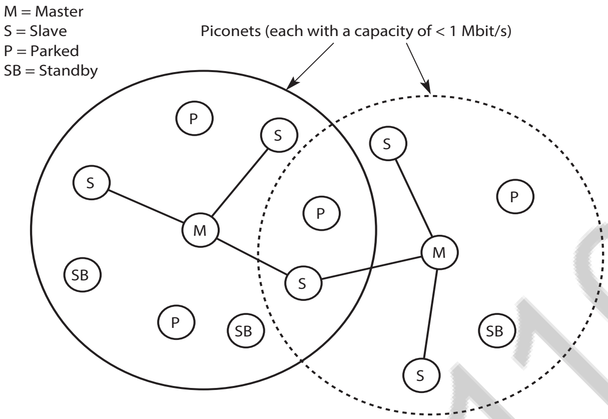


Figure 1.43 Bluetooth scatternet

A master can also leave its piconet and act as a slave in another piconet. It is clearly not possible for a master of one piconet to act as the master of another piconet as this would lead to identical behavior (both would have the same hopping sequence, which is determined by the master per definition). As soon as a master leaves a piconet, all traffic within this piconet is suspended until the master returns.

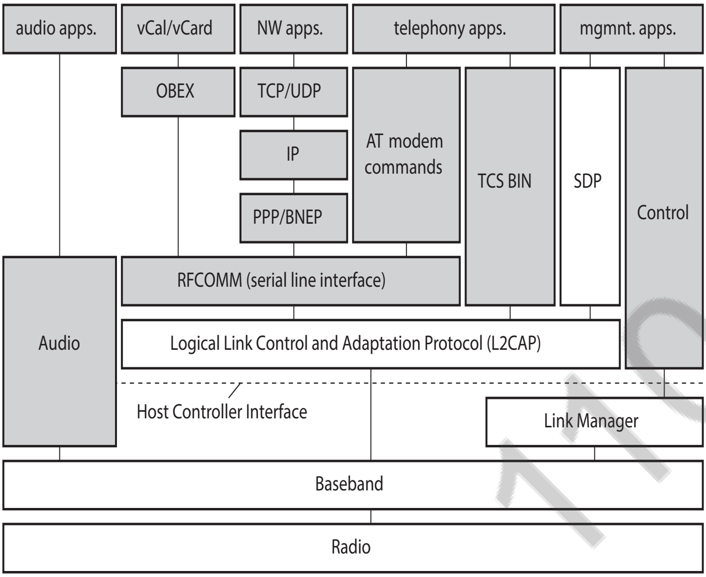
Communication between different piconets takes place by devices jumping back and forth between theses nets. If this is done periodically, for instance, isochronous data streams can be forwarded from one piconet to another. However, scatternets are not yet supported by all devices.

Protocol stack

As Figure 1.44 shows, the Bluetooth specification already comprises many protocols and components. The Bluetooth protocol stack can be divided into a **core specification** which describes the protocols from physical layer to the data link control together with management functions, and **profile specifications**.

The **core protocols** of Bluetooth comprise the following elements:

- **Radio:** Specification of the air interface, i.e., frequencies, modulation, and transmit power.
- **Baseband:** Description of basic connection establishment, packet formats, timing, and basic QoS parameters.



AT: attention sequence
OBEX: object exchange
TCS BIN: telephony control protocol specification – binary
BNEP: Bluetooth network encapsulation protocol
SDP: service discovery protocol
RFCOMM: radio frequency comm.

Figure 1.44 Bluetooth protocol stack

- **Link manager protocol:** Link set-up and management between devices including security functions and parameter negotiation
- **Logical link control and adaptation protocol (L2CAP):** Adaptation of higher layers to the baseband.
- **Service discovery protocol:** Device discovery in close proximity plus querying of service characteristics.

On top of L2CAP is the **cable replacement protocol** RFCOMM that emulates a serial line interface following the EIA-232 (formerly RS-232) standards. This allows for a simple replacement of serial line cables and enables many legacy applications and protocols to run over Bluetooth. RFCOMM supports multiple serial ports over a single physical channel. The **telephony control protocol specification binary** (TCS BIN) describes a bit-oriented protocol that defines call control signaling for the establishment of voice and data calls between Bluetooth devices. It also describes mobility and group management functions.

The **host controller interface** (HCI) between the baseband and L2CAP provides a command interface to the baseband controller and link manager, and access to the hardware status and control registers. The HCI can be seen as the hardware/software boundary.

Many **protocols** have been **adopted** in the Bluetooth standard. Classical Internet applications can still use the standard TCP/IP stack running over PPP or use the more efficient Bluetooth network encapsulation protocol (BNEP). Telephony applications can use the AT modem commands as if they were using a standard modem. Calendar and business card objects (vCalendar/vCard) can be exchanged using the object exchange protocol (OBEX) as common with IrDA interfaces.

A real difference to other protocol stacks is the support of **audio**. Audio applications may directly use the baseband layer after encoding the audio signals.

Radio layer

The radio specification is a rather short document (less than ten pages) and only defines the carrier frequencies and output power. Several limitations had to be taken into account when Bluetooth's radio layer was designed. Bluetooth devices 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. Worldwide operation also requires a frequency which is available worldwide. The combined use for data and voice transmission has to be reflected in the design, i.e., Bluetooth has to support multi-media data.

Bluetooth uses the license-free frequency band at 2.4 GHz allowing for worldwide operation with some minor adaptations to national restrictions. A frequency-hopping/time-division duplex scheme is used for transmission, with a fast hopping rate of 1,600 hops per second. The time between two hops is called a slot, which is an interval of 625 μs. Each slot uses a different frequency. Bluetooth uses 79 hop carriers equally spaced with 1 MHz. After worldwide harmonization, Bluetooth devices can be used (almost) anywhere.

Bluetooth transceivers use Gaussian FSK for modulation and are available in three 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, and minimum power is 0.25 mW (typ. 10 m range without obstacles). Power control is optional.
- **Power class 3:** Maximum power is 1 mW.

Baseband layer

The functions of the baseband layer are quite complex as it not only performs frequency hopping for interference mitigation and medium access, but also defines physical links and many packet formats. Figure 1.45 shows several examples of frequency selection during data transmission. Remember that each device participating in a certain piconet hops at the same time to the same carrier frequency , for example, the master sends data at f_k, then a slave may answer at f_{k+1} . This scenario shows another feature of bluetooth. TDD is used for separation of the transmission directions. The upper part of Figure 1.45 shows so-called **1-slot** **packets** as the data transmission uses one 625 μs slot. Within each slot the master or one out of seven slaves may transmit data in an alternating fashion. The control of medium access will be described later. Bluetooth also defines **3-slot** and **5-slot** packets for higher data rates (multi-slot packets). If a master or a slave sends a packet covering three or five slots, the radio transmitter remains on the same frequency. No frequency hopping is performed within packets. After transmitting the packet, the radio returns to the frequency required for its hopping sequence. The reason for this is quite simple: not every slave might receive a transmission (hidden terminal

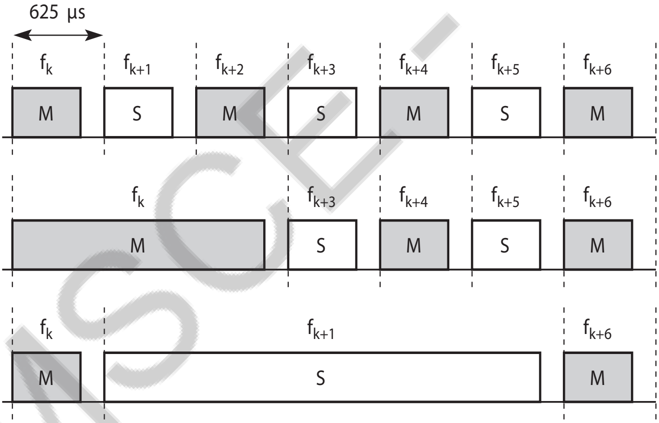


Figure 1.45 Frequency selection during data transmission (1, 3, 5 slot packets)

problem) and it can not react on a multi-slot transmission. Those slaves not involved in the transmission will continue with the hopping sequence. This behavior is important so that all devices can remain synchronized, because the piconet is uniquely defined by having the same hopping sequence with the same phase. Shifting the phase in one device would destroy the piconet.

Figure 1.46 shows the components of a Bluetooth packet at baseband layer. The packet typically consists of the following three fields:

- **Access code:** This first field of a packet is needed for timing synchronization and piconet identification (channel access code, CAC). It may represent special codes during paging (device access code, DAC) and inquiry. The access code consists of a 4 bit **preamble**, a **synchronization** field, and a **trailer** (if a packet header follows). The 64-bit synchronization field is derived from the lower 24 bit of an address (lower address part, LAP). If the access code is used for channel access (i.e., data transmission between a master and a slave or vice versa), the LAP is derived from the master's globally unique 48-bit address. In case of paging (DAC) the LAP of the paged device is used. If a Bluetooth device wants to discover other (arbitrary) devices in transmission range (general inquiry procedure) it uses a special reserved LAP. Special LAPs can be defined for inquiries of dedicated groups of devices.

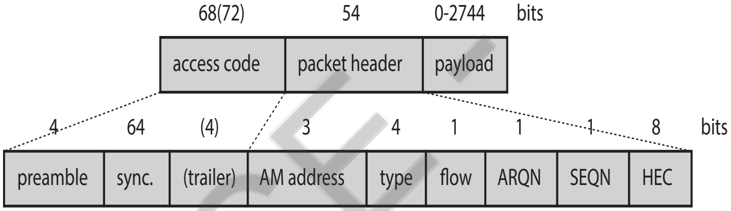
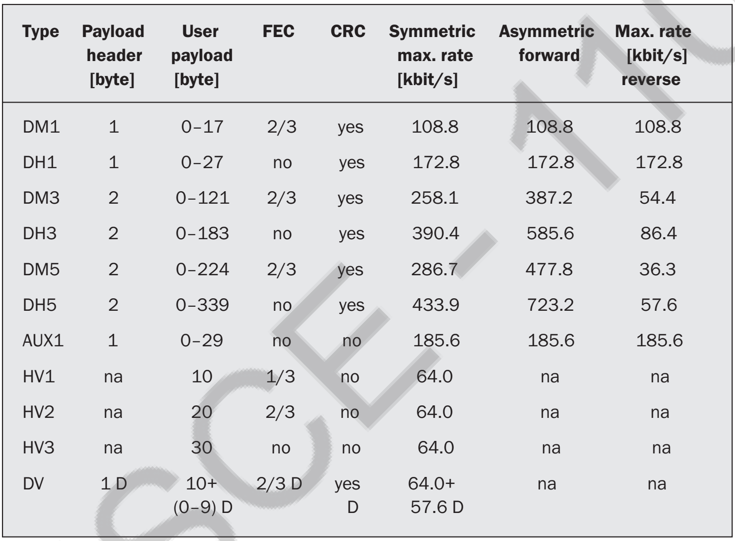


Figure 1.46 Bluetooth Baseband data Rules

- **Packet header:** This field contains typical layer 2 features: address, packet type, flow and error control, and checksum. The 3-bit **active member address** represents the active address of a slave. Active addresses are temporarily assigned to a slave in a piconet. If a master sends data to a slave the address is interpreted as receiver address. If a slave sends data to the master the address represents the sender address. As only a master may communicate with a slave this scheme works well. Seven addresses may be used this way. The zero value is reserved for a broadcast from the master to all slaves. The 4-bit **type** field determines the type of the packet. Examples for packet types are given in Table 1.6. Packets may carry control, synchronous, or asynchronous data. A simple flow control mechanism for asynchronous traffic uses the 1-bit **flow** field. If a packet is received with flow=0 asynchronous data, transmission must stop. As soon as a packet with flow=1 is received, transmission may resume. If an acknowledgement of packets is required, Bluetooth sends this in the slot following the data (using its time

Table 1.6 Bluetooth baseband data rules



Type	Payload header [byte]	User payload [byte]	FEC	CRC	Symmetric max. rate [kbit/s]	Asymmetric forward	Max. rate [kbit/s] reverse
DM1	1	0–17	2/3	yes	108.8	108.8	108.8
DH1	1	0–27	no	yes	172.8	172.8	172.8
DM3	2	0–121	2/3	yes	258.1	387.2	54.4
DH3	2	0–183	no	yes	390.4	585.6	86.4
DM5	2	0–224	2/3	yes	286.7	477.8	36.3
DH5	2	0–339	no	yes	433.9	723.2	57.6
AUX1	1	0–29	no	no	185.6	185.6	185.6
HV1	na	10	1/3	no	64.0	na	na
HV2	na	20	2/3	no	64.0	na	na
HV3	na	30	no	no	64.0	na	na
DV	1 D	10+ (0–9) D	2/3 D	yes D	64.0+ 57.6 D	na	na

division duplex scheme). A simple alternating bit protocol with a single bit sequence number **SEQN** and acknowledgement number **ARQN** can be used. An 8-bit **header error check** (HEC) is used to protect the packet header. The

packet header is also protected by a one-third rate forward error correction (FEC) code because it contains valuable link information and should survive bit errors. Therefore, the 18-bit header requires 54 bits in the packet.

- **Payload:** Up to 343 bytes payload can be transferred. The structure of the payload field depends on the type of link and is explained in the following sections.

Physical links

Bluetooth offers two different types of links, a synchronous connection-oriented link and an asynchronous connectionless link:

- **Synchronous connection-oriented link (SCO):** Classical telephone (voice) connections require symmetrical, circuit-switched, point-to-point connections. For this type of link, the master reserves two consecutive slots (forward and return slots) at fixed intervals. A master can support up to three simultaneous SCO links to the same slave or to different slaves. A slave supports up to two links from different masters or up to three links from the same master. Using an SCO link, three different types of single-slot packets can be used . Each SCO link carries voice at 64 kbit/s, and no **forward error correction** (FEC), 2/3 FEC, or 1/3 FEC can be selected. The 1/3 FEC is as strong as the FEC for the packet header and triples the amount of data. Depending on the error rate of the channel, different FEC schemes can be applied. FEC always causes an overhead, but avoids retransmission of data with a higher probability. However, voice data over an SCO is never retransmitted. Instead, a very robust voice-encoding scheme, **continuous variable slope delta (CVSD)**, is applied .

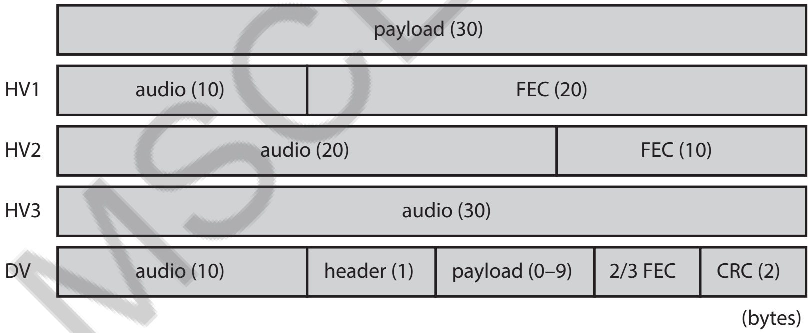


Figure 1.47 SCO payload types

- **Asynchronous connectionless link (ACL):** Typical data applications require symmetrical or asymmetrical (e.g., web traffic), packet-switched, point-to-multipoint transfer scenarios (including broadcast). Here the master uses a polling scheme. A slave may only answer if it has been addressed in the preceding slot. Only one ACL link can exist between a master and a slave. For ACLs carrying data, 1-slot, 3-slot or 5-slot packets can be used . Additionally, data can be protected using a 2/3 FEC scheme. This

FEC protection helps in noisy environments with a high link error rate. However, the overhead introduced by FEC might be too high. Bluetooth therefore offers a fast automatic repeat request (ARQ) scheme for reliable transmission. The **payload header** (1 byte for 1-slot packets, 2 bytes for multi-slot packets) contains an identifier for a logical channel between L2CAP entities, a flow field for flow control at L2CAP level, and a length field indicating the number of bytes of data in the payload, excluding payload header and CRC. Payload is always CRC protected except for the AUX1 packet.

Table 1.6 lists Bluetooth's ACL and SCO packets. Additionally, control packets are available for polling slaves, hopping synchronization, or acknowledgement. The ACL types DM1 (data medium rate) and DH1 (data high rate) use a single slot and a one byte header. DM3 and DH3 use three slots, DM5 and DH5 use five slots. Medium rates are always FEC protected, the high rates rely on CRC only for error detection. The highest available data rates for Bluetooth devices are 433.9 kbit/s (symmetric) or 723.3/57.6 kbit/s (asymmetric). High quality voice (HV) packets always use a single slot but differ with respect to the amount of redundancy for FEC. DV (data and voice) is a combined packet where CRC, FEC, and payload header are valid for the data part only.

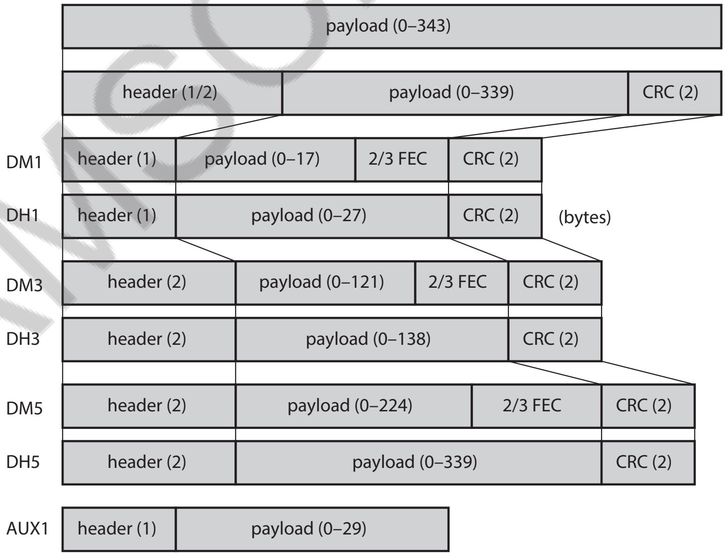


Figure 1.48 ACL payload types

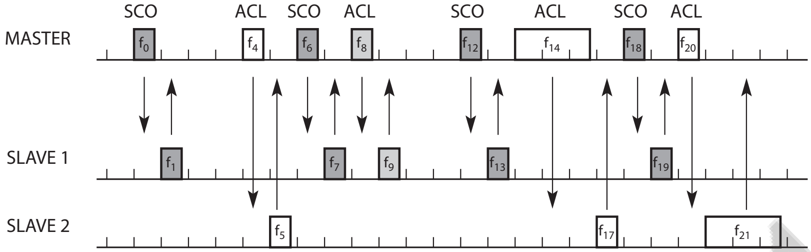


Figure 1.49 Example data transmission

Figure 1.49 shows an example transmission between a master and two slaves. The master always uses the even frequency slots, the odd slots are for the slaves. In this example every sixth slot is used for an SCO link between the master and slave 1. The ACL links use single or multiple slots providing asymmetric bandwidth for connectionless packet transmission. This example again shows the hopping sequence which is independent of the transmission of packets.

The robustness of Bluetooth data transmissions is based on several technologies. FH-CDMA separates different piconets within a scatternet. FHSS mitigates interference from other devices operating in the 2.4 GHz ISM band. Additionally, FEC can be used to correct transmission errors. Bluetooth's 1/3 FEC simply sends three copies of each bit. The receiver then performs a majority decision: each received triple of bits is mapped into whichever bit is in majority. This simple scheme can correct all single bit errors in these triples. The 2/3 FEC encoding detects all double errors and can correct all single bit errors in a codeword.

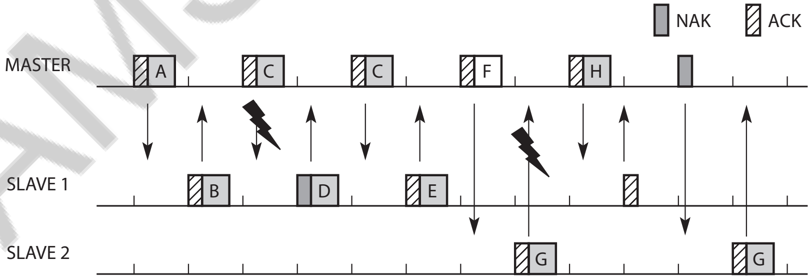


Figure 1.50 Error recovery

ACL links can additionally be protected using an ARQ scheme and a checksum. Each packet can be acknowledged in the slot following the packet. If a packet is lost, a sender can retransmit it immediately in the next slot after the negative acknowledgement, so it is called a fast ARQ scheme.

This scheme hardly exhibits any overheads in environments with low error rates, as only packets which are lost or destroyed have to be retransmitted. Retransmission is triggered by a negative acknowledgement or a time-out.

Link manager protocol

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

- **Authentication, pairing, and encryption:** Although basic authentication is handled in the baseband, LMP has to control the exchange of random numbers and signed responses. The pairing service is needed to establish an initial trust 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 encryption mode (no encryption, point-to-point, or broadcast), key size, and random speed.
- **Synchronization:** Precise synchronization is of major importance within a Bluetooth network. The clock offset is updated each time a packet is received from the master. Additionally, special synchronization packets can be received. Devices can also exchange timing information related to the time differences (slot boundaries) between two adjacent piconets.
- **Capability negotiation:** Not only the version of the LMP can be exchanged but also information about the supported features. Not all Bluetooth devices will support all features that are described in the standard, so devices have to agree the usage of, e.g., multi-slot packets, encryption, SCO links, voice encoding, park/sniff/hold mode (explained below), HV2/HV3 packets etc.
- **Quality of service negotiation:** Different parameters control the QoS of a Bluetooth device at these lower layers. The poll interval, i.e., the maximum time between transmissions from a master to a particular slave, controls the latency and transfer capacity.

Depending on the quality of the channel, DM or DH packets may be used (i.e., 2/3 FEC protection or no protection). The number of repetitions for broadcast packets can be controlled. A master can also limit the number of slots available for slaves' answers to increase its own bandwidth.

- **Power control:** A Bluetooth device can measure the received signal strength. Depending on this signal level the device can direct the sender of the measured signal to increase or decrease its transmit power.

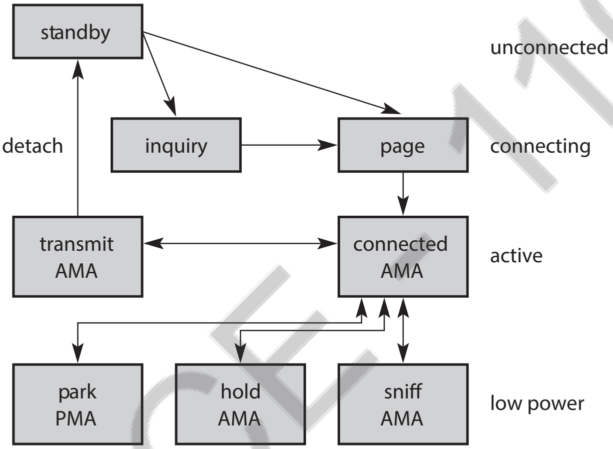


Figure 1.51 Major baseband states of a Bluetooth device

- **Link supervision:** LMP has to control the activity of a link, it may set up new SCO links, or it may declare the failure of a link.
- **State and transmission mode change:** Devices might switch the master/slave role, detach themselves from a connection, or change the operating mode.

With transmission power of up to 100 mW, Bluetooth devices can have a range of up to 100 m. Having this power and relying on batteries, a Bluetooth device cannot be in an active transmit mode all the time. Bluetooth defines several low-power states for a device.

Figure 1.51 shows the major states of a Bluetooth device and typical transitions.

Every device, which is currently not participating in a piconet (and not switched off), is in **standby** mode. This is a low-power mode where only the native clock is running. The next step towards the **inquiry** mode can happen in two different ways. Either a device wants to establish a piconet or a device just wants to listen to see if something is going on.

- ✷ A device wants to establish a piconet: A user of the device wants to scan for other devices in the radio range. The device starts the inquiry procedure by sending an inquiry access code (IAC) that is common to all Bluetooth devices. The IAC is broadcast over 32 so-called wake-up carriers in turn.
- ✷ Devices in standby that listen periodically: Devices in standby may enter the inquiry mode periodically to search for IAC messages on the wake-up carriers. As soon as a device detects an inquiry it returns a packet containing its device address and timing information required by the master to initiate a connection. From that moment on, the device acts as slave.

If the inquiry was successful, a device enters the page mode. The inquiry phase is not coordinated; inquiry messages and answers to these messages may collide, so it may take a while before the inquiry is successful. After a while , a Bluetooth device sees all the devices in its radio range.

During the **page** state two different roles are defined. After finding all required devices the master is able to set up connections to each device, i.e., setting up a piconet. Depending on the device addresses received the master calculates special hopping sequences to contact each device individually. The slaves answer and synchronize with the master's clock, i.e., start with the hopping sequence defined by the master. The master may continue to page more devices that will be added to the piconet. As soon as a device synchronizes to the hopping pattern of the piconet it also enters the connection state.

The connection state comprises the active state and the low power states park, sniff, and hold. In the **active** state the slave participates in the piconet by listening, transmitting, and receiving. ACL and SCO links can be used. A master periodically synchronizes with all slaves. All devices being active must have the 3-bit **active member address** (AMA). Within the active state devices either transmit data or are simply connected. A device can enter standby again, via a detach procedure

To save battery power, a Bluetooth device can go into one of three low power states:

- **Sniff state:** The sniff state has the highest power consumption of the low power states. Here, the device listens to the piconet at a reduced rate (not on every other slot as is the case in the active state). The interval for listening into the medium can be programed and is application dependent. The master designates a reduced number of slots for transmission to slaves in sniff state. However, the device keeps its AMA.

- **Hold state:** The device does not release its AMA but stops ACL transmission. A slave may still exchange SCO packets. If there is no activity in the piconet, the slave may either reduce power consumption or participate in another piconet.

- **Park state:** In this state the device has the lowest duty cycle and the lowest power consumption. The device releases its AMA and receives a parked member address (PMA). The device is still a member of the piconet, but gives room for another device to become active (AMA is only 3 bit, PMA 8 bit). Parked devices are still FH synchronized and wake up at certain beacon intervals for re-synchronization. All PDUs sent to parked slaves are broadcast.

Table 1.7 Example power consumption (CSR, 2002)

Operating mode	Average current [mA]
SCO, HV1	53
SCO, HV3, 1 s interval sniff mode	26
ACL, 723.2 kbit/s	53
ACL, 115.2 kbit/s	15.5
ACL, 38.4 kbit/s, 40 ms interval sniff mode	4
ACL, 38.4 kbit/s, 1.28 s interval sniff mode	0.5
Park mode, 1.28 s beacon interval	0.6
Standby (no RF activity)	0.047

The effect of the low power states is shown in Table 1.7. This table shows the typical average power consumption of a Bluetooth device (BlueCore2, CSR, 2002). It is obvious that higher data rates also require more transmission power. The intervals in sniff mode also influence power consumption. Typical IEEE 802.11b products have an average current in the order of 200 mA while receiving, 300 mA while sending, and 20 mA in standby.

L2CAP

The **logical link control and adaptation protocol (L2CAP)** is a data link control protocol on top of the baseband layer offering logical channels between Bluetooth devices with QoS properties. L2CAP is available for ACLs only. Audio applications using SCOs have to use the baseband layer directly. L2CAP provides three different types of logical channels that are transported via the ACL between master and slave:

- **Connectionless:** These unidirectional channels are typically used for broadcasts from a master to its slave(s).
- **Connection-oriented:** Each channel of this type is bi-directional and supports QoS flow specifications for each direction. These flow specs follow RFC 1363 and define average/peak data rate, maximum burst size, latency, and jitter.
- **Signaling:** This third type of logical channel is used to exchanging signaling messages between L2CAP entities.

Each channel can be identified by its **channel identifier (CID)**. Signaling channels always use a CID value of 1, a CID value of 2 is reserved for connectionless channels. For connection-oriented channels a unique CID (>= 64) is dynamically assigned at each end of the channel to identify the connection (CIDs 3 to 63 are reserved). Figure 1.52 gives an example for logical channels using the ACL link between master and slave. The master has a bi-directional signaling channel to each slave. The CID at each end is 1. Additionally, the master maintains a connectionless, unidirectional channel to both slaves. The CID at the slaves is 2, while the CID at the beginning of the connectionless channel is dynamically assigned. L2CAP provides mechanisms to add slaves to, and remove slaves from, such a multicast group. The master has one connection oriented channel to the left slave and two to the right slave. All CIDs for these channels are dynamically assigned (between 64 and 65535).

Figure 1.53 shows the three packet types belonging to the three logical channel types. The **length** field indicates the length of the payload (plus PSM for connectionless PDUs). The **CID** has the multiplexing/demultiplexing function . For connectionless PDUs a **protocol/service multiplexor (PSM)** field is needed to identify the higher layer recipient for the payload. For connection-oriented PDUs the CID already fulfills this function. Several PSM values have been defined, e.g., 1 (SDP), 3 (RFCOMM), 5 (TCS-

BIN). Values above 4096 can be assigned dynamically. The payload of the signaling PDU contains one or more **commands**. Each command has its own **code** (e.g., for command reject, connection request, disconnection response etc.) and an **ID** that matches a request with its reply. The **length** field indicates the length of the **data** field for this command.

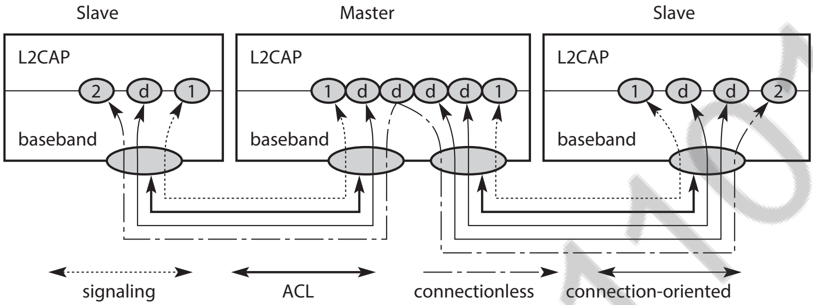
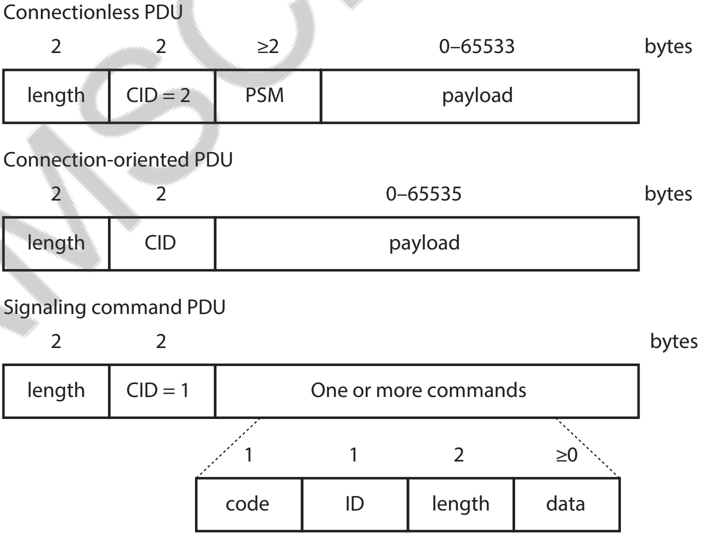


Figure 1.52 Logical channels between devices

Besides protocol multiplexing, flow specification, and group management, the L2CAP layer also provides segmentation and reassembly functions. Depending on the baseband capabilities, large packets have to be chopped into smaller segments. DH5 links, for example, can carry a maximum of 339 bytes while the L2CAP layer accepts up to 64 kbyte.



Connectionless PDU

2	2	≥2	0–65533	bytes
length	CID = 2	PSM	payload	

Connection-oriented PDU

2	2	0–65535	bytes
length	CID	payload	

Signaling command PDU

2	2		bytes
length	CID = 1	One or more commands	

1	1	2	≥0
code	ID	length	data

Figure 1.53 L2CAP packet formats

Security

A radio interface is by nature easy to access. Bluetooth devices can transmit pri-vate data, e.g., schedules between a PDA and a mobile phone. A user clearly does not want another person to eavesdrop the data transfer. Just imagine a scenario where two Bluetooth enabled PDAs in suitcases 'meet' on the conveyor belt of an airport exchanging personal information! Bluetooth offers mechanisms for authentication and encryption on the MAC layer, which must be implemented in the same way within each device.

The main security features offered by Bluetooth include a challengeresponse routine for authentication, a stream cipher for encryption, and a session key generation. Each connection may require a one-way, two-way, or no authentication using the challenge-response routine. All these schemes have to be implemented in silicon, and higher layers should offer stronger encryption if needed. The security features included in Bluetooth only help to set up a local domain of trust between devices.

The security algorithms use the public identity of a device, a secret private user key, and an internally generated random key as input parameters. For each transaction, a new random number is generated on the Bluetooth chip. Key management is left to higher layer software.

Figure 7.54 shows several steps in the security architecture of Bluetooth. The illustration is simplified and the interested reader is referred to Bluetooth (2001a) for further details. The first step, called **pairing**, is necessary if two Bluetooth devices have never met before. To set up trust between the two devices a user can enter a secret PIN into both devices. This PIN can have a length of up to 16 byte. Unfortunately, most devices limit the length to four digits or, even worse, program the devices with the fixed PIN '0000' rendering the whole security concept of Bluetooth questionable at least. Based on the PIN, the device address, and random numbers, several keys can be computed which can be used as link key for **authentication**. Link keys are typically stored in a persistent storage. The authentication is a challenge-response process based on the link key, a random number generated by a verifier (the device that requests authentication), and the device address of the claimant (the device that is authenticated).

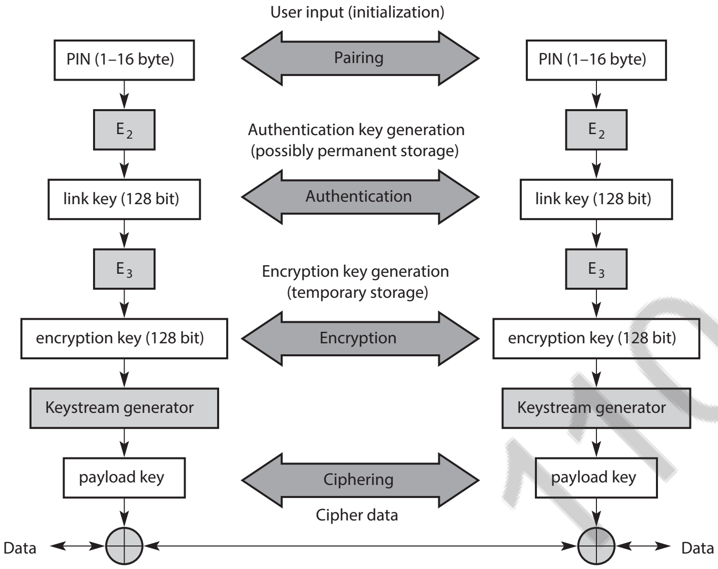


Figure 1.54 Bluetooth security components and protocols

Based on the link key, values generated during the authentication, and again a random number an encryption key is generated during the **encryption** stage of the security architecture. This key has a maximum size of 128 bits and can be individually generated for each transmission. Based on the encryption key, the device address and the current clock a payload key is generated for ciphering user data. The payload key is a stream of pseudo-random bits. The **ciphering** process is a simple XOR of the user data and the payload key.

Compared to WEP in 802.11, Bluetooth offers a lot more security. However, Bluetooth, too, has some weaknesses when it comes to real implementations. The PINs are quite often fixed. Some of the keys are permanently stored on the devices and the quality of the random number generators has not been specified. If Bluetooth devices are switched on they can be detected unless they operate in the non-discoverable mode (no answers to inquiry requests). Either a user can use all services as intended by the Bluetooth system, or the devices are hidden to protect privacy. Either roaming profiles can be established, or devices are hidden and, thus many services will not work. If a lot of people carry Bluetooth devices (mobile phones, PDAs etc.) this could give, e.g., department stores, a lot of information regarding consumer behavior.

SDP

Bluetooth devices should work together with other devices in unknown environments in an ad-hoc fashion. It is essential to know what devices, or more specifically what services, are available in radio proximity. To find new services, Bluetooth defined the **service discovery protocol (SDP)**. SDP defines only the discovery of services, not their usage. Discovered services can be cached and gradual discovery is possible. Devices that want to offer a service have to instal an SDP server. For all other devices an SDP client is sufficient.

Table 1.8 Example attributes for an SDP service record

Attribute name	Attribute ID	Attribute value type	Example
ServiceRecordHandle	0000	32-bit unsigned integer	1f3e4723
ServiceClassIDList	0001	Data element sequence (UUIDs)	ColorPostscriptPrinterService ClassID, PostscriptPrinterService ClassID, PrinterServiceClassID
ProtocolDescriptorList	0004	Data element sequence	((L2CAP , PSM=RFCOMM), (RFCOMM, CN=2), (PPP), (IP), (TCP), (IPP))
DocumentationURL	000A	URL	www.xy.zz/print/srvs.html
IconURL	000C	URL	www.xy.zz/print/ico.png
ServiceName	0100	String	Color Printer

All the information an SDP server has about a service is contained in a **service record**. This consists of a list of service attributes and is identified by a 32-bit service record handle. SDP does not inform clients of any added or removed services. There is no service access control or service brokerage. A **service attribute** consists of an attribute ID and an attribute value. The 16-bit **attribute ID** distinguishes each service attribute from other service attributes within a service record. The attribute ID also identifies the semantics of the associated attribute value. The **attribute value** can be an integer, a UUID (universally unique identifier), a string, a Boolean, a URL (uniform resource locator) etc. Table 1.8 gives some example attributes. The service handle as well as the ID list must be present. The ID list contains the UUIDs of the service classes in increasing generality (from the specific color postscript printer to printers in general). The protocol descriptor list comprises the protocols needed to access this service. Additionally, the URLs for service documentation, an icon for the service and a service name which can be displayed together with the icon are stored in the example service record.

Profiles

Profiles represent default solutions for a certain usage model. They use a selection of protocols and parameter set to form a basis for interoperability. Protocols can be seen as horizontal layers while profiles are vertical slices (as illustrated in Figure 7.55). The following **basic profiles** have been specified: generic access, service discovery, cordless telephony, intercom, serial port, headset, dialup networking, fax, LAN access, generic object exchange, object push, file transfer, and synchronization. **Additional profiles** are: advanced audio distribution, PAN, audio video remote control, basic printing, basic imaging, extended service discovery, generic audio video distribution, hands-free, and hardcopy cable replacement. Each profile selects a set of protocols. For example, the serial port profile needs RFCOMM, SDP, LMP, L2CAP. Baseband and radio are always required. The profile further defines all interoperability requirements, such as RS232 control signals for RFCOMM or configuration options for L2CAP (QoS, max. transmission unit).

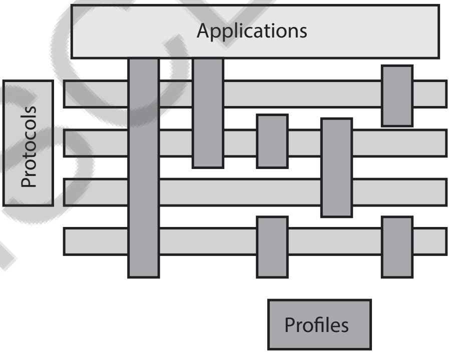


Figure 1.55 Bluetooth profiles.

20. Explain IEEE 802.15 (WPAN) Architecture in detail.

In 1999 the IEEE established a working group for wireless personal area net-works (WPAN) with similar goals to Bluetooth. The working group was divided into several subgroups focusing on different aspects of WPANs (IEEE, 2002c). The following gives a quick overview and presents the standard for low-rate WPANs, 802.15.4, in some more detail:

- **IEEE 802.15.1:** This group standardizes the lower layers of **Bluetooth** together with the Bluetooth consortium. IEEE LANs focus only on the physical and data link layer, while the Bluetooth standard also comprises higher layers, application profiles, service description etc. as explained above.

- **IEEE 802.15.2:** The **coexistence** of wireless personal area networks (WPAN) and wireless local area networks (WLAN) is the focus of this group. One task is to quantify mutual interference and to develop algorithms and protocols for coexistence. Without additional mechanisms, Bluetooth/802.15.1 may act like a rogue member of an IEEE 802.11 network. Bluetooth is not aware of gaps, inter-frame spacing, frame structures etc. Figure 1.56 illustrates the problem. WLANs following the IEEE 802.11b standard may use three non-overlapping channels that are chosen during installation of the access points. Bluetooth/802.15.1 networks use a frequency hopping pattern to separate different piconets 79 channels can be used. Without additional mechanisms, the hopping pattern of Bluetooth is independent of 802.11b's channel selection. Both systems work in the 2.4 GHz ISM band and might interfere with each other. Figure 7.56 shows two hopping sequences of two piconets interfering with several data packets, acknowledgements, and inter-frame spacings of 802.11b. The real effects of the interference range from 'almost no effect' to 'complete breakdown of the WLAN'. Publications on this issue differ depending on the test scenario, traffic load, signal power, propagation conditions etc. However, it seems that Bluetooth with its FHSS scheme is more robust than 802.11b with CSMA/CA. To overcome the interference problems between 802.11b and 802.15.1, however severe they might be, the 802.15.2 working group proposes **adaptive frequency hopping**. This coexistence mechanism is non-collaborative in the sense that Bluetooth devices do not have to interact with the WLAN. However, the WPAN devices can check for the occupied channels and exclude them from their list of channels used for hopping. This mechanism avoids hopping into a channel occupied by 802.11b, but still offers enough channels for FHSS. The lower number of FHSS channels increases the interference among the WPANs due to a higher probability of collisions. However, if not too many piconets overlap this effect will be negligible. This type of interference in the crowded 2.4 GHz band is a strong argument for 5 GHz WLANs.

- **IEEE 802.15.3:** A **high-rate** study group looks for a standard providing data rates of 20 Mbit/s or greater while still working with low-power at low-cost. The standard should support isochronous data delivery, ad-hoc peer-to-peer networking, security features, and should meet the demanding requirements of portable consumer imaging and multi-media applications.
- **IEEE 802.15.4:** The fourth working group goes in the opposite direction for data rates. This group standardizes **low-rate wireless personal area networks (LR-WPAN)**, which are explained in the following section in more detail. The ZigBee consortium tries to standardize the higher layers of 802.15.4 similar to the activities of the Bluetooth consortium for 802.15.1 (ZigBee, 2002).

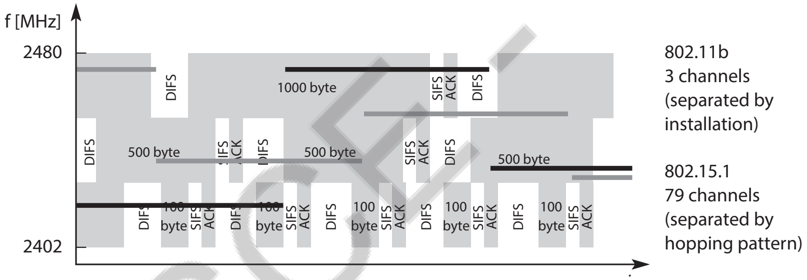


Figure 1.56 Possible interference between 802.15.1 (Bluetooth) and 802.11b

IEEE 802.15.4 Low-rate WPANs

The reason for having low data rates is the focus of the working group on extremely low power consumption enabling multi-year battery life . Compared to 802.11 or Bluetooth, the new system should have a much lower complexity making it suitable for low-cost wireless communication (remember that Bluetooth started with similar goals with respect to the idea of cable replacement). Example **applications** include industrial control and monitoring, smart badges, interconnection of environmental sensors, interconnection of peripherals (also an envisaged application area for Bluetooth!), remote controls etc. The new standard should offer data rates between 20 and 250 kbit/s as maximum and latencies down to 15 ms. This is enough for many home automation and consumer electronics applications.

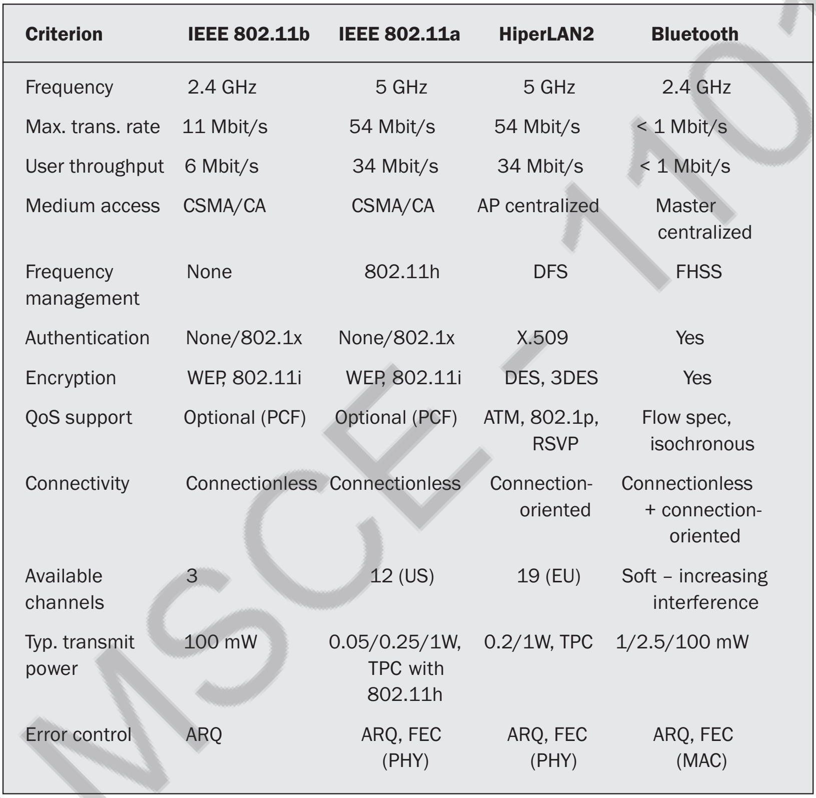
IEEE 802.15.4 offers two different PHY options using DSSS. The **868/915 MHz PHY** operates in Europe at 868.0-868.6 MHz and in the US at 902-928 MHz. At 868 MHz one channel is available offering a data rate of 20 kbit/s. At 915 MHz 10 channels with 40 kbit/s per channel are available (in Europe GSM uses these frequencies). The advantages of the lower frequencies are better propagation conditions. However, there is also interference in these bands as many analog transmission systems use them. The **2.4 GHz PHY** operates at 2.4-2.4835 GHz and offers 16 channels with 250 kbit/s per channel. This PHY offers worldwide operation but suffers from interference in the 2.4 GHz ISM band and higher propagation loss. Typical devices with 1 mW output power are expected to cover a 10-20 m range. All PHY PDUs start with a 32 bit preamble for synchronization. After a start-of-packet delimiter, the PHY header indicates the length of the payload (maximum 127 bytes).

Compared to Bluetooth the **MAC layer** of 802.15.4 is much simpler. For example, no synchronous voice links are supported. MAC frames start with a 2-byte frame control field, which specifies how the rest of the frame looks and what it contains. The following 1-byte sequence number is needed to match acknowledgements with a previous data transmission. The variable address field (0-20 bytes) may contain source and/or destination addresses in various formats. The payload is variable in length; however, the whole MAC frame may not exceed 127 bytes in length. A 16-bit FCS protects the frame. Four different MAC frames have been defined: beacon, data, acknowledgement, and MAC command.

Optionally, this LR-WPAN offers a **superframe mode**. In this mode, a PAN coordinator transmits beacons in predetermined intervals (15 ms-245 s). With the help of beacons, the medium access scheme can have a period when contention is possible and a period which is contention free. Furthermore, with beacons a slotted **CSMA/CA** is available. Without beacons standard CSMA/CA is used for medium access. Acknowledgement frames confirming a previous transmission do not use the CSMA mechanism. These frames are sent immediately following the previous packet.

IEEE 802.15.4 specifies three levels of **security**: no security, access control lists, and symmetric encryption using AES-128. Key distribution is not specified further. Security is a must for home automation or industry control applications. Up to now, the success of this standard is unclear as it is squeezed between Bluetooth, which also aims at cable replacement, and enhanced RFIDs/RF controllers.

21. Compare different Wireless Networks in detail.



Criterion	IEEE 802.11b	IEEE 802.11a	HiperLAN2	Bluetooth
Frequency	2.4 GHz	5 GHz	5 GHz	2.4 GHz
Max. trans. rate	11 Mbit/s	54 Mbit/s	54 Mbit/s	< 1 Mbit/s
User throughput	6 Mbit/s	34 Mbit/s	34 Mbit/s	< 1 Mbit/s
Medium access	CSMA/CA	CSMA/CA	AP centralized	Master centralized
Frequency management	None	802.11h	DFS	FHSS
Authentication	None/802.1x	None/802.1x	X.509	Yes
Encryption	WEP, 802.11i	WEP, 802.11i	DES, 3DES	Yes
QoS support	Optional (PCF)	Optional (PCF)	ATM, 802.1p, RSVP	Flow spec, isochronous
Connectivity	Connectionless	Connectionless	Connection-oriented	Connectionless + connection-oriented
Available channels	3	12 (US)	19 (EU)	Soft – increasing interference
Typ. transmit power	100 mW	0.05/0.25/1W, TPC with 802.11h	0.2/1W, TPC	1/2.5/100 mW
Error control	ARQ	ARQ, FEC (PHY)	ARQ, FEC (PHY)	ARQ, FEC (MAC)

Table1.9 Comparison of wireless networks

UNIT - 2

MOBILE NETWORK LAYER

PART - A

1. **What is a Mobile IP?**
   Mobile IP is a protocol developed to allow internetwork mobility for wireless nodes without them having to change their IP addresses.

2. **What are the entities of Mobile IP?**
   Mobile Node (MN) Correspondent Node (CN) Home Network (HN) Foreign Network (FN) Foreign Agent (FA) Home Agent (HA)

3. **What are the benefits of Mobile IP?**
   The major benefit of Mobile IP is that it frees the user from a fixed location. Mobile IP makes invisible the boundaries between attachment points, it is able to track and deliver information to mobile devices without needing to change the device's long-term Internet Protocol (IP) address.

4. **What is Care-Of Address (COA)?**
   The Care of Address defines the current location of the MN from an IP point of view. All IP packets sent to the MN are delivered to the COA, not directly to the subnet.

5. **What is agent advertisement?**
   Home Agent (HA) and Foreign Agent (FA) advertise their presence periodically using agent advertisement messages. These advertisement messages can be seen as a beacon broadcast into the subnet.

6. **What is the need for registration?**
   The main purpose of the registration is to inform the HA of the current location for correct forwarding of packets.

7. Define – Encapsulation and Decapsulation

Encapsulation is the mechanism of taking a packet consisting of packet header and data and putting it into the data part of a new packet. The reverse operation, taking a packet out of the data part of another packet, is called decapsulation.

8. What is triangular routing?

Tunneling in its simplest form has all packets to Home Network and then sent to MN via a tunnel. The inefficient behavior of a non-optimized mobile IP is called triangular routing.

9. What is DHCP?

The Dynamic Host Configuration Protocol (DHCP) is based on the bootstrap protocol (BOOTP), which provides the framework for passing configuration information to hosts on a TCP/IP network. DHCP adds the automatically allocate reusable network addresses and configuration options to internet hosts.

10. What is SIP?

The Session Initiation Protocol (SIP) is an application-layer control (signaling) protocol for creating, modifying and terminating sessions with one or more participants. It is a IETF (Internet Standard) RFC 3261 protocol.

11. What are the functions of Session Initiation Protocol (SIP)?

SIP has following major functions SIP allows for the establishment of user location SIP provides a mechanism for call management SIP provides feature negotiation, so that all the parties in the call can agree to the features supported among them.

12. What are the characteristics of MANET? (M/J - 12)

The characteristics of MANET are Dynamic Topologies Bandwidth Constraints and Variable Capacity Links Energy Constrained Operations Limited Physical Security

13. Differentiate an ad hoc network and a cellular network with respect to a) Bandwidth usage b) Cost effectiveness (N/D - 12)

Parameter	Cellular Network	Ad HOC Network
Bandwidth usage	1. Easier to employ bandwidth reservation 2. Guaranteed bandwidth (designed for voice traffic)	Bandwidth reservation requires complex medium access control protocols 2. Shared radio channel (more suitable for best-effort data traffic)
Cost effectiveness	Cost of network maintenance is high (backup power source, staffing, etc.)	Self-organization and maintenance properties are built into the network. Hence the cost of network maintenance is less.

14. What are the challenging issues in ad hoc network maintenance? (M/J - 12)

The challenging issues in ad hoc network are Medium access scheme Routing Multicast routing Transport layer protocol Pricing Schemes Quality of Service Provisioning Self-Organization Security Addressing and Service Discovery Energy Management Scalability Deployment considerations.

15. Why are ad hoc networks needed? (M/J - 12)

Ad hoc networking is often needed where an infrastructure network cannot be deployed and managed. The presence of dynamic and adaptive routing protocols enables quick formation of ad hoc networks and is suitable for emergency situations like natural disasters, spontaneous meetings or military conflicts.

16. List out the applications of ad hoc networks.

Ad hoc networks are widely used in Military applications and battlefields Collaborative and distributed computing Emergency search and rescue operations Wireless sensor and mesh networks

17. Give the classifications of routing protocol in MANET.

The classifications of routing protocol in MANET are a) Proactive protocols: This protocol attempt to evaluate continuously the routes within the network, so that when a packet needs to be forwarded, the router is already known and can be immediately used.

b) Reactive protocols: This protocol invoke a route determination procedure only on demand. The routing protocols may also be categorized as follows: Table-driven protocols.

Source-initiated on-demand protocols.

18. List the Source-initiated On-Demand Routing Protocols. The Source-initiated On-Demand Routing Protocols are Ad-hoc On-Demand Distance Vector Routing (AODV)
Dynamic Source Routing (DSR)

Temporarily Ordered Routing Algorithm (TORA)

Associatively Based Routing (ABR)

Signal Stability Based Routing (SSR)

19. Differentiate proactive and reactive routing protocols. Write examples for each. (M/J - 12)

S.No.	Proactive	Reactive
1	Route is pre-established	Route establishment is on-demand
2	Continuously discover the routes	Route discovery by some global search
3	Updates topology information (table) periodically	No information update is done
4	No latency in route discovery	Longer delay due to latency of route discovery
5	Large capacity is needed to update network information	Large capacity is not needed
6	A lot of routing information may never be used	May not be appropriate for real-time communication
7	Eg: DSDV, WRP	Eg: AODV, ABR

20. What is DSDV?
Distance-Vector Routing (DSDV) is a table driven routing scheme for ad-hoc mobile networks. The main contribution of the algorithm was to solve the routing loop problem.

21. List out the advantages and disadvantages of DSDV routing protocols.

The advantages and disadvantages of DSDV routing protocols are Advantages Less Delay is involved in route setup process. DSDV protocol guarantees loop free paths. Incremental updates with sequence number tags make the existing wired network protocols adaptable to ad-hoc wireless networks.

Count to infinity problem is reduced in DSDV. Path Selection: DSDV maintains only the best path instead of maintaining multiple paths to every destination. With this, the amount of space in routing table is reduced.

Disadvantages Updates due to broken links lead to heavy control overhead during mobility. The control overhead is directly proportional to the number of nodes. Small network with high mobility or large network with low mobility can choke the available bandwidth. Wastage of bandwidth due to unnecessary advertising of routing information even if there is no change in the network topology. Delay in obtaining information about a node could result in stale routing at the nodes.

PART B

1. Explain the Goals, assumptions and requirements of Mobile IP in detail.

Mobile computing is clearly the paradigm of the future. A host sends an IP packet with the header containing a destination address with other fields. The destination address not only determines the receiver of the packet, but also the physical subnet of the receiver. For example, the destination address 129.13.42.99 shows that the receiver must be connected to the physical subnet with the network prefix 129.13.42. Routers in the internet now look at the destination addresses of incoming packets and forward them according to internal look-up tables. To avoid an explosion of routing tables, only prefixes are stored and further optimizations are applied. A router would otherwise have to store the addresses of all computers in the internet, which is obviously not feasible. As long as the receiver can be reached within its physical subnet, it gets the packets; as soon as it moves outside the subnet, a packet will not reach it. A host needs a so-called **topologically correct address.**

Quick 'solutions'

One might think that a quick solution to this problem would be to assign to the computer a new, topologically correct IP address. This is what many users do with the help of DHCP. So moving to a new location would mean assigning a new IP address. The problem is that nobody knows about this new address. It is almost impossible to find a (mobile) host on the internet which has just changed its address.

With the help of dynamic DNS an update of the mapping logical name IP address is possible. This is what many computer users do if they have a dynamic IP address and still want to be permanently reachable using the same logical computer name. It is important to note that these considerations, indeed most of mobile IP's motivation, are important if a user wants to offer services from a mobile node, i.e., the node should act as server. Typically, the IP address is of no special interest for service usage: in this case DHCP is sufficient. Another motivation for permanent IP addresses is emergency communication with permanent and quick reachability via the same IP address.

The problem is that the domain name system (DNS) needs some time before it updates the internal tables necessary to map a logical name to an IP address. This approach does not work if the mobile node moves quite often. The internet and DNS have not been built for frequent updates. Just imagine millions of nodes moving at the same time. DNS could never present a consistent view of names and addresses, as it uses caching to improve scalability. It is simply too expensive to update quickly.

There is a severe problem with higher layer protocols like TCP which rely on IP addresses. Changing the IP address while still having a TCP connection open means breaking the connection. A TCP connection is identified by the tuple (source IP address, source port, destination IP address, destination port), also known as a **socket pair** (a socket consists of address and port). Therefore, a TCP connection cannot survive any address change. Breaking TCP connections is not an option, using even simple programs like telnet would be impossible. The mobile node would also have to notify all communication partners about the new address.

Another approach is the creation of specific routes to the mobile node. Routers always choose the best-fitting prefix for the routing decision. If a router now has an entry for a prefix 129.13.42 and an address 129.13.42.99, it would choose the port associated with the latter for forwarding, if a packet with the destination address 129.13.42.99 comes in. While it is theoretically possible to change routing tables all over the world to create specific routes to a mobile node, this does not scale at all with the number of nodes in the internet. Routers are built for extremely fast forwarding, but not for fast updates of routing tables. While the first is done with special hardware support, the latter is typically a piece of software which cannot handle the burden of frequent updates. Routers are the 'brains' of the internet, holding the whole net together. No service provider or system administrator would allow changes to the routing tables, probably sacrificing stability, just to provide mobility for individual users.

Requirements

Since the quick 'solutions' obviously did not work, a more general architecture had to be designed. Many field trials and proprietary systems finally led to mobile IP as a standard to enable mobility in the internet. Several requirements accompanied the development of the standard:

- **Compatibility:** The installed base of Internet computers, i.e., computers running TCP/IP and connected to the internet, is huge. A new standard cannot introduce changes for applications or network protocols already in use. People still want to use their favorite browser for www and do not want to change applications just for mobility, the same holds for operating systems. Mobile IP has to be integrated into existing operating systems or at least work with them (today it is available for many platforms). Routers within the internet should not necessarily require other software. While it is possible to enhance the capabilities of some routers to support mobility, it is almost impossible to change all of them. Mobile IP has to remain compatible with all lower layers used for the standard, non-mobile, IP. Mobile IP must not require special media or MAC/LLC protocols, so it must use the same interfaces and mechanisms to access the lower layers as IP does. Finally, end-systems enhanced with a mobile IP implementation should still be able to communicate with fixed systems without mobile IP. Mobile IP has to ensure that users can still access all the other servers and systems in the internet. But that implies using the same address format and routing mechanisms.

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

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

- **Security:** Mobility poses many security problems. The minimum requirement is that of all the messages related to the management of Mobile IP are authenticated. The IP layer must be sure that if it forwards a packet to a mobile host that this host receives the packet. The IP layer can only guarantee that the IP address of the receiver is correct. There are no ways of preventing fake IP addresses or other attacks. According to Internet philosophy, this is left to higher layers (keep the core of the internet simple, push more complex services to the edge).

The goal of a mobile IP can be summarized as: 'supporting end-system mobility while maintaining scalability, efficiency, and compatibility in all respects with existing applications and Internet protocols'.

2. Mention the different Entities and terminology in a Mobile IP.

The following defines several entities and terms needed to understand mobile IP as defined in RFC 3344 . Figure 2.1 illustrates an example scenario.

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

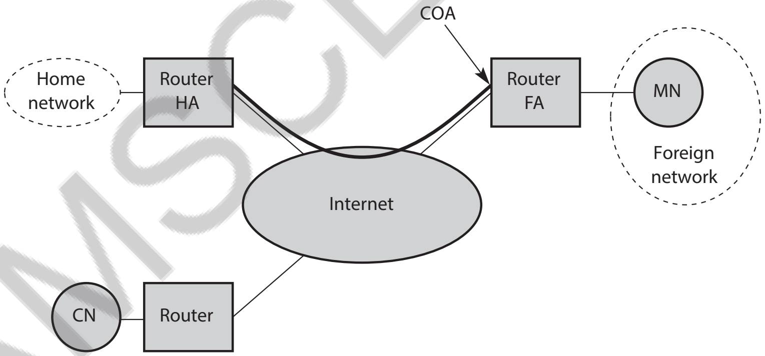


Figure 2.1 Mobile IP example network

- **Correspondent node (CN):** At least one partner is needed for communication. In the following the CN represents this partner for the MN. The CN can be a fixed or mobile node.
- **Home network:** The home network is the subnet the MN belongs to with respect to its IP address. No mobile IP support is needed within the home network.

- **Foreign network:** The foreign network is the current subnet the MN visits and which is not the home network.

- **Foreign agent (FA):** The FA can provide several services to the MN during its visit to the foreign network. The FA can have the COA (defined below), acting as tunnel endpoint and forwarding packets to the MN. The FA can be the default router for the MN. FAs can also provide security services because they belong to the foreign network as opposed to the MN which is only visiting. For mobile IP functioning, FAs are not necessarily needed. Typically, an FA is implemented on a router for the subnet the MN attaches to.

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

  - **Foreign agent COA:** The COA could be located at the FA, i.e., the COA is an IP address of the FA. The FA is the tunnel endpoint and forwards packets to the MN. Many MN using the FA can share this COA as common COA.

  - **Co-located COA:** The COA is co-located if the MN temporarily acquired an additional IP address which acts as COA. This address is now topologically correct, and the tunnel endpoint is at the MN. Co-located addresses can be acquired using services such as DHCP.. One problem associated with this approach is the need for additional addresses if MNs request a COA. This is not always a good idea considering the scarcity of IPv4 addresses.

- **Home agent (HA):** The HA provides several services for the MN and is locatedin the home network. The tunnel for packets toward the MN starts at the HA. The HA maintains a location registry, i.e., it is informed of the MN's location by the current COA. Three alternatives for the implementation of an HA exist. ● The HA can be implemented on a router that is responsible for the home network. This is obviously the best position, because without optimizations to mobile IP, all packets for the MN have to go through the router anyway.

- If changing the router's software is not possible, the HA could also be implemented on an arbitrary node in the subnet. One disadvantage of this solution is the double crossing of the router by the packet if the MN is in a foreign network. A packet for the MN comes in via the router; the HA sends it through the tunnel which again crosses the router.
- Finally, a home network is not necessary at all. The HA could be again on the 'router' but this time only acting as a manager for MNs belonging to a virtual home network. All MNs are always in a foreign network with this solution.

The example network in Figure 2.1 shows the following situation: A CN is connected via a router to the internet, as are the home network and the foreign network. The HA is implemented on the router connecting the home network with the internet, an FA is implemented on the router to the foreign network. The MN is currently in the foreign network. The tunnel for packets toward the MN starts at the HA and ends at the FA, for the FA has the COA in this example.

3. What is mobile IP? Describe the mobile IP protocol. Explain with a diagram, how IP packets are transmitted between nodes. Also explain how packet delivery mechanism in the mobile IP protocol differs from IP

(OR)

(i) Explain how the mobile node discover that the foreign node has moved?

(ii) Access the IP packet delivery to and from mobile node with a neat figure

IP packet delivery

Figure 2.2 illustrates packet delivery to and from the MN using the example network of Figure 2.1. A correspondent node CN wants to send an IP packet to the MN. One of the requirements of mobile IP was to support hiding the mobility of the MN. CN does not need to know anything about the MN's current location and sends the packet as usual to the IP address of MN (step 1). This means that CN sends an IP packet with MN as a destination address and CN as a source address. The internet, not having information on the current location of MN, routes the packet to the router responsible for the home network of MN. This is done using the standard routing mechanisms of the internet.

The HA now intercepts the packet, knowing that MN is currently not in its home network. The packet is not forwarded into the subnet as usual, but encapsulated and tunnelled to the COA. A new header is put in front of the old IP header showing the COA as new destination and HA as source of the encapsulated packet (step 2). The foreign agent now decapsulates the packet, i.e., removes the additional header, and forwards the original packet with CN as source and MN as destination to the MN (step 3). Again, for the MN mobility is not visible. It receives the packet with the same sender and receiver address as it would have done in the home network.

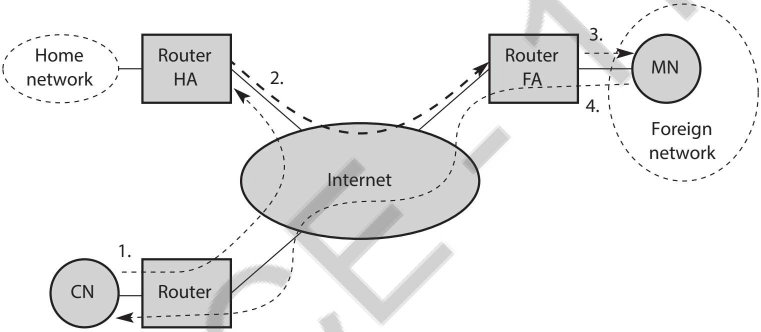


Figure 2.2 Packet delivery to and from the mobile node

At first glance, sending packets from the MN to the CN is much simpler. The MN sends the packet as usual with its own fixed IP address as source and CN's address as destination (step 4). The router with the FA acts as default router and forwards the packet in the same way as it would do for any other node in the foreign network. As long as CN is a fixed node the remainder is in the fixed internet as usual. If CN were also a mobile node residing in a foreign network, the same mechanisms as described in steps 1 through 3 would apply now in the other direction.

Agent discovery

One initial problem of an MN after moving is how to find a foreign agent. How does the MN discover that it has moved? For this purpose mobile IP describes two methods: agent advertisement and agent solicitation, which are in fact router discovery methods plus extensions.

Agent advertisement

For the first method, foreign agents and home agents advertise their presence periodically using special **agent advertisement** messages. These advertisement messages can be seen as a beacon broadcast into the subnet. For these advertisements Internet control message protocol (ICMP) messages according to RFC 1256 are used with some mobility extensions. Routers in the fixed network implementing this standard also advertise their routing service periodically to the attached links.

The agent advertisement packet according to RFC 1256 with the extension for mobility is shown in Figure 2.3. The upper part represents the ICMP packet while the lower part is the extension needed for mobility. The fields necessary on lower layers for the agent advertisement are not shown in this figure. Clearly, mobile nodes must be reached with the appropriate link layer address. The TTL field of the IP packet is set to 1 for all advertisements to avoid forwarding them. The IP destination address according to standard router advertisements can be either set to 224.0.0.1, which is the multicast address for all systems on a link or to the broadcast address 255.255.255.255.

The fields in the ICMP part are defined as follows. The **type** is set to 9, the **code** can be 0, if the agent also routes traffic from non-mobile nodes, or 16, if it does not route anything other than mobile traffic. Foreign agents are at least required to forward packets from the mobile node. The number of addresses advertised with this packet is in **#addresses** while the **addresses** themselves follow as shown. **Lifetime** denotes the length of time this advertisement is valid. **Preference** levels for each address help a node to choose the router that is the most eager one to get a new node.

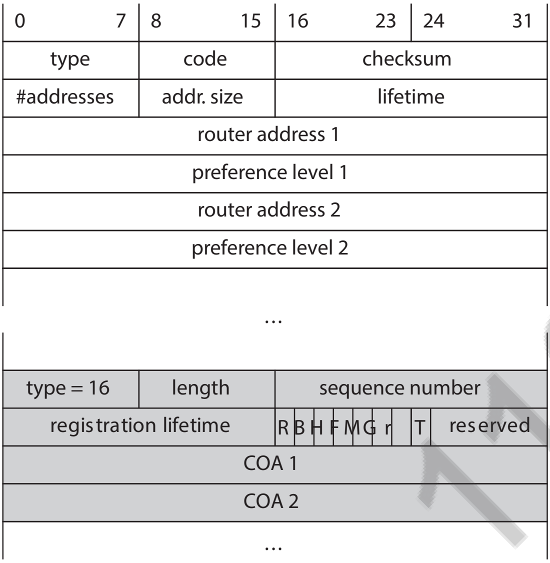


Figure 2.3 Agent advertisement packet (RFC 1256 + mobility extension)

The difference compared with standard ICMP advertisements is what happens after the router addresses. This extension for mobility has the following fields defined: **type** is set to 16, **length** depends on the number of COAs provided with the message and equals 6 + 4*(number of addresses). An agent shows the total number of advertisements sent since initialization in the **sequence number**. By the **registration lifetime** the agent can specify the maximum lifetime in seconds a node can request during registration . The following bits specify the characteristics of an agent in detail. The **R** bit (registration) shows, if a registration with this agent is required even when using a colocated COA at the MN. If the agent is currently too busy to accept new registrations it can set the **B** bit. The following two bits denote if the agent offers services as a home agent (**H**) or foreign agent (**F**) on the link where the advertisement has been sent. Bits M and G specify the method of encapsulation used for the tunnel. While IP-in-IP encapsulation is the mandatory standard, **M** can specify minimal encapsulation and **G** generic routing encapsulation. In the first version of mobile IP (RFC 2002) the **V** bit specified the use of header compression according to RFC 1144 . Now the field **r** at the same bit position is set to zero and must be ignored. The new field **T** indicates that reverse tunneling is supported by the FA. The following fields contain the **COAs** advertised. A foreign agent setting the F bit must advertise at least one COA. Further details and special extensions can be found in Perkins and RFC 3220. A mobile node in a subnet can now receive agent advertisements from either its home agent or a foreign agent. This is one way for the MN to discover its location.

Agent solicitation

If no agent advertisements are present or the inter-arrival time is too high, and an MN has not received a COA by other means, the mobile node must send **agent solicitations**. These solicitations are again based on RFC 1256 for router solicitations. Care must be taken to ensure that these solicitation messages do not flood the network, but basically an MN can search for an FA endlessly sending out solicitation messages. Typically, a mobile node can send out three solicitations, one per second, as soon as it enters a new network. It should be noted that in highly dynamic wireless networks with moving MNs and probably with applications requiring continuous packet streams even one second intervals between solicitation messages might be too long. Before an MN even gets a new address many packets will be lost without additional mechanisms.

If a node does not receive an answer to its solicitations it must decrease the rate of solicitations exponentially to avoid flooding the network until it reaches a maximum interval between solicitations (typically one minute). Discovering a new agent can be done anytime, not just if the MN is not connected to one. Consider the case that an MN is looking for a better connection while still sending via the old path. This is the case while moving through several cells of different wireless networks.

After these steps of advertisements or solicitations the MN can now receive a COA, either one for an FA or a co-located COA. The MN knows its location (home network or foreign network) and the capabilities of the agent (if needed). The next step for the MN is the registration with the HA if the MN is in a foreign network as described in the following.

Registration

Having received a COA, the MN has to register with the HA. The main purpose of the registration is to inform the HA of the current location for correct forwarding of packets. Registration can be done in two different ways depending on the location of the COA.

- If the COA is at the FA, registration is done as illustrated in Figure 2.4 (left). The MN sends its registration request containing the COA (see Figure 2.5) to the FA which is forwarding the request to the HA. The HA now sets up a **mobility binding** containing the mobile node's home IP address and the current COA. Additionally, the mobility binding contains the lifetime of the registration which is negotiated during the registration process. Registration expires automatically after the lifetime and is deleted; so, an MN should reregister before expiration. This mechanism is necessary to avoid mobility bindings which are no longer used. After setting up the mobility binding, the HA sends a reply message back to the FA which forwards it to the MN.
- If the COA is co-located, registration can be simpler, as shown in Figure 2.4 (right). The MN may send the request directly to the HA and vice versa. This, by the way, is also the registration procedure for MNs returning to their home network. Here they also register directly with the HA. However, if the MN received an agent advertisement from the FA it should register via this FA if the R bit is set in the advertisement.

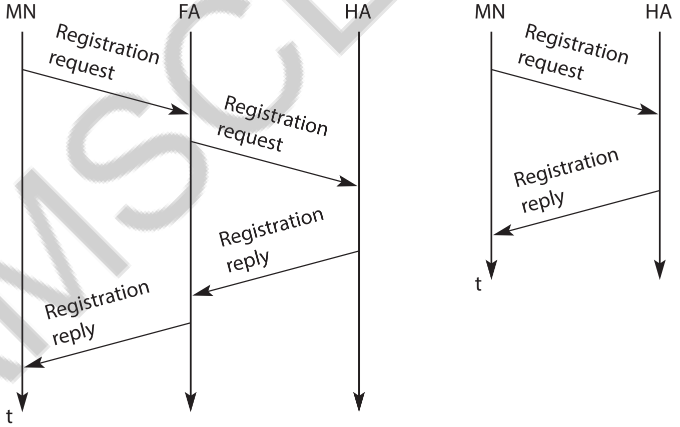


Figure 2.4 Registration of a mobile node via the FA or directly with the HA

0 7	8 15	16 23	24 31
type 1	S B D M G r T x	lifetime	
home address			
home agent			
COA			
identification			
extensions ...			

Figure 2.5 Registration request

UDP packets are used for **registration requests**. The IP source address of the packet is set to the interface address of the MN, the IP destination address is that of the FA or HA (depending on the location of the COA). The UDP destination port is set to 434. UDP is used because of low overheads and better performance compared to TCP in wireless environments . The fields relevant for mobile IP registration requests follow as UDP data . The fields are defined as follows.

The first field **type** is set to 1 for a registration request. With the **S** bit an MN can specify if it wants the HA to retain prior mobility bindings. This allows for simultaneous bindings. The following bits denote the requested behavior for packet forwarding. Setting the **B** bit generally indicates that an MN also wants to receive the broadcast packets which have been received by the HA in the home network. A more detailed description of how to filter broadcast messages which are not needed by the MN can be found in Perkins (1997). If an MN uses a co-located COA, it also takes care of the decapsulation at the tunnel endpoint. The **D** bit indicates this behavior. As already defined for agent advertisements, the following bits **M** and **G** denote the use of minimal encapsulation or generic routing encapsulation, respectively. **T** indicates reverse tunneling, **r** and **x** are set to zero.

0 7	8 15	16 31
type = 3	code	lifetime
home address		
home agent		
identification		
extensions ...		

Figure 2.6 Registration reply

Lifetime denotes the validity of the registration in seconds. A value of zero indicates deregistration; all bits set indicates infinity. The **home address** is the fixed IP address of the MN, **home agent** is the IP address of the HA, and **COA** represents the tunnel endpoint. The 64 bit **identification** is generated by the MN to identify a request and match it with registration replies. This field is used for protection against replay attacks of registrations. The **extensions** must at least contain parameters for authentication.

A **registration reply**, which is conveyed in a UDP packet, contains a **type** field set to 3 and a **code** indicating the result of the registration request. Table 2.1 gives some example codes.

Table 2.1 Example registration reply codes

Registration	Code	Explanation
successful	0	registration accepted
	1	registration accepted, but simultaneous mobility bindings unsupported
denied by FA	65	administratively prohibited
	66	insufficient resources
	67	mobile node failed authentication
	68	home agent failed authentication
	69	requested lifetime too long
denied by HA	129	administratively prohibited
	130	insufficient resources
	131	mobile node failed authentication
	132	foreign agent failed authentication
	133	registration identification mismatch
	135	too many simultaneous mobility bindings

The **lifetime** field indicates how many seconds the registration is valid if it was successful. **Home address** and **home agent** are the addresses of the MN and the HA, respectively. The 64-bit **identification** is used to match registration requests with replies. The value is based on the identification field from the registration and the authentication method. Again, the **extensions** must at least contain parameters for authentication.

4. **Demonstrate how tunneling works in general and especially for mobile IP using IP-IP, minimal, and generic routing encapsulation, respectively.**

Tunneling and encapsulation

The following describes the mechanisms used for forwarding packets between the HA and the COA, as shown in Figure 2.2, step 2. A **tunnel** establishes a virtual pipe for data packets between a tunnel entry and a tunnel endpoint. Packets entering a tunnel are forwarded inside the tunnel and leave the tunnel unchanged. Tunneling, i.e., sending a packet through a tunnel, is achieved by using encapsulation.

Encapsulation is the mechanism of taking a packet consisting of packet header and data and putting it into the data part of a new packet. The reverse operation, taking a packet out of the data part of another packet, is called **decapsulation**. Encapsulation and decapsulation are the operations typically performed when a packet is transferred from a higher protocol layer to a lower layer or from a lower to a higher layer respectively. Here these functions are used within the same layer.

This mechanism is shown in Figure 2.7 and describes exactly what the HA at the tunnel entry does. The HA takes the original packet with the MN as destination, puts it into the data part of a new packet and sets the new IP header in such a way that the packet is routed to the COA. The new header is also called the **outer header** for obvious reasons. Additionally, there is an **inner header** which can be identical to the original header as this is the case for IP-in-IP encapsulation, or the inner header can be computed during encapsulation.

IP-in-IP encapsulation

There are different ways of performing the encapsulation needed for the tunnel between HA and COA. Mandatory for mobile IP is **IP-in-IP encapsulation** as specified in RFC 2003. Figure 2.8 shows a packet inside the tunnel. The fields follow the standard specification of the IP protocol as defined

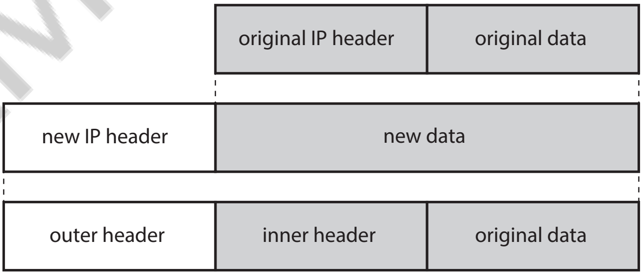
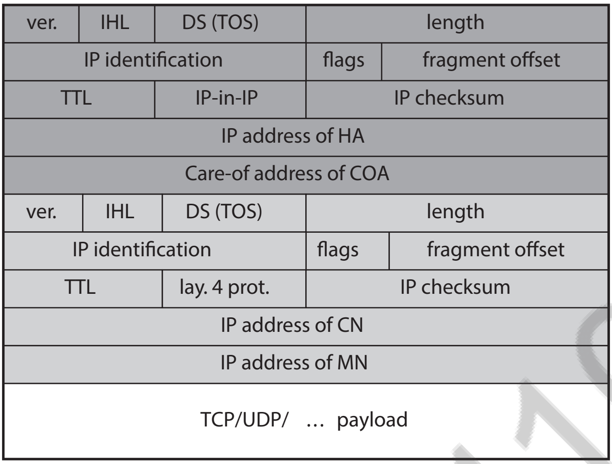


Figure 2.7 IP encapsulation



<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>IP-in-IP</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>

Figure 2.8 IP-in-IP encapsulation

in RFC 791 and the new interpretation of the former TOS, now DS field in the context of differentiated services . The fields of the outer header are set as follows. The version field **ver** is 4 for IP version 4, the internet header length (**IHL**) denotes the length of the outer header in 32 bit words. **DS(TOS)** is just copied from the inner header, the **length** field covers the complete encapsulated packet. The fields up to TTL have no special meaning for mobile IP and are set according to RFC 791. **TTL** must be high enough so the packet can reach the tunnel endpoint. The next field, here denoted with **IP-in-IP**, is the type of the protocol used in the IP payload. This field is set to 4, the protocol type for IPv4 because again an IPv4 packet follows after this outer header. IP **checksum** is calculated as usual. The next fields are the tunnel entry as source address (the **IP address of the HA**) and the tunnel exit point as destination address (the **COA**).

If no options follow the outer header, the inner header starts with the same fields as just explained. This header remains almost unchanged during encapsulation, thus showing the original sender CN and the receiver MN of the packet. The only change is TTL which is decremented by 1. This means that the whole tunnel is considered a single hop from the original packet's point of view. This is a very important feature of tunneling as it allows the MN to behave as if it were attached to the home network. No matter how many real hops the packet has to take in the tunnel, it is just one (logical) hop away for the MN. Finally, the payload follows the two headers.

Minimal encapsulation

As seen with IP-in-IP encapsulation, several fields are redundant. For example, TOS is just copied, fragmentation is often not needed etc. Therefore, **minimal encapsulation** (RFC 2004) as shown in Figure 2.9 is an optional encapsulation method for mobile IP (Perkins, 1996c). The tunnel entry point and endpoint are specified. In this case, the field for the type of the following header contains the value 55 for the minimal encapsulation protocol. The inner header is different for minimal encapsulation. The type of the following protocol and the address of the MN are needed. 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.

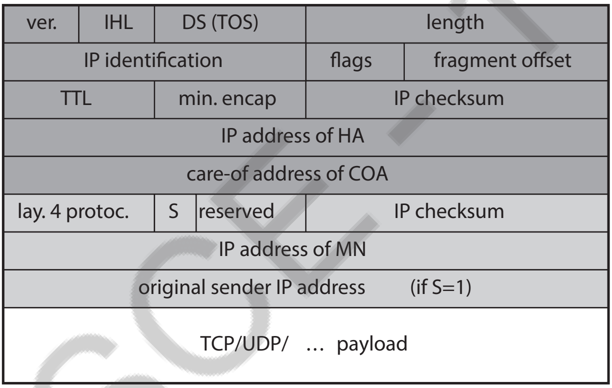


Figure 2.9 Minimal encapsulation

Generic routing encapsulation

While IP-in-IP encapsulation and minimal encapsulation work only for IP, the following encapsulation scheme also supports other network layer protocols in addition to IP. **Generic routing encapsulation** (GRE) allows the encapsulation of packets of one protocol suite into the payload portion of a packet of another protocol suite (Hanks, 1994). Figure 2.10 shows this procedure. The packet of one protocol suite with the original packet header and data is taken and a new GRE header is prepended. Together this forms the new data part of the new packet. Finally, the header of the second protocol suite is put in front.

Figure 2.11 shows on the left side the fields of a packet inside the tunnel between home agent and COA using GRE as an encapsulation scheme according to RFC 1701. The outer header is the standard IP header with HA as source address and COA as destination address. The protocol type used in this outer IP header is 47 for GRE. The other fields of the outer packet, such as TTL and TOS, may be copied from the original IP header. However, the TTL must be decremented by 1 when the packet is decapsulated to prevent indefinite forwarding.

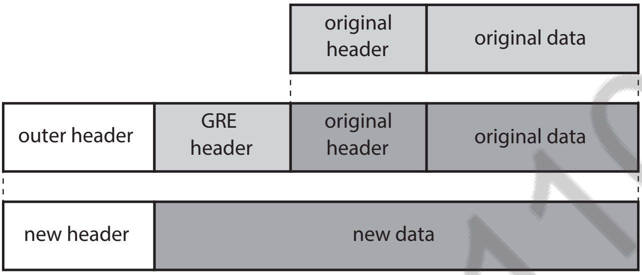


Figure 2.10 Generic routing encapsulation

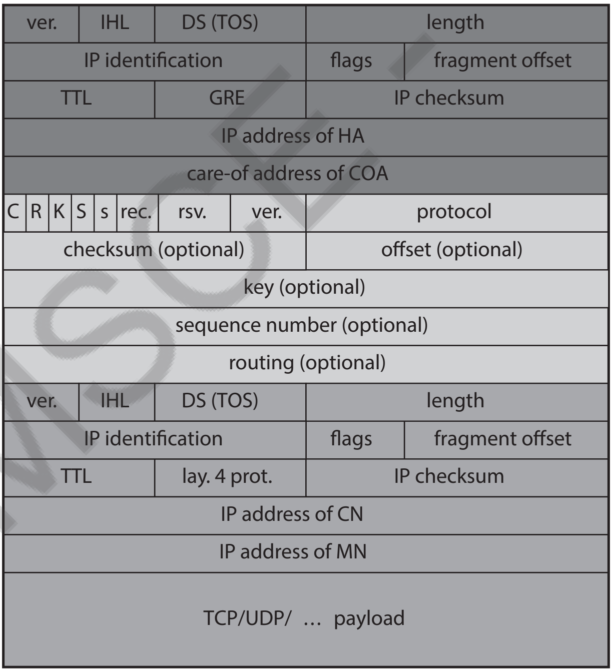


Figure 2.11 Protocol fields for GRE according to RFC 1701

The GRE header starts with several flags indicating if certain fields are present or not. A minimal GRE header uses only 4 bytes; nevertheless, GRE is flexible enough to include several mechanisms in its header. The **C** bit indicates if the checksum field is present and contains valid information.

If **C** is set, the **checksum** field contains a valid IP checksum of the GRE header and the payload. The **R** bit indicates if the offset and routing fields are present and contain valid information. The **offset** represents the offset in bytes for the first source **routing** entry. The routing field, if present, has a variable length and contains fields for source routing. If the C bit is set, the offset field is also present and, vice versa, if the R bit is set, the checksum field must be present. The only reason for this is to align the following fields to 4 bytes. The checksum field is valid only if C is set, and the offset field is valid only if R is set respectively.

GRE also offers a **key** field which may be used for authentication. If this field is present, the **K** bit is set. However, the authentication algorithms are not further specified by GRE. The sequence number bit **S** indicates if the **sequence** number field is present, if the s bit is set, strict source routing is used. Sequence numbers may be used by a decapsulator to restore packet order. This can be important, if a protocol guaranteeing in-order transmission is encapsulated and transferred using a protocol which does not guarantee in-order delivery, e.g., IP. Now the decapsulator at the tunnel exit must restore the sequence to maintain the characteristic of the protocol.

<table>
<tr><td>C</td><td>reserved0</td><td>ver.</td><td>protocol</td></tr>
<tr><td colspan="3">checksum (optional)</td><td>reserved1 (=0)</td></tr>
</table>

Figure 2.12 Protocol fields for GRE according to RFC 2784

The **recursion control** field (rec.) is an important field that additionally distinguishes GRE from IP-in-IP and minimal encapsulation. This field represents a counter that shows the number of allowed recursive encapsulations. As soon as a packet arrives at an encapsulator it checks whether this field equals zero. If the field is not zero, additional encapsulation is allowed the packet is encapsulated and the field decremented by one. Otherwise the packet will most likely be discarded. This mechanism prevents indefinite recursive encapsulation which might happen with the other schemes if tunnels are set up improperly (e.g., several tunnels forming a loop). The default value of this field should be 0, thus allowing only one level of encapsulation.

The following **reserved** fields must be zero and are ignored on reception. The **version** field contains 0 for the GRE version. The following 2 byte **protocol** field represents the protocol of the packet following the GRE header. Several values have been defined, e.g., 0 × 6558 for transparent

Ethernet bridging using a GRE tunnel. In the case of a mobile IP tunnel, the protocol field contains 0 × 800 for IP.

The standard header of the original packet follows with the source address of the correspondent node and the destination address of the mobile node. Figure 2.12 shows the simplified header of GRE following RFC 2784 , which is a more generalized version of GRE compared to RFC 1701. This version does not address mutual encapsulation and ignores several protocol-specific nuances on purpose. The field **C** indicates again if a checksum is present. The next 5 bits are set to zero, then 7 reserved bits follow. The **version** field contains the value zero. The **protocol** type, again, defines the protocol of the payload following RFC 3232 . If the flag C is set, then **checksum** field and a field called reserved1 follows. The latter field is constant zero set to zero follow. RFC 2784 deprecates several fields of RFC 1701, but can interoperate with RFC 1701-compliant implementations.

5. (i) Why is routing in multi-hop ad-hoc networks complicated, what are the special challenges?

(ii) How would you solve the problem of triangular routing?

Optimizations

Imagine the following scenario. A Japanese and a German meet at a conference on Hawaii. Both want to use their laptops for exchanging data, both run mobile IP for mobility support. Now recall Figure 8.2 and think of the way the packets between both computers take.

If the Japanese sends a packet to the German, his computer sends the data to the HA of the German, i.e., from Hawaii to Germany. The HA in Germany now encapsulates the packets and tunnels them to the COA of the German laptop on Hawaii. This means that although the computers might be only meters away, the packets have to travel around the world! This inefficient behavior of a nonoptimized mobile IP is called **triangular routing**. The triangle is made of the three segments, CN to HA, HA to COA/MN, and MN back to CN.

With the basic mobile IP protocol all packets to the MN have to go through the HA. This can cause unnecessary overheads for the network between CN and HA, but also between HA and COA, depending on the current location of the MN. As the example shows, latency can increase dramatically. This is particularly unfortunate if the MNs and HAs are separated by, e.g., transatlantic links.

One way to optimize the route is to inform the CN of the current location of the MN. The CN can learn the location by caching it in a **binding cache** which is a part of the local routing table for the CN. The appropriate entity to inform the CN of the location is the HA. The optimized mobile IP protocol needs four additional messages.

- **Binding request:** Any node that wants to know the current location of an MN can send a binding request to the HA. The HA can check if the MN has allowed dissemination of its current location. If the HA is allowed to reveal the location it sends back a binding update.
- **Binding update:** This message sent by the HA to CNs reveals the current location of an MN. The message contains the fixed IP address of the MN and the COA. The binding update can request an acknowledgement.
- **Binding acknowledgement:** If requested, a node returns this acknowledgement after receiving a binding update message.
- **Binding warning:** If a node decapsulates a packet for an MN, but it is not the current FA for this MN, this node sends a binding warning. The warning contains MN's home address and a target node address, i.e., the address of the node that has tried to send the packet to this MN. The recipient of the warning then knows that the target node could benefit from obtaining a fresh binding for the MN. The recipient can be the HA, so the HA should now send a binding update to the node that obviously has a wrong COA for the MN.

Figure 2.13 explains these additional four messages together with the case of an MN changing its FA. The CN can request the current location from the HA. If allowed by the MN, the HA returns the COA of the MN via an update message. The CN acknowledges this update message and stores the mobility binding. Now the CN can send its data directly to the current foreign agent FA_{old}. FA_{old} forwards the packets to the MN. This scenario shows a COA located at an FA. Encapsulation of data for tunneling to the COA is now done by the CN, not the HA.

The MN might now change its location and register with a new foreign agent, FA_{new}. This registration is also forwarded to the HA to update its location database. Furthermore, FA_{new} informs FA_{old} about the new registration of MN. MN's registration message contains the address of FA_{old} for this purpose. Passing this information is achieved via an update message, which is acknowledged by FA_{old}. Registration replies are not shown in this scenario. Without the information provided by the new FA, the old FA would not get to know anything about the new location of MN. In this case, CN does not know anything about the new location, so it still tunnels its packets for MN to the old FA, FA_{old}. This FA now notices packets with destination MN, but also knows that it is not the current FA of MN. FA_{old} might now forward these packets to the new COA of MN which is FA_{new} in this example. This forwarding of packets is another optimization of the basic Mobile IP providing **smooth handovers**. Without this optimization, all packets in transit would be lost while the MN moves from one FA to another. With TCP as the higher layer protocol this would result in severe performance degradation .

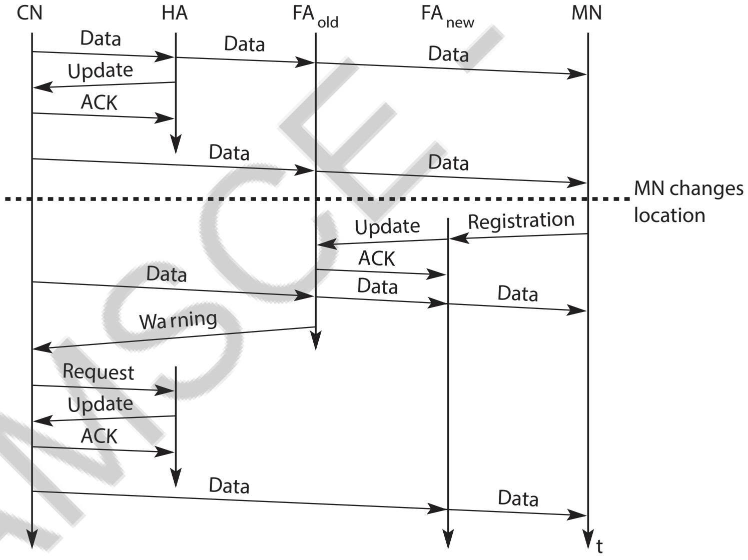


Figure 2.13 Change of the foreign agent with an optimized mobile IP

To tell CN that it has a stale binding cache, FA_{old} sends, in this example, a binding warning message to CN. CN then requests a binding update. (The warning could also be directly sent to the HA triggering an update). The HA sends an update to inform the CN about the new location, which is acknowledged. Now CN can send its packets directly to FA_{new}, again avoiding triangular routing. Unfortunately, this optimization of mobile IP to avoid triangular routing causes several security problems (e.g., tunnel hijacking) as discussed in Montenegro (1998). Not all users of mobile communication systems want to reveal their current 'location' (in the sense of an IP subnet) to a communication partner.

6. What is the main idea of reverse tunneling ?

Reverse tunneling

At first glance, the return path from the MN to the CN shown in Figure 2.2 looks quite simple. The MN can directly send its packets to the CN as in any other standard IP situation. The destination address in the packets is that of CN. But there are several severe problems associated with this simple solution.

- **Firewalls:** Almost all companies and many other institutions secure their internal networks (intranet) connected to the internet with the help of a fire-wall. All data to and from the intranet must pass through the firewall. Besides many other functions, firewalls can be set up to filter out malicious addresses from an administrator's point of view. Quite often firewalls only allow packets with topologically correct addresses to pass. This provides at least a first and simple protection against misconfigured systems of unknown addresses. However, MN still sends packets with its fixed IP address as source which is not topologically correct in a foreign network. Firewalls often filter packets coming from outside containing a source address from computers of the internal network. This avoids other computers that could use internal addresses and claim to be internal computers. However, this also implies that an MN cannot send a packet to a computer residing in its home network. Altogether, this means that not only does the destination address matter for forwarding IP packets, but also the source address due to security concerns. Further complications arise through the use of private addresses inside the intranet and the translation into global addresses when communicating with the internet. This **network address translation** is used by many companies to hide internal resources (routers, computers, printers etc.) and to use only some globally available addresses .
- **Multi-cast:** Reverse tunnels are needed for the MN to participate in a multicast group. While the nodes in the home network might

participate in a multi-cast group, an MN in a foreign network cannot transmit multi-cast packets in a way that they emanate from its home network without a reverse tunnel. The foreign network might not even provide the technical infrastructure for multi-cast communication (multi-cast backbone, Mbone).

- ✸ **TTL:** Consider an MN sending packets with a certain TTL while still in its home network. The TTL might be low enough so that no packet is transmitted outside a certain region. If the MN now moves to a foreign network, this TTL might be too low for the packets to reach the same nodes as before. Mobile IP is no longer transparent if a user has to adjust the TTL while moving. A reverse tunnel is needed that represents only one hop, no matter how many hops are really needed from the foreign to the home network.

All these considerations led to RFC 2344 defining reverse tunneling as an extension to mobile IP. The new RFC 3024 (Montenegro, 2001) renders RFC 2344 obsolete but comprises only some minor changes for the original standard. The RFC was designed backwards-compatible to mobile IP and defines topologically correct reverse tunneling as necessary to handle the problems described above. Reverse tunneling was added as an option to mobile IP in the new standard (RFC 3344).

Reverse tunneling creates a triangular routing problem in the reverse direction. All packets from an MN to a CN go through the HA. RFC 3024 does not offer a solution for this reverse triangular routing, because it is not clear if the CN can decapsulate packets. Remember that mobile IP should work together with all traditional, non-mobile IP nodes. Therefore, one cannot assume that a CN is able to be a tunnel endpoint.

Reverse tunneling also raises several security issues which have not been really solved up to now. For example, tunnels starting in the private network of a company and reaching out into the internet could be hijacked and abused for sending packets through a firewall. It is not clear if companies would allow for setting up tunnels through a firewall without further checking of packets. It is more likely that a company will set up a special virtual network for visiting mobile nodes outside the firewall with full connectivity to the internet. This allows guests to use their mobile equipment, and at the same time, today's security standards are maintained. Initial architectures integrating mobility and security aspects within firewalls exist

7. (i) What facts show the advantages of IPv6 offer for mobility? (4)

(ii) Enumerate the three prominent approaches to address micro-mobility problems. (12)

While mobile IP was originally designed for IP version 4, IP version 6 makes life much easier. Several mechanisms that had to be specified separately for mobility support come free in IPv6. . One issue is security with regard to authentication, which is now a required feature for all IPv6 nodes. No special mechanisms as add-ons are needed for securing mobile IP registration. Every IPv6 node masters address autoconfiguration the mechanisms for acquiring a COA are already built in. Neighbor discovery as a mechanism mandatory for every node is also included in the specification; special foreign agents are no longer needed to advertise services. Combining the features of autoconfiguration and neighbor discovery means that every mobile node is able to create or obtain a topologically correct address for the current point of attachment.

Every IPv6 node can send binding updates to another node, so the MN can send its current COA directly to the CN and HA. These mechanisms are an integral part of IPv6. A soft handover is possible with IPv6. The MN sends its new COA to the old router servicing the MN at the old COA, and the old router encapsulates all incoming packets for the MN and forwards them to the new COA.

Altogether, mobile IP in IPv6 networks requires very few additional mechanisms of a CN, MN, and HA. The FA is not needed any more. A CN only has to be able to process binding updates, i.e., to create or to update an entry in the routing cache. The MN itself has to be able to decapsulate packets, to detect when it needs a new COA, and to determine when to send binding updates to the HA and CN. A HA must be able to encapsulate packets. However, IPv6 does not solve any firewall or privacy problems. Additional mechanisms on higher layers are needed for this.

IP micro-mobility support

Mobile IP exhibits several problems regarding the duration of handover and the scalability of the registration procedure. Assuming a large number of mobile devices changing networks quite frequently, a high load on the home agents as well as on the networks is generated by registration and binding update messages. IP micro-mobility protocols can complement mobile IP by offering fast and almost seamless handover control in limited geographical areas.

Consider a client arriving with his or her laptop at the customer's premises. The home agent only has to know an entry point to the customer's network, not the details within this network. The entry point acts as the current location. Changes in the location within the customer's network should be handled locally to minimize network traffic and to speed-up local handover. The basic underlying idea is the same for all micro-mobility protocols: Keep the frequent updates generated by local changes of the points of attachment away from the home network and only inform the home agent about major changes, i.e., changes of a region. In some sense all micro-mobility protocols establish a hierarchy. However, the debate is still going on if micro-mobility aspects should really be handled on the IP layer or if layer 2 is the better place for it. Layer 2 mobility support would comprise, e.g., the inter access point protocol (IAPP) of 802.11 WLANs or the mobility support mechanisms of mobile phone systems .

The following presents three of the most prominent approaches, which should be seen neither as standards nor as final solutions of the micro-mobility problems. Campbell (2002) presents a comparison of the three approaches.

Cellular IP

Cellular IP provides local handovers without renewed registration by instaling a single **cellular IP gateway (CIPGW)** for each domain, which acts to the outside world as a foreign agent . Inside the cellular IP domain, all nodes collect routing information for accessing MNs based on the origin of packets sent by the MNs towards the CIPGW. Soft handovers are achieved by allowing simultaneous forwarding of packets destined for a mobile node along multiple paths. A mobile node moving between adjacent cells will temporarily be able to receive packets via both old and new **base stations (BS)** if this is supported by the lower protocol layers.

Concerning the manageability of cellular IP, it has to be noted that the approach has a simple and elegant architecture and is mostly self-configuring. However, mobile IP tunnels could be controlled more easily if the CIPGW was integrated into a firewall, but there are no detailed specifications in regarding such integration. Cellular IP requires changes to the basic mobile IP protocol and is not transparent to existing systems. The foreign network's routing tables are changed based on messages sent by mobile nodes. These should not be trusted blindly even if they have been authenticated. This could be exploited by systems in the foreign network for wiretapping packets destined for an MN by sending packets to the CIPGW with the source address set to the MN's address. In enterprise scenarios requiring basic communications security, this may not be acceptable.

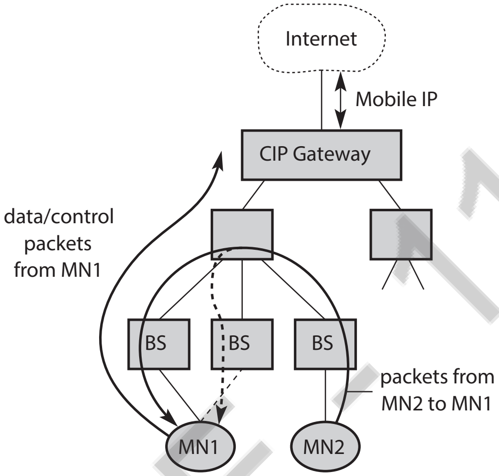


Figure 2.14 Basic architecture of cellular IP

Advantage

- Manageability: Cellular IP is mostly self-configuring, and integration of the CIPGW into a firewall would facilitate administration of mobility-related functionality. This is, however, not explicitly specified in (Campbell, 2000).

Disadvantages

- Efficiency: Additional network load is induced by forwarding packets on multiple paths.
- Transparency: Changes to MNs are required.
- Security: Routing tables are changed based on messages sent by mobile nodes. Additionally, all systems in the network can easily obtain a copy of all packets destined for an MN by sending packets with the MN's source address to the CIPGW.

Hawaii

HAWAII (Handoff-Aware Wireless Access Internet Infrastructure) tries to keep micro-mobility support as transparent as possible for both home agents and mobile nodes (which have to support route optimization). Its concrete goals are performance and reliability improvements and support for quality of service mechanisms. On entering an HAWAII domain, a mobile node obtains a co-located COA (step 1) and registers with the HA (step 2). Additionally, when moving to another cell inside the foreign domain, the MN sends a registration request to the new base station as to a foreign agent (step 3), thus mixing the concepts of co-located COA and foreign agent COA. The base station intercepts the registration request and sends out a handoff update message, which reconfigures all routers on the paths from the old and new base station to the so-called crossover router (step 4). When routing has been reconfigured successfully, the base station sends a registration reply to the mobile node, again as if it were a foreign agent.

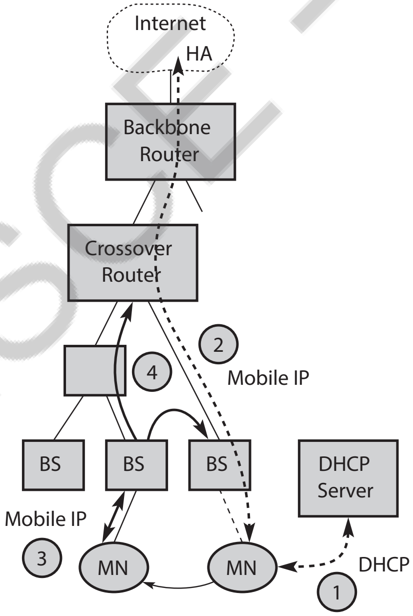


Figure 2.15 Basic architecture of HAWAII

The use of challenge-response extensions for authenticating a mobile node is mandatory. In contrast to cellular IP, routing changes are always initiated by the foreign domain's infrastructure, and the corresponding messages could be authenticated, e.g., by means of an IPSec authentication header , reducing the risk of malicious rerouting of traffic initiated by bogus mobile hosts. HAWAII claims to be mostly transparent to mobile nodes, but this claim has to be regarded with some caution as the requirement to support a co-located care-ofaddress as well as to interact with foreign agents could cause difficulties with some mobile nodes.

Advantages

✷ ● Security: Challenge-response extensions are mandatory. In contrast to Cellular IP, routing changes are always initiated by the foreign domain's infrastructure.

✷ Transparency: HAWAII is mostly transparent to mobile nodes.

Disadvantages

✷ Security: There are no provisions regarding the setup of IPSec tunnels.

✷ Implementation: No private address support is possible because of colocated COAs.

Hierarchical mobile IPv6 (HMIPv6)

As introducing hierarchies is the natural choice for handling micro-mobility issues, several proposals for a 'hierarchical' mobile IP exist.

HMIPv6 provides micro-mobility support by installing a **mobility anchor point (MAP)**, which is responsible for a certain domain and acts as a local HA within this domain for visiting MNs . The MAP receives all packets on behalf of the MN, encapsulates and forwards them directly to the MN's current address (link COA, **LCOA**). As long as an MN stays within the domain of a MAP, the globally visible COA (regional COA, **RCOA**) does not change. A MAP domain's boundaries are defined by the **access routers (AR)** advertising the MAP information to the attached MNs. A MAP assists with local handovers and maps RCOA to LCOA. MNs register their RCOA with the HA using a binding update. When a MN moves locally it must only register its new LCOA with its MAP. The RCOA stays unchanged. To support smooth handovers between MAP domains, an MN can send a binding update to its former MAP.

It should be mentioned as a security benefit that mobile nodes can be provided with some kind of limited location privacy because LCOAs on lower levels of the mobility hierarchy can be hidden from the outside world.

However, this applies only to micro mobility, that is, as long as the mobile node rests in the same domain. A MN can also send a binding update to a CN who shares the same link. This reveals its location but optimizes packet flow (direct routing without going through the MAP). MNs can use their RCOA as source address. The extended mode of HMIPv6 supports both mobile nodes and mobile networks.

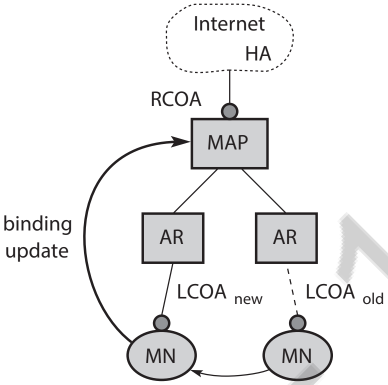


Figure 2.16 Internet Basic architecture of hierarchical mobile IP

Advantages

- ✷ Security: MNs can have (limited) location privacy because LCOAs can be hidden.
- ✷ Efficiency: Direct routing between CNs sharing the same link is possible

Disadvantages

- ✷ ● Transparency: Additional infrastructure component (MAP).
- ✷ ● Security: Routing tables are changed based on messages sent by mobile nodes. This requires strong authentication and protection against denial of service attacks. Additional security functions might be necessary in MAPs

8. (i) What ideas justify the use of DHCP for mobility and support of mobile IP? (8)

(ii) Explain the benefits of Mobile ad-hoc networks (8)

Dynamic host configuration protocol

The dynamic host configuration protocol is mainly used to simplifiy the installation and maintenance of networked computers. If a new computer is connected to a network, DHCP can provide it with all the necessary information for full system integration into the network, e.g., addresses of a DNS server and the default router, the subnet mask, the domain name, and an IP address. Providing an IP address, makes DHCP very attractive for mobile IP as a source of care-of-addresses. While the basic DHCP mechanisms are quite simple, many options are available as described in RFC 2132 .

DHCP is based on a client/server model as shown in Figure 2.17. DHCP clients send a request to a server to which the server responds. A client sends requests using MAC broadcasts to reach all devices in the LAN. A DHCP relay might be needed to forward requests across inter-working units to a DHCP server.

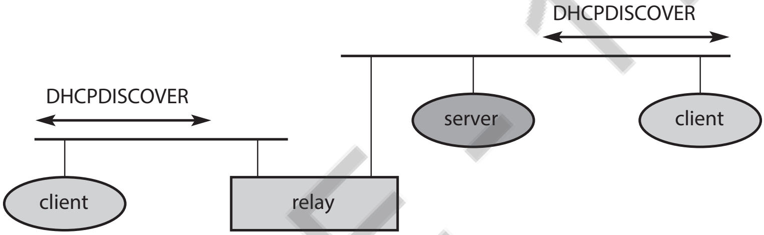


Figure 2.17 Basic DHCP configuration

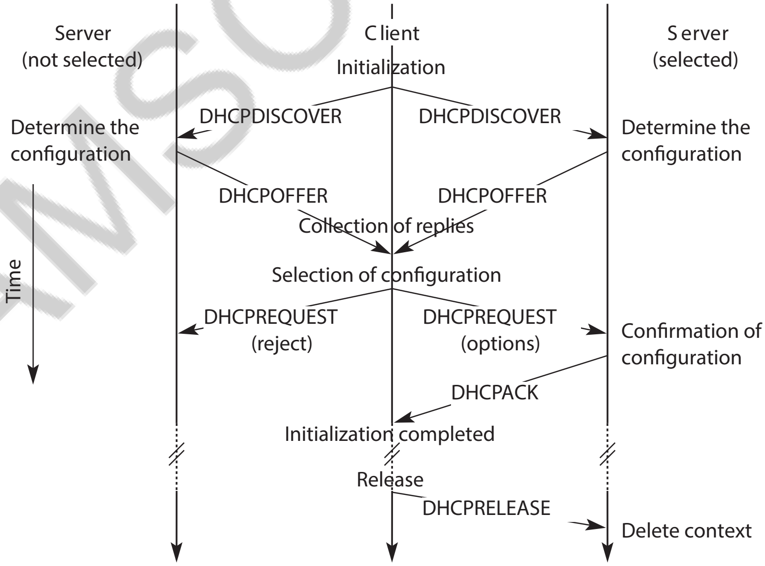


Figure 2.18 Client initialization via DHCP

A typical initialization of a DHCP client is shown in Figure 2.18. The figure shows one client and two servers. As described above, the client broadcasts a DHCPDISCOVER into the subnet. There might be a relay to forward this broadcast. In the case shown, two servers receive this broadcast and determine the configuration they can offer to the client. One example for this could be the checking of available IP addresses and choosing one for the client. Servers reply to the client's request with DHCPOFFER and offer a list of configuration parameters. The client can now choose one of the configurations offered. The client in turn replies to the servers, accepting one of the configurations and rejecting the others using DHCPREQUEST. If a server receives a DHCPREQUEST with a rejection, it can free the reserved configuration for other possible clients. The server with the configuration accepted by the client now confirms the configuration with DHCPACK. This completes the initialization phase.

If a client leaves a subnet, it should release the configuration received by the server using DHCPRELEASE. Now the server can free the context stored for the client and offer the configuration again. The configuration a client gets from a server is only leased for a certain amount of time, it has to be reconfirmed from time to time. Otherwise the server will free the configuration. This timeout of configuration helps in the case of crashed nodes or nodes moved away without releasing the context.

DHCP is a good candidate for supporting the acquisition of care-ofaddresses for mobile nodes. The same holds for all other parameters needed, such as addresses of the default router, DNS servers, the timeserver etc. A DHCP server should be located in the subnet of the access point of the mobile node, or at least a DHCP relay should provide forwarding of the messages. RFC 3118 specifies authentication for DHCP messages which is needed to protect mobile nodes from malicious DHCP servers. Without authentication, the mobile node cannot trust a DHCP server, and the DHCP server cannot trust the mobile node.

9. i). How does mobile adhoc network differ from IEEE802.11 standard? With an example, explain DSDV algorithm.

ii) Explain the benefits of Mobile ad-hoc networks (8)

(OR)

Explain about Mobile ad-hoc networks in detail.

Examples for the use of such mobile, wireless, multi-hop ad-hoc networks, which are only called ad-hoc networks here for simplicity, are:

- **Instant infrastructure:** Unplanned meetings, spontaneous interpersonal communications etc. cannot rely on any infrastructure. Infrastructures need planning and administration. It would take too long to set up this kind of infrastructure; therefore, ad-hoc connectivity has to be set up.
- **Disaster relief:** Infrastructures typically break down in disaster areas. Hurricanes cut phone and power lines, floods destroy base stations, fires burn servers. Emergency teams can only rely on an infrastructure they can set up themselves. No forward planning can be done, and the set-up must be extremely fast and reliable. The same applies to many military activities, which is, to be honest, one of the major driving forces behind mobile ad-hoc networking research.
- **Remote areas:** Even if infrastructures could be planned ahead, it is sometimes too expensive to set up an infrastructure in sparsely populated areas. Depending on the communication pattern, ad-hoc networks or satellite infrastructures can be a solution.
- **Effectiveness:** Services provided by existing infrastructures might be too expensive for certain applications. If, for example, only connectionoriented cellular networks exist, but an application sends only a small status information every other minute, a cheaper ad-hoc packet-oriented network might be a better solution. Registration procedures might take too long, and communication overheads might be too high with existing networks. Application-tailored ad-hoc networks can offer a better solution.

Over the last few years ad-hoc networking has attracted a lot of research interest. This has led to creation of a working group at the IETF that is focussing on **mobile ad-hoc networking**, called **MANET** .Figure 2.19 shows the relation of MANET to mobile IP and DHCP. While mobile IP and DHCP handle the connection of mobile devices to a fixed infrastructure, MANET comprises mobile routers, too. 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. MANET research is responsible for developing protocols and components to enable ad-hoc networking between mobile devices. It should be noted that the separation of end system and router is only a logical separation. Typically, mobile nodes in an adhoc scenario comprise routing and end system functionality.

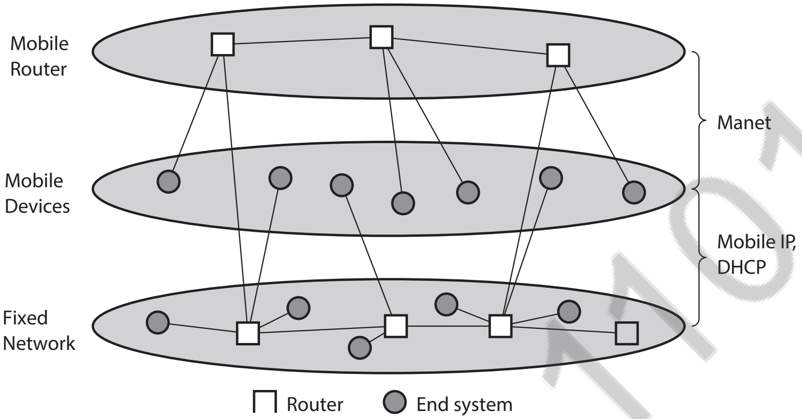


Figure 2.19 MANETs and mobile IP

The reason for having a special section about ad-hoc networks within a chapter about the network layer is that routing of data is one of the most difficult issues in ad-hoc networks.

One of the first ad-hoc wireless networks was the packet radio network started by ARPA in 1973. It allowed up to 138 nodes in the ad-hoc network and used IP packets for data transport. This made an easy connection possible to the ARPAnet, the starting point of today's Internet. Twenty radio channels between 1718.4-1840 MHz were used offering 100 or 400 kbit/s. The system used DSSS with 128 or 32 chips/bit.

A variant of distance vector routing was used in this ad-hoc network . In this approach, each node sends a routing advertisement every 7.5 s. These advertisements contain a neighbor table with a list of link qualities to each neighbor. Each node updates the local routing table according to the distance vector algorithm based on these advertisements. Received packets also help to update the routing table. A sender now transmits a packet to its first hop neighbor using the local neighbor table. Each node forwards a packet received based on its own local neighbor table.

10. (i) How Dynamic source routing eases traffic with an example? (8)

(ii) Show the method by which MN registers with HA? (8)

(OR)

What is the motivation behind dynamic source routing compared to other routing algorithms from fixed network and how does dynamic source routing handle routing? (16)

Routing

While in wireless networks with infrastructure support a base station always reaches all mobile nodes, this is not always the case in an ad-hoc network. A destination node might be out of range of a source node transmitting packets. Routing is needed to find a path between source and destination and to forward the packets appropriately. In wireless networks using an infrastructure, cells have been defined. Within a cell, the base station can reach all mobile nodes without routing via a broadcast. In the case of ad-hoc networks, each node must be able to forward data for other nodes. This creates many additional problems that are discussed in the following paragraphs.

Figure 2.20 gives a simple example of an ad-hoc network. At a certain time t_1 the network topology might look as illustrated on the left side of the figure. Five nodes, N_1 to N_5, are connected depending on the current transmission characteristics between them. In this snapshot of the network, N_4 can receive N_1 over a good link, but N_1 receives N_4 only via a weak link. Links do not necessarily have the same characteristics in both directions. The reasons for this are, e.g., different antenna characteristics or transmit power. N_1 cannot receive N_2 at all, N_2 receives a signal from N_1.

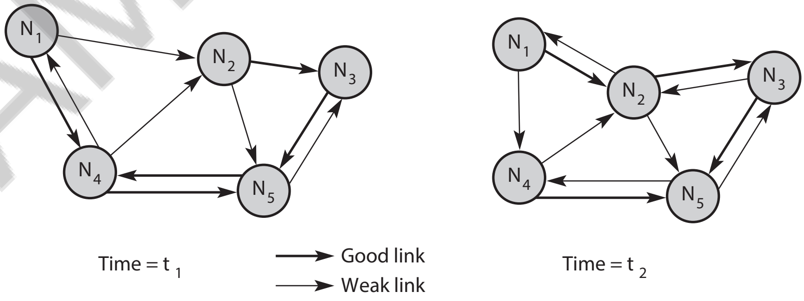


Figure 2.20 Example ad-hoc network N_1

This situation can change quite fast as the snapshot at t2 shows. N1 cannot receive N_4 any longer, N_4 receives N_1 only via a weak link. But now N_1 has an asymmetric but bi-directional link to N_2 that did not exist before.

This very simple example already shows some fundamental differences between wired networks and ad-hoc wireless networks related to routing.

- **Asymmetric links:** Node A receives a signal from node B. But this does not tell us anything about the quality of the connection in reverse. B might receive nothing, have a weak link, or even have a better link than the reverse direction. Routing information collected for one direction is of almost no use for the other direction. However, many routing algorithms for wired networks rely on a symmetric scenario.

- **Redundant links:** Wired networks, too, have redundant links to survive link failures. However, there is only some redundancy in wired networks, which, additionally, are controlled by a network administrator. In ad-hoc networks nobody controls redundancy, so there might be many redundant links up to the extreme of a completely meshed topology. Routing algorithms for wired networks can handle some redundancy, but a high redundancy can cause a large computational overhead for routing table updates.

- **Interference:** In wired networks links exist only where a wire exists, and connections are planned by network administrators. This is not the case for wireless ad-hoc networks. Links come and go depending on transmission characteristics, one transmission might interfere with another, and nodes might overhear the transmissions of other nodes. Interference creates new problems by 'unplanned' links between nodes: if two close-by nodes forward two transmissions, they might interfere and destroy each other. On the other hand, interference might also help routing. A node can learn the topology with the help of packets it has overheard.

- **Dynamic topology:** The greatest problem for routing arises from the highly dynamic topology. The mobile nodes might move as shown in Figure 8.20 or medium characteristics might change. This results in frequent changes in topology, so snapshots are valid only for a very short period of time. In adhoc networks, routing tables must somehow reflect these frequent changes in topology, and routing algorithms have to be adapted. Routing algorithms used in wired networks would either react much too slowly or generate too many updates to reflect all changes in topology. Routing table updates in fixed networks, for example, take place every 30 seconds. This updating frequency might be too low to be useful for ad-hoc networks. Some algorithms rely on a complete picture of the whole network. While this works in wired networks where changes are rare, it fails completely in ad-hoc networks. The topology changes during the distribution of the 'current' snapshot of the network, rendering the snapshot useless.

Assume that node N_1 wants to send data to N_3 and needs an acknowledgement. If N_1 had a complete overview of the network at time t_1, which is not always the case in ad-hoc networks, it would choose the path N_1, N_2, N_3, for this requires only two hops. Acknowledgements cannot take the same path, N_3 chooses N_3, N_5, N_4, N_1. This takes three hops and already shows that routing also strongly influences the function of higher layers. TCP, for example, makes round trip measurements assuming the same path in both directions

At time t_2, the topology has changed. Now N_3 cannot take the same path to send acknowledgements back to N_1, while N_1 can still take the old path to N_3. Although already more complicated than fixed networks, this example still assumes that nodes can have a complete insight into the current situation. 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. While even in fixed networks traffic flows are not exactly predictable, for ad-hoc networks link capacities are additionally unknown. The capacity of each link can change from 0 to the maximum of the transmission technology used. In real ad-hoc networks no node knows all these factors, and establishing up-to-date snapshots of the network is almost impossible.

Ad-hoc networks using mobile nodes face additional problems due to hardware limitations. Using the standard routing protocols with periodic updates wastes battery power without sending any user data and disables sleep modes. Periodic updates waste bandwidth and these resources are already scarce for wireless links.

An additional problem is interference between two or more transmissions that do not use the same nodes for forwarding. If, for example, a second transmission from node N_4 to N_5 (see Figure 2.20) takes place at the same time as the transmission from N_1 to N_3, they could interfere. Interference could take place at N_2 which can receive signals from N_1 and N_4, or at N_5 receiving N_4 and N_2. If shielded correctly, there is no interference between two wires.

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.

- Traditional routing algorithms known from wired networks will not work efficiently or fail completely. These algorithms have not been designed with a highly dynamic topology, asymmetric links, or interference in mind.
- Routing in wireless ad-hoc networks cannot rely on layer three knowledge alone. Information from lower layers concerning connectivity or interference can help routing algorithms to find a good path.
- Centralized approaches will not really work, because it takes too long to collect the current status and disseminate it again. Within this time the topology has already changed.
- Many nodes need routing capabilities. While there might be some without, at least one router has to be within the range of each node. Algorithms have to consider the limited battery power of these nodes.
- The notion of a connection with certain characteristics cannot work properly. Ad-hoc networks will be connectionless, because it is not possible to maintain a connection in a fast changing environment and to forward data following this connection. Nodes have to make local decisions for forwarding and send packets roughly toward the final destination.
- A last alternative to forward a packet across an unknown topology is flooding. This approach always works if the load is low, but it is very inefficient. A hop counter is needed in each packet to avoid looping, and the diameter of the ad-hoc network, i.e., the maximum number of hops, should be known.

Hierarchical clustering of nodes might help. If it is possible to identify certain groups of nodes belonging together, clusters can be established. While individual nodes might move faster, the whole cluster can be rather stationary. Routing between clusters might be simpler and less dynamic .

The following sections give two examples for routing algorithms that were historically at the beginning of MANET research, DSDV and DSR, and useful metrics that are different from the usual hop counting. An overview of protocols follows. This is subdivided into the three categories: flat, hierarchical, and geographic-position-assisted routing .

11. How does mobile adhoc network differ from IEEE802.11 standard? With an example, explain DSDV algorithm.

(OR)

List the differences between AODV and the standard distance vector algorithm. (8)

Destination sequence distance vector

Destination sequence distance vector (DSDV) routing is an enhancement to distance vector routing for ad-hoc networks. DSDV can be considered historically, however, an on-demand version (ad-hoc on-demand distance vector, AODV) is among the protocols currently discussed . Distance vector routing is used as routing information protocol (RIP) in wired networks. It performs extremely poorly with certain network changes due to the count-to-infinity problem. Each node exchanges its neighbor table periodically with its neighbors. Changes at one node in the network propagate slowly through the network (step-by-step with every exchange). The strategies to avoid this problem which are used in fixed networks (poisoned-reverse/ splithorizon) do not help in the case of wireless ad-hoc networks, due to the rapidly changing topology. This might create loops or unreachable regions within the network.

DSDV now adds two things to the distance vector algorithm:

- **Sequence numbers:** Each routing advertisement comes with a sequence number. Within ad-hoc networks, advertisements may propagate along many paths. Sequence numbers help to apply the advertisements in correct order. This avoids the loops that are likely with the unchanged distance vector algorithm.

- **Damping:** Transient changes in topology that are of short duration should not destabilize the routing mechanisms. 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.

The routing table for N_1 in Figure 2.20 would be as shown in Table 2.2.

For each node N_1 stores the next hop toward this node, the metric (here number of hops), the sequence number of the last advertisement for this node, and the time at which the path has been installed first. The table contains flags and a settling time helping to decide when the path can be assumed stable. Router advertisements from N_1 now contain data from the first, third, and fourth column: destination address, metric, and sequence number. Besides being loop-free at all times, DSDV has low memory requirements and a quick convergence via triggered updates.

Dynamic source routing

Imagine what happens in an ad-hoc network where nodes exchange packets from time to time, i.e., the network is only lightly loaded, and DSDV or one of the traditional distance vector or link state algorithms is used for updating routing tables. Although only some user data has to be transmitted, the nodes exchange routing information to keep track of the topology. These algorithms maintain routes between all nodes, although there may currently be no data exchange at all. This causes unnecessary traffic and prevents nodes from saving battery power.

Table 2.2 Part of a routing table for DSDV

Destination	Next hop	Metric	Sequence no.	Instal time
N_1	N_1	0	S_1-321	T_4-001
N_2	N_2	1	S_2-218	T_4-001
N_3	N_2	2	S_3-043	T_4-002
N_4	N_4	1	S_4-092	T_4-001
N_5	N_4	2	S_5-163	T_4-002

Dynamic source routing (DSR), therefore, divides the task of routing into two separate problems (Johnson, 1996), (Johnson, 2002a):

- **Route discovery:** A node only tries to discover a route to a destination if it has to send something to this destination and there is currently no known route.
- **Route maintenance:** If a node is continuously sending packets via a route, it has to make sure that the route is held upright. As soon as a node detects problems with the current route, it has to find an alternative.

The basic principle of source routing is also used in fixed networks, e.g. token rings. Dynamic source routing eliminates all periodic routing updates and works as follows. If a node needs to discover a route, it broadcasts a route request with a unique identifier and the destination address as parameters. Any node that receives a route request does the following.

- 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.

Using this approach, the route request collects a list of addresses representing a possible path on its way towards the destination. As soon as the request reaches the destination, it can return the request packet containing the list to the receiver using this list in reverse order. One condition for this is that the links work bi-directionally. If this is not the case, and the destination node does not currently maintain a route back to the initiator of the request, it has to start a route discovery by itself. The destination may receive several lists containing different paths from the initiator. It could return the best path, the first path, or several paths to offer the initiator a choice.

Applying route discovery to the example in Figure 8.20 for a route from N_1 to N_3 at time t_1 results in the following.

- N_1 broadcasts the request ((N_1), id = 42, target = N_3), N_2 and N_4 receive this request.
- N_2 then broadcasts ((N_1, N_2), id = 42, target = N_3), N_4 broadcasts ((N_1, N_4), id = 42, target = N_3). N_3 and N_5 receive N_2's broadcast, N_1, N_2, and N_5 receive N_4's broadcast.

- N_3 recognizes itself as target, N_5 broadcasts ((N_1, N_2, N_5), id = 42, target = N_3). N_3 and N_4 receive N_5's broadcast. N_1, N_2, and N_5 drop N_4's broadcast packet, because they all recognize an already received route request (and N_2's broadcast reached N_5 before N_4's did).
- N_4 drops N_5's broadcast, N_3 recognizes (N_1, N_2, N_5) as an alternate, but longer route.
- N_3 now has to return the path (N_1, N_2, N_3) to N_1. This is simple assuming symmetric links working in both directions. N_3 can forward the information using the list in reverse order.

The assumption of bi-directional links holds for many ad-hoc networks. However, if links are not bi-directional, the scenario gets more complicated. The algorithm has to be applied again, in the reverse direction if the target does not maintain a current path to the source of the route request.

- N_3 has to broadcast a route request ((N_3), id = 17, target = N_1). Only N_5 receives this request.
- N_5 now broadcasts ((N_3, N_5), id = 17, target = N_1), N_3 and N_4 receive the broadcast.
- N_3 drops the request because it recognizes an already known id. N_4 broadcasts ((N_3, N_5, N_4), id = 17, target = N_1), N_5, N_2, and N_1 receive the broadcast.
- N_5 drops the request packet, N_1 recognizes itself as target, and N_2 broadcasts ((N_3, N_5, N_4, N_2), id = 17, target = N_1). N_3 and N_5 receive N_2's broadcast. ● N_3 and N_5 drop the request packet.

Now N_3 holds the list for a path from N_1 to N_3, (N_1, N_2, N_3), and N_1 knows the path from N_3 to N_1, (N_3, N_5, N_4, N_1). But N_1 still does not know how to send data to N_3! The only solution is to send the list (N_1, N_2, N_3) with the broadcasts initiated by N_3 in the reverse direction. This example shows clearly how much simpler routing can be if links are symmetrical.

The basic algorithm for route discovery can be optimized in many ways.

- To avoid too many broadcasts, each route request could contain a counter. Every node rebroadcasting the request increments the counter by one. Knowing the maximum network diameter (take the number of nodes if nothing else is known), nodes can drop a request if the counter reaches this number. ● A node can cache path fragments from recent requests. These fragments can now be used to answer other route requests much faster.

- A node can also update this cache from packet headers while forwarding other packets.
- If a node overhears transmissions from other nodes, it can also use this information for shortening routes.

After a route has been discovered, it has to be maintained for as long as the node sends packets along this route. Depending on layer two mechanisms, different approaches can be taken:

- If the link layer uses an acknowledgement (as, for example, IEEE 802.11) the node can interpret this acknowledgement as an intact route.
- If possible, the node could also listen to the next node forwarding the packet, so getting a passive acknowledgement.
- A node could request an explicit acknowledgement.

Again, this situation is complicated if links are not bi-directional. If a node detects connectivity problems, it has to inform the sender of a packet, initiating a new route discovery starting from the sender. Alternatively, the node could try to discover a new route by itself.

Although dynamic source routing offers benefits compared to other algorithms by being much more bandwidth efficient, problems arise if the topology is highly dynamic and links are asymmetrical.

Alternative metrics

The examples shown in this chapter typically use the number of hops as routing metric. Although very simple, especially in wireless ad-hoc networks, this is not always the best choice. Even for fixed networks, e.g., bandwidth can also be a factor for the routing metric. Due to the varying link quality and the fact that different transmissions can interfere, other metrics can be more useful.

One other metric, called **least interference routing** (LIR), takes possible interference into account. Figure 2.21 shows an ad-hoc network topology. Sender S_1 wants to send a packet to receiver R_1, S_2 to R_2. Using the hop count as metric, S_1 could choose three different paths with three hops, which is also the minimum. Possible paths are (S_1, N_3, N_4, R_1), (S_1, N_3,

N_2, R_1), and (S_1, N_1, N_2, R_1). S_2 would choose the only available path with only three hops (S_2, N_5, N_6, R_2). Taking interference into account, this picture changes. To calculate the possible interference of a path, each node calculates its possible interference (interference is defined here as the number of neighbors that can overhear a transmission). Every node only needs local information to compute its interference.

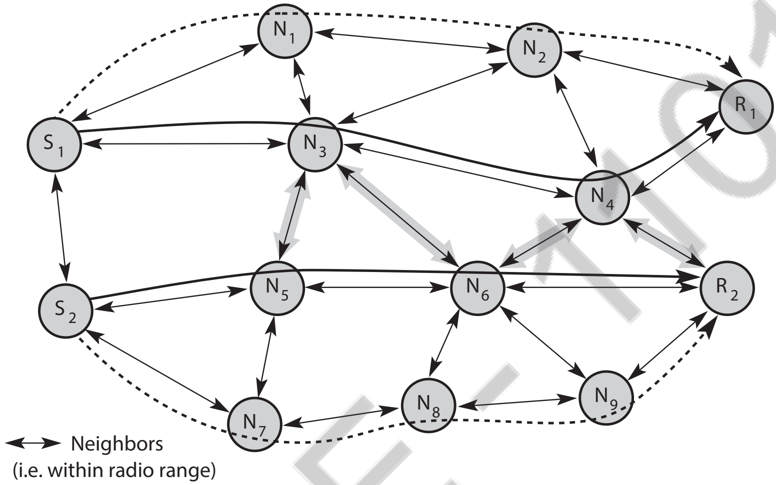


Figure 2.21 Example for least interference routing

In this example, the interference of node N_3 is 6, that of node N_4 is 5 etc. Calculating the costs of possible paths between S_1 and R_1 results in the following:

C1 = cost(S_1, N_3, N_4, R_1) = 16,

C_2 = cost(S_1, N_3, N_2, R_1) = 15,

and C_3 = cost(S_1, N_1, N_2, R_1) = 12.

All three paths have the same number of hops, but the last path has the lowest cost due to interference. Thus, S_1 chooses (S_1, N_1, N_2, R_1). S_2 also computes the cost of different paths, examples are C_4 = cost(S_2, N_5, N_6, R_2) = 16 and C_5 = cost(S_2, N_7, N_8, N_9, R_2) = 15. S_2 would, therefore, choose the path (S_2, N_7, N_8, N_9, R_2), although this path has one hop more than the first one.

With both transmissions taking place simultaneously, there would have been interference between them as shown in Figure 2.21. In this case, least interference routing helped to avoid interference. Taking only local decisions and not knowing what paths other transmissions take, this scheme can just lower the probability of interference. Interference can only be avoided if all senders know of all other transmissions (and the whole routing topology) and base routing on this knowledge.

Routing can take several metrics into account at the same time and weigh them. Metrics could be the number of hops *h*, interference *i*, reliability *r*, error rate *e* etc. The cost of a path could then be determined as:

$$\text{cost} = \alpha h + \beta i + \gamma r + \delta e + \ldots$$

It is not at all easy (if even possible) to choose the weights *α*, *β*, *γ*, *δ*,... to achieve the desired routing behavior.

Overview of ad-hoc routing protocols

Ad-hoc networking has attracted a lot of research over the last few years. This has led to the development of many new routing algorithms. They all come with special pros and cons. It can be divided into three categories: flat routing, hierarchical routing, and geographic-position-assisted routing.

Flat ad-hoc routing

Flat ad-hoc routing protocols comprise those protocols that do not set up hierarchies with clusters of nodes, special nodes acting as the head of a cluster, or different routing algorithms inside or outside certain regions. All nodes in this approach play an equal role in routing. The addressing scheme is flat.

This category again falls into two subcategories: proactive and reactive protocols.

Proactive protocols set up tables required for routing regardless of any traffic that would require routing functionality. DSDV, is a classic member of this group. Many protocols belonging to this group are based on a link-state algorithm as known from fixed networks. Link-state algorithms flood their information about neighbors periodically or event triggered . In mobile ad-hoc environments this method exhibits severe drawbacks: either updating takes place often enough to reflect the actual configuration of the network or it tries to minimize network load. Both goals cannot be achieved at the same time without additional mechanisms. **Fisheye state routing** and **fuzzy sighted link-state** attack this problem by making the update period dependent on the distance to a certain hop. Routing entries corresponding to a faraway destination are propagated with lower frequency than those corresponding to nearby destinations. The result are routing tables that reflect the proximity of a node very precisely, while imprecise entries may exist for nodes further away. Other link-state protocols that try to reduce the traffic caused by link-state information dissemination are **topology broadcast based on reverse path forwarding** and **optimized link-state routing** . A general **advantage** of proactive protocols is that they can give QoS guarantees related to connection set-up, latency or other realtime requirements. As long as the topology does not change too fast, the routing tables reflect the current topology with a certain precision. The propagation characteristics (delay, bandwidth etc.) of a certain path between a sender and a receiver are already known before a data packet is sent. A big **disadvantage** of proactive schemes are their overheads in lightly loaded networks. Independent of any real communication the algorithm continuously updates the routing tables. This generates a lot of unnecessary traffic and drains the batteries of mobile devices.

Reactive protocols try to avoid this problem by setting up a path between sender and receiver only if a communication is waiting. The two most prominent members of this group are **dynamic source routing (DSR)** and **ad-hoc on-demand distance vector** (AODV), an on-demand version of DSDV. AODV acquires and maintains routes only on demand like DSR does. Both protocols, DSR and AODV, are the leading candidates for standardization in the IETF.

A clear **advantage** of on-demand protocols is scalability as long as there is only light traffic and low mobility. Mobile devices can utilize longer low-power periods as they only have to wake up for data transmission or route discovery. However, these protocols also exhibit **disadvantages**. The initial search latency may degrade the performance of interactive applications and the quality of a path is not known *a priori*. Route caching, a mechanism typically employed by on-demand protocols, proves useless in high mobility situations as routes change too frequently.

Hierarchical ad-hoc routing

Algorithms such as DSDV, AODV, and DSR only work for a smaller number of nodes and depend heavily on the mobility of nodes. For larger networks, clustering of nodes and using different routing algorithms between and within clusters can be a scalable and efficient solution. The motivation behind this approach is the locality property, meaning that if a cluster can be established, nodes typically remain within a cluster, only some change clusters. If the topology within a cluster changes, only nodes of the cluster have to be informed. Nodes of other clusters only need to know how to reach the cluster. The approach basically hides all the small details in clusters which are further away.

From time to time each node needs to get some information about the topology. Again, updates from clusters further away will be sent out less frequently compared to local updates. Clusters can be combined to form super clusters etc., building up a larger hierarchy. Using this approach, one or more nodes can act as clusterheads, representing a router for all traffic to/from the cluster. All nodes within the cluster and all other clusterheads use these as gateway for the cluster.

Figure 2.22 shows an ad-hoc network with interconnection to the internet via a base station. This base station transfers data to and from the cluster heads. In this example, one cluster head also acts as head of the super cluster, routing traffic to and from the super cluster. Different routing protocols may be used inside and outside clusters.

Clusterhead-Gateway Switch Routing (CGSR) is a typical representative of hierarchical routing algorithms based on distance vector (DV) routing. Compared to DV protocols, the hierarchy helps to reduce routing tables tremendously. However, it might be difficult to maintain the cluster structure in a highly mobile environment. An algorithm based on the link-state (LS) principle is **hierarchical state routing** . This applies the principle of clustering recursively, creating multiple levels of clusters and clusters of clusters etc. This recursion is also reflected in a hierarchical addressing scheme. A typical hybrid hierarchical routing protocol is the **zone routing protocol** (ZRP). Each node using ZRP has a predefined zone with the node as the center. The zone comprises all other nodes within a certain hop-limit. Proactive routing is applied within the zone, while on-demand routing is used outside the zone.

Due to the established hierarchy, HSR and CGSR force the traffic to go through certain nodes which may be a bottleneck and which may lead to suboptimal paths. Additionally, maintaining clusters or a hierarchy of clusters causes additional overheads. ZRP faces the problem of flat on-demand schemes as soon as the network size increases as many destinations are then outside the zone.

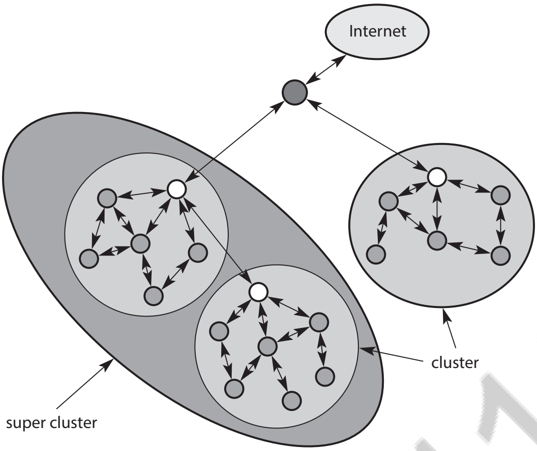


Figure 2.22 Building hierarchies in ad-hoc networks

Geographic-position-assisted ad-hoc routing

If mobile nodes know their geographical position this can be used for routing purposes. This improves the overall performance of routing algorithms if geographical proximity also means radio proximity (which is typically, but not always, the case just think of obstacles between two close-by nodes). One way to acquire position information is via the global positioning system (GPS). Mauve gives an overview of several position-based ad-hoc routing protocols.

GeoCast allows messages to be sent to all nodes in a specific region. This is done using addresses based on geographic information instead of logical numbers. Additionally, a hierarchy of geographical routers can be employed which are responsible for regions of different scale. The **locationaided routing** protocol is similar to DSR, but limits route discovery to certain geographical regions. Another protocol that is based on location information is **greedy perimeter stateless routing** (GPSR). This uses only the location information of neighbors that are exchanged via periodic beacon messages or via piggybacking in data packets. The main scheme of the protocol, which is the greedy part, is quite simple. Packets are always forwarded to the neighbor that is geographically closest to the destination. Additional mechanisms are applied if a dead end is reached (no neighbor is closer to the destination than the node currently holding the data packet to be forwarded).

UNIT - 3

MOBILE TRANSPORT LAYER

PART - A

1. **Define Traditional TCP?**

   The Traditional Control Protocol (TCP) is the most widely used transport protocol in the internet architecture

2. **What are the services of TCP?**

   TCP provides connection-oriented, reliable, byte-stream service that is both flow and congestion control to the upper layers.

3. **What are the algorithms used for congestion control in TCP?**

   The congestion control functionality of TCP is provided by four main algorithms namely Slow start Congestion avoidance Fast transmit Fast recovery

4. **What is slow start mechanism?**

   Slow start is a mechanism used by the sender to control the transmission rate. The sender always calculates a congestion window for a receiver. The start size of the congestion window is one TCP packet.

5. **What is Fast Retransmit algorithm in TCP?**

   During TCP congestion control, when three or more duplicate ACKs are received, the sender does not even wait for a retransmission timer to expire before retransmitting the segment. This process is called the Fast Retransmit Algorithm.

6. **What is Congestion Avoidance algorithm?**

   In the Congestion Avoidance algorithm a retransmission timer expiring or the reception of duplicate ACKs can implicitly signal the sender that a network congestion situation is going on.

   The sender immediately sets its transmission window to one half of the current window size, but to at least two segments. If congestion was indicated by a timeout, the congestion window is reset to one segment, which automatically puts the sender into Slow Start mode.

7. **What are the techniques for classical improvements?**

With the goal of increasing TCPs performance in wireless and mobile environments several scheme were proposed, Some of them are: 1. Indirect TCP 2. Mobile TCP 3. Snooping TCP 4. Fast Transmit/ Fast Recovery 5. Transmission/ time-out freezing 6. Selective Retransmission

8. **Write a short note on I- TCP.**

Indirect TCP is a split connection solution that utilizes the resources of Mobility Support Routers (MSRs) to transport layer communication between mobile hosts and fixed hosts.

It uses the standard TCP for its connection over the wireless hop and like other spit connection protocols, attempts to separate loss recovery over the wireless link from the wired link.

9. **What are the advantages and disadvantages of I – TCP?**

Advantages:

I-TCP does not require any changes in the TCP Protocol

Transmission errors on the wireless link cannot propagate into the fixed network.

Optimizing new mechanisms is quite simple because they only cover one single hop.

Disadvantages:

The loss of the end-to-end semantics of TCP might cause problems if the foreign agent partitioning the connection crashes.

10. **What are the advantages and disadvantages of Mobile TCP?**

Advantages

M-TCP maintains the TCP end-to-end semantics. The Supervisory Host (SH) does not send any ACK itself but forwards the ACKS from the MH. If the MH is detached, it avoids useless transmissions, slow starts or breaking connections by simply shrinking the sender's window to zero.

11. **What is Snooping TCP?**

The main drawback of I-TCP is the segmentation of the single TCP connection into two TCP connections, which losses the original end-to-end TCP semantics. A new enhancement which leaves the TCP intact and is completely transparent, is Snooping TCP. The main function is to buffer data close to the mobile hast to perform fast local retransmission in the case of packet loss.

12. What is time-out freezing?

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

TCP then stops sending and freezes the current state of its congestion window and further timers. When the MAC layer notices the upcoming interruption early enough, both the mobile and correspondent host can be informed.

13. What are the advantages and disadvantages of time out freezing?

Advantages:

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

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

Disadvantages

TCP on mobile host has to be changed, mechanism depends on MAC layer.

Need resynchronization after interruption.

14. What is Selective Retransmission?

TCP acknowledgements are collective. They acknowledge in-order receipt of packets upto certain packets. Even if a single packet is lost, the sender has to retransmit everything starting from the lost packet.

To overcome this problem, TCP can indirectly request a selective retransmission of packets. The receiver may acknowledge single packets and also trains of in-sequence packets.

15. What are the applications of TCP?

Some applications of TCP are www-browsing-mail and FTP

PART - B

1. Explain the various T-TCP (Traditional TCP) in detail.

(OR)

Summarize slow start in TCP operation? Explain its working. How does slow start help improve the performance of TCP?

(OR)

(i) Differentiate traditional TCP with Mobile TCP. (8)

(OR)

(ii) Write short notes on selective retransmission. (8)

Traditional TCP

Congestion control

A transport layer protocol such as TCP has been designed for fixed networks with fixed end-systems. Data transmission takes place using network adapters, fiber optics, copper wires, special hardware for routers etc. This hardware typically works without introducing transmission errors. If the software is mature enough, it will not drop packets or flip bits, so if a packet on its way from a sender to a receiver is lost in a fixed network, it is not because of hardware or software errors. The probable reason for a packet loss in a fixed network is a temporary overload some point in the transmission path, i.e., a state of congestion at a node.

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. Although it is not guaranteed that all packets of the TCP connection take the same way through the network, this assumption holds for most of the packets. 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. This cooperation of TCP connections in the internet is one of the main reasons for its survival as it is today. Using UDP is not a solution, because the throughput is higher compared to a TCP connection just at the beginning. As soon as everyone uses UDP, this advantage disappears. After that, congestion is standard and data transmission quality is unpredictable. Even under heavy load, TCP guarantees at least sharing of the bandwidth.

Slow start

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

The sender always calculates a **congestion window** for a receiver. The start size of the congestion window is one segment (TCP packet). The sender sends one packet and waits for acknowledgement. If this acknowledgement arrives, the sender increases the congestion window by one, now sending two packets (congestion window = 2). After arrival of the two corresponding acknowledgements, the sender again adds 2 to the congestion window, one for each of the acknowledgements. Now the congestion window equals 4. This scheme doubles the congestion window every time the acknowledgements come back, which takes one round trip time (RTT). This is called the exponential growth of the congestion window in the slow start mechanism.

It is too dangerous to double the congestion window each time because the steps might become too large. The exponential growth stops at the **congestion threshold**. As soon as the congestion window reaches the congestion threshold, further increase of the transmission rate is only linear by adding 1 to the congestion window each time the acknowledgements come back.

Linear increase continues until a time-out at the sender occurs due to a missing acknowledgement, or until the sender detects a gap in transmitted data because of continuous acknowledgements for the same packet. In either case the sender sets the congestion threshold to half of the current congestion window. The congestion window itself is set to one segment and the sender starts sending a single segment. The exponential growth (as described above) starts once more up to the new congestion threshold, then the window grows in linear fashion.

Fast retransmit/fast recovery

Two things lead to a reduction of the congestion threshold. One is a sender receiving continuous acknowledgements for the same packet. This informs the sender of two things. One is that the receiver got all packets up to the acknowledged packet in sequence. In TCP, a receiver sends acknowledgements only if it receives any packets from the sender. Receiving acknowledgements from a receiver also shows that the receiver continuously receives something from the sender. The gap in the packet stream is not due to severe 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.**.

The receipt of acknowledgements shows that there is no congestion to justify a slow start. The sender can continue with the current congestion window. The sender performs a **fast recovery** from the packet loss. This mechanism can improve the efficiency of TCP dramatically.

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

Implications on mobility

While slow start is one of the most useful mechanisms in fixed networks, it drastically decreases the efficiency of TCP if used together with mobile receivers or senders. From a missing acknowledgement, 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 retransmissions (ARQ) or error correction (FEC). 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 endsystem. 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 (missing acknowledgement due to a transmission error) is misused for congestion control (missing acknowledgement 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. RFC 3155 states that Explicit congestion notification (ECN) cannot be used as surrogate for explicit transmission error notification. 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 internetoperable. 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.

2. Explain in detail, the various classical TCP improvement mechanisms.

(OR)

(i) With neat diagram, describe snooping TCP and mobile TCP.

(ii) List the various mechanisms of TCP that influence the efficiency of TCP in mobile communication and explain.

(OR)

Identify the transmission mechanism of Indirect TCP and Snooping TCP (16)

(OR)

Can you explain the advancements of Mobile TCP and Snooping TCP (16)

Classical TCP improvements

Indirect TCP

Two competing insights led to the development of indirect TCP (I-TCP) . One is that TCP performs poorly together with wireless links; the other is that TCP within the fixed network cannot be changed. I-TCP segments a TCP connection into a fixed part and a wireless part. Figure 3.1 shows an example with a mobile host connected via a wireless link and an access point to the 'wired' internet where the correspondent host resides. The correspondent node could also use wireless access. The following would then also be applied to the access link of the correspondent host.

Standard TCP is used between the fixed computer and the access point. No computer in the internet recognizes any changes to TCP. Instead of the mobile host, the access point now terminates the standard TCP connection, acting as a proxy. This means that the access point is now seen as the mobile host for the fixed host and as the fixed host for the mobile host. Between the access point and the mobile host, a special TCP, adapted to wireless links, is used. However, changing TCP for the wireless link is not a requirement. Even an unchanged TCP can benefit from the much shorter round trip time, starting retransmission much faster. A good place for segmenting the connection between mobile host and correspondent host is at the foreign agent of mobile IP . The foreign agent controls the mobility of the mobile host anyway and can also hand over the connection to the next foreign agent when the mobile host

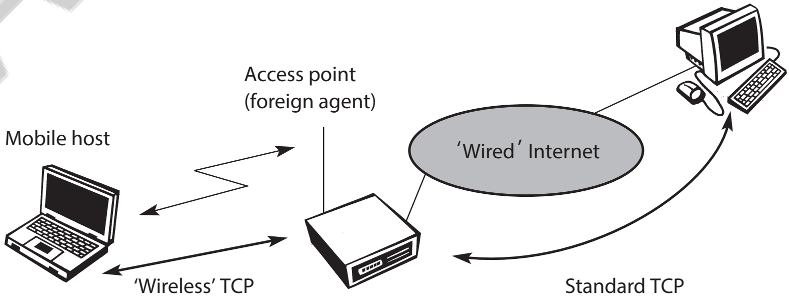


Figure 3.1 Indirect TCP segments a TCP connection into two parts

moves on. However, one can also imagine separating the TCP connections at a special server, e.g., at the entry point to a mobile phone network (e.g., IWF in GSM, GGSN in GPRS).

The correspondent host in the fixed network does not notice the wireless link or the segmentation of the connection. The foreign agent acts as a proxy and relays all data in both directions. If the correspondent host sends a packet, the foreign agent acknowledges this packet and tries to forward the packet to the mobile host. If the mobile host receives the packet, it acknowledges the packet. However, this acknowledgement is only used by the foreign agent. If a packet is lost on the wireless link due to a transmission error, the correspondent host would not notice this. In this case, the foreign agent tries to retransmit this packet locally to maintain reliable data transport.

Similarly, if the mobile host sends a packet, the foreign agent acknowledges this packet and tries to forward it to the correspondent host. If the packet is lost on the wireless link, the mobile hosts notice this much faster due to the lower round trip time and can directly retransmit the packet. Packet loss in the wired network is now handled by the foreign agent.

I-TCP requires several actions as soon as a handover takes place. As Figure 3.2 demonstrates, not only the packets have to be redirected using, e.g., mobile IP. In the example shown, the access point acts as a proxy buffering packets for retrans- mission. After the handover, the old proxy must forward buffered data to the new proxy because it has already acknowledged the data. As explained in chapter 8, after registration with the new foreign agent, this new foreign agent can inform the old one about its location to enable packet forwarding. Besides buffer content, the sockets of the proxy, too, must migrate to the new foreign agent located in the access point. The socket reflects the current state of the TCP con- nection, i.e., sequence number, addresses, ports etc. No new connection may be established for the mobile host, and the correspondent host must not see any changes in connection state.

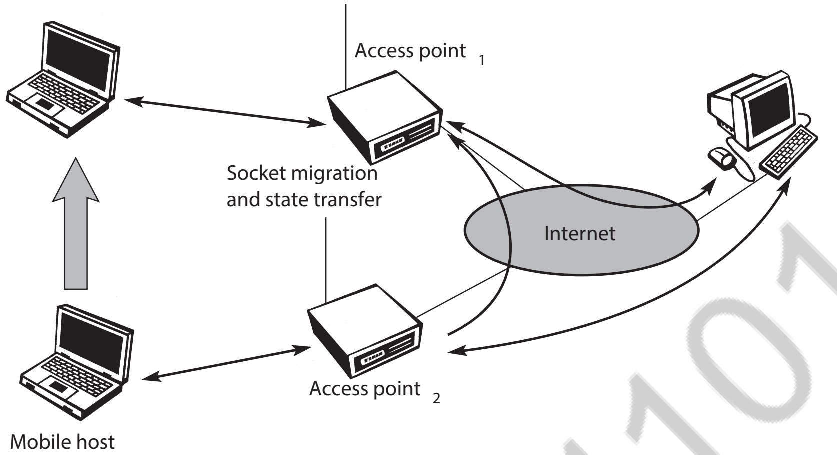


Figure 3.2 Socket and state migration after handover of a mobile host

There are several advantages with I-TCP:

- I-TCP does not require any changes in the TCP protocol as used by the hosts in the fixed network or other hosts in a wireless network that do not use this optimization. All current optimizations for TCP still work between the foreign agent and the correspondent host.

- Due to the strict partitioning into two connections, transmission errors on the wireless link, i.e., lost packets, cannot propagate into the fixed network. Without partitioning, retransmission of lost packets would take place between mobile host and correspondent host across the whole network. Now only packets in sequence, without gaps leave the foreign agent.

- It is always dangerous to introduce new mechanisms into a huge network such as the internet without knowing exactly how they will behave. However, new mechanisms are needed to improve TCP performance (e.g., disabling slow start under certain circumstances), but with I-TCP only between the mobile host and the foreign agent. Different solutions can be tested or used at the same time without jeopardizing the stability of the internet. Furthermore, optimizing of these new mechanisms is quite simple because they only cover one single hop.

- ● The authors assume that the short delay between the mobile host and foreign agent could be determined and was independent of other traffic streams. An optimized TCP could use precise time-outs to guarantee retrans- mission as fast as possible. Even standard TCP could benefit from the short round trip time, so recovering faster from packet loss. Delay is much higher in a typical wide area wireless network than in wired networks due to FEC and MAC. GSM has a delay of up to 100 ms circuit switched, 200 ms and more packet switched (depending on packet size and current traffic). This is even higher than the delay on transatlantic links.

- Partitioning into two connections also allows the use of a different transport layer protocol between the foreign agent and the mobile host or the use of compressed headers etc. The foreign agent can now act as a gateway to translate between the different protocols.

But the idea of segmentation in I-TCP also comes with some **disadvantages**:

- The loss of the end-to-end semantics of TCP might cause problems if the foreign agent partitioning the TCP connection crashes. If a sender receives an acknowledgement, it assumes that the receiver got the packet. Receiving an acknowledgement now only means (for the mobile host and a correspon- dent host) that the foreign agent received the packet. The correspondent node does not know anything about the partitioning, so a crashing access node may also crash applications running on the correspondent node assuming reliable end-to-end delivery.

- In practical use, increased handover latency may be much more problematic. All packets sent by the correspondent host are buffered by the foreign agent besides forwarding them to the mobile host (if the TCP connection is split at the foreign agent). The foreign agent removes a packet from the buffer as soon as the appropriate acknowledgement arrives. If the mobile host now performs a handover to another foreign agent, it takes a while before the old foreign agent can forward the buffered data to the new for eign agent. During this time more packets may arrive. All these packets have to be forwarded to the new foreign agent first, before it can start for- warding the new packets redirected to it.

- The foreign agent must be a trusted entity because the TCP connections end at this point. If users apply end-to-end encryption, e.g., according to RFC 2401 , the foreign agent has to be integrated into all security mechanisms.

Snooping TCP

One of the drawbacks of I-TCP is the segmentation of the single TCP connection into two TCP connections. This loses the original end-to-end TCP semantic. The following TCP enhancement works completely transparently and leaves the TCP end-to-end connection intact. The main function of the enhancement is to buffer data close to the mobile host to perform fast local retransmission in case of packet loss. A good place for the enhancement of TCP could be the foreign agent in the Mobile IP context .

In this approach, the foreign agent buffers all packets with **destination mobile host** and additionally 'snoops' the packet flow in both directions to rec- ognize acknowledgements . The reason for buffering packets toward the mobile node is to enable the foreign agent to per- form a local retransmission in case of packet loss on the wireless link. The foreign agent buffers every packet until it receives an acknowledgement from the mobile host. If the foreign agent does not receive an acknowledgement from the mobile host within a certain amount of time, either the packet or the acknowledgement has been lost. Alternatively, the foreign agent could receive a duplicate ACK which also shows the loss of a packet. Now the foreign agent retransmits the packet directly from the buffer, performing a much faster retransmission compared to the correspondent host. The time out for acknowl- edgements can be much shorter, because it reflects only the delay of one hop plus processing time.

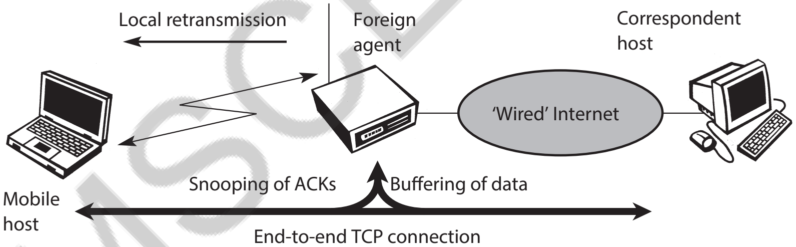


Figure 3.3 Snooping TCP as a transparent TCP extension

To remain transparent, the foreign agent must not acknowledge data to the correspondent host. This would make the correspondent host believe that the mobile host had received the data and would violate the end-to-end semantic in case of a foreign agent failure. However, the foreign agent can filter the dupli- cate acknowledgements to avoid unnecessary retransmissions of data from the correspondent host. If the foreign agent now crashes, the time-out of the corre- spondent host still works and triggers a retransmission. The foreign agent may discard duplicates of packets already retransmitted locally and acknowledged by the mobile host. This avoids unnecessary traffic on the wireless link.

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

Extending the functions of a foreign agent with a ‘snooping’ TCP has several **advantages**:

- The end-to-end TCP semantic is preserved. No matter at what time the for- eign agent crashes (if this is the location of the buffering and snooping mechanisms), neither the correspondent host nor the mobile host have an inconsistent view of the TCP connection as is possible with I-TCP. The approach automatically falls back to standard TCP if the enhancements stop working.
- The correspondent host does not need to be changed; most of the enhance- ments are in the foreign agent. Supporting only the packet stream from the correspondent host to the mobile host does not even require changes in the mobile host.
- It does not need a handover of state as soon as the mobile host moves to another foreign agent. Assume there might still be data in the buffer not transferred to the next foreign agent. All that happens is a time-out at the correspondent host and retransmission of the packets, possibly already to the new care-of address.
- It does not matter if the next foreign agent uses the enhancement or not. If not, the approach automatically falls back to the standard solution. This is one of the problems of I-TCP, since the old foreign agent may have already signaled the correct receipt of data via acknowledgements to the correspon- dent host and now has to transfer these packets to the mobile host via the new foreign agent.

However, the simplicity of the scheme also results in some **disadvantages:**

- Snooping TCP does not isolate the behavior of the wireless link as well as I-TCP. Assume, for example, that it takes some time until the foreign agent can successfully retransmit a packet from its buffer due to problems on the wireless link (congestion, interference). Although the time-out in the foreign agent may be much shorter than the one of the correspondent host, after a while the time-out in the correspondent host triggers a retransmission. The problems on the wireless link are now also visible for the correspondent host and not fully isolated. The quality of the isolation, which snooping TCP offers, strongly depends on the quality of the wireless link, time-out values, and further traffic characteristics. It is problematic that the wireless link exhibits very high delays compared to the wired link due to error correction on layer 2 (factor 10 and more higher). This is similar to I- TCP. If this is the case, the timers in the foreign agent and the correspondent host are almost equal and the approach is almost ineffective.
- Using negative acknowledgements between the foreign agent and the mobile host assumes additional mechanisms on the mobile host. This approach is no longer transparent for arbitrary mobile hosts.
- All efforts for snooping and buffering data may be useless if certain encryption schemes are applied end-to-end between the correspondent host and mobile host. Using IP encapsulation security payload the TCP protocol header will be encrypted - snooping on the sequence numbers will no longer work. Retransmitting data from the for- eign agent may not work because many security schemes prevent replay attacks - retransmitting data from the foreign agent may be misinterpreted as replay. Encrypting end-to-end is the way many applications work so it is not clear how this scheme could be used in the future. If encryption is used above the transport layer (e.g., SSL/TLS) snooping TCP can be used.

Mobile TCP

Dropping packets due to a handover or higher bit error rates is not the only phenomenon of wireless links and mobility - the occurrence of lengthy and/or frequent disconnections is another problem. Quite often mobile users cannot connect at all. One example is islands of wireless LANs inside buildings but no coverage of the whole campus. What happens to standard TCP in the case of disconnection?

A TCP sender tries to retransmit data controlled by a retransmission timer that doubles with each unsuccessful retransmission attempt, up to a maximum of one minute (the initial value depends on the round trip time). This means that the sender tries to retransmit an unacknowledged packet every minute and will give up after 12 retransmissions. What happens if connectivity is back earlier than this? No data is successfully transmitted for a period of one minute! The retransmission time-out is still valid and the sender has to wait. The sender also goes into slow-start because it assumes congestion.

What happens in the case of I-TCP if the mobile is disconnected? The proxy has to buffer more and more data, so the longer the period of disconnection, the more buffer is needed. If a handover follows the disconnection, which is typical, even more state has to be transferred to the new proxy. The snooping approach also suffers from being disconnected. The mobile will not be able to send ACKs so, snooping cannot help in this situation.

The **M-TCP (mobile TCP)**[1] approach has the same goals as I-TCP and snooping TCP: to prevent the sender window from shrinking if bit errors or dis- connection but not congestion cause current problems. M-TCP wants to improve overall throughput, to lower the delay, to maintain end-to-end seman- tics of TCP, and to provide a more efficient handover. Additionally, M-TCP is especially adapted to the problems arising from lengthy or frequent disconnec- tions.

M-TCP splits the TCP connection into two parts as I-TCP does. An unmodi- fied TCP is used on the standard host-**supervisory host (SH)** connection, while an optimized TCP is used on the SH-MH connection. The supervisory host is responsible for exchanging data between both parts similar to the proxy in I- TCP . The M-TCP approach assumes a relatively low bit error rate on the wireless link. Therefore, it does not perform caching/retransmission of data via the SH. If a packet is lost on the wireless link, it has to be retransmitted by the original sender. This maintains the TCP end-to-end semantics.

The SH monitors all packets sent to the MH and ACKs returned from the MH. If the SH does not receive an ACK for some time, it assumes that the MH is disconnected. It then chokes the sender by setting the sender’s window size to 0. Setting the window size to 0 forces the sender to go into **persistent mode**, i.e., the state of the sender will not change no matter how long the receiver is discon- nected. This means that the sender will not try to retransmit data. As soon as the SH (either the old SH or a new SH) detects connectivity again, it reopens the window of the sender to the old value. The sender can continue sending at full speed. This mechanism does not require changes to the sender's TCP.

The wireless side uses an adapted TCP that can recover from packet loss much faster. This modified TCP does not use slow start, thus, M-TCP needs a **bandwidth manager** to implement fair sharing over the wireless link.

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

- It maintains the TCP end-to-end semantics. The SH does not send any ACK itself but forwards the ACKs from the MH.
- If the MH is disconnected, it avoids useless retransmissions, slow starts or breaking connections by simply shrinking the sender's window to 0.
- Since it does not buffer data in the SH as I-TCP does, it is not necessary to forward buffers to a new SH. Lost packets will be automatically retransmit- ted to the new SH.

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

- As the SH does not act as proxy as in I-TCP, packet loss on the wireless link due to bit errors is propagated to the sender. M-TCP assumes low bit error rates, which is not always a valid assumption.
- A modified TCP on the wireless link not only requires modifications to the MH protocol software but also new network elements like the bandwidth manager.

Fast retransmit/fast recovery

As soon as the mobile host registers at a new foreign agent using mobile IP, it starts sending duplicated acknowledgements to correspondent hosts. The proposal is to send three dupli- cates. This forces the corresponding host to go into fast retransmit mode and not to start slow start, i.e., the correspondent host continues to send with the same rate it did before the mobile host moved to another foreign agent.

As the mobile host may also go into slow start after moving to a new for- eign agent, this approach additionally puts the mobile host into fast retransmit. The mobile host retransmits all unacknowledged packets using the current con- gestion window size without going into slow start.

The **advantage** of this approach is its simplicity. Only minor changes in the mobile host's software already result in a performance increase. No foreign agent or correspondent host has to be changed.

The main **disadvantage** of this scheme is the insufficient isolation of packet losses. Forcing fast retransmission increases the efficiency, but retransmitted packets still have to cross the whole network between correspondent host and mobile host. If the handover from one foreign agent to another takes a longer time, the correspondent host will have already started retransmission. The approach focuses on loss due to handover: packet loss due to problems on the wireless link is not considered. This approach requires more cooperation between the mobile IP and TCP layer making it harder to change one without influencing the other.

3. Why is timeout freezing required in case of mobile nodes?What are the modifications done in TCP layer to enforce timeout freezing?

(OR)

Explain the working of Freeze-TCP.

Transmission/time-out freezing

The MAC layer has noticed connection problems, before the connection is actually interrupted. Additionally, the MAC layer knows the real reason for the interruption and does not assume congestion, as TCP would. The MAC layer can inform the TCP layer of an upcoming loss of connection or that the current interruption is not caused by congestion. TCP can now stop sending and 'freezes' the current state of its congestion window and further timers. If the MAC layer notices the upcoming interruption early enough, both the mobile and correspondent host can be informed. With a fast interruption of the wireless link, additional mechanisms in the access point are needed to inform the correspondent host of the reason for interruption. Otherwise, the correspondent host goes into slow start assuming congestion and finally breaks the connection.

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

The **advantage** of this approach is that it offers a way to resume TCP con- nections even after longer interruptions of the connection. It is independent of any other TCP mechanism, such as acknowledgements or sequence numbers, so it can be used together with encrypted data. However, this scheme has some severe **disadvantages**. Not only does the software on the mobile host have to be changed, to be more effective the correspondent host cannot remain unchanged. All mechanisms rely on the capability of the MAC layer to detect future interruptions. Freezing the state of TCP does not help in case of some encryption schemes that use time-dependent random numbers. These schemes need resynchronization after interruption.

4. Write short notes on Selective retransmission.

(OR)

What is the motive behind Selective Retransmission in TCP.

A very useful extension of TCP is the use of selective retransmission. TCP acknowledgements are cumulative, i.e., they acknowledge in-order receipt of packets up to a certain packet. If a single packet is lost, the sender has to retrans- mit everything starting from the lost packet (go-back-n retransmission). This obviously wastes bandwidth, not just in the case of a mobile network, but for any network (particularly those with a high path capacity, i.e., bandwidth- delay-product).

Using RFC 2018 TCP can indirectly request a selective retrans- mission of packets. The receiver can acknowledge single packets, not only trains of in-sequence packets. The sender can now determine precisely which packet is needed and can retransmit it.

The **advantage** of this approach is obvious: a sender retransmits only the lost packets. This lowers bandwidth requirements and is extremely helpful in slow wireless links. The gain in efficiency is not restricted to wireless links and mobile environments. Using selective retransmission is also beneficial in all other networks. However, there might be the minor **disadvantage** of more com- plex software on the receiver side, because now more buffer is necessary to resequence data and to wait for gaps to be filled. But while memory sizes and CPU performance permanently increase, the bandwidth of the air interface remains almost the same. Therefore, the higher complexity is no real disadvan- tage any longer as it was in the early days of TCP.

5. Describe Transaction-oriented TCP. How does the integration of connection establishment,data transmission and connection termination functions help for TCP communications?

(OR)

Explain the features of Transaction-oriented TCP.

Assume an application running on the mobile host that sends a short request to a server from time to time, which responds with a short message. If the application requires reliable transport of the packets, it may use TCP (many applications of this kind use UDP and solve reliability on a higher, application-oriented layer).

Using TCP now requires several packets over the wireless link. First, TCP uses a three-way handshake to establish the connection. At least one additional packet is usually needed for transmission of the request, and requires three more packets to close the connection via a three-way handshake. Assuming connec- tions with a lot of traffic or with a long duration, this overhead is minimal. But in an example of only one data packet, TCP may need seven packets altogether. Figure 3.4 shows an example for the overhead introduced by using TCP over GPRS in a web scenario. Web services are based on HTTP which requires a reli- able transport system. In the internet, TCP is used for this purpose. Before a

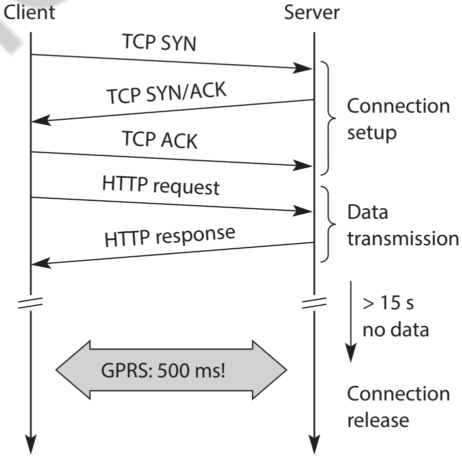


Figure 3.4 Example TCP connection setup overhead

HTTP request can be transmitted the TCP connection has to be established. This already 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 con- nection already takes far more than a second.

This led to the development of a transaction-oriented TCP (T/TCP, RFC 1644 . T/TCP can combine packets for connection establishment and connection release with user data packets. This can reduce the number of packets down to two instead of seven. Similar considerations led to the develop- ment of a transaction service in WAP .

The obvious **advantage** for certain applications is the reduction in the over- head which standard TCP has for connection setup and connection release. However, T/TCP is not the original TCP anymore, so it requires changes in the mobile host and all correspondent hosts, which is a major **disadvantage**. This solution no longer hides mobility. Furthermore, T/TCP exhibits several security problems .

Table 3.1 Overview of

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, security problems
Snooping TCP	Snoops data and acknowledgements, local retransmission	Transparent for end-to-end connection, MAC integration possible	Insufficient isolation of wireless link, security problems
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, security problems
Fast retransmit/ fast recovery	Avoids slow-start after roaming	Simple and efficient	Mixed layers, not transparent
Transmission/ time-out freezing	Freezes TCP state at disconnection, resumes after reconnection	Independent of content, works for longer interruptions	Changes in TCP required, MAC dependent
Selective retransmission	Retransmits only lost data	Very efficient	Slightly more complex receiver software, more buffer space needed
Transaction-oriented TCP	Combines connection setup/release and data transmission	Efficient for certain applications	Changes in TCP required, not transparent, security problems

Table 3.1 shows an overview of the classical mechanisms presented together with some advantages and disadvantages. The approaches are not all exclusive, but can be combined. Selective retransmission, for example, can be used together with the others and can even be applied to fixed networks.

An additional scheme that can be used to reduce TCP overhead is **header compression** (Degermark, 1997). Using tunneling schemes as in mobile IP together with TCP, results in protocol headers of 60 byte in case of IPv4 and 100 byte for IPv6 due to the larger addresses. Many fields in the IP and TCP header remain unchanged for every packet. Only just transmitting the dif- ferences is often sufficient. Especially delay sensitive applications like, e.g., interactive games, which have small packets benefit from small headers. However, header compression experiences difficulties when error rates are high due to the loss of the common context between sender and receiver.

With the new possibilities of wireless wide area networks (WWAN) and their tremendous success, the focus of research has shifted more and more towards these 2.5G/3G networks. Up to now there are no final solutions to the problems arising when TCP is used in WWANs. However, some guidelines do exist.

6. Explain in detail, the TCP over 2.5/3G wireless networks.

(OR)

What conclusion can you draw about TCP over 3G wireless networks.

(OR)

Explain the various issues in 2.5G/3G wireless networks.

The current internet draft for TCP over 2.5G/3G wireless networks describes a profile for optimizing TCP over today's and tomorrow's wireless WANs such as GSM/GPRS, UMTS, or cdma2000. The configuration optimizations recommended in this draft can be found in most of today's TCP implementations so this draft does not require an update of millions of TCP stacks. The focus on 2.5G/3G for transport of internet data is important as already more than 1 billion people use mobile phones and it is obvious that the mobile phone systems will also be used to transport arbitrary internet data.

The following characteristics have to be considered when deploying applications over 2.5G/3G wireless links:

- **Data rates:** While typical data rates of today's 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 down- load 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. To support multiple users within a radio cell, a scheduler may have to repeatedly allocate and deallo- cate resources for each user. This may lead to a periodic allocation and release of a high-speed channel.

- **Latency:** All wireless systems comprise elaborated algorithms for error correction and protection, such as forward error correction (FEC), check summing, and interleaving. FEC and interleaving let the round trip time (RTT) grow to several hundred milliseconds up to some seconds. The current GPRS standard specifies an average delay of less than two seconds for the transport class with the highest quality.

- **Jitter:** Wireless systems suffer from large delay variations or 'delay spikes'. 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 virtually seamless 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).

- **Packet loss:** Packets might be lost during handovers or due to corruption.Thanks to link-level retransmissions the loss rates of 2.5G/3G systems due to corruption are relatively low (but still orders of magnitude higher than, e.g., fiber connections!). However, recovery at the link layer appears as jitter to the higher layers.

Based on these characteristics, suggests the following con- figuration **parameters** to adapt TCP to wireless environments:

- **Large windows:** TCP should support large enough window sizes based on the bandwidth delay product experienced in wireless systems. With the help of the windows scale option (RFC 1323) and larger buffer sizes this can be accomplished (typical buffer size settings of 16 kbyte are not enough). A larger initial window (more than the typical one segment) of 2 to 4 seg- ments may increase performance particularly for short transmissions (a few segments in total).
- **Limited transmit:** This mechanism, defined in RFC 3042 (Allman, 2001) is an extension of Fast Retransmission/Fast Recovery (Caceres, 1995) and is particularly useful when small amounts of data are to be transmitted (stan- dard for, e.g., web service requests).
- **Large MTU:** The larger the MTU (Maximum Transfer Unit) the faster TCP increases the congestion window. Link layers fragment PDUs for trans- mission anyway according to their needs and large MTUs may be used to increase performance. MTU path discovery according to RFC 1191 (IPv4) or RFC 1981 (IPv6) should be used to employ larger segment sizes instead of assuming the small default MTU.
- **Selective Acknowledgement (SACK):** SACK (RFC 2018) allows the selective retransmission of packets and is almost always beneficial compared to the standard cumulative scheme.
- **Explicit Congestion Notification (ECN):** ECN as defined in RFC 3168 (Ramakrishnan, 2001) allows a receiver to inform a sender of congestion in the network by setting the ECN-Echo flag on receiving an IP packet that has experienced congestion. This mechanism makes it easier to distinguish packet loss due to transmission errors from packet loss due to congestion. However, this can only be achieved when ECN capable routers are deployed in the network.
- **Timestamp:** TCP connections with large windows may benefit from more frequent RTT samples provided with timestamps by adapting quicker to changing network conditions. With the help of

timestamps higher delay spikes can be tolerated by TCP without experiencing a spurious timeout. The effect of bandwidth oscillation is also reduced.

- **No header compression:** As the TCP header compression mechanism according to RFC 1144 does not perform well in the presence of packet losses this mechanism should not be used. Header compression according to RFC 2507 or RFC 1144 is not compatible with TCP options such as SACK or timestamps.

It is important to note that although these recommendations are still at the draft-stage, they are already used in i-mode running over FOMA as deployed in Japan and are part of the WAP 2.0 standard (aka TCP with wireless profile).

7. Explain about the Performance enhancing proxies in detail.

RFC 3135 'Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations' lists many proxy architectures that can also be beneficial for wire- less and mobile internet access (Border, 2001). Some initial proxy approaches, such as snooping TCP and indirect TCP have already been discussed. In prin- ciple, proxies can be placed on any layer in a communication system. However, the approaches discussed in RFC 3135 are located in the transport and applica- tion layer. One of the key features of a proxy is its transparency with respect to the end systems, the applications and the users.

Transport layer proxies are typically used for local retransmissions, local acknowledgements, TCP acknowledgement filtering or acknowledgement handling in general. Application level proxies can be used for content filtering, content-aware compression, picture downscaling etc. Prominent examples are internet/WAP gateways making at least some of the standard web content access- ible from WAP devices . Figure 3.5 shows the general architecture of a wireless system connected via a proxy with the internet.

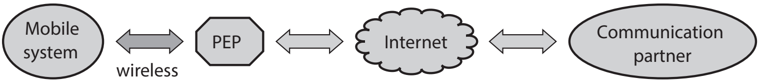


Figure 3.5 Performance enhancing proxy

However, all proxies share a common problem as they break the end-to-end semantics of a connection. According to RFC 3135, the most detrimental negative implication of breaking the end-to-end semantics is that it disables end-to-end use of IP security (RFC 2401). Using IP security with ESP (encapsulation security payload) the major part of the IP packet including the TCP header and applica- tion data is encrypted so is not accessible for a proxy. For any application one has to choose between using a performance enhancing proxy and using IP security. This is a killer criterion in any commercial environment as the only 'solution' would mean the integration of the proxy into the security association between the end systems. Typically this is not feasible as the proxy does not belong to the same organisation as the mobile node and the corresponding node.

RFC 3155 'End-to-end performance implications of **links with errors**' discusses the implications of the use of wireless links for internet access on the performance of TCP (Dawkins, 2001b). Among others, it is stated that it is not possible to use the explicit congestion notification (RFC 2481) as a surrogate for explicit transmission error notification. Such a mechanism is still lacking in the internet.

It is easy to see that it is not easy to adjust TCP behavior according to the current environment. Users may roam between WLANs, 2.5G/3G cellular sys- tems and other wireless/wired technologies. Each technology may exhibit a special behavior which can be classified as 'link with error', 'slow link' etc. Without permanent adaptation, TCP's performance will be poor as will the per- formance of all protocols built on top of TCP (such as HTTP, SOAP). All the problems related to the relatively high connection set-up time due to a three- way handshake still remain if a stream-oriented protocol such as TCP is used in a transaction-oriented manner. Very short lived connections and TCP still do not go together very well.

An unchanged TCP faces even more problems when used over satellite links or in general links to a spacecraft (ranging from an LEO to interplanetary deep- space probes). The main problems are the extremely high RTT, error-prone links, limited link capacity, intermittent connectivity, and asymmetric channels (up to 1,000:1). Asymmetric channels with, for example, a high bandwidth from the spacecraft to ground control, limit throughput due to the limited capacity for the acknowledgements on the return path. (Durst, 1997) presents a set of TCP enhancements, primarily a **selective negative acknowledgement (SNACK)** option, that adapt TCP to the requirements in space communication. The set of protocols developed for space communication is known as **space communica- tions protocol standards (SCPS)**, the extended TCP is called **SCPS-transport protocol (SCPS-TP)**. RFC 2488 (Allman, 1999a) specifies the best current practise for enhancing TCP over satellite channels using standard mechanisms already available in TCP. Choosing the right parameter settings enables TCP to more effectively utilize the available capacity of the network path.

Many questions on the transport layer are still unsolved. Parameters like RTT are difficult to estimate due to high jitter. This influences many timeout values in TCP like the retransmission timer. For an initial estimation of **TCP's performance** the following formulas can be used (Karn, 2002). Both formulas assume long running connections, large enough receiver windows, and Reno TCP according to RFC 2581 (Allman, 1999b). The upper bound on the bandwidth (*BW*) of a TCP connection is given by

$$\mathrm{BW} = \frac{0.93 \cdot \mathrm{MSS}}{\mathrm{RTT} \cdot \sqrt{p}} \text{ (Mathis, 1997).}$$

RTT is the average end-to-end round trip time of the TCP connection. The max- imum segment size (*MSS*) is the segment size being used by the TCP connection. *p* denotes the packet loss probability for the path.

This simple formula neglects retransmissions due to errors. If error rate is above one per cent these retransmissions have to be considered. This leads to a more complicated formula:

$$\mathrm{BW} = \frac{\mathrm{MSS}}{\mathrm{RTT} \cdot \sqrt{1.33p} + \mathrm{RTO} \cdot p \cdot \left(1 + 32 \cdot p^2\right) \cdot \min\left(1, 3\sqrt{0.75p}\right)}$$

This formula also integrates the retransmission timeout (*RTO*), which TCP bases on the RTT. Typically, the simplification *RTO* = 5 *RTT* can be made. For short living connections (less than 10 packets) TCP performance is completely driven by the TCP slow start algorithm without additional enhancements.

UNIT - 4
WIRELESS WIDE AREA NETWORK

PART – A (2 MARKS)

1. **What are the applications of 3G?**
   Applications for a 3G wireless network range from simple voice-only communications to simultaneous video, data, voice and other multimedia applications.

2. **Name some of the wireless technology services?**
   Some of the wireless technology services are General Packet Radio Services (GPRS) Enhanced Data for GSM evolution (EDGE) service Wideband Code Division Multiple access (WCDMA) Universal Mobile telecommunications Services (UMTS) High-Speed Downlink Packet Access (HSDPA)

3. **What is UMTS?**
   Universal Mobile telecommunications Services (UMTS) is a new radio access network based on 5 MHz WCDMA and optimized for efficient support of 3G services. UMTS can be used in both new and existing spectra.

4. **What are the layers of UMTS?**
   The UMTS terrestrial radio access network (UTRAN) has an access layer and non access layer. The access layer includes air interface and provides functions related to OSI layer 1, layer 2, and the lower part of layer 3. The non-access layer deals with communication between user equipment (UE) and core network (CN) and includes OSI layer 3 (upper part) to layer 7.

5. **What is radio resource control (RRC)?**
   The radio resource control (RRC) layer broadcasts system information, handles radio resources such as code allocation, handover, admission control, measurement/control report.

6. **What are the duties of Radio network control (RNC)?**
   The major duties of RNC are Intra UTRAN handover Macro diversity combining/ splitting of Iub data systems. Outer loop power control IU interface user plane setup Serving RNS (SRNS) relocation Radio resource allocation.

7. **What are the planes of UTRAN?**
   There are three planes Control plane User plane Transport network control plane.

8. **What are the functions provided by 3G-MSC?**
   The following functionality is provided by the 3G-MSC: Mobility management Call management Supplementary services Short message services (SMS) OAM (operation, administration, and maintenance) agent functionality.

9. **What is Transport Network Control Plane (TNCP)?**
   Transport Network Control Plane (TNCP) carries information for the control of transport network used within UCN.

10. **What is 3G-SGSN?**
    The 3G-SGSN (serving GPRS Support Node) provides the appropriate signaling and data interface that includes connection to an IP-based network toward the 3G-GGSN, SS7 towards the HLR/EIR/AUC and TCP/IP or SS7 toward the UTRAN.

11. **What is 3G-GGSN?**
    The GGSN (Gateway GPRS Support Node) is connected with SGSN via an IP-based network. It provides interworking with the external PS network.

12. **What are the functions provided by 3G-GGSN?**
    Macro-Mobility (maintaining information locations at SGSN level) Gateway between UMTS packet network and external data networks Initiate mobile terminate route Mobile Terminated Packets User data screening/security.

13. **What is SMS-GMSC?**
    The SMS-GMSC (gateway MSC) is a MSC which can receive a terminated short message.

PART - B

1. Explain the evolution of 3G Networks?

Third-generation (3G) wireless systems [2,3,9] offer access to services anywhere from a single terminal; the old boundaries between telephony, information, and entertainment services are disappearing. Mobility is built into many of the services currently considered as fixed, especially in such areas as high speed access to the Internet, entertainment, information, and electronic commerce (e-commerce) services.

The distinction between the range of services offered via wireline or wireless is becoming less and less clear and, as the evolution toward 3G mobile services speeds up, these distinctions will disappear in the first decade of the new millennium.

Applications for a 3G wireless network range from simple voice-only communications to simultaneous video, data, voice, and other multimedia applications.

One of the main benefits of 3G is that it allows a broad range of wireless services to be provided efficiently to many different users.

Packet-based Internet Protocol (IP) technology is at the core of the 3G services. Users have continuous access to on-line information. E-mail messages arrive at hand-held terminals nearly instantaneously and business users are able to stay permanently connected to the company intranet. Wireless users are able to make video conference calls to the office and surf the Internet simultaneously, or play computer games interactively with friends in other locations.

Figure shows the data rate requirement for various services.

In 1997, the TIA/EIA IS-136 community through the Universal Wireless Consortium (UWC) and the Telecommunications Industry Association (TIA) TR 45.3 adopted a three-part strategy for evolving its IS-136 TDMA-based networks to 3G wireless networks to satisfy International Mobile Telephony-2000 (IMT-2000) requirements.

The strategy consists of Enhancing the voice and data capabilities of the existing 30 kHz carrier (IS-136_) Adding a 200 kHz carrier for high-speed data (384 kbps) in high mobility applications Introducing a 1.6 MHz carrier for very high-speed data (2 Mbps) in low-mobility applications.

The highlight of UWC strategy was the global convergence of IS-136 time division multiple access (TDMA) with a global system for mobile communications (GSM) through the evolution of the 200 kHz GSM carrier for supporting high-speed data applications (384 kbps) while also improving a 30 kHz carrier for voice and mid-speed data applications.

Two 3G radio access schemes are identified to support the different spectrum scenarios:

1. Enhanced data rates for GSM evolution (EDGE) with high-level modulation in a 200 kHz TDMA channel is based on plug-in transceiver equipment, thereby allowing the migration of existing bands in small spectrum segments.
2. Universal mobile telecommunications services (UMTS) is a new radio access network based on 5 MHz WCDMA and optimized for efficient support of 3G services. UMTS can be used in both new and existing spectra.

From a network point of view, 3G capabilities implies the addition of packet switched (PS) services, Internet access, and IP connectivity. With this approach, the existing mobile networks reuse the elements of mobility support, user authentication service handling, and circuit switched (CS) services. With packet switched services, IP connectivity can then be added to provide a mobile multimedia core network by evolving the existing mobile network.

GSM is moving to develop enhanced cutting-edge, customer-focused solutions to meet the challenges of the new millennium and 3G mobile services

When GSM was first introduced, no one could have predicted the dramatic growth of the Internet and the rising demand for multimedia services. These developments have brought about new challenges to the world of GSM. For GSM operators, the emphasis is now rapidly changing from that of instigating and driving the development of technology to fundamentally enabling mobile data transmission to that of improving speed, quality, simplicity, coverage, and reliability in terms of tools and services that will boost mass market take-up.

Users are increasingly looking to gain access to information and services wherever they are and whenever they want. GSM should provide that connectivity.

Internet access, web browsing and the whole range of mobile multimedia capability are the major drivers for development of higher data speed technologies.

Current data traffic on most GSM networks is modest, less than 5% of total GSM traffic. But with the new initiatives coming to fruition during the course of the next two to three years, exponential growth in data traffic is forecast. The use of messaging-based applications may reach up to about 90% by the year 2008. GSM data transmission using high-speed circuit switched data (HSCSD) and GPRS may reach a penetration of about 80% by 2008

GSM operators have two nonexclusive options for evolving their networks to 3G wideband multimedia operation:

(1) using GPRS and EDGE in the existing radio spectrum, and in small amounts of the new spectrum; or

(2) Using WCDMA in the new 2 GHz bands, or in large amounts of the existing spectrum.

Both approaches offer a high degree of investment flexibility because roll-out can proceed in line with market demand with the extensive reuse of existing network equipment and radio sites.

In the new 2 GHz bands, 3G capabilities are delivered using a new wideband radio interface that offers much higher user data rates than are available today — 384 kbps in the wide area and up to 2 Mbps in the local area. Of equal importance for such services is the high-speed packet switching provided by GPRS and its connection to public and private IP networks.

GSM and digital (D)AMPS (IS-136) operators can use existing radio bands to deliver some of the 3G services, even without the new wideband spectrum by evolving current networks and deploying GPRS and EDGE technologies.

In the early years of 3G service deployment, a large proportion of wireless traffic will still be voice-only and low-rate data. So whatever the ultimate capabilities of 3G networks, efficient and profitable ways of delivering more basic wireless services are still needed.

The significance of EDGE for today's GSM operators is that it increases data rates up to 384 kbps and potentially even higher in a good quality radio environment using current GSM spectrum and carrier structures more efficiently. EDGE is both a complement and an alternative to new WCDMA coverage. EDGE also has the effect of unifying the GSM, D-AMPS and WCDMA services through the use of dual-mode terminals.

2. Explain General packet radio system?

The general packet radio service (GPRS) [6,7] enhances GSM data services significantly by providing end-to-end packet switched data connections. This is particularly efficient in Internet/intranet traffic, where short bursts of intense data communications are actively interspersed with relatively long periods of inactivity.

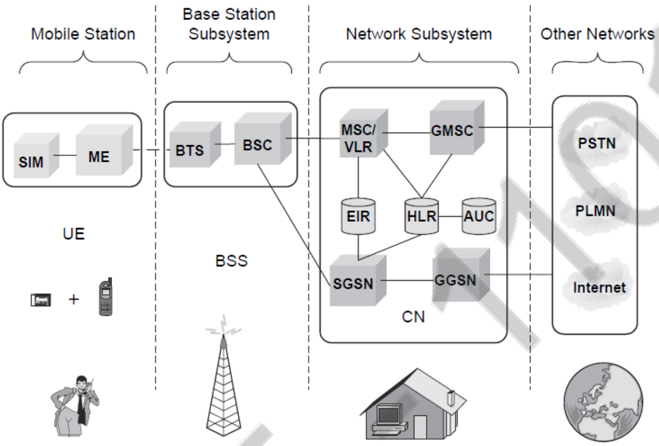
Because there is no real end-to-end connection to be established, setting up a GPRS call is almost instantaneous and users can be continuously on-line. Users have the additional benefits of paying for the actual data transmitted, rather than for connection time. Because GPRS does not require any dedicated end-to-end connection, it only uses network resources and bandwidth when data is actually being transmitted. This means that a given amount of radio bandwidth can be shared efficiently among many users simultaneously.

The next phase in the high-speed road map is the evolution of current short message service (SMS), such as smart messaging and unstructured supplementary service data (USSD), toward the new GPRS, a packet data service using TCP/IP and X.25 to offer speeds up to 115 kbps. GPRS has been standardized to optimally support a wide range of applications ranging from very frequent transmissions of medium to large data volume. Services of GPRS have been developed to reduce connection set-up time and allow an optimum usage of radio resources.

GPRS provides a packet data service for GSM where time slots on the air interface can be assigned to GPRS over which packet data from several mobile stations is multiplexed.

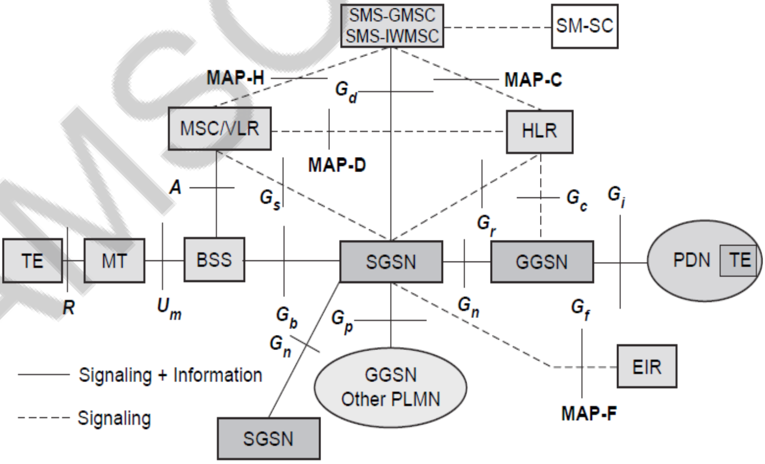
GPRS provides a core network platform for current GSM operators not only to expand the wireless data market in preparation for the introduction of 3G services, but also a platform on which to build IMT-2000 frequencies should they acquire them.

The implementation of GPRS has a limited impact on the GSM core network. It simply requires the addition of new packet data switching and gateway nodes, and an upgrade to existing nodes to provide a routing path for packet data between the wireless terminal and a gateway node. The gateway node provides interworking with external packet data networks for access to the Internet, intranet, and databases.



SGSN: Serving GPRS Support Node
GGSN: Gateway GPRS Support Node
HLR: Home Location Register
VLR: Visitor Location Register

ARICHITECTURE OF GPRS IN GSM



SGSN: Serving GPRS Support Node
GGSN: Gateway GPRS Support Node
MAP: Mobile Application Part
HLR: Home Location Register
VLR: Visitor Location Register
MSC: Mobile Switching Center

protocols, including IP, so it is possible to connect with any data source from anywhere in the world using a GPRS mobile terminal. GPRS supports applications ranging from low-speed short messages to high-speed corporate LAN communications.

However, one of the key benefits of GPRS — that it is connected through the existing GSM air interface modulation scheme — is also a limitation, restricting its potential for delivering higher data rates than 115 kbps. To build even higher rate data capabilities into GSM, a new modulation scheme is needed.

GPRS can be implemented in the existing GSM systems. Changes are required in an existing GSM network to introduce GPRS. The base station subsystem (BSS) consists of a base station controller (BSC) and packet control unit (PCU). The PCU supports all GPRS protocols for communication over the air interface. Its function is to set up, supervise, and disconnect packet switched calls. The packet control unit supports cell change, radio resource configuration, and channel assignment. The base station transceiver (BTS) is a relay station without protocol functions. It performs modulation and demodulation.

The GPRS standard introduces two new nodes, the serving GPRS support node (SGSN) and the gateway GPRS support node (GGSN). The home location register (HLR) is enhanced with GPRS subscriber data and routing information.

Two types of services are provided by GPRS:

Point-to-point (PTP)

Point-to-multipoint (PTM)

Independent packet routing and transfer within the public land mobile network (PLMN) is supported by a new logical network node called the GPRS support node (GSN). The GGSN acts as a logical interface to external packet data networks. Within the GPRS networks, protocol data units (PDUs) are encapsulated at the originating GSN and decapsulated at the destination GSN. In between the GSNs, IP is used as the backbone to transfer PDUs. This whole process is referred to as tunnelling in GPRS. The GGSN also maintains routing information used to tunnel the PDUs to the SGSN that is currently serving the mobile station (MS).

All GPRS user related data required by the SGSN to perform the routing and datatransfer functionality is stored within the HLR. In GPRS, a user may have multiple data sessions in operation at one time. These sessions are called packet data protocol (PDP) contexts. The number of PDP contexts that are open for a user is only limited by the user's subscription and any operational constraints of the network.

The main goal of the GPRS-136 architecture is to integrate IS-136 and GSM GPRS as much as possible with minimum changes to both technologies. In order to provide subscription roaming between GPRS-136 and GSM GPRS networks, a separate functional GSM GPRS HLR is incorporated into the architecture in addition to the IS-41 HLR.

The European Telecommunication Standards Institute (ETSI) has specifi ed GPRS as an overlay to the existing GSM network to provide packet data services. In order to operate a GPRS over a GSM network, new functionality has been introduced into existing GSM network elements (NEs) and new NEs are integrated into the existing service provider's GSM network.

The BSS of GSM is upgraded to support GPRS over the air interface. The BSS works with the GPRS backbone system (GBS) to provide GPRS service in a similar manner to its interaction with the switching subsystem for the circuit-switched services. The GBS manages the GPRS sessions set up between the mobile terminal and the network by providing functions such as admission control, mobility management (MM), and service management (SM). Subscriber and equipment information is shared between GPRS and the switched functions of GSM by the use of a common HLR and coordination of data between the visitor location register (VLR) and the GPRS support nodes of the GBS.

The GBS is composed of two new NEs — the SGSN and the GGSN. The SGSN serves the mobile and performs security and access control functions.The SGSN is connected to the BSS via frame-relay. The SGSN provides packet routing, mobility management, authentication, and ciphering to and from all GPRS subscribers located in the SGSN service area. A GPRS subscriber may be served by any SGSN in the network, depending on location. The traffic is routed from the SGSN to the BSC and to the mobile terminal via a BTS. At GPRS attach, the SGSN establishes a mobility management context containing information

about mobility and security for the mobile. At PDP context activation, the

SGSN establishes a PDP context which is used for routing purposes with the GGSN that the GPRS subscriber uses. The SGSN may send in some cases location information to the MSC/VLR and receive paging requests.

The GGSN provides the gateway to the external IP network, handling security and accounting functions as well as the dynamic allocation of IP addresses.

The GGSN contains routing information for the attached GPRS users. The routing information is used to tunnel PDUs to the mobile's current point of attachment, SGSN. The GGSN may be connected with the HLR via optional interface Gc. The GGSN is the first point of public data network (PDN) interconnection with a GSM PLMN supporting GPRS. From the external IP network's point of view, the GGSN is a host that owns all IP addresses of all subscribers served by the GPRS network.

The PTM-SC handles PTM traffic between the GPRS backbone and the HLR. The nodes are connected by an IP backbone network. The SGSN and GGSN functions may be combined in the same physical node or separated — even residing in different mobile networks.

3. Explain in detail about UMTS NETWORK?

A UMTS system can be divided into a set of domains and the reference points that interconnect them. A simplified mapping of functional entities to the domain model is shown in Figure 15.14. Note that this is a reference model and does not represent any physical architecture. The Iu is split functionally into two logical interfaces, Iups connecting the packet switched domain to the access network and the Iucs connecting the circuit switched domain to the access network. The standards do not dictate that these are physically separate, but the user plane for each is different and the control plane may be different. The Iur logically connects radio network controllers (RNCs) but could be physically realized by a direct connection between RNCs or via the core network

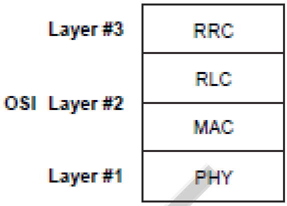
The UMTS terrestrial radio access network (UTRAN) [10–22] has an access stratum and nonaccess stratum. The access stratum includes air interface and provides functions related to OSI layer 1, layer 2, and the lower part of layer 3. The non-access stratum deals with communication between user equipment (UE) and core network (CN) and includes OSI layer 3 (upper part) to layer 7.

The radio interface, Uu, is the interface between UE and UTRAN. It consists of three protocol layers: physical layer, data link layer, and network layer .The radio interface provides physical channels to carry data over the radio path and logical channels to carry a particular type of data. There are two types of logical channels: signaling and control, and traffic channel.

The physical layer in UTRAN performs the following functions:

Forward error correction, bit-interleaving, and rate matching

Signal measurements Micro-diversity distribution/combining and soft handoff execution Multiplexing/mapping of services on dedicated physical codes



Layer #1:
- Physical Layer

Layer #2:
- Upper part: Radio Link Control (RLC)
- Lower part: Medium Access Control (MAC)

Layer #3:
- Lower part: Radio Resource Control (RRC)

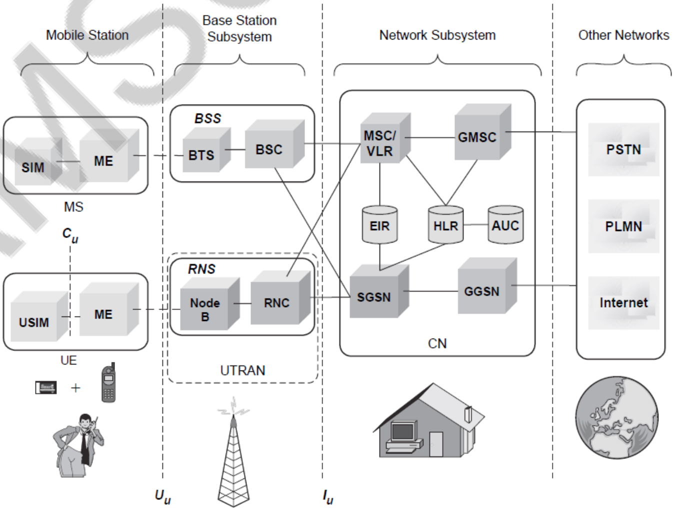
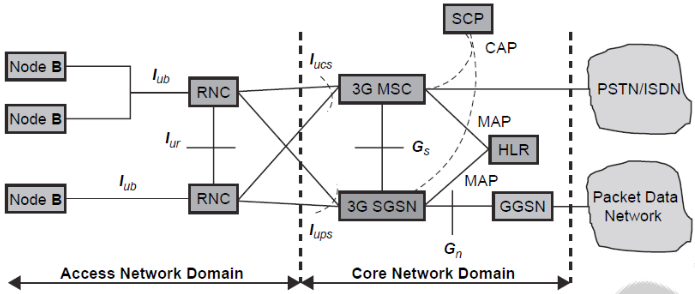


Figure 15.13 UMTS—3G reference architecture.



RNC: Radio Link Control
Node B: Radio Base Station
SCP: Signal Control Point
HLR: Home Location Register
MAP: Mobile Application Part
CAP: CAMEL Application Part

Modulation, spreading, demodulation, despreading of physical channels Frequency and time (chip, bit, slot, frame) synchronization

Fast closed-loop power control

Power weighting and combining of physical channels

Radio frequency (RF) processing.

The medium access control sublayer is responsible for efficiently transferring data for both real-time (CS) and non-real-time (PS) services to the physical layer. MAC offers services to the radio link control (RLC) sublayer and higher layers.

The MAC layer provides data transfer services on logical channels. MAC is responsible for Selection of appropriate transport format (basically bit rate) within a predefined set, per information unit delivered to the physical layer

Service multiplexing on random access channel (RACH), forward access channel (FACH), and dedicated channel (DCH)

Priority handling between data flow of a user as well as between data flows from several users

Access control on RACH and FACH

Contention resolution on RACH

Radio link control (RLC) sets up a logical link over the radio interface and is responsible for fulfilling QoS requirements. RLC responsibilities include:

Segmentation and assembly of the packet data unit

Transfer of user data

Error correction through retransmission

Sequence integrity

Duplication information detection

Flow control of data

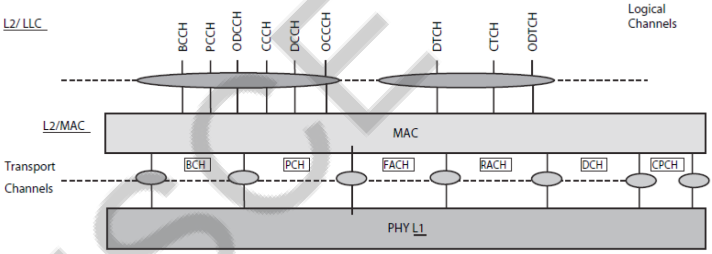
The radio resource control (RRC) layer broadcasts system information,

handles radio resources (i.e., code allocation, handover, admission control, and measurement/control report), and controls the requested QoS. The RRC layer offers the following services to the core network:

General control (GC) service used as an information broadcast service Notification (Nt) service used for paging and notification of a selected UE

Dedicated control (DC) service used to establish/release a connections and transfer messages

The channels in UTRAN are physical, transport, and logical



- Physical channels: identified by code, frequency, phase (*I*/*Q*) (UL only) and time slot (TDD only)
- Transport channels: offer information transfer services to the MAC layer
- Logical channels: are used by MAC layer to provide data transport services

BCCH: Broadcast Control Channel PCCH: Paging Control Channel DCCH: Dedicated Control Channel
ODCCH: ODMA Dedicated Control Channel OCCCH: ODMA Common Control Channel DTCH: Dedicated Traffic Channel
ODTCH: ODMA Dedicated Traffic Channel CTCH: Common Traffic Channel CCCH: Common Control Channel

The functions of logical control channels and logical traffic channels in UTRAN are listed in Tables.

In UTRAN transport channels can be either common (i.e., shared between users) or dedicated channels. They offer information transfer services to the MAC sublayer. Dedicated transport channels are dedicated channel (DCH) (up link, UL and down link, DL), fast uplink signaling channel (FAUSCH), and ODMA (opportunity driven multiple access) dedicated channel (ODCH).

The common DL transport channels are listed

The mapping between logical and transport channels is given below

BCCH is connected to BCH

PCCH is connected to PCH

CCCH is connected to RACH and FACH

Table: Logical control channel in UTRAN.

Channel	Function
Broadcast control channel (BCCH)	DL channel for broadcasting system and control information
Paging control channel (PCCH)	DL channel to transfer page information, used when: (1) network does not know the location of cell of the mobile, and (2) mobile is in cell connected state (using sleep mode)
Common control channel (CCCH)	Bidirectional channel to transfer control information between network and mobile, it is used: (1) by mobile without RRC connection with the network, and (2) by mobile using common transport channel to access a new cell after cell resection
Dedicated control channel (DCCH)	Point-to-point bidirectional channel to transmit dedicated information between a mobile and network. The channel is established through RRC connection setup procedure
ODMA common control channel (OCCCH)	Bidirectional channel to transmit control information between mobiles
ODMA dedicated control channel (ODCCH)	Point-to-multipoint bidirectional channel to transmit dedicated control information between mobiles. This channel is established through RRC connection setup procedure

Table: Logical control channel in UTRAN.

Channel	Function
Broadcast channel (BCH)	DL channel used to broadcast system- and cell-specific information, transmitted over the entire cell with low fixed bit rate
Forward access channel (FACH)	DL channel transmitted over the entire or only a part of cell using beam-forming antennas, uses slow power control
Paging channel (PCH)	DL channel transmitted over the entire cell, transmission of PCH is associated with the transmission of a physical layer signal, the paging indicator, to support efficient sleep mode procedure
Random access channel (RACH)	UL channel received over the entire cell, characterized by a limited size data field, a collision risk, and by use of open loop power control
Common packet channel (CPCH)	UL channel, contention-based random access channel used for transmission of bursty data traffic, associated with a DCH on DL, which provides power control for the UL CPCH
Downlink shared channel (DSCH)	DL channel shared by several mobiles, associated with a DCH

DTCH can be connected to either RACH and FACH and DSCH, to DCH and DSCH, to a DCH, to a CPCH (FDD mode only) CTCH can be connected to DSCH, FACH, or BCH DCCH can be connected to either RACH and FACH, to RACH and DSCH, to DCH and DSCH, to a DCH, a CPCH (TDD mode only), to FAUSCH, CPCH (FDD mode only).

In UTRAN, the basic physical resource is a physical channel identified by code and frequency. Physical channels consist of radio frames and time slots (). The length of a radio frame is 10 ms and one frame consists of 15 time slots. For DL channels two codes are used, one to identify the cell and the other to identify a particular channel within that cell. For UL a long code is used to identify the channel. The UL channel uses different data streams transmitted on the I and Q branch. A physical channel corresponds to a specific carrier frequency, code(s), and on UL a relative phase (0, _/2).

The UL dedicated physical channel (DPCH) is a user dedicated, point-topoint channel between UE and node B. These channels carry dedicated channels at various rates up to 2 Mbps. The UL-dedicated physical data channels are I/Q (i.e., DPDCH on I-branch and DPCCH on Q-branch) and code multiplexed.

There are two types of DPCH:

(1) Dedicated physical data channel (DPDCH) to carry user data and signaling information generated at layer 2 (there may be none, one, or several DPDCHs); and

(2) Dedicated physical control channel (DPCCH) to carry control information generated at layer 1 (pilot bits, transmit power control (TPC) commands, feedback information (FBI) commands, and optional transport format combination indicator (TFCI)). For each layer 1 connection, there is only one UL DPCCH. DPCCH rate and power remain constant. The UL dedicated physical channel carries 10 _ 2k (k _ 0, 1, . . . , 6) bits per slot and may have a spreading factor (SF) from 256 and 4.

The UL common physical channels are physical random access channel (PRACH) used to carry the RACH and fast uplink signaling channel (FAUSCH) and physical common packet channel (PCPCH) to carry CPCH.

The DL dedicated physical data channel is time multiplexed. The DL dedicated physical channel carries 10 _ 2k (k _ 0, 1, . . . , 7) bits per slot and may have spreading factor (SF _ 512/2k) from 512 to 4 (see Figure 15.18).

The DL common physical channels include the following channels:

Primary common control physical channel (PCCPCH) carries BCH, rate 30 kbps, SF _ 256; continuous transmission; no power control Secondary common control physical channel (SCCPCH) carries FACH and PCH, transmitted when data is available; SF range is from 256 to 4 Synchronization channel (SCH) is used for cell search. It consists of two sub-channels: primary SCH transmits a modulated code of 256 chips once every slot, and secondary SCH transmits repeatedly 15 codes of 256 chips.

4. Explain in detail on High Speed Downlink Packet Access(HSDPA)

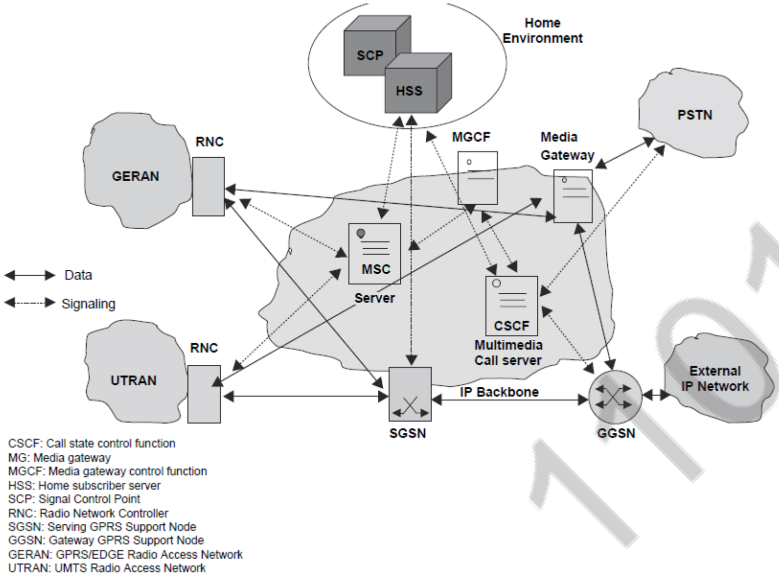
HSDPA is based on the same set of technologies as high data rate (HDR) to improve spectral efficiency for data services — such as shared downlink packet data channel and high peak data rates — using high-order modulation and adaptive modulation and coding, hybrid ARQ (HARQ) retransmission schemes, fast scheduling and shorter frame sizes.

HSDPA marks a similar boost for WCDMA that EDGE does for GSM. It provides a two-fold increase in air interface capacity and a five-fold increase in data speeds in the downlink direction. HSDPA also shortens the round-trip time between the network and terminals and reduces variance in downlink transmission delay.

The improvements in performance are achieved by:

Bringing some key functions, such as scheduling of data packet transmission and processing of retransmissions (in case of transmission errors) into the base station — that is, closer to the air interface.

Using a short frame length to further accelerate packet scheduling for transmission. Employing incremental redundancy for minimizing the air-interface load caused by retransmissions.


Home
Environment
SCP
HSS
PSTN
RNC
GERAN
MGCF
Media
Gateway
Data
Signaling
MSC
Server
CSCF
Multimedia
Call server
RNC
UTRAN
IP Backbone
External
IP Network
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

Adopting a new transport channel type, known as high-speed downlink shared channel (HS-DSCH) to facilitate air interface channel sharing between several users.

Adapting the modulation and coding scheme according to the quality of theradio link The primary objective behind HSDPA is to provide a cost-effective, highbandwidth, low-delay, packet-oriented service within UMTS. Backward compatibility is critical, so the HSDPA architecture adheres to an evolutionary philosophy.

From an architectural perspective, HSDPA is a straightforward enhancement of 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 be supported in a network supporting HSDPA, since HSDPA mobile terminals (UEs) are designed to coexist with R99 UEs.

HSDPA is particularly suited to extremely asymmetrical data services, which require significantly higher data rates for the transmission from the network to the UE, than they do for the transmission from the UE to the network.

HSDPA introduces enablers for the high-speed transmission at the physical layer like the use of a shorter transmission time interval (TTI) (2 ms), and the use of adaptive modulation and coding. HS-DPCCH is used to carry the acknowledgment signals 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 to reduce trip time, to increase the granularity in the scheduling process, and to track the time varying radio channel better.

The basic operational principles behind HSDPA are relatively simple. The RNC 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 mobile terminal over the air interface by matching the user's priority and estimated channel operating environment with an appropriately chosen coding and modulation scheme (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, and so 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 data packet was lost and retransmits it. HSDPA continuously strives, with some modest constraints, to give the maximal bandwidth to the user with the best channel conditions. The data rates achievable with HSDPA are more than adequate for supporting multimediastreaming services

Although conceptually simple, HSDPA's implementation within the context of 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 radio tower and the power cabinet. For indoor deployments the radio cabinet may be a simple rack, while in outdoor deployments it may be an environmental-control unit. The guts of the radio cabinet are an antenna interface section (filters, power amplifiers, and the like), core processing chassis (RF transceivers, combiner, highperformance channel cards, network interface and system controller card, timing card, back-plane, and so on), plus mechatronics (power supply, fans, cables, etc.) and other miscellaneous elements. The core processing chassis is the cornerstone of 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 be modified to support HSDPA.

The typical UMTS channel card comprises a general-purpose processor that handles the miscellaneous control tasks, a pool of digital signal processor (DSP) resources to handle symbol-rate processing and chip-rate assist functions, and a pool of specialized ASIC (application specific intergrated circuit) devices to handle intensive chip-rate operations such as spreading, scrambling, modulation, rake receiving, 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 modulation schemes and new downlink slot formats associated with HSDPA. In addition, the downlink symbol-rate processing section must be modified to support HSDPA 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 the RNC. This is the most intrusive augmentation to the channel card because it requires the introduction of a programmable processing entity together with a retransmission buffer. Since the channel card already contains both a general-purpose processor and a DSP, one can make convincing arguments that the MAC-hs could be effectively realized using either of the two types of devices. Nonetheless, many designers are finding that, because of the close ties between the MAC-hs function and the lower-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 75 or so users.

The new channels introduced in HSDPA are high-speed downlink shared channel (HS-DSCH), high-speed shared control channel (HS-SCCH), and highspeed dedicated physical control channel (HS-DPCCH). The HS-DSCH is the primary radio bearer. Its resources can be shared among all users in a particular sector. The primary channel multiplexing occurs in a time domain, where each TTI 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 code multiplexing, with a maximum of 15 parallel codes allocated to HS-DSCH. Codes may all be assigned to one user, or may be split across several users. The number of codes allocated to each user depends on cell loading, QoS requirements, and UE code capabilities (5, 10, or 15 codes).

5. Explain about different 3G Networks?

3G-MSC

The 3G-MSC is the main CN element to provide CS services. The 3G-MSC also provides the necessary control and corresponding signaling interfaces including SS7, MAP, ISUP (ISDN user part), etc. The 3G MSC provides the interconnection to 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 such as call waiting, etc.

CS data services:

The IWF provides rate adaptation and message translation for circuit mode data services, such as fax. Vocoding SS7, MAP and RANAP interfaces: The 3G-MSC is able to complete originating or terminating calls in the network in interaction with other entities of a 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 traffic across the Iu interface. Higher rate CS data rates may be supported using a different adaptation layer.

Short message services (SMS): This functionality allows the user to send and receive SMS data to and from the SMS-GMSC/SMS-IWMSC (Inter working MSC).

VLR functionality: The VLR is a database that may be located within the 3G-MSC and can serve as intermediate storage for subscriber data in order to support subscriber mobility. IN and CAMEL.

OAM (operation, administration, and maintenance) agent functionality 3G-SGSN 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 also provides the appropriate signaling and data interfaces including connection to an IP-based network toward the 3G-GGSN, SS7 toward the HLR/EIR/AUC and TCP/IP or SS7 toward the UTRAN. The 3G-SGSN provides the following functions: Session management: Handles session set-up messages from/to the UE and the GGSN and operates Admission Control and QoS mechanisms Iu and Gn MAP interface: The 3G-SGSN is able to complete originating or terminating sessions in the network by interaction with other entities of a mobile network, e.g., GGSN, HLR, AUC. It also controls/communicates with UTRAN using RANAP. ATM/AAL5 physical connection to the UTRAN for transportation of user data plane traffic across the Iu interface using GPRS tunneling protocol (GTP). Connection across the Gn interface toward the GGSN for transportation of user plane traffic using GTP. Note that no physical transport layer is defined for 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 HLR and SRNS relocation, and intersystem handover. Subscriber database functionality: This database (similar to the VLR) is located within the 3G-SGSN and serves as intermediate storage for subscriber data to support subscriber mobility.

Charging: The SGSN collects charging information related to radio network usage by the user. OAM agent functionality.

3G-GGSN

The GGSN provides interworking with the external PS network. It is connected with SGSN via an IP-based network. The GGSN may optionally support an SS7 interface with the HLR to handle mobile terminated packet sessions. The 3G-GGSN provides the following functions: 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 (depending on 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. OAM functionality

SMS-GMSC/SMS-IWMSC

The overall requirement for these two nodes is to handle the SMS from point to point. The functionality required can be split into two parts. The SMS-GMSC is an MSC capable of receiving a terminated short message from a service center, interrogating an HLR for routing information and SMS information, and delivering the short message to the SGSN of the recipient UE. The SMS-GMSC provides the 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 the routing information The SMS-IWMSC is an MSC capable of receiving an originating short message 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 or 3G-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, although the MSC and the service center may be integrated.

UNIT - 5

4G NETWORKS

PART – A

1. **What are the main functions of Cognitive Radio?**
   The main functions of Cognitive Radio are Spectrum Sensing, Dynamic Spectrum Management and Adaptive Communications.

2. **Define Cognitive Radio**
   The Federal Communications Commission FCC defined Cognitive Radio as “A radio that can change its transmitter parameters based on interaction with the environment in which it operates.

3. **Write a short note on time slot scheduler.**
   The time slot scheduler shares the spectrum efficiently between users by satisfying the QoS requirements. When the channel quality for each radio link can be predicted for a short duration into the future and accessible by the link layer, then ARQ with an adaptive modulation and coding system can be selected for each user to satisfy the Bit Error Rate(BER) requirement and offer throughput.

4. **What is meant by MIMO?**
   MIMO means Multiple Input and Multiple Output that represents multiple individual, parallel data streams that are carried on the air interface.

5. **What are the benefits of Smart Antenna Technology?**
   The benefits of Smart Antenna Technology are: a. Reduction in Co – Channel Interference b. Range Improvement c. Increase in Capacity d. Reduction in Transmitted Power e. Reduction in Handoff

6. **What is meant by receiver diversity?**
   The Single Input Multiple Output (SIMO) configuration of the radio channel is known as receiver diversity. The input the channel is single transmitter signal that feeds two receiver paths. Depending

on multipath fading and the correlation between two receiver gain is achieved in the form of fading resistance.

7. **What is Smart Antenna?**
A Smart Antenna is a multi- element antenna where the signals received at each antenna element are intelligently combined to improve the performance of the wireless system.

8. **Define Multi Carrier Modulation (MCM)**
Multi Carrier Modulation (MCM) is a baseband process that uses parallel equal bandwidth sub channels to transmit information and is normally implemented with Fast Fourier Techniques (FFT) techniques.

9. **What are the types of MCM that are likely preferred for 4G?**
The two different types of MCM that are likely preferred for 4G are:

a. Multi Carrier Code Division Multiple Access b. Orthogonal Frequency Division Multiplexing (OFDM) using TDMA

10. **What are the advantages of MCM?**
The advantages of MCM are

a. Better performance in the Inter Symbol Interference environment b. Avoidance of single frequency interference

11. **What are the technologies used in 4G?**
The technologies used in 4G are a. Multi Carrier Modulation (MCM) b. Smart Antenna Techniques c. OFDM – MIMO Systems d. Adaptive Modulation and Coding with Time Slot Scheduler e. Cognitive Radio

12. **List out the applications of 4G technologies.**
The applications of 4G technologies are

a. Virtual Presence b. Virtual Navigation c. Tele-Medicine d. Tele-Geo-Processing applications e. Gaming f. Cloud Computing g. Crisis detection and prevention h. Education

13. **What are the techniques to improve network survivability in different layers?**
The techniques to improve network survivability in different layers are

a. Prevention b. Network design and capacity allocation c. Traffic Management and restoration

14. What are the challenges of 4G?

The main challenges are

a. Multimode user terminals b. Wireless System Discovery and Selection c. Terminal Mobility d. Network Infrastructure and QoS Support e. Security and Privacy f. Fault tolerance and Survivability g. Multiple Operators and Billing Systems h. Personal Mobility

15. What are the main issues in terminal mobility of 4G?

The two main issues in terminal mobility are

a. Location Management b. Handoff Management With location management, the system tracks and locates a mobile terminal fir possible connection Handoff management maintains ongoing communications when the terminal roams.

16. Define 4G

4G can defined as MAGIC

MAGIC

a. Mobile Multimedia b. Anytime Anywhere c. Global Mobility Support d. Integrated Wireless Solution e. Customized Personal Services f. Also known as Mobile Broadband Everywhere

17. What are the goals of 4G?

The ambitious goal of 4G is to allow everyone to access the Internet anytime and everywhere. The provided connection to Internet will allow users to access all types of services including text, databases and multimedia. Unlike 3G, 4G is IP based, that is every user connected to the Internet will have an IP address.

18. Compare 3G with 4G.

Content

3G

4G

Switching Technique

Packet Switching

Packet Switching, Message Switching

Peak Download Rate

100 Mbps

1Gbps

Frequency Band

1.8 – 2.5 MHz

2-8 GHz

Access

Wideband CDMA

Multi-Carrier – CDMA or OFDM

19. What are the features of 4G Wireless Systems?

The features of 4G Wireless Systems are

a. Support interactive multimedia, voice, video, wireless internet and other broadband services. b. High speed, high capacity and low cost per bit. c. Global mobility, service portability, scalable mobile networks d. Seamless switching, variety of services based on Quality of Services requirements.

20. Write a short note on security challenges in 4G.

The security challenges with IP network is one of the most significant factors that slows down the further adoption of network technologies. An end to end system approach to security is required in next generation wireless networks, including:

a. Platform hardening

b. User/Operator authentication, authorization and auditing

c. Secure protocols, communication and data storage

d. Software and configuration integrity

PART – B

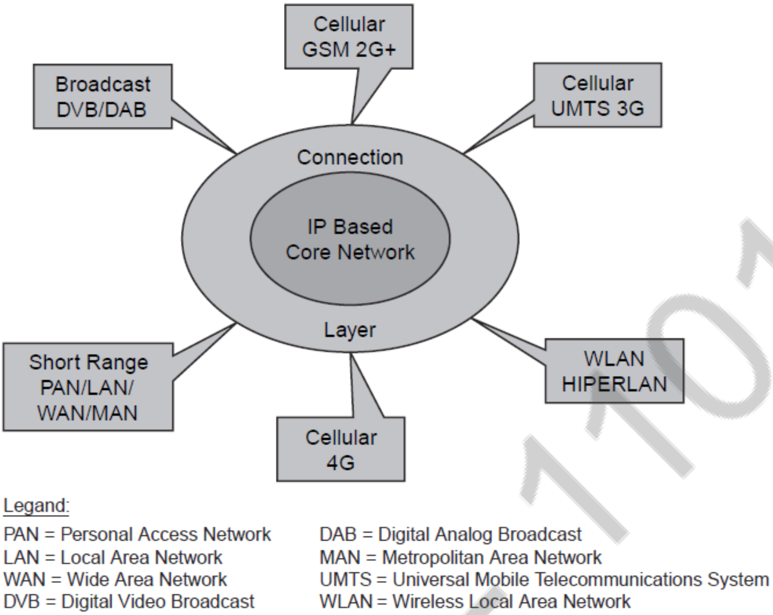
1. Explain the Fourth Generation System(4G) ?

With the rapid development of wireless communication networks, it is expected that fourth-generation (4G) mobile systems will be launched within a decade. 4G mobile systems focus on seamless integration of existing wireless technologies including WWAN, WLAN, and Bluetooth (see Figure 23.1). This is in contrast with 3G, which merely focuses on developing new standards and hardware.

The recent convergence of the Internet and mobile radio has accelerated the demand for “Internet in the pocket,” as well as for radio technologies that increase data throughput and reduce the cost per bit. Mobile networks are going multimedia, potentially leading to an explosion in throughput from a few bytes for the short message service (SMS) to a few kilobits per second (kbps) for the multimedia messaging service (MMS), to several 100 kbps for video content. In addition to wide area cellular systems, diverse wireless transmission technologies are being deployed, including digital audio broadcast (DAB) and digital video broadcast (DVB) for wide-area broadcasting, local multipoint distribution service (LMDS),and multichannel multipoint distribution service (MMDS) for fixed wireless access.

IEEE 802.11b, 11a, 11g, 11n, and 11h standards for wireless local area networks (WLANs) are extending from the enterprise world into public and residential domains. Because they complement cellular networks, these new wireless network technologies and their derivatives may well prove to be the infrastructure components of the future 4G mobile networks when multistandard terminals become widely available. This is already the case for WiFi in the public “hotspots,” which is being deployed by mobile operators around the world with the aim to offer seamless mobility with wireless wide-area networks.

The 4G systems 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 be interoperable with 2G and 3G systems, as well as with digital (broadband) broadcasting systems. The 4G intends to integrate from satellite broadband to high altitude platform to cellular 2G and 3G systems to wireless local loop (WLL) and broadband wireless access (BWA) to WLAN, and wireless personal area networks (WPANs), all with IP as the integrating


Cellular
GSM 2G+
Broadcast
DVB/DAB
Cellular
UMTS 3G
Connection
IP Based
Core Network
Layer
Short Range
PAN/LAN/
WAN/MAN
Cellular
4G
WLAN
HIPERLAN
Legand:
PAN = Personal Access Network
LAN = Local Area Network
WAN = Wide Area Network
DVB = Digital Video Broadcast
DAB = Digital Analog Broadcast
MAN = Metropolitan Area Network
UMTS = Universal Mobile Telecommunications System
WLAN = Wireless Local Area Network

2. Explain the Comparison between of key parameters of 4G with 3G.

Details	3G including 2.5G (EDGE)	4G
Major requirement driving Architecture Network architecture Speeds Frequency band Switching design basis Access technologies Forward error correction Component design Internet protocol (IP) Mobile top speed	Predominantly voice driven, data was always add on Wide area cell-based 384 kbps to 2 Mbps Dependent on country or continent (1.8 to 2.4 GHz) 5 to 20 MHz Circuit and packet WCDMA, cdma2000 Convolutional codes rate 1/2, 1/3 Optimized antenna design, multiband adapters Number of airlink protocol including IPv5.0 200 km/h	Converge data and voice over IP Hybrid-integration of WLAN (WiFi, Bluetooth) and wireless wide-area networks 20 to 100 Mbps in mobile Mode Higher frequency bands (2 to 8 GHz) 100 MHz or more All digital with packetized Voice OFDM and multicarrier (MC)-CDMA Concatenated coding Schemes Smart antenna, softwaredefined multiband and wideband radios All IP (IPv6.0) 200 km/h

3. Explain the features and challenges of 4G Network?

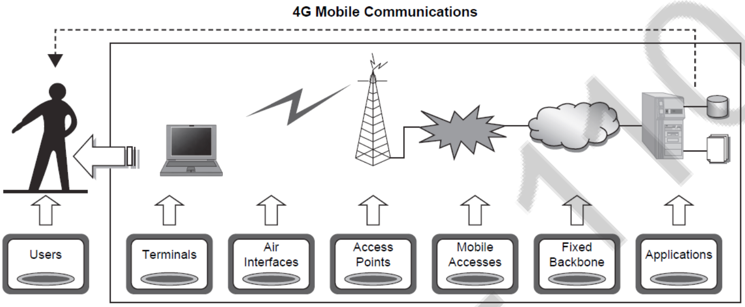
Some key features of 4G mobile networks are as follows

1) High usability: anytime, anywhere, and with any technology

2) Support for multimedia services at low transmission cost

3) Personalization

4) Integrated services



4G network

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.

To migrate current systems to 4G with the above-mentioned features, we have to face a number of challenges. Table 23.2 lists the key challenges and their proposed solutions. Figure 23.4 shows the carriers migration from 3.5G to 4G systems.

Applications of 4G

The following are some of the applications of the 4G system:

1)Virtual presence — 4G will provide user services at all times, even if the user is off-site.

2)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.

3)Tele-medicine — 4G will support the remote health monitoring of patients via video conference assistance for a doctor at anytime and anywhere.

4)Tele-geo-processing applications — 4G will combine geographical information systems (GIS) and global positioning systems (GPS) in which a user will get location querying.

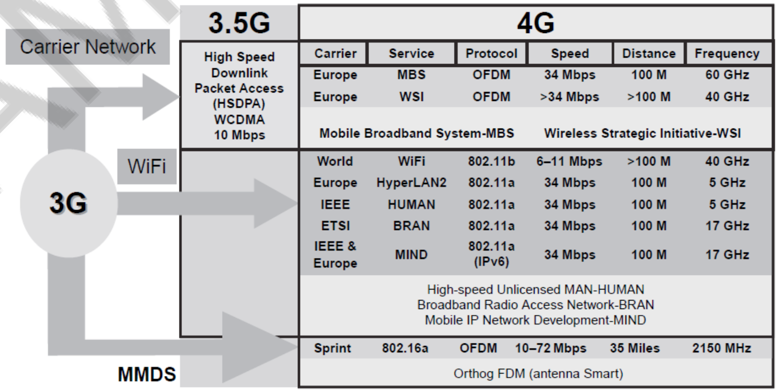
5)Education — 4G will provide a good opportunity to people anywhere in the world to continue their education on-line in a cost-effective manner.

4G Technologies

1) Multicarrier Modulation

Multicarrier modulation (MCM) is a derivative of frequency-division multiplexing.

It is not a new technology. Forms of multicarrier systems are currently used in DSL modems and digital audio/video broadcast (DAB/DVB). MCM is a baseband



Carrier	Service	Protocol	Speed	Distance	Frequency
Europe	MBS	OFDM	34 Mbps	100 M	60 GHz
Europe	WSI	OFDM	>34 Mbps	>100 M	40 GHz
World	WiFi	802.11b	6–11 Mbps	>100 M	40 GHz
Europe	HyperLAN2	802.11a	34 Mbps	100 M	5 GHz
IEEE	HUMAN	802.11a	34 Mbps	100 M	5 GHz
ETSI	BRAN	802.11a	34 Mbps	100 M	17 GHz
IEEE & Europe	MIND	802.11a (IPv6)	34 Mbps	100 M	17 GHz
Sprint	802.16a	OFDM	10–72 Mbps	35 Miles	2150 MHz

Mobile Broadband System-MBS Wireless Strategic Initiative-WSI

High-speed Unlicensed MAN-HUMAN
Broadband Radio Access Network-BRAN
Mobile IP Network Development-MIND

Orthog FDM (antenna Smart)

process that uses parallel equal bandwidth subchannels to transmit information and is normally implemented with fat 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 a cyclic extension 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 channnel'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:

$$(SNR)_{loss} = 10 \log \frac{L_b + L_c - 1}{L_b} (dB)$$

Two different types of MCM are likely candidates for 4G. These include 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 receivedata. 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) (where M _ 4 to 256).

However, to optimize overall performance, adaptive modulation can be used, where the level of quadrature amplitude modulation (QAM) for all subcarriers is chosen based on measured parameters. In OFDM the subcarrier pulse shape is a square wave. The task of pulse forming and modulation is performed by a simple inverse fast Fourier transform (IFFT) which can be implemented very efficiently. To decode the transmission, a receiver needs only to implement FFT.

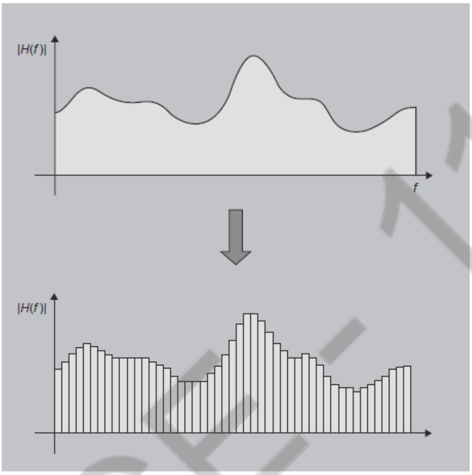


Figure 1 A broadband channel divided into many parallel narrowband channels.

The OFDM divides a broadband channel into many parallel subchannels.

The subchannel pulse shape is a square wave The OFDM receiver senses the channel and corrects distortion on each subchannel 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 subchannels overlap, they do not interfere with each other

4. Explain the different 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 receiver to increase the capacity of the wireless channel. 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. What this means is that high capacities are theoretically possible, unless there is a direct line of-sight between the transmitter and receiver.

The number of transmitting antennas is *M*, and the number of receiving antennas is *N*, where *N* _ *M*. We examine four cases:

Single-Input, Single-Output (SISO)

Single-Input, Multiple-Output (SIMO)

Multiple-Input, Single-Output (MISO)

Multiple-Input, Multiple-Output (MIMO)

Single-input, single-output:

The channel bandwidth is *B*, the transmitter power is *Pt*, the signal at the receiver has an average signal-to-noise ratio of SNR0, then the Shannon limit on channel capacity *C* is

$$C = B \log 2\ (1\ \_\ \mathrm{SNR}\ 0\ )$$

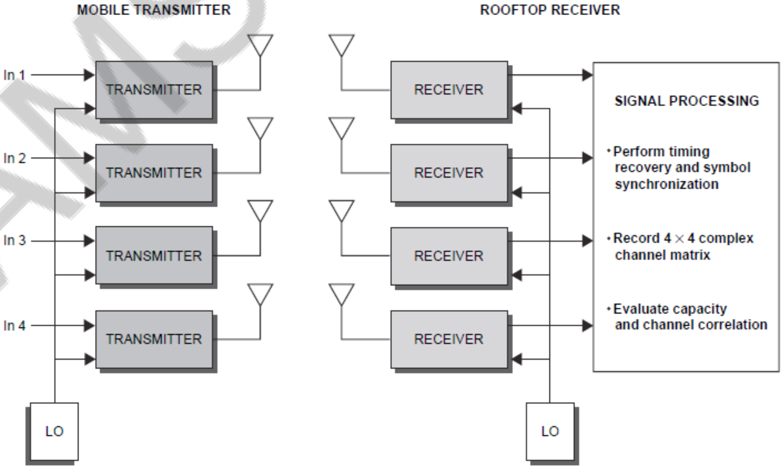


Fig: MIMO system

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:

$$\mathrm{SNR} \approx \frac{N^2 \times (\text{signal power})}{N \times (\text{noise})} = N \times \mathrm{SNR}_0$$

The capacity for this channel is approximately equal to

$$C \approx B \log_2 \left[1 + N \times \mathrm{SNR}_0\right]$$

Multiple-input, single-output: We have 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 some as SISO. The overall increase in SNR as approximately

$$\mathrm{SNR} \approx \frac{M^2 \cdot \left[(\text{signal power})/M\right]}{(\text{noise})} = M \times \mathrm{SNR}_0$$

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 SNR0 giving a channel capacity equal to

$$C \approx B \log_2 \left(1 + M \times N \times \mathrm{SNR}_0\right)$$

Assuming *N* _ *M*, we can send different signals using the same bandwidth and still be able to decode correctly at the receiver. Thus, we are creating a channel for each one of the transmitters. The capacity of each one of these channels is roughly equal to

$$C_{\text{single}} \approx B \log_2 \left(1 + \frac{N}{M} \times \mathrm{SNR}_0\right)$$

Since we have *M* of these channels (*M* transmitting antennas), the total capacity of the system is

$$C \approx MB \log_2 \left(1 + \frac{N}{M} \times \mathrm{SNR}_0\right)$$

We get a linear increase in capacity with respect to the transmitting antennas. As an example we assume SNR0 is equal to 10 dB, *M* _ 4, *N* _ 5 and bandwidth *B*(MHz) and list the system capacity for each channel type in Table

OFDM-MIMO Systems

OFDM and MIMO techniques can be combined to achieve high spectral efficiency and increased throughput. The OFDM-MIMO system transmits independent OFDM modulated data from multiple antennas simultaneously. At the receiver, after OFDM demodulation, MIMO decodes each subchannel to extract data from all transmit antennas on all the subchannels.

Adaptive Modulation and Coding with Time-Slot Scheduler

In general, TCP/IP is designed for a highly reliable transmission medium in wired networks where packet losses are seldom and are interpreted as congestion in the network. On the other hand, a wireless network uses a time varying channel where packet losses may be common 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 retransmi ssion. The time-slot scheduler shares the spectrum efficiently between users while satisfying the QoS requirements.

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 about individual data streams (along with predicted values of different radio links and selected modulation and coding systems by the link layer) and distributes the time slots among the users.

5. Explain the Bell Labs Layered Space Time (BLAST) System ?

BLAST is a space division multiplexing (SDM)-based MIMO system. It provides the best trade-off between system performance (spectral efficiency and capacity) and system implementation complexity. The spectral efficiency of BLAST ranges from 20 to 40 bps/Hz. It uses a zero-forcing (ZF) nonlinear detection algorithm based on a spatial nulling process combined with symbol cancellation to improve system performance. The BLAST exploits multipath by using scattering characteristics of the propagation environment to enhance transmission accuracy.

Figure shows the architecture of the BLAST system.

Transmitter: The data stream of a user is divided into multiple substreams. An array of transmit antennas (*M*) is used to simultaneously launch parallel data substreams. Each substream is mapped to a symbol by the same constellation and sent to its transmit antenna. All substreams are transmitted in the same frequency band and are independent of one another. Effective transmission rate is increased roughly in proportion to the number of transmit antennas used. The individual transmitter power is scaled by 1/*M*, so that the total power remains constant independent of the number of transmitters.

Receiver: An array of antennas (*N* _ *M*) is used to receive multiple transmitted substreams and their scattered images. Since substreams originate from different transmit antennas, they are located at different points in space. Using sophisticated signal processing, the substreams are identified and recovered.

Model: Each time sequence s_j (t), j= 1, 2, ..., M is referred to as layer. At the receiver, the signal $r_i(t)$ is received at time t. It is a noisy superposition of the M transmitted signal repectively corrupted by noise $n_i(t)$:

$$r_i(t) = \sum_{j=1}^{M} h_{ij}(t) s_i(t) + n_i(t) \qquad i=1, 2, 3, ..., N$$

Where $h_{ij}(t)$ is the channel gain (complex transfer function) from transmit antenna j to receive antenna i at any time t.

We make the following assumptions:

Quasi-static flat fading channel. That is, channel gain $h_{ij}(t)$ remains constant over a block of time, and then changes block by block in an independent manner.

Channel is rich scattering. This is true if antenna spacing is sufficient (i.e., several times of wavelength). This condition provides a large number of local scatters around transmitter or receiver and supports that the channel gains are complex Gaussian and independent of one another.

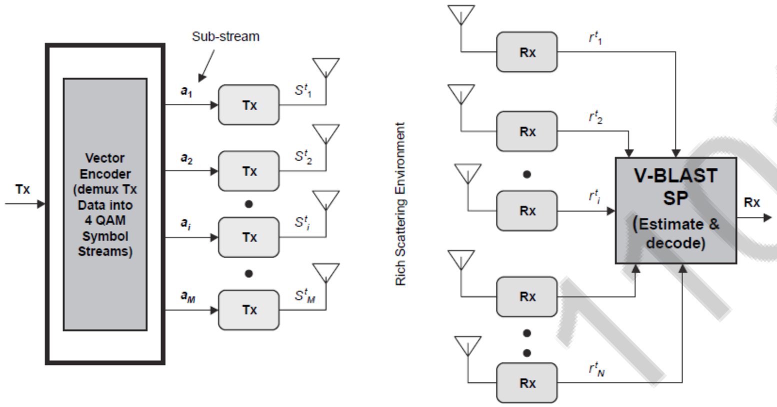

Tx
Vector Encoder (demux Tx Data into 4 QAM Symbol Streams)
Sub-stream
a_1
a_2
a_i
a_M
Tx
Tx
Tx
Tx
s^t_1
s^t_2
s^t_i
s^t_M
Rich Scattering Environment
Rx
Rx
Rx
Rx
Rx
r^t_1
r^t_2
r^t_i
r^t_N
V-BLAST SP (Estimate & decode)
Rx

Figure Architecture of BLAST system.

$$[r] = [H] \cdot [s] + [n] \tag{23.11}$$

where:

$$[r] = \begin{bmatrix} r_1 \\ r_2 \\ \bullet \\ \bullet \\ r_N \end{bmatrix}; [n] = \begin{bmatrix} n_1 \\ n_2 \\ \bullet \\ \bullet \\ n_N \end{bmatrix}; [s] = \begin{bmatrix} s_1 \\ s_2 \\ \bullet \\ \bullet \\ s_M \end{bmatrix} \text{ and } [H] = \begin{bmatrix} h_{11} & h_{12} & \bullet & \bullet & h_{1M} \\ h_{21} & h_{22} & \bullet & \bullet & h_{2M} \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ h_{N1} & h_{N2} & \bullet & \bullet & h_{NM} \end{bmatrix}$$

Signal Processing Algorithm:

At the bank of the receiving antennas, highspeed signal processors look at signals from all the receiving antennas simultaneously, first extracting the strongest substream from the morass, then proceeding with the remaining weaker signals, which are easier to recover once the stronger signals have been removed as a source of interference. Maximum-Likelihood (ML) detection is optimal for BLAST, but it is too complex to implement.

As an example, with six transmit antennas and QPSK modulation, a total of 46 _ 4096 comparisons have to be made for each transmitted symbol. A low complexity suboptimal detection algorithm, called ZF is used. At each symbol time, the strongest layer (transmitted signal) is first detected and its effect is cancelled for each received signal. We then proceed to detect the strongest of the remaining layers, and so on.

The ZF algorithm consists of four recursive steps:

1. *Ordering*: Determine the optimal detection order.
2. *Nulling*: Choose the nulling vector to null out all the weaker transmit signals and obtain the strongest transmit signal.
3. *Slicing*: Detect the estimated value of the strongest signal by slicing to the nearest value in the signal constellation.
4. *Cancellation*: Cancel the effect of the strongest signal from the received signal vector to reduce the detection complexity for the remaining transmit signal. Go to step 2 — nulling process.

6. Explain the Cognitive Radio ?

With the CR paradigm, spectrum can be efficiently shared in a more flexible fashion by a number of operators/users/systems. 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 incumbent 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 wireless networks.

The CR can be regarded as an extension of SDR. In 2003, the IEEE

Committee on Communications and Information Policy (CCIP) recommended CR for consideration by the FCC as a means to conserve valuable spectrum utilization. The CR focuses on applying software capabilities that have been developed to support algorithm control across a wide spectrum of signal processing technologies to add smarts to the software that allows it to determine when frequencies are free to use and then use them in the most efficient manner possible.

Most of the research work currently is focusing on spectrum sensing cognitive radio — particularly on the utilization of TV bands for communication. The essential problem of spectrum sensing CR is the design of high quality sensing devices and algorithms for exchanging spectrum sensing data between nodes. It has been shown in that a simple energy detector cannot guarantee accurate detection of signal presence. This calls for more sophisticated spectrum sensing techniques and requires that information about spectrum sensing be exchanged between nodes regularly.

It is not implicit that a CR must be software-defined radio. It is possible to implement CR features — the ability to detect and avoid (protect) incumbent users — while using relatively conventional radio transmitter/receiver architectures and techniques. The goal of CR is to relieve radio spectrum overcrowding, which actually translates to a lack of access to full radio spectrum utilization.