PURBANCHAL UNIVERSITY

WIRELESS NETWORKS AND MOBILE COMPUTING

(Compiled Notes)

YEAR-IV (BCA454WN - ELECTIVE-II) SEMESTER-II

Jeevan Poudel

BCA-VIII, 2077

Purbanchal University, Nepal

Wireless Networks and Mobile Computing (BCA454WN)

(Compiled Notes)

BCA-VIII

Jeevan Poudel

श्री गोमेन्द्र बहुमुखी महाविद्यालय
विर्तामोड, झापा
चैत ६, २०७७
(2021)

यो पाठ्य सामग्री तयार पार्न साथ, सहयोग र हौसला प्रदान
गर्नुहुने आदरणीय गुरु श्री मदन उप्रेती र मेरा सबै साथीहरुप्रति
हार्दिक आभार प्रकट गर्दछु।

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER

1

WIRELESS NETWORKS

1.1 Wireless Network

Wireless is a term used to describe telecommunications in which electromagnetic waves carry the signal. A wireless network is defined as technology that allows two or more computers to communicate, using standard protocol but without the use of network cabling. A wireless LAN (or WLAN) is one in which a mobile user can connect to a local area network through a wireless (radio) connection. The IEEE 802.11 group of standards specifies the technologies for wireless LANs.

1.2 Wireless Network Architecture

WLAN consist of two main components:

- An Access Point (AP) and wireless adapters.
- An access point looks like an external modem with two small antennae.
- Radio cards are also called WLAN cards.
- The range of an 802.11b WLAN is typically 100 feet and can be extended to several hundred feet by using an antenna.

- Wireless bridges are similar to the wired bridges and are used to connect two WLANs.

See Section 2.3.1 for explanation.

1.3 Wireless Switching Technology

Switching is process to forward packets coming in from one port to a port leading towards the destination. When data comes on a port it is called ingress, and when data leaves a port or goes out it is called egress. A communication system may include number of switches and nodes.

1.3.1 Circuit Switching

- When two nodes communicate with each other over a dedicated communication path, it is called circuit switching.
- There is a need of pre-specified route from which data travels and no other data is permitted.
- In circuit switching, to transfer the data, circuit must be established so that the data transfer can take place.
- Circuit switching was designed for voice applications.
- Telephone is the best suitable example of circuit switching. Before a user can make a call, a virtual path between caller and call receiver person is established over the network.

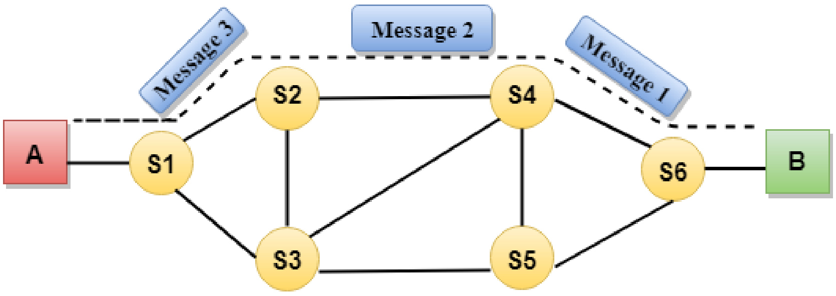


Figure 1.1: Circuit switching

Circuits can be permanent or temporary. Applications which use circuit switching may have to go through three phases:

a. Establish a circuit

b. Transfer the data

c. Disconnect the circuit

Advantages

- In the case of Circuit Switching technique, the communication channel is dedicated.
- It has fixed bandwidth.

Disadvantages

- Once the dedicated path is established, the only delay occurs in the speed of data transmission.
- It takes a long time to establish a connection approx 10 seconds during which no data can be transmitted.
- It is more expensive than other switching techniques as a dedicated path is required for each connection.
- It is inefficient to use because once the path is established and no data is transferred, then the capacity of the path is wasted.
- In this case, the connection is dedicated therefore no other data can be transferred even if the channel is free.

1.3.2 Message Switching

- This technique was somewhere in the middle of circuit switching and packet switching.
- In message switching, the whole message is treated as a data unit and is switching / transferred in its entirety.
- A switch working on message switching, first receives the whole message and buffers it until there are resources available to transfer it to the next hop.
- If the next hop is not having enough resource to accommodate large size message, the message is stored and switch waits.
- This technique was considered substitute to circuit switching.
- As in circuit switching, the whole path is blocked for two entities only.
- Message switching is replaced by packet switching.

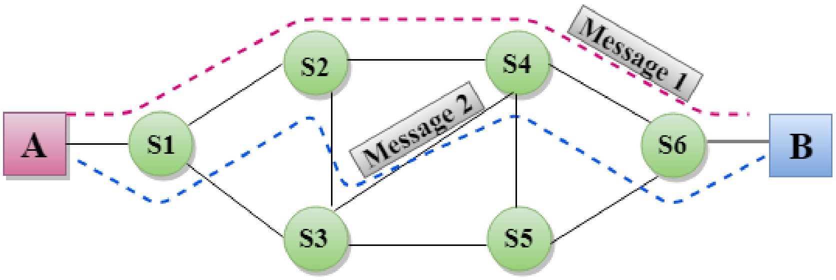


Figure 1.2: Message switching

Advantages

- Data channels are shared among the communicating devices that improve the efficiency of using available bandwidth.
- Traffic congestion can be reduced because the message is temporarily stored in the nodes.
- Message priority can be used to manage the network.
- The size of the message which is sent over the network can be varied. Therefore, it supports the data of unlimited size.

Disadvantages

- The message switches must be equipped with sufficient storage to enable them to store the messages until the message is forwarded.
- The Long delay can occur due to the storing and forwarding facility provided by the message switching technique.

1.3.3 Packet Switching

- Shortcomings of message switching gave birth to an idea of packet switching.
- The entire message is broken down into smaller chunks called packets.
- The switching information is added in the header of each packet and transmitted independently.
- It is easier for intermediate networking devices to store small size packets, and they do not take much resources either on carrier path or in the internal memory of switches. Packet switching enhances line efficiency as packets from multiple applications can be multiplexed

over the carrier.

- The Internet uses packet switching technique. Packet switching enables the user to differentiate data streams based on priorities.

- Packets are stored and forwarded according to their priority to provide quality of service.

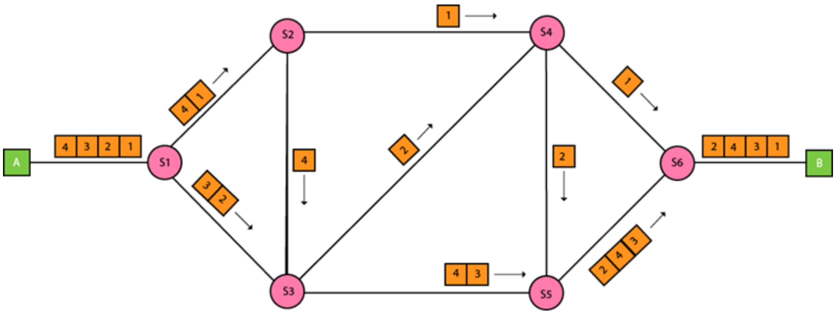


Figure 1.3: Packet switching

Advantages

- *Cost-effective*: In packet switching technique, switching devices do not require massive secondary storage to store the packets, so cost is minimized to some extent. Therefore, we can say that the packet switching technique is a cost-effective technique.

- *Reliable*: If any node is busy, then the packets can be rerouted. This ensures that the Packet Switching technique provides reliable communication.

- *Efficient*: Packet Switching is an efficient technique. It does not require any established path prior to the transmission, and many users can use the same communication channel simultaneously, hence makes use of available bandwidth very efficiently.

Disadvantages

- Packet Switching technique cannot be implemented in those applications that require low delay and high-quality services.

- The protocols used in a packet switching technique are very complex and requires high implementation cost.

- If the network is overloaded or corrupted, then it requires retransmission of lost packets. It can also lead to the loss of critical information if errors are not recovered.

Table 1.1: Differences between circuit switching and packet switching

Circuit Switching	**Packet Switching**
Physical path between source and destination	No physical path
All packets use same path	Packets travels independently
Bandwidth wastage	No bandwidth wastage
More reliable	Less reliable
No store and forward transmission	Supports store and forward transmission

1.4 Wireless Communication Problem

1.4.1 Interference

- Radio transmission cannot be protected against interference using shielding as this is done in coaxial cable or shielded twisted pair.

- For example, electrical engines and lightning cause severe interference and result in higher loss rates for transmitted data or higher bit error rates respectively.

1.4.2 Regulations and Spectrum

Frequencies have to be coordinated, and unfortunately, only a very limited amount of frequencies are available

1.4.3 Low Bandwidth

Transmission rates are very low for wireless devices compared to desktop systems.

1.4.4 High Delays

A serious problem for communication protocols used in today's Internet (TCP/IP) is the big variation in link characteristics.

1.4.5 Lower Security

Not only can portable devices be stolen more easily, but the radio interface is also prone to the dangers of eavesdropping.

1.4.6 Shared Medium

Radio access is always realized via a shared medium. As it is impossible to have a separate wire between a sender and each receiver, different competitors have to 'fight' for the medium.

1.5 Wireless Network Reference Model

The architecture of a network defines the protocols and components necessary to satisfy application requirements. One popular standard for illustrating the architecture is the seven-layer Open System Interconnect (OSI) Reference Model, developed by the International Standards Organization (ISO). OSI specifies a complete set of network functions, grouped into layers, which reside within each network component. The OSI Reference Model is also a handy model for representing the various standards and interoperability of a wireless network.

The OSI layers provide the following network functionality:

Layer 7 - Application Layer

- Establishes communications among users and provides basic communications services such as file transfer and e-mail.

- Examples of software that runs at this layer include Simple Mail Transfer Protocol (SMTP), HyperText Transfer Protocol (HTTP) and File Transfer Protocol (FTP).

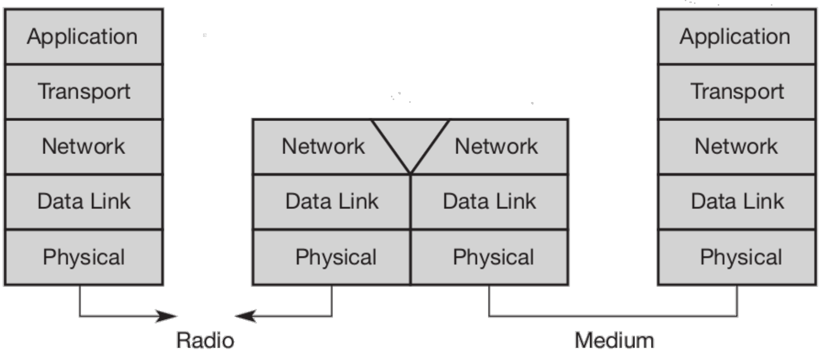


Figure 1.4: Wireless Network Reference Model

Layer 6 - Presentation Layer

- Negotiates data transfer syntax for the application layer and performs translations between different data formats, if necessary.
- For example, this layer can translate the coding that represents the data when communicating with a remote system made by a different vendor.

Layer 5 - Session layer

- Establishes, manages, and terminates sessions between applications. Wireless middleware and access controllers provide this form of connectivity over wireless networks.
- If the wireless network encounters interference, the session layer functions will suspend communications until the interference goes away.

Layer 4 - Transport Layer

- Provides mechanisms for the establishment, maintenance, and orderly termination of virtual circuits, while shielding the higher layers from the network implementation details.
- In general, these circuits are connections made between network applications from one end of the communications circuit to another (such as between the web browser on a laptop to a web page on a server).
- Protocols such as Transmission Control Protocol (TCP) operate at this layer.

Layer 3 - Network Layer

- Provides the routing of packets though a network from source to destination.
- This routing ensures that data packets are sent in a direction that leads to a particular destination.
- Protocols such as Internet Protocol (IP) operate at this layer.

Layer 2 - Data link Layer

- Ensures medium access, as well as synchronization and error control between two entities.
- With wireless networks, this often involves coordination of access to the common air medium and recovery from errors that might occur in the data as it propagates from source to destination.
- Most wireless network types have a common method of performing data link layer functions independent of the actual means of transmission.

Layer 1 - Physical Layer

- Provides the actual transmission of information through the medium.
- Physical layers include radio waves and infrared light.

The combined layers of a network architecture define the functionality of a wireless network, but wireless networks directly implement only the lower layers of the model. A wireless NIC, for example, implements the data link layer and physical layer functions. Other elements of the network (such as wireless middleware), however, offer functions that the session layer implements. In some cases, the addition of a wireless network might impact only the lower layers, but attention to higher layers is necessary to ensure that applications operate effectively in the presence of wireless network impairments.

Each layer of the OSI model supports the layers above it. In fact, the lower layers often appear transparent to the layers above. For example, TCP operating at the transport layer establishes connections with applications at a distant host computer, without awareness that lower layers are taking care of synchronization and signaling.

1.6 Wireless Networking Issues & Standards

1.6.1 Wireless Networking Issues

- **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, and higher 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. However, these additional features only work in a homogeneous environment, i. e. , when adapters from the same vendors are used for all wireless nodes.
- **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. 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 precautions 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.

1.6.2 Wireless Networking Standards

802.11 represents the IEEE designation for wireless networking. Several wireless networking specifications exist under the 802.11 banner. The 802.11 wireless standards can differ in terms of speed, transmission ranges, and frequency used, but in terms of actual implementation they are similar. All standards can use either an infrastructure or ad hoc network design, and each can use the same security protocols.

IEEE 802.11

- There were actually two variations on the initial 802.11 wireless standard.
- Both offered 1 or $2Mbps$ transmission speeds and the same RF of $2.4GHz$.
- The difference between the two was in how data traveled through the RF media.
  - one used FHSS (Frequency Hopping Spread Spectrum), and the
  - other used DSSS (Direct Sequence Spread Spectrum).
- The original 802.11 standards are far too slow for modern networking needs and are now no longer deployed.

IEEE 802.11a

- In terms of speed, the 802.11a standard was far ahead of the original 802.11 standards.
- 802.11a specified speeds of up to $54Mbps$ in the $5GHz$ band.
- 802.11a is incompatible with the 802.11b and 802.11g wireless standards.

IEEE 802.11b

- The 802.11b standard provides for a maximum transmission speed of $11Mbps$.
- Devices are designed to be backward-compatible with previous 802.11 standards
- 802.11b uses a $2.4GHz$ RF range and is compatible with 802.11g.

IEEE 802.11g

- 802.11g is a popular wireless standard today.
- 802.11g offers wireless transmission over distances of $150feet$ and speeds up to $54Mbps$
- 802.11g operates in the $2.4GHz$ range and therefore is compatible with it.

IEEE 802.11n

- The newest of the wireless standards listed is 802.11n.
- The goal of the 802.11n standard is to significantly increase throughput in both the $2.4GHz$ and the $5GHz$ frequency range.
- The baseline goal of the standard was to reach speeds of $100Mbps$.
- Given the right conditions, speeds might reach a $600Mbps$.
- In practical operation, 802.11n speeds is much slower.

1.7 Mobile Computing

Mobile Computing is a technology that allows transmission of data, voice and video via a computer or any other wireless enabled device without having to be connected to a fixed physical link.

1.7.1 Mobile Communication

Mobile Communication is a wireless form of communication in which voice and data information is emitted, transmitted and received via microwaves. This type of communication allows individuals to converse with one another and/or transmit and receive data while moving from place to place. Some examples include: cellular and digital cordless telephones, pagers, telephone answering devices, air-to-ground telecommunications and satellite-based communications.

1.7.2 Principles of Mobile Computing

- **Portability**: Devices/nodes connected within the mobile computing system should facilitate mobility. These devices may have limited device capabilities and limited power supply, but should have a sufficient processing capability and physical portability to operate in a movable environment.
- **Connectivity**: This defines the quality of service (QoS) of the network connectivity. In a mobile computing system, the network availability is expected to be maintained at a high level with the minimal amount of lag/downtime without being affected by the mobility of the connected nodes.

- **Interactivity**: The nodes belonging to a mobile computing system are connected with one another to communicate and collaborate through active transactions of data.

- **Individuality**: A portable device or a mobile node connected to a mobile network often denote an individual; a mobile computing system should be able to adopt the technology to cater the individual needs and also to obtain contextual information of each node.

1.7.3 Mobile Computing Architecture

The network-centric mobile computing architecture uses three-tier architecture as shown in Figure 1.5.

1. *Tier-1* : Presentation Tier / User interface
2. *Tier-2* : Application Tier / Process Management Tier
3. *Tier-2* : Data Tier / Database Management Tier

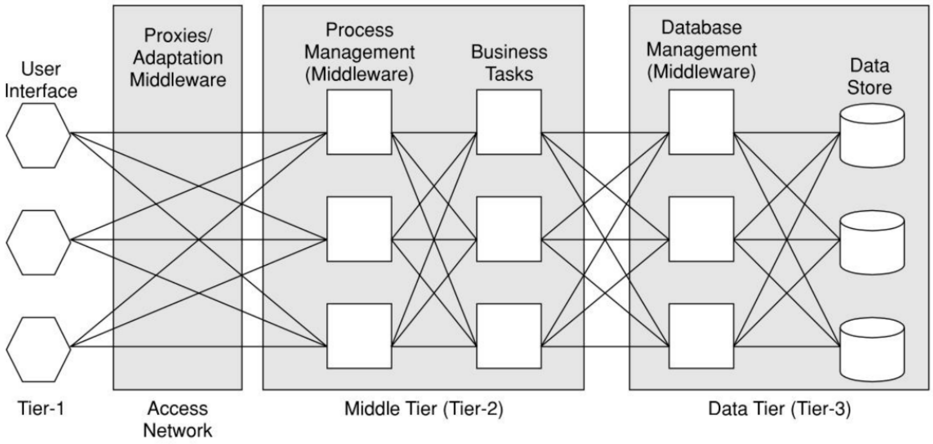


Figure 1.5: Three-tier architecture for mobile computing

1.7.3.1 Presentation Tier

- This is the user facing system in the first tier.

- This is the layer of agent applications and systems which run on the client device and offer all the user interfaces.

- These applications run on the client device and offer all the user interfaces.

- This tier is responsible for presenting the information to the end user.
- This tier includes web browsers and customized client programs.

1.7.3.2 Application Tier

- The application tier or middle tier is the "engine" of a ubiquitous application.
- It performs the business logic of processing user input, obtaining data, and making decisions.
- In certain cases, this layer will do the transcoding of data for appropriate rendering in the Presentation Tier.

In addition to the business logic there are quite a few additional management functions that need to be performed.

Different functions are implemented using different middleware software.

Middleware

A middleware framework is defined as a layer of software, which sits in the middle between the operating system and the user facing software.

Function relate to decisions on:

- rendering,
- network management,
- security,
- datastore access, etc.

Middleware covers a wide range of software systems. We can group middleware into the following major categories:

- Message-oriented Middleware.
- Transaction Processing Middleware.
- Database Middleware.
- Communication Middleware.
- Distributed Object and Components.
- Transcoding Middleware.

1.7.3.3 Data Tier

- The Data Tier is used to store data needed by the application and acts as a repository for both temporary and permanent data.
- The data can be stored in any form of datastore or database.
- These can range from so-

phisticated relational database, legacy hierarchical database, to even simple text files.

- The data can also be stored in XML format for interoperability with other systems and data sources.

Database Middleware

Database middleware allows the business logic to be independent and transparent of the database technology and the database vendor.

1. Database middleware runs between the application program and the database.
2. These are sometimes called database connectors as well.
3. Examples of such middleware will be ODBC, JDBC, etc.
4. Using these middleware, the application will be able to access data from any data source

1.8 Mobile Devices

Mobile hardware includes mobile devices or device components that receive or access the service of mobility. They would range from portable laptops, smartphones, tablet PCs, Personal Digital Assistants. These devices will have a receptor medium that is capable of sensing and receiving signals. These devices are configured to operate in full-duplex, whereby they are capable of sending and receiving signals at the same time. They don't have to wait until one device has finished communicating for the other device to initiate communications.

Categories of mobile devices:

Sensor

A very simple wireless device is represented by a sensor transmitting state information.

Embedded Controllers

Many appliances already contain a simple or sometimes more complex controller. Keyboards, mice, headsets, washing machines, coffee machines, hair dryers and TV sets are just some examples.

Pager

As a very simple receiver, a pager can only display short text messages, has a tiny display, and cannot send any messages. Pagers can even be integrated into watches.

Mobile phones

The traditional mobile phone only had a simple black and white text display and could send/receive voice or short messages. Mobile phones with full color graphic display, touch screen, and Internet browser are available.

Personal digital assistant

PDAs typically accompany a user and offer simple versions of office software (calendar, notepad, mail).

Pocket computer

The next steps toward full computers are pocket computers offering tiny keyboards, color displays, and simple versions of programs found on desktop computers (text processing, spreadsheets etc.).

Tablet

A tablet computer is a mobile device, a mobile operating system and touchscreen display processing, rechargeable battery in a single, thin and flat package

Notebook/Laptop

Finally, Laptops offer more or less the same performance as standard desktop computers; they use the same software — the only technical difference being size, weight, and the ability to run on a battery.

1.9 Mobile System Networks

- Cellular Networks
- WLAN Networks
- Mobile IP
- Ad Hoc Networks

Cellular Networks

- A cell is the coverage area of a base station, connected to other stations via wire or fibre or wirelessly through switching centres.
- The coverage area defines a cell and its boundaries.
- Each cell base station functions as an access point for the mobile service.
- Each mobile device connects to the base station of the cell which covers the current location of the device.
- All the mobile devices within the range of a given base station communicate with each other through that base station only.

WLAN Networks

- WLAN is used for connectivity between the Internet, two LANs, mobile devices, and computers
- Mobile device connects to an access point.
- The access point, in turn, connects to a host LAN which links up to the Internet through a router

Mobile IP

Mobile IP communication protocol refers to the forwarding of Internet traffic with a fixed IP address even outside the home network. It allows users having wireless or mobile devices to use the Internet remotely.

Mobile IP is mostly used in WAN networks, where users need to carry their mobile devices across different LANs with different IP addresses. Mobile IP is not a wireless protocol. However, it could be employed for the IP infrastructure of cellular networks.

Ad Hoc Networks

- A wireless ad hoc network (WANET) or MANET (Mobile ad hoc network) is a decentralized type of wireless network.
- The network is ad hoc because it does not rely on a pre-existing infrastructure, such as routers in wired networks or access points in managed (infrastructure) wireless networks.

- Instead, each node participates in routing by forwarding data for other nodes, so the determination of which nodes forward data is made dynamically on the basis of network connectivity and the routing algorithm in use.

- Wireless mobile ad hoc networks are self-configuring, dynamic networks in which nodes are free to move.

- Wireless networks lack the complexities of infrastructure setup and administration, enabling devices to create and join networks “on the fly” — anywhere, anytime.

1.10 Mobility Management

Mobility management is a functionality that facilitates mobile device operations in Universal Mobile Telecommunications System (UMTS) or Global System for Mobile Communications (GSM) networks. Mobility management is used to trace physical user and subscriber locations to provide mobile phone services, like calls and Short Message Service (SMS).

Mobility management contains two components:

1. location management and

2. handoff management

Location Management

Location management enables the system to track the attachment points of MTs (Mobile Terminals) between consecutive communications.

Handoff

Handoff (or handover) management enables the network to maintain a user’s connection as the MT continues to move and change its access point to the network. Moreover, when a user is in the coverage area of multiple wireless networks, for example, in heterogeneous wireless environments, handoff management provides always best connectivity to the user by connecting the user to the best available network.

Mobility in wireless networks can take different forms, such as:

- **Terminal mobility**: the ability for a user terminal to continue to access the network when the terminal moves;

- **User mobility**: the ability for a user to continue to access network services from different terminals under the same user identity when the user moves;
- **Service mobility**: the ability for a user to access the same services regardless of where the user is.

CHAPTER

2 WIRELESS LAN

Some advantages of WLAN 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.).
- **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.
- **Design**: Wireless networks allow for the design of small, independent devices which can for example be put into a pocket. Wireless senders and receivers can be hidden in historic buildings, e. g. , 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.
- **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.

2.1 Infrared Vs Radio Transmission

Two different basic transmission technologies can be used to set up WLANs. One technology is based on the transmission of *infrared 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.

2.1.1 Infrared Transmission

- Infrared 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.

Advantages

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

Disadvantages

- Disadvantages of infrared 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.
- However, their main disadvantage is that infrared is quite easily shielded.

- Infrared transmission cannot penetrate walls or other obstacles.
- Typically, for good transmission quality and high data rates a LOS, i. e. direct connection, is needed.

2.1.2 Radio Transmission

HIPERLAN and Bluetooth rely on radio transmission.

Advantages

- 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 than infrared.

The main advantage is also a big disadvantage of radio transmission.

Disadvantages

- 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.

2.2 Infrastructure and Ad-hoc Network

2.2.1 Infrastructure Network

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 2.1), 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 2.1 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.

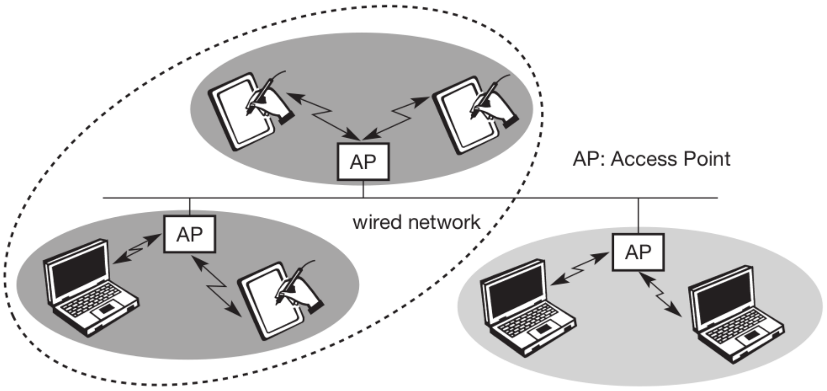


Figure 2.1: Example of three infrastructure-based wireless networks

Advantages

- 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.

- 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.

Disadvantage(s)

- 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.

2.2.2 Ad-hoc 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 2.2 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 2.2 cannot, therefore, communicate with each other if they are not within the same radio range.

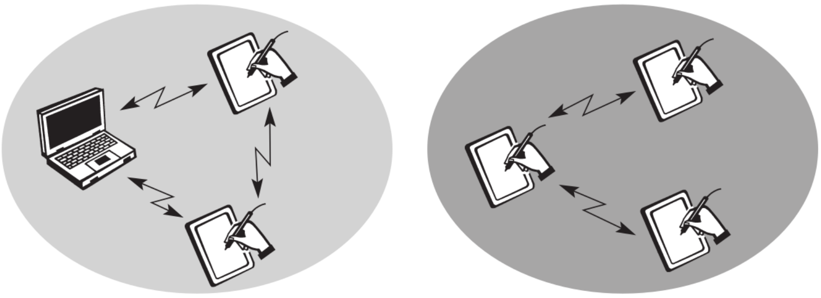


Figure 2.2: Example of two ad-hoc wireless networks

Advantage

- 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.

Disadvantage

- 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.

2.3 IEEE 802.11

The IEEE standard 802.11 specifies the most famous family of WLANs in which many products are available. This standard belongs to the group of 802.x LAN standards, e. g. , 802.3 Ethernet or 802.5 Token Ring.

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 infrared 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.

2.3.1 System Architecture

Wireless networks can exhibit two different basic system architectures as shown in section 2.2: *infrastructure-based* or *ad-hoc*. Figure 2.3 shows the components of an infrastructure and a wireless part as specified for IEEE 802.11.

2.3.1.1 Access Point (AP)

Several nodes, called **stations** (STA_i), are connected to **access points (AP)**.

2.3.1.2 Stations (STA)

Stations are terminals with access mechanisms to the wireless medium and radio contact to the AP.

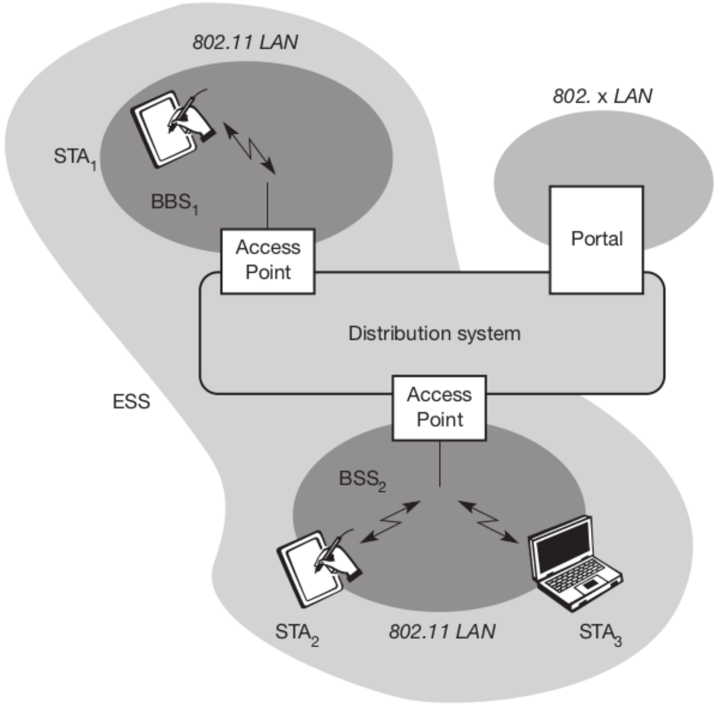


Figure 2.3: Architecture of an infrastructure-based IEEE 802.11

2.3.1.3 Basic Service Set (BSS)

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.

2.3.1.4 Extended Service Set (ESS)

A distribution system connects several BSSs via the AP to form a single network and thereby extends the wireless coverage area. This network is now called an **extended service set (ESS)** and has its own identifier, the ESSID. The ESSID is the 'name' of a network and is used to separate different networks. Without knowing the ESSID (and assuming no hacking) it should not be possible to participate in the WLAN.

2.3.1.5 Portal

The distribution system connects the wireless networks via the APs with a portal, which forms the interworking unit to other LANs.

The architecture of the distribution system is not specified further in IEEE 802.11. It could consist of bridged IEEE LANs, wireless links, or any other networks. However, **distribution system** services are defined in the standard

2.3.1.6 Distribution System

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 2.4. 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 STA_2 but not with STA_5. Several IBSSs can either be formed via the distance between the IBSSs (see Figure 2.4) or by using different carrier frequencies (then the IBSSs could overlap physically).

2.3.2 Protocol Architecture

IEEE 802.11 fits seamlessly into the other 802.x standards for wired LANs. Figure 2.5 shows the most common scenario:

1. An IEEE 802.11 wireless LAN connected to a switched IEEE 802.3 Ethernet via a bridge.
2. Applications should not notice any difference apart from the lower bandwidth and perhaps higher access time from the wireless LAN.
3. The WLAN behaves like a slow wired LAN.
4. Consequently, the higher layers (application, TCP, IP) look the same for wireless nodes as for wired nodes.

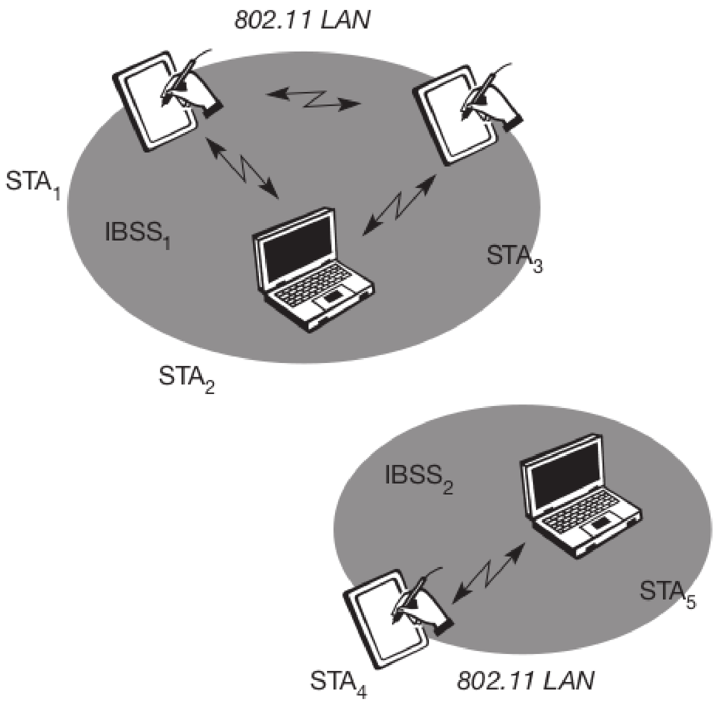
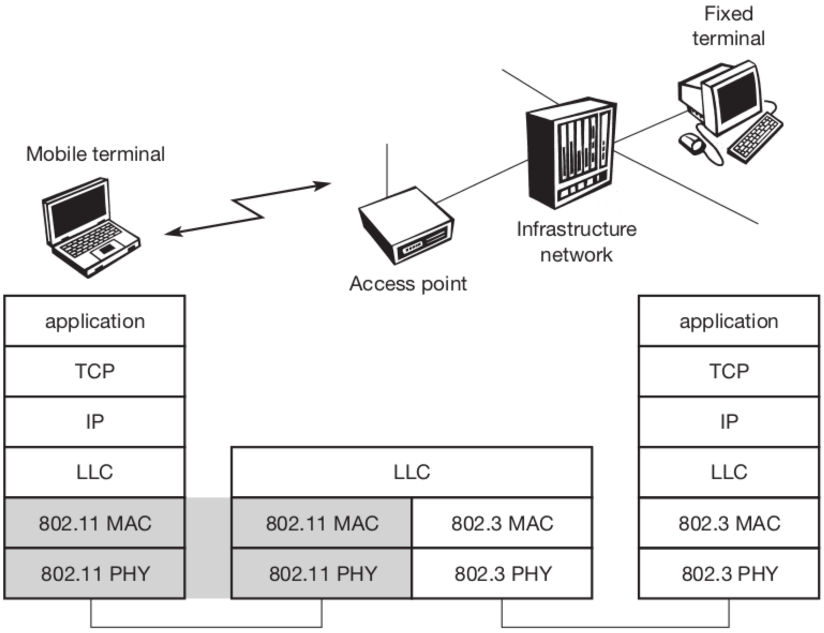


Figure 2.4: Architecture of IEEE 802.11 ad-hoc wireless LANs



application			application
TCP			TCP
IP			IP
LLC	LLC		LLC
802.11 MAC	802.11 MAC	802.3 MAC	802.3 MAC
802.11 PHY	802.11 PHY	802.3 PHY	802.3 PHY

Figure 2.5: IEEE 802.11 protocol architecture and bridging

5. 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.

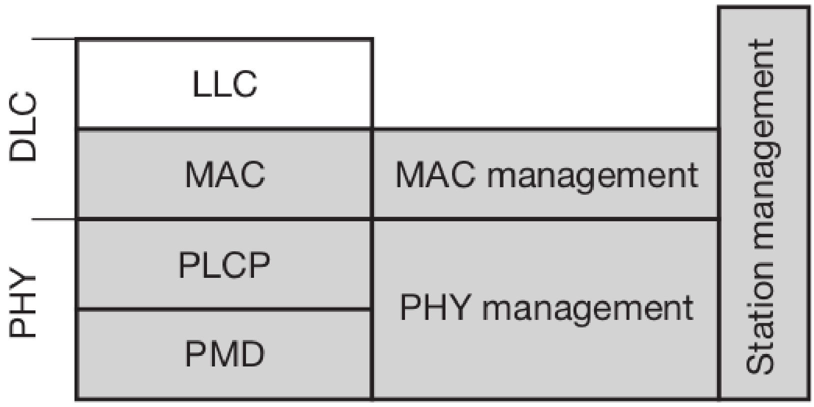


Figure 2.6: Detailed IEEE 802.11 protocol architecture and management

2.3.2.1 Physical Layer

The physical layer is subdivided into:

1. **physical layer convergence protocol (PLCP)** and
2. **physical medium dependent** (**PMD**) sublayer (see Figure 2.6).

2.3.2.1.1 PLCP

- 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.

2.3.2.2 MAC Layer

The basic tasks of the MAC layer comprise:

- medium access,
- fragmentation of user data, and
- encryption.

Apart from the protocol sublayers, the standard specifies **management layers** and the **station management**.

MAC 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).

PHY Management

The main tasks of the PHY management include channel tuning and PHY MIB maintenance.

Station Management

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).

2.3.3 802.11b

Soon after the first commercial 802.11 products came on the market some companies offered proprietary solutions. To avoid market segmentation, a common standard, **IEEE 802.11b**[1] 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.
- Depending on the interference and the distance between sender and receiver 802.11b systems offer 11, 5.5, 2, or $1Mbit/s$.
- Maximum user data rate is approx $6Mbit/s$.
- 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.
- The standards operates on certain frequencies in the 2.4 GHz ISM band. These depend on national regulations.
- Devices using 802.11b experience interference from other products operating in the 2.4 GHz band.
- **Pros of 802.11b**: lowest cost; signal range is good and not easily obstructed.
- **Cons of 802.11b**: slowest maximum speed; home appliances may interfere on the unregulated frequency band.

2.3.4 802.11a

While 802.11b was in development, IEEE created a second extension to the original 802.11 standard called 802.11a. Because 802.11b gained in popularity much faster than did 802.11a.

- Uses the same data link layer protocol and frame format as the original standard, but an OFDM[2] based air interface (physical layer).

[1] Do not get confused about the fact that 802.11b hit the market before 802.11a. The standards are named according to the order in which the respective study groups have been established.

[2] orthogonal frequency-division multiplexing (OFDM) is a type of digital transmission and a method of encoding digital data on multiple carrier frequencies.

- It operates in the 5 GHz band with a maximum net data rate of $54 Mbit/s$.
- The higher frequency also means 802.11a signals have more difficulty penetrating walls and other obstructions.
- Because 802.11a and 802.11b utilize different frequencies, the two technologies are incompatible with each other.
- **Pros of 802.11a**: fast maximum speed; regulated frequencies prevent signal interference from other devices.
- **Cons of 802.11a**: highest cost; shorter range signal that is more easily obstructed.

2.3.5 Newer Developments

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.

- **802.11e (MAC enhancements)**: 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 standard currently only describes the basic architecture of 802.11 networks and their components.
- **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.
- **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 (as also specified for the European HiperLAN2 standard)

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.

2.3.6 HIPERLAN 1

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 service offered by a HIPERLAN 1 is compatible with the standard MAC services known from IEEE 802.x LANs

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.

HiperLAN features:

- range $100m$
- slow mobility ($1.4m/s$)
- supports asynchronous and synchronous traffic
- Bit rate - $23.59Mbit/s$

2.3.7 HIPERLAN 2

While HIPERLAN 1 did not succeed HiperLAN2 might have a better chance. (This is also written as HIPERLAN/2, HiperLAN/2, H/2) Standardized by ETSI this wireless network works at 5 GHz 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.

Features:

High-throughput transmission

- 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.) remain almost constant over a wide range of user packet sizes and Data rates.

Connection-oriented

Prior to data transmission HiperLAN2 networks establish logical connections between a sender and a receiver.

Quality of service support

With the help of connections, support of QoS is much simpler. Each connection has its own set of QoS parameters.

Dynamic frequency selection

- HiperLAN2 does not require frequency planning.
- All access points have built-in support which automatically selects an appropriate frequency within their coverage area.

Security support

Authentication as well as encryption is supported by HiperLAN2.

Mobility support

- Mobile terminals can move around while transmission always takes place between the terminal and the access point with the best radio signal.
- Handover between access points is performed automatically.

2.4 Bluetooth

Compared to the WLAN technologies, the 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.

Swedish IT-company Ericsson initiated some studies in 1994 around a so-called multi-communicator link. The project was renamed 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, whose goal was the development of mobile phones, laptops, notebooks, headsets etc. including Bluetooth technology, by the end of 1999.

In 2001, the first products hit the mass market, and many mobile phones, laptops, PDAs, video cameras etc. are equipped with Bluetooth technology today.

2.4.1 User Scenarios

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 (schedules, sales figures etc.). 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.

- **Bridging of networks**: Using wireless piconets, a mobile phone can be connected to a PDA or laptop in a simple way. Mobile phones have a Bluetooth chip. The mobile phone can then act as a bridge between the local piconet and, e. g. , the global GSM network.

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.

2.4.2 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.

Networking

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.

A very important term in the context of Bluetooth is a **piconet**.

2.4.2.1 Piconet

A piconet is a collection of Bluetooth devices which are synchronized to the same hopping sequence. Figure 2.7 shows a collection of devices with different roles.

1. One device in the piconet can act as **master** (M) all other devices connected to the master must act as **slaves** (S)

2. The master determines the hopping pattern in the piconet and the slaves have to synchronize to this pattern.

3. Each piconet has a unique hopping pattern. If a device wants to participate it has to synchronize to this.

4. Two additional types of devices are shown:

    - **parked devices** (P) can not actively participate in the piconet (e. g. , 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.

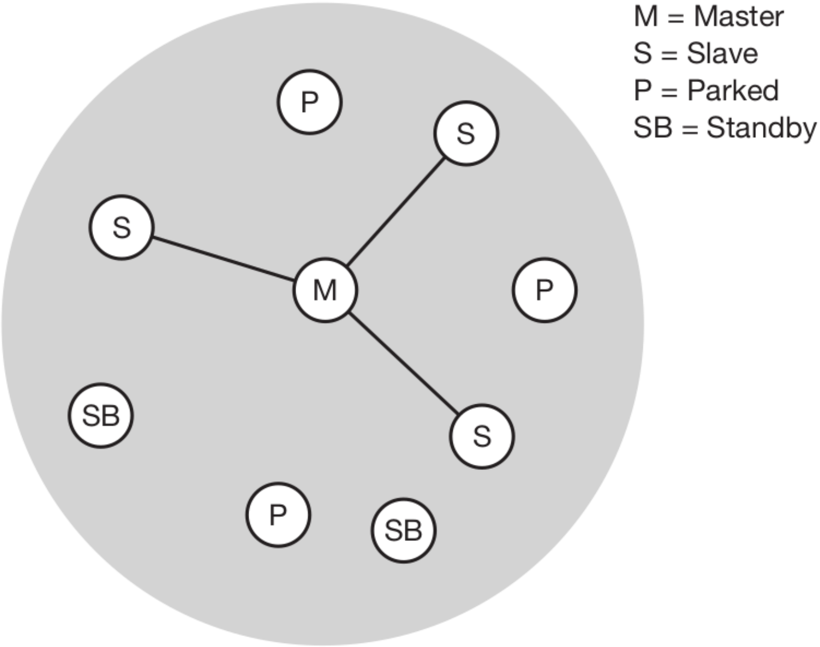


Figure 2.7: Simple Bluetooth piconet

2.4.2.2 Scatternet

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. This 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,

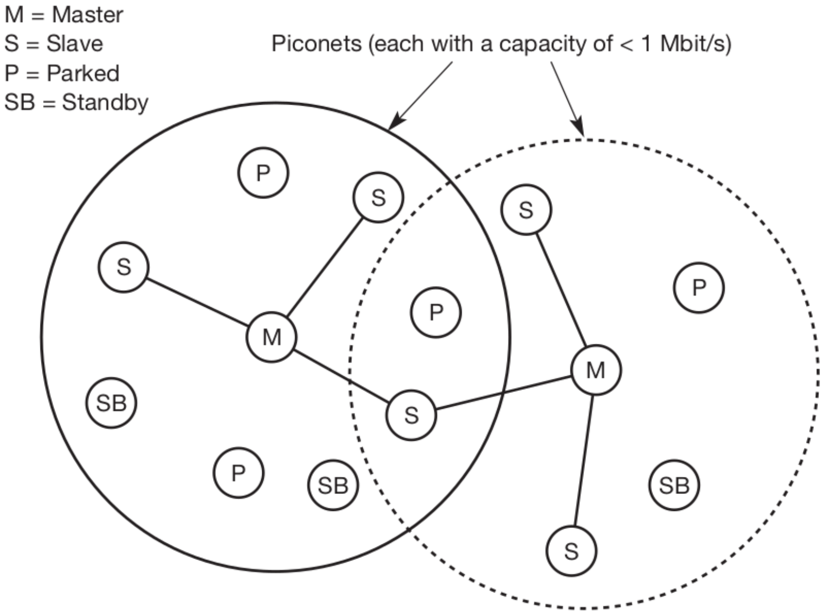


Figure 2.8: Bluetooth scatternet

1. the scatternet consists of two piconets, in which one device participates in two different piconets.
2. Both piconets use a different hopping sequence, always determined by the master of the piconet.
3. In an average sense, all piconets can share the total of 80 MHz bandwidth available.
4. Adding more piconets leads to a graceful performance degradation of a single piconet because more and more collisions may occur.
5. A collision occurs if two or more piconets use the same carrier frequency at the same time.

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.
- 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.

A master can also leave its piconet and act as a slave in another piconet. 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.

2.4.3 Protocol Stack

Bluetooth protocol stack can be thought of as a combination of multiple application specific stacks.

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 latter describes many protocols and functions needed to adapt the wireless Bluetooth technology to legacy and new applications.

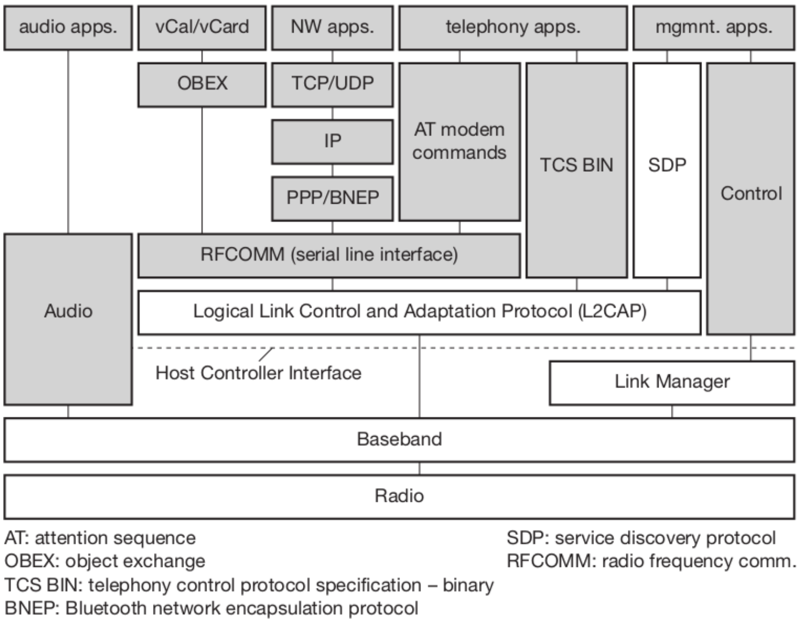

audio apps.
vCal/vCard
NW apps.
telephony apps.
mgmnt. apps.
OBEX
TCP/UDP
IP
PPP/BNEP
AT modem commands
TCS BIN
SDP
Control
RFCOMM (serial line interface)
Audio
Logical Link Control and Adaptation Protocol (L2CAP)
Host Controller Interface
Link Manager
Baseband
Radio
AT: attention sequence
OBEX: object exchange
TCS BIN: telephony control protocol specification – binary
BNEP: Bluetooth network encapsulation protocol
SDP: service discovery protocol
RFCOMM: radio frequency comm.

Figure 2.9: Bluetooth protocol stack

Bluetooth protocol stack (see figure 2.9) can be divided into different components according to their functions.

2.4.3.1 Core Protocols

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

- **Radio**: Specification of the air interface, e. g. , frequencies, modulation, and transmit power.
- **Baseband**: Description of basic connection establishment, packet formats, timing, and basic QoS parameters.
- **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 (connectionless and connection-oriented services).
- **Service discovery protocol**: Device discovery in close proximity plus querying of service characteristics.

2.4.3.2 Cable Replacement Protocol

On top of L2CAP is the **cable replacement protocol** RFCOMM that emulates a serial line interface.

- 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.

2.4.3.3 Telephony Control Protocol

- 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.

2.4.3.4 Adopted Protocols

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.

2.4.3.5 Host Controller Interface (HCI)

- The 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.

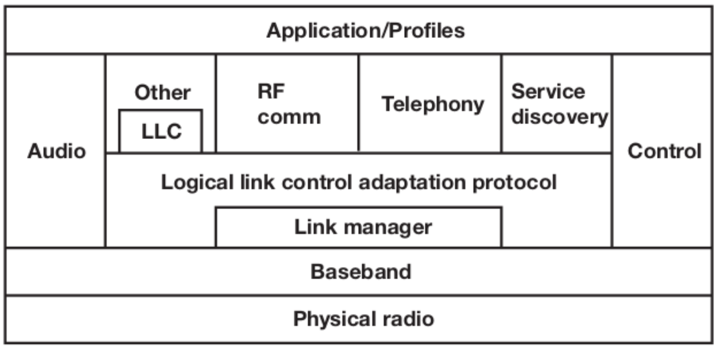
- The HCI can be seen as the hardware/software boundary.

2.4.3.6 Audio

- 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.

ब्लुटुथ प्रोटोकल स्ट्याकको सरल चित्र। माथि चित्रण गरिएको चित्रको साटो यो बनाएपनि हुन्छ। (Simplified figure of Bluetooth protocol stack.)



CHAPTER

3

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS (GSM)

Global System For Mobile Communication (GSM) is the most successful digital mobile telecommunication system in the world. It is used by billions of people around the world.

In the early 1980s, Europe had numerous coexisting analog mobile phone systems, which were often based on similar standards, but ran on slightly different carrier frequencies. To avoid this situation for a second generation fully digital system, the **groupe spéciale mobile (GSM)** was founded in 1982. This system was soon named the *global system for mobile communications (GSM)*, with the specification process lying in the hands of (European Telecommunications Standards Institute (ETSI).

The primary goal of GSM was to provide a mobile phone system that allows users to roam throughout Europe and provides voice services compatible to Integrated Services Digital Network (ISDN) and other Public Switched Telephone Network (PSTN) systems.

- GSM is a typical second generation system, replacing the first generation analog systems.
- GSM has initially been deployed in Europe using 890–915 MHz for

uplinks and 935–960 MHz for downlinks.

- There are multiple versions of GSM system.
  - **GSM 900**: GSM at 900 MHz.
  - **Digital Cellular Service (DCS)** 1800: GSM at 1800 MHz.
  - **Personal Communication Service (PCS) 19000**: GSM at 1900 MHz.

3.1 Mobile Services

GSM permits the integration of different voice and data services and the interworking with existing networks. Services make a network interesting for customers. GSM has defined three different categories of services:

- bearer services
- tele services and
- supplementary services

3.1.1 Bearer Services

Bearer services are telecommunication services that are used to transfer user data and control signals between two pieces of equipment. Bearer services permit

- transparent and non-transparent,
- synchronous or asynchronous data transmission.

Transparent Bearer Services

- Transparent bearer services only use the functions of the physical layer (layer 1) to transmit data.
- Data transmission has a constant delay and throughput if no transmission errors occur.
- Transparent bearer services do not try to recover lost data.

Non-transparent Bearer Services

Non-transparent bearer services use protocols of layers two and three to implement error correction and flow control.

3.1.2 Tele services

GSM mainly focuses on voice-oriented tele services. These comprise encrypted voice transmission, message services, and basic data communication with terminals as known from the PSTN or ISDN.

3.1.2.1 Telephony

The primary goal of GSM was the provision of high-quality digital voice transmission.

3.1.2.2 Emergency Number

This service is mandatory for all providers and free of charge. This connection also has the highest priority, possibly pre-empting other connections, and will automatically be set up with the closest emergency center.

3.1.2.3 SMS

Short Message Service (SMS) offers transmission of messages of up to 160 characters. It's successor is Enhanced Message Service (EMS).

3.1.2.4 EMS

EMS offers a *larger message size* (e. g. , 760 characters), *formatted text*, and the *transmission of animated pictures*, *small images* and *ring tones* in a standardized way.

3.1.2.5 MMS

EMS never really took off. Multimedia Message Service (MMS) offers the transmission of *larger pictures*, *short video clips* etc.

3.1.2.6 Group 3 Fax

Group 3 fax is a non-voice tele service. Fax data is transmitted as digital data over the analog telephone network using modems.

3.1.3 Supplementary services

In addition to tele and bearer services, GSM providers can offer **supplementary services**. These services offer various enhancements for the standard telephony service. Typical services are:

- user identification
- call redirection
- closed user groups and
- multi party communication

Closed user groups allow, for example, a company-specific GSM subnetwork, to which only members of the group have access.

3.2 System Architecture

Figure 3.1 gives an overview of the GSM system. A GSM system consists of three subsystems:

1. *Radio Subsystem (RSS)*
2. *Network Switching Subsystem (NSS)* and
3. *Operation Subsystem (OSS)*

नोटः GSM Architecture with interface भनेमा Figure 3.1 मा देखाइएको सबै बनाउने।

3.2.1 Radio Subsystem (RSS)

- RSS comprises all radio specific entities, i. e. *Mobile Station (MS)* and *Base Station Service (BSS).*
- Figure 3.1 shows the connection between the RSS and the NSS via the **A interface** (solid lines) and the connection to the OSS via the **O interface** (dashed lines).

3.2.1.1 Base Station Subsystem (BSS)

- A GSM network comprises many BSSs, each controlled by a Base Station Connector (BSC).

- The BSS performs all functions necessary to maintain:
  - radio connections to an MS,
  - coding/decoding of voice, and
  - rate adaptation to/from the wireless network part.
- Besides a BSC, the BSS contains several BTSs.

3.2.1.2 Base Transceiver Station (BTS)

- A Base Transceiver Receiver (BTS) comprises all radio equipment, i. e. , *antennas*, *signal processing*, *amplifiers* necessary for radio transmission.
- A BTS:
  - can form a radio cell using sectorized antennas.
  - is connected to MS via the U_m interface, and

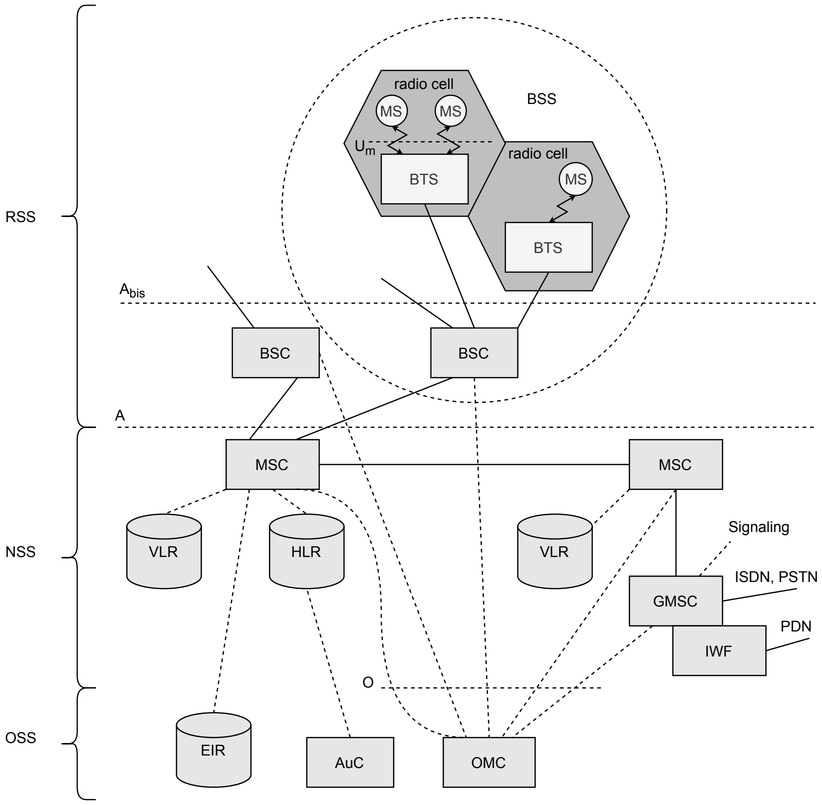


Figure 3.1: Functional architecture of a GSM system

  - connected to the BSC via the A_{bis} interface.

- The U_m interface contains all the mechanisms necessary for wireless transmission.
- A GSM cell can measure between some $100m$ and $35km$ depending on the environment.

Tasks of The BTS Within a BSS

- Frequency hopping
- Channel coding and decoding
- Rate adaptation
- Encryption and decryption
- Paging
- Uplink signal measurement

3.2.1.3 Base Station Controller (BSC)

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

Tasks of The BSC Within a BSS

- Management of radio channels
- Frequency hopping
- Management of terrestrial channels
- Mapping of terrestrial onto radio channels
- Encryption and decryption
- Paging
- Traffic measurement
- Authentication
- Location registry, location update
- Handover management

3.2.1.4 Mobile Station(MS)

- The MS comprises all user equipment and software needed for communication with a GSM network.
- An MS consists of:

  - user independent hardware and software and
  - the *Subscriber Identity Module (SIM)*, which stores all user-specific data that is relevant to GSM[1].

- An MS can be identified via the *International Mobile Equipment Identity (IMEI)*.

A user can personalize any MS using his or her SIM. The SIM card contains many identifiers and tables, such as:

- card-type
- serial number
- a list of subscribed services
- a *Personal Identification Number (PIN)*
- a *PIN Unblocking Key (PUK)*
- an authentication key K_i , and
- the *International Mobile Subscriber Identity (IMSI)*

The PIN is used to unlock the MS. Using the wrong PIN three times will lock the SIM. In such cases, the PUK is needed to unlock the SIM. The MS stores dynamic information while logged onto the GSM system, such as, e. g. , the cipher key K_c and the location information consisting of a *Temporary Mobile Subscriber Identity (TMSI)* and the *Location Area Identification (LAI)*.

3.2.2 Network and Switching Subsystem (NSS)

- The *heart* of the GSM system is formed by the *NSS*.
- The NSS:
  - connects the wireless network with standard public networks,
  - performs handovers between different BSSs,
  - comprises functions for worldwide localization of users and
  - supports charging, accounting, and roaming of users between different providers in different countries.

The NSS consists of the following switches and databases:

[1] Many additional items can be stored on the mobile device. However, this is irrelevant to GSM.

3.2.2.1 Mobile Services Switching Center (MSC)

- Mobile Station Connector (MSC)s are high-performance digital ISDN switches.
- They set up connections to other MSCs and to the BSCs via the *A* interface, and form the fixed backbone network of a GSM system.
- Typically, an MSC manages several BSCs in a geographical region.
- A Gateway MSC (GMSC) has additional connections to other fixed networks, such as PSTN and ISDN.
- Using additional Interworking Functions (IWF), an MSC can also connect to Public Data Networks (PDN) such as $X.25$.
- An MSC handles all signaling needed for connection setup, connection release and handover of connections to other MSCs.
- An MSC also performs all functions needed for supplementary services such as call forwarding, multi-party calls, etc.

3.2.2.2 Home Location Register (HLR)

- The Home Location Register (HLR) is the most important database in a GSM system as it stores all user-relevant information.
- This comprises static information, such as:
  - *Mobile Subscriber ISDN Number (MSISDN)*,
  - subscribed services (e. g. , call forwarding, roaming restrictions, GPRS), and
  - *IMSI*
- Dynamic information is also needed:
  - *Location Area (LA)* of the MS,
  - *Mobile Subscriber Roaming Number (MSRN)*,
  - current Visitor Location Register (VLR) and MSC.
- As soon as an MS leaves its current LA, the information in the HLR is updated.

- All these user-specific information elements only exist once for each user in a single HLR.
- Responsible for charging and accounting.
- HLR contains highly specialized databases to perform various specialized tasks such as answering requests within certain time-bounds.

3.2.2.3 Visitor Location Register (VLR)

- The VLR associated to each MSC is a dynamic database.
- VLR stores all important information needed for the MS users currently in the LA that is associated to the MSC.
- If a new MS comes into an LA the VLR is responsible for, it copies all relevant information for this user from the HLR.
- This hierarchy of VLR and HLR avoids frequent HLR updates and long-distance signaling of user information.

3.2.3 Operation Subsystem (OSS)

The third part of a GSM system, the OSS contains the necessary functions for network operation and maintenance. The following entities have been defined:

3.2.3.1 Operation and Maintenance Center (OMC)

The Operation and Maintenance Center (OMC) monitors and controls all other network entities via the *O* interface. Typical OMC management functions are:

- traffic monitoring
- status reports of network entities
- subscriber and security management
- accounting and billing

3.2.3.2 Authentication Center (AuC)

- The Authentication Centre (AuC) contains the algorithms for authentication as well as the keys for encryption and
- Generates the values needed for user authentication in the HLR.
- The AuC may be situated in a special protected part of the HLR.

3.2.3.3 Equipment Identity Register (EIR)

- The Equipment Identity Register (EIR) is a database for all IMEIs.
- The EIR has a *blacklist* of stolen (or locked) devices.
- The EIR also contains a list of valid IMEIs (*white list*), and a list of malfunctioning devices (*gray list*).

3.3 Protocols

GSM architecture is a layered model that is designed to allow communications between two different systems. The lower layers assure the services of the upper-layer protocols. Each layer passes suitable notifications to ensure the transmitted data has been formatted, transmitted, and received accurately.

Figure 3.2 shows the protocol architecture of GSM with signaling protocols, interfaces, as well as the entities already shown in Figure 3.1.

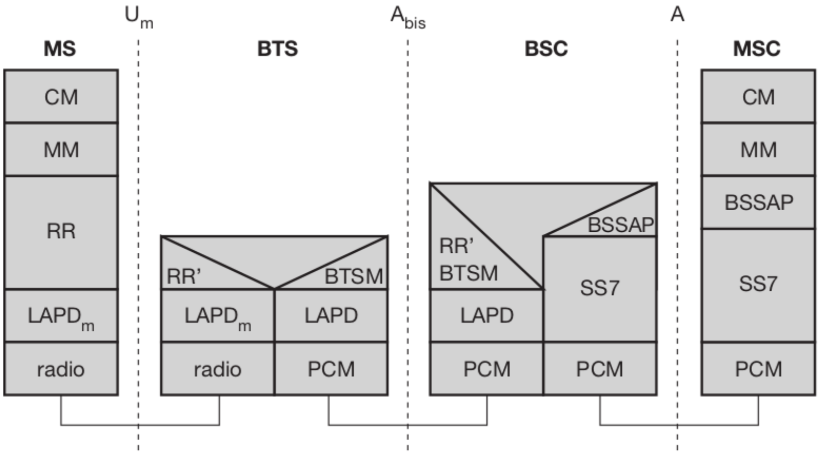


Figure 3.2: Protocol architecture for signaling

3.3.1 MS Protocols

3.3.1.1 Layer 1: Physical Layer

Layer 1, the *physical layer*, handles all *radio specific* functions.

- channel coding
- error detection/correction

3.3.1.2 Layer 2: Data Link Layer

- In layer 2, at the U_m interface, $\mathbf{LAPD_m}$ has been defined for the purpose of signaling between entities in a GSM network.
- $\mathbf{LAPD_m}$ has been derived from link access procedure for the D-channel (LAPD) in ISDN systems, which is a version of HDLC.
- $\mathbf{LAPD_m}$ is a lightweight LAPD because it does not need synchronization flags or checksumming for error detection.

3.3.1.3 Layer 3: Network Layer

The network layer in GSM, layer three, comprises several sublayers as shown in Figure 3.2.

1. Radio Resource Management (RR)
2. Mobility Management (MM)
3. Connection Management (CM)

3.3.2 MS to BTS Protocols

RR

The lowest sublayer is the radio resource management (RR). The main tasks of RR are:

- *setup*, *maintenance*, and *release of radio channels.*
- RR also directly accesses the physical layer for radio information and offers a reliable connection to the next higher layer.

MM

The MM layer is stacked above the RR layer.

- Mobility management (MM) contains functions for *registration*, *authentication*, *identification*, *location updating*, and the *provision of a TMSI* that replaces the IMSI.
  - TMSI hides the real identity of an MS user over the air interface.
  - TMSI is valid only in the current location area of a VLR.
- MM offers a reliable connection to the next higher layer.

CM

The CM layer is the topmost layer of the GSM protocol stack. The call management (CM) layer contains three entities:

1. call control (CC),
2. short message service (SMS),

and

3. supplementary service (SS).

SMS allows for message transfer.

CC provides a point-to-point connection between two terminals and is used by higher layers for call establishment, call clearing and change of call parameters.

3.3.3 BSC Protocols

- The BSC uses a different set of protocols after receiving the data from the BTS.
- The A_{bis} interface is used between the BTS and BSC.
- At this level, the radio resources at the lower portion of Layer 3 are changed from the RR to the Base Transceiver Station Management (BTSM).
- The BTS management layer is a relay function at the BTS to the BSC.

The RR protocols are responsible for the allocation and reallocation of traffic channels between the MS and the BTS.

To transit from the BSC to the MSC, the BSS mobile application part or the direct application part is used, and SS7 protocols is applied by the relay.

3.3.4 MSC Protocols

- At the MSC, starting from the BSC, the information is mapped across the *A* interface to the MTP Layers 1 through 3.
- Here, Base Station System Management Application Part (BSS MAP) is said to be the equivalent set of radio resources.
- The relay process is finished by the layers that are stacked on top of Layer 3 protocols, they are BSS MAP/DTAP, MM, and CM.
- This completes the relay process.
- To find and connect to the users across the network, MSCs interact using the control-signaling network.

3.4 Localization and Calling

One fundamental feature of the GSM system is the automatic, worldwide localization of users. The system always knows where a user currently is, and the same phone number is valid worldwide. To provide this service, GSM performs periodic location updates even if a user does not use the mobile station (provided that the MS is still logged into the GSM network and is not completely switched off).

The HLR always contains information about the current location, and the VLR currently responsible for the MS informs the HLR about location changes. As soon as an MS moves into the range of a new VLR (a new location area), the HLR sends all user data needed to the new VLR. **Changing VLRs with uninterrupted availability of all services is also called roaming**. Roaming can take place within the network of one provider, between two providers in one country, but also between different providers in different countries (international roaming).

3.4.1 Localization

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

3.4.1.1 MSISDN

MSISDN is the phone number of a GSM user. It follows the standard for addresses as it is also used in fixed ISDN networks. This number consists of the:

- **country code** (CC) (e. g. `+977 024-123456` with 977 for Nepal)
- the **national destination code (NDC)** (i. e. , the address of the network provider), and
- the **subscriber number (SN)**.

3.4.1.2 IMSI

- GSM uses the IMSI for internal unique identification of a subscriber.
- IMSI consists of:
  - a **mobile country code (MCC)** (e. g. , 429 for Nepal),
  - the **mobile network code (MNC)** (i. e. , the code of the network provider. e. g. , 01 for Nepal Telecom, 02 for Ncell), and finally
  - the **mobile subscriber identification number (MSIN)**.

3.4.1.3 TMSI

- TMSI hides the IMSI to prevent leaking identity of the user signaling over the air interface.
- GSM uses the 4 byte TMSI for local subscriber identification.
- TMSI is selected by the current VLR.
- TMSI is only valid temporarily and within the location area of the VLR.
- VLR may change the TMSI periodically.

3.4.1.4 Mobile Station Roaming Number (MSRN)

- MSRN is slo a temporary address that hides the identity and location of a subscriber.
- The VLR generates this address on request from the MSC, and the address is also stored in the HLR.
- MSRN contains:
  - the current **visitor country code (VCC)**,

  - the **visitor national destination code (VNDC)**,
  - the identification of the current MSC together with the subscriber number.

- The MSRN helps the HLR to find a subscriber for an incoming call.

All these numbers are needed to find a subscriber and to maintain the connection with a mobile station. The interesting case is the **mobile terminated call (MTC)**.

3.4.2 Calling

There are two approaches to call setup:

- *Mobile originated call (MOC)*: Mobile subscriber initiates the call.
- *Mobile terminated call (MTC)*: Subscriber Receives the call.

3.4.2.1 MTC

MTC is a situation in which a station calls a mobile station (the calling station could be outside the GSM network or another mobile station). Figure 3.3 shows the basic steps needed to connect the calling station with the mobile user.

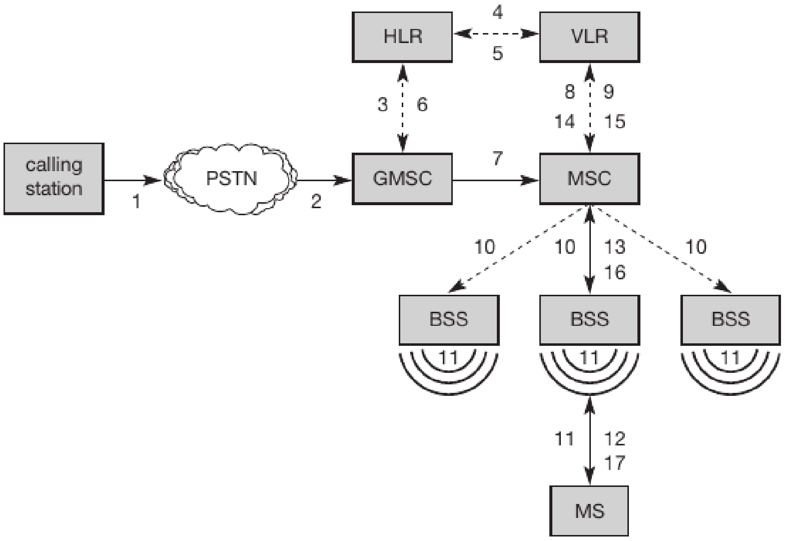


Figure 3.3: Mobile terminated call (MTC)

Following are the steps involved during MTC:

Step 1: A user dials the phone number of a GSM subscriber.

Step 2: The PSTN forwards the call setup to the GMSC.

Step 3: The GMSC identifies the HLR for the subscriber and signals the call setup to the HLR.

Step 4: The HLR now checks whether the number exists and whether the user has subscribed to the requested services, and requests an MSRN from the current VLR.

Step 5: HLR Receives an MSRN from the VLR.

Step 6: The HLR can determine the MSC responsible for the MS and forwards this information to the GMSC.

Step 7: The GMSC can now forward the call setup request to the MSC indicated.

From this point on, the MSC is responsible for all further steps.

Step 8: Requests the current status of the MS from the VLR.

Step 9: Receives the status of the MS.

Step 10: MSC initiates paging in all cells it is responsible for, if the MS is available.

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

Step 12: MS answers to the BSS.

Step 13: BSS answers to the MSC.

Step 14: The VLR has to perform security checks (set up encryption etc.).

Step 15: The VLR then signals to the MSC to set up a connection.

Step 16: MSC then signals BSS to set up a connection.

Step 17: BSS then signals MS to set up a connection.

3.4.2.2 MOC

It is much simpler to perform a MOC compared to a MTC (see Figure 3.4). Following steps are involved during MOC:

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

Step 2: The BSS forwards this request to the MSC.

Step 3: The MSC request VLR to checks if this user is allowed to set up a call with the requested service.

Step 4: VLR confirms request made in *Step 3*: and signals to the MSC.

Step 5: MSC request GMSC for the availability of resources through GSM network.

Step 6: GMSC then request PSTN for resources.

Step 7: PSTN responds to GMSC.

Step 8: GMSC responds back to MSC.

Step 9: MSC sets up connection with BSS.

Step 10: BSS and MSC then sets up connection between MS and the fixed network.

MOC लाई बुँदागत रुपमा नलेखेर सङ्क्षेपमा यस प्रकारले पनि लेख्न सकिन्छ

It is much simpler to perform a mobile originated call (MOC) compared to a MTC. The MS transmits a request for a new connection (1), the BSS forwards this request to the MSC (2). The MSC then checks if this user is allowed to set up a call with the requested service (3 and 4) and checks the availability of resources through the GSM network and into the PSTN. If all resources are available, the MSC sets up a connection between the MS and the fixed network.

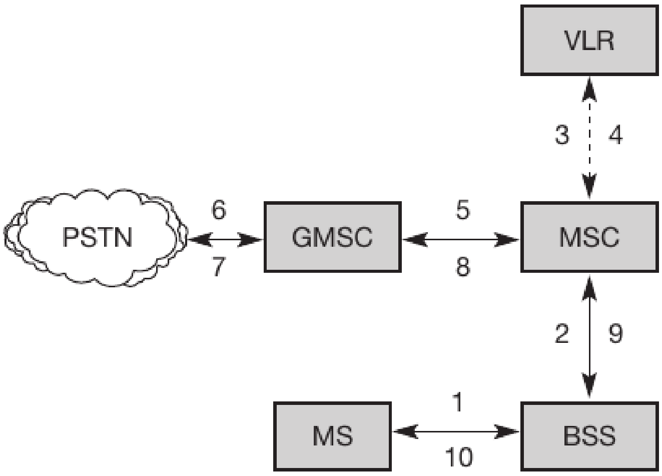


Figure 3.4: Mobile originated call (MOC)

3.5 Handover

Cellular systems require *handover* procedures, as single cells do not cover the whole service area, but, e. g. , only up to 35 km around each antenna on the countryside and some hundred meters in cities. The smaller the cell size and the faster the movement of a mobile station through the cells (up to $250 km/h$ for GSM), the more handovers of ongoing calls are required. However, a handover should not cause a cut-off, also called call drop. GSM aims at maximum handover duration of $60 ms$.

3.5.1 Basic Reasons For Handover

There are two basic reasons for a handover:

3.5.1.1 MS moves out of the range of a BTS

- The mobile station moves out of the range of a BTS or a certain antenna of a BTS respectively.
- The received *signal level* decreases continuously until it falls below the minimal requirements for communication.
- *The error rate* may grow due to interference, the distance to the BTS may be too high

- All these effects may diminish the *quality of the radio link* and make radio transmission impossible in the near future.

3.5.1.2 Too high traffic in one cell

- The wired infrastructure (MSC, BSC) may decide that the *traffic in one cell is too high* and shift some MS to other cells with a lower load (if possible).

- Handover may be due to *load balancing*.

3.5.2 Handover Scenarios

Figure 3.5 shows four possible handover scenarios in GSM:

- *Intra-cell handover*.

- *Inter-cell, intra-BSC handover*

- *Inter-BSC, intra-MSC handover*:

- *Inter MSC handover*

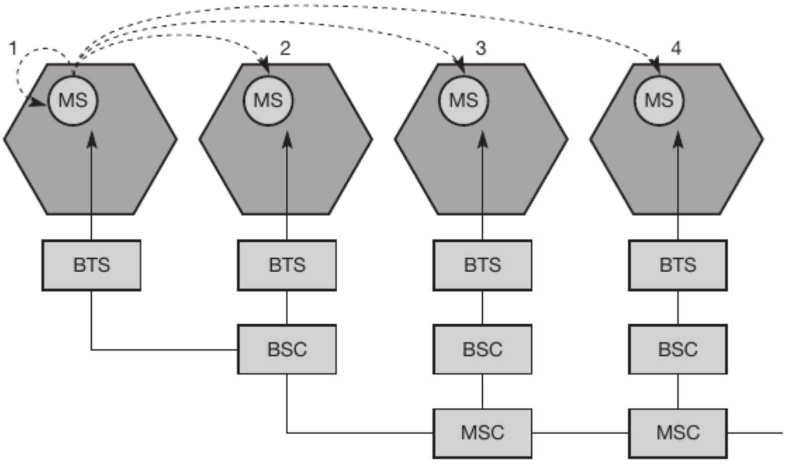


Figure 3.5: Types of handover in GSM

3.5.2.1 Scenario 1: Intra-cell Handover

- Within a cell, narrow-band interference could make transmission at a certain frequency impossible.
- The BSC could then decide to change the carrier frequency.

3.5.2.2 Scenario 2: Inter-cell, Intra-BSC Handover

- This is a typical handover scenario.
- The mobile station moves from one cell to another, but stays within the control of the same BSC.
- The BSC then performs a handover, assigns a new radio channel in the new cell and releases the old one.

3.5.2.3 Scenario 3: Inter-BSC, Intra-MSC Handover

- As a BSC only controls a limited number of cells; GSM also has to perform handovers between cells controlled by different BSCs.
- This handover then has to be controlled by the MSC.

3.5.2.4 Scenario 4: Inter MSC handover

- A handover could be required between two cells belonging to different MSCs.
- Now both MSCs perform the handover together.

Figure 3.6 shows the typical signal flow during an inter-BSC, intra-MSC handover.

- The MS sends its periodic measurements reports, the BTS_{old} forwards these reports to the BSC_{old} together with its own measurements.
- Based on these values and, the BSC_{old} may decide to perform a handover and sends the message `HO_required` to the MSC.
- The task of the MSC then comprises the request of the resources needed for the handover from the new BSC, BSC_{new}.
- This BSC checks if enough resources are available and activates a physical channel at the BTS_{new} to prepare for the arrival of the MS.

- The BTS_{new} acknowledges the successful channel activation, BSC_{new} acknowledges the handover request.

- The MSC then issues a handover command that is forwarded to the MS.

- The MS now breaks its old radio link and accesses the new BTS.

- The next steps include the establishment of the link (this includes layer two link establishment and handover complete messages from the MS).

- Basically, the MS has then finished the handover, but it is important to release the resources at the old BSC and BTS and to signal the successful handover using the handover and clear complete messages as shown.

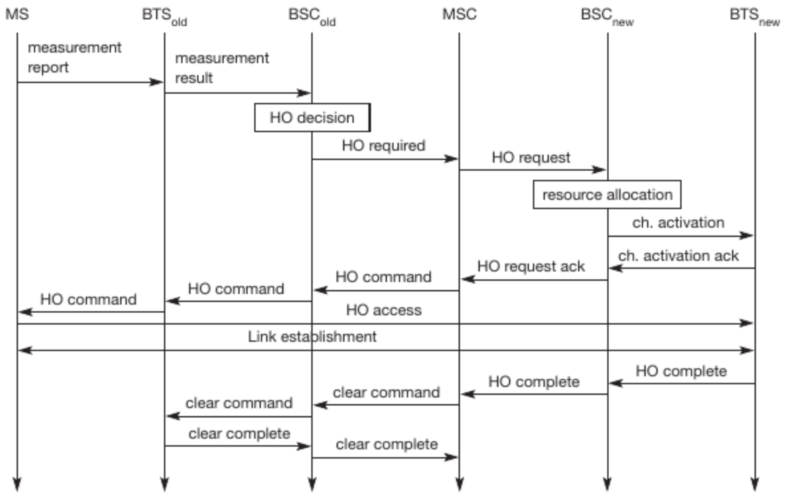


Figure 3.6: Intra-MSC handover

3.6 Security

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

3.6.1 GSM Security Services

The security services offered by GSM are:

- *Access control and authentication*
- *Confidentiality*
- *Anonymity*

3.6.1.1 Access Control and Authentication

- The first step includes the authentication of a valid user for the SIM. The user needs a secret PIN to access the SIM.
- The next step is the subscriber authentication.
- This step is based on a challenge-response scheme as presented in Section 3.6.2.

3.6.1.2 Confidentiality

- All user-related data is encrypted.
- After authentication, BTS and MS apply encryption to voice, data, and signaling as shown in Section 3.6.3.
- This confidentiality exists only between MS and BTS, but it does not exist end-to-end or within the whole fixed GSM/telephone network.

3.6.1.3 Anonymity

- To provide user anonymity, all data is encrypted before transmission, and user identifiers are not used over the air.
- Instead, GSM transmits a temporary identifier (TMSI), which is newly assigned by the VLR after each location update.
- Additionally, the VLR can change the TMSI at any time.

Three algorithms have been specified to provide security services in GSM:

- Algorithm $A3$ is used for *authentication*,
- $A5$ for *encryption*, and

- $A8$ for the *generation of a cipher key.*

In the GSM standard only algorithm $A5$ was publicly available, whereas $A3$ and $A8$ were secret, but standardized with open interfaces. Both $A3$ and $A8$ are no longer secret, but were published on the internet in 1998.

Algorithms $A3$ and $A8$ (or their replacements) are located on the SIM and in the AuC and can be proprietary. Only $A5$ which is implemented in the devices has to be identical for all providers.

3.6.2 Authentication

Before a subscriber can use any service from the GSM network, he or she must be authenticated. Authentication is based on the SIM, which stores the:

1. **individual authentication key** K_i,
2. **user identification IMSI**, and
3. algorithm used for authentication **A3**.

Authentication uses a challenge-response method:

Step 1: The access control AC generates a random number **RAND** as challenge, and

Step 2: the SIM within the MS answers with **SRES** (signed response) as response (see Figure 3.7).

Step 3: The AuC performs the basic generation of random values $RAND$, signed responses $SRES$, and cipher keys K_c for each IMSI, and then forwards this information to the HLR.

Step 4: The current VLR requests the appropriate values for $RAND$, $SRES$, and K_c from the HLR.

Step 5: The VLR sends the random value $RAND$ to the SIM.

Step 6: Both sides, network and subscriber module, perform the same operation with $RAND$ and the key K_i, called $A3$.

Step 7: The MS sends back the $SRES$ generated by the SIM; the VLR can now compare both values.

Step 8: If they are the same, the VLR accepts the subscriber, otherwise the subscriber is rejected.

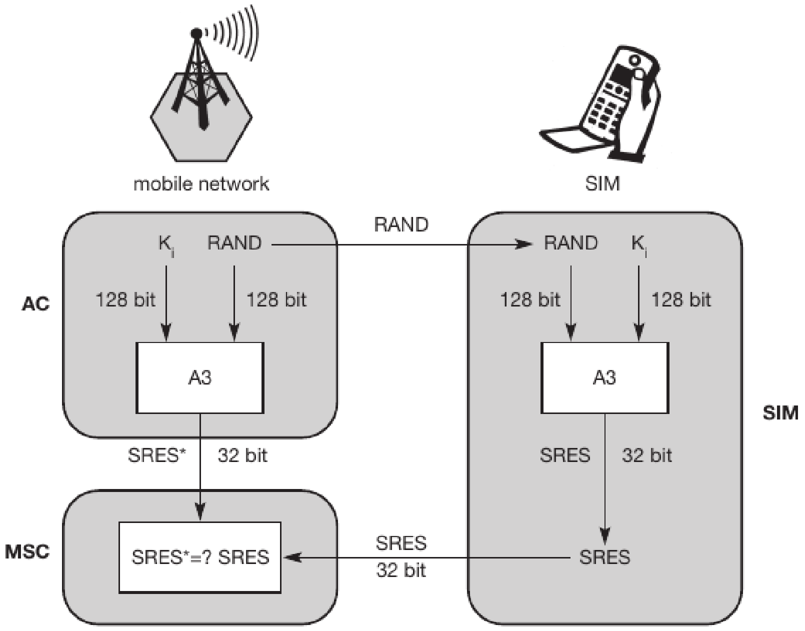


Figure 3.7: Subscriber authentication

3.6.3 Encryption

To ensure privacy, all messages containing user-related information are encrypted in GSM over the air interface.

- After authentication, MS and BSS can start using encryption by applying the cipher key K_c.
- K_c is generated using the individual key K_i and a random value by applying the algorithm $A8$.
- SIM in the MS and the network both calculate the same K_c based on the random value RAND.
- The key K_c itself is not transmitted over the air interface.

MS and BTS can now encrypt and decrypt data using the algorithm A5 and the cipher key K_c. As Figure 3.8 shows, K_c should be a 64 bit key — which is not very strong, but is at least a good protection against simple eavesdropping.

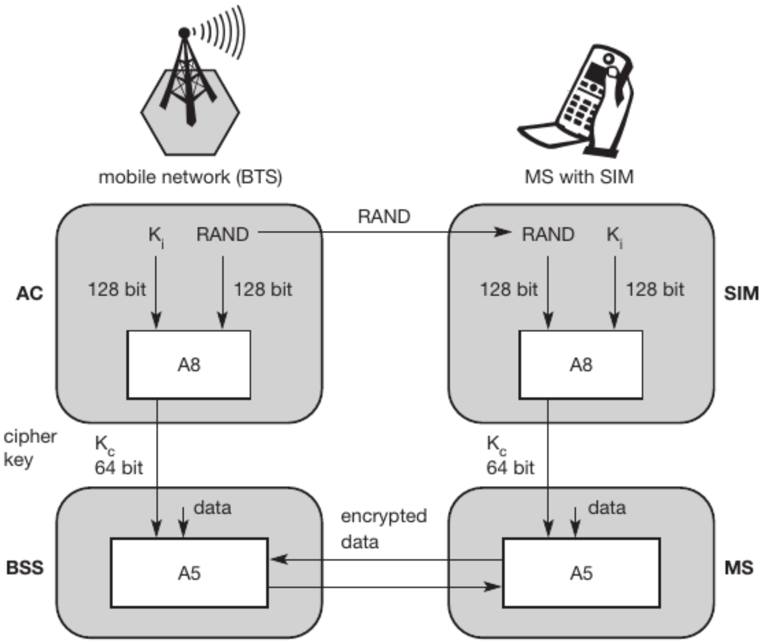


Figure 3.8: Data encryption

3.7 GPRS

The general packet radio service (GPRS) provides packet mode transfer for applications that exhibit traffic patterns such as frequent transmission of small volumes (e. g. , typical web requests) or infrequent transmissions of small or medium volumes (e. g. , typical web responses) according to the requirement specification.

The key element of GPRS technology is that it uses packet switched data rather than circuit switched data, and this technique makes much more efficient use of the available capacity. The data is split into packets and tags inserted into the packet to provide the destination address. Packets from several sources can then be transmitted over the link.

Goal

The provision of a more efficient and, thus, cheaper packet transfer service for typical internet applications that usually rely solely on packet transfer.

Benefit

The main benefit for users of GPRS is the ‘always on’ characteristic — no connection has to be set up prior to data transfer.

Motivation for Development

GPRS was driven by the tremendous success of the packet-oriented internet, and by the new traffic models and applications.

3.7.1 GPRS System Architecture

3.7.1.1 GPRS Support Node (GSN)

GPRS support nodes (GSN) and are in fact routers. All GSNs are integrated into the standard GSM architecture, and many new interfaces have been defined (see Figure 3.9).

3.7.1.2 Gateway GPRS Support Node (GGSN)

The **gateway GPRS support node (GGSN)** is the interworking unit between the GPRS network and external **packet data networks (PDN)**. This node

- contains routing information for GPRS users,
- performs address conversion, and
- tunnels data to a user via encapsulation.

The GGSN is connected to external networks (e. g. , IP or X.25) via the G_i interface and transfers packets to the SGSN via an IP-based GPRS backbone network (G_n interface).

3.7.1.3 Serving GPRS Support Node

The **serving GPRS support node (SGSN)** which supports the MS via the G_b interface. The SGSN, for example,

- requests user addresses from the **GPRS register (GR)**,
- keeps track of the individual MSs’ location, is responsible for collecting billing information (e. g. , counting bytes), and
- performs several security functions such as access control.

The SGSN is connected to a BSC via frame relay and is basically on the same hierarchy level as an MSC. The GR, which is typically a part of the HLR, stores all GPRS-relevant data. GGSNs and SGSNs can be compared with home and foreign agents, respectively, in a mobile IP network.

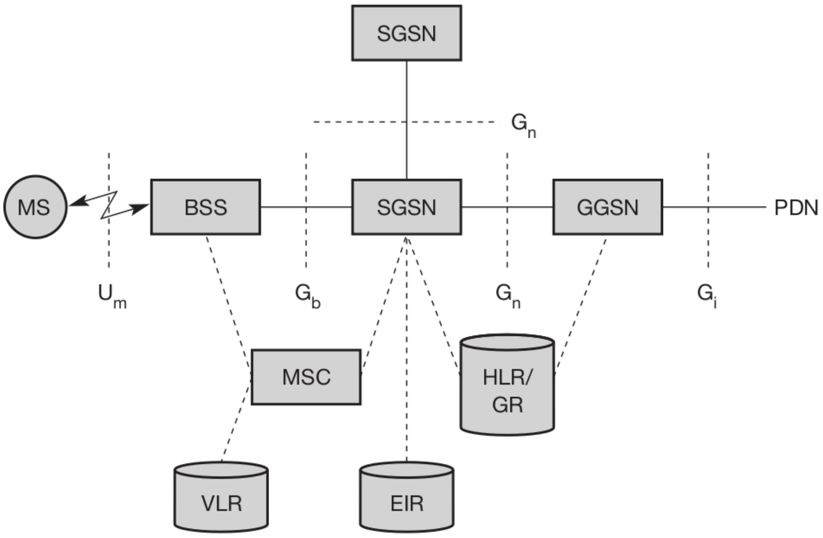


Figure 3.9: GPRS architecture reference model

As shown in Figure 3.9, packet data is transmitted from a PDN, via the GGSN and SGSN directly to the BSS and finally to the MS. The MSC, which is responsible for data transport in the traditional circuit-switched GSM, is only used for signaling in the GPRS scenario.

Before sending any data over the GPRS network,

- An MS must attach to it, following the procedures of the **mobility management**.
- The attachment procedure includes assigning a temporal identifier, called a **temporary logical link identity (TLLI)**, and a **ciphering key sequence number (CKSN)** for data encryption.
- For each MS, a **GPRS context** is set up and stored in the MS and in the corresponding SGSN. This context comprises:
  - the status of the MS (which can be ready, idle, or standby),
  - the CKSN,
  - a flag indicating if compression is used, and

  - routing data (TLLI, the routing area RA, a cell identifier, and a packet data channel, PDCH, identifier)

- Besides attaching and detaching, mobility management also comprises functions for:
  - authentication,
  - location management, and
  - ciphering.

- In **idle** mode an MS is not reachable and all context is deleted.

- In the **standby** state only movement across routing areas is updated to the SGSN but not changes of the cell.

Permanent updating would waste battery power, no updating would require system-wide paging. The update procedure in standby mode is a compromise. Only in the **ready** state every movement of the MS is indicated to the SGSN.

3.8 UMTS

The Universal Mobile Telecommunications System (UMTS) is a third generation mobile cellular system for networks based on the GSM standard. Developed and maintained by the 3GPP (3^{rd} Generation Partnership Project), UMTS is a component of the Standard International Union all IMT-2000 telecommunications and compares it with the standard set for CDMA2000 networks based on competition cdmaOne technology. UMTS uses wideband code division multiple access (W-CDMA) radio access technology to provide greater spectral efficiency and bandwidth mobile network operators.

3.8.1 UMTS System Architecture

Figure 3.10 shows the very simplified UMTS reference architecture.

- The **UTRA network (UTRAN)** handles cell level mobility and comprises several **radio network subsystems (RNS)**.

- The functions of the RNS include radio channel ciphering and deciphering, handover control, radio resource management etc.

- The UTRAN is connected to the **user equipment (UE)** via the radio interface U_u (which is comparable to the U_m interface in GSM) via the I_u interface (which is similar to the A interface in GSM), UTRAN communicates with the **core network (CN)**.
- The CN contains functions for inter-system handover, gateways to other networks (fixed or wireless), and performs location management if there is no dedicated connection between UE and UTRAN.

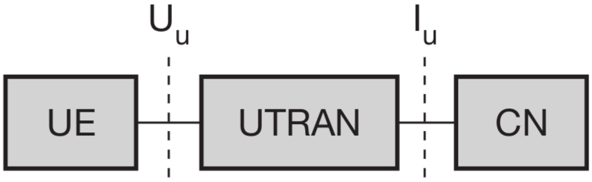


Figure 3.10: Main components of the UMTS reference architecture

The UMTS network architecture can be divided following elements. Figure 3.11 shows the UMTS architecture.

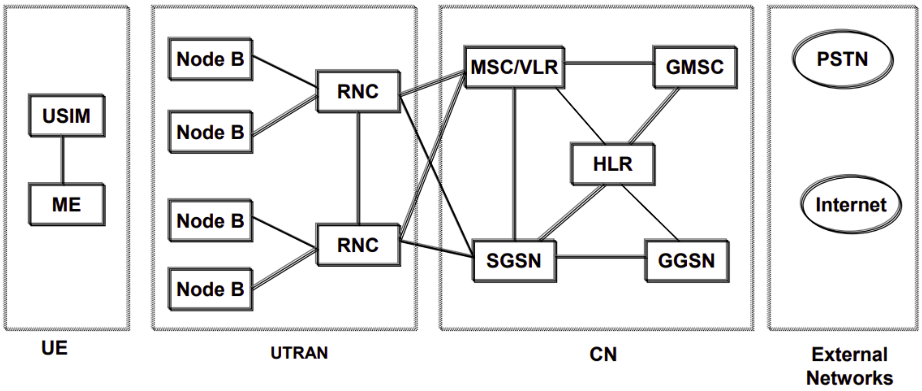


Figure 3.11: UMTS architecture

UMTS system composed of three main subsystems:

- **UE (User Equipment)** that interfaces with the user
- **UTRAN (UMTS Terrestrial Radio Access Network)** handles all radio related functionality — WCDMA is radio interface standard here.
- **CN (Core Network)** is responsible for transport functions such as switching and routing calls and data, tracking users.

3.8.1.1 UE

It includes:

- ME (Mobile Equipment) and
- USIM (UMTS Subscriber Identity Module)

3.8.1.1.1 ME ME is the single or multimode terminal used for radio communication

3.8.1.1.2 USIM USIM is a smart card that holds the subscriber identity, subscribed services, authentication and encryption keys.

3.8.1.2 UTRAN

It includes:

- Node B
- RNC (Radio Network Controller)

3.8.1.2.1 Node B Node B is equivalent to BTS in GSM/GPRS.

- It performs the air interface processing (channel coding, rate adaptation, spreading, synchronization, power control).
- It can operate a group of antennas/radios.

3.8.1.2.2 RNC RNC is equivalent to GSM BSC.

- It is responsible for radio resource management and control of the Node Bs.
- Also responsible for handoff decisions, congestion control, power control, encryption, admission control, protocol conversion, etc.

3.8.1.3 CN

3.8.1.3.1 HLR

- HLR is a database located in the user's home system that stores the master copy of the user's service profile.
- The HLR also stores the UE location on the level of MSC and SGSN.

3.8.1.3.2 3G MSC / VLR

- MSC/VLR are Switch and database that serves the UE in its current location for Circuit Switched (CS) services.
- The MSC function is used to switch the CS transactions, and VLR function holds a copy of the visiting user's service profile, as well as more precise information on the UE's location within the serving system

3.8.1.3.3 3G GMSC

- GMSC is a switch at the point where UMTS is connected to external CS networks.
- All incoming and outgoing CS connections go through GMSC.

3.8.1.3.4 3G SGSN

- It is similar to that of MSC / VLR but is used for Packet Switched (PS) services.
- The part of the network that is accessed via the SGSN is often referred to as the PS domain.
- Upgrade version of serving GPRS support node.

3.8.1.3.5 3G GGSN

- GGSN Functionality is close to that of GMSC but is in the relation to PS services.
- Upgraded version of gateway GPRS support Node.

The Core Network (CN) and the Interface I_u are separated into two logical domains:

1. **Circuit Switched Domain (CSD)**
    - Circuit switched service including signaling
    - Resource reservation at connection setup
    - 3G versions of GSM components (MSC, GMSC, VLR, HLR)
    - I_uCS
2. **Packet Switched Domain (PSD)**

- Handles all packet data services
- 3G versions of GPRS components (SGSN, GGSN)
- I_uPS

3.9 LTE

3.9.1 Long Term Evolution

LTE stands for Long Term Evolution and is a registered trademark owned by ETSI (European Telecommunications Standards Institute) for the wireless data communications technology and a development of the GSM/UMTS standards.

LTE was the 4G successor to the 3G UMTS system which was developed to provide a further evolution of the mobile telecommunications system available.

Providing much higher data speeds and greatly improved performance as well as lower operating costs, the scheme started to be deployed in its basic form around 2008.

Initial deployments gave little improvement over 3G HSPA and were sometimes dubbed 3.5G or 3.99G, but soon the full capability of LTE was realized it provided a full 4G level of performance.

The first deployments were simply known as LTE, but later deployments were designated 4G LTE Advanced and later still 4G LTE Pro.

Not only was the radio access network improved for 4G LTE, but the network architecture was overhauled enabling lower latency and much better interconnectivity between elements of the radio access network, RAN.

LTE is commonly marketed as 4G LTE & Advance 4G, but it does not meet the technical criteria of a 4G wireless service.

3.9.2 4G

4G is a collection of fourth generation cellular data technologies. It succeeds 3G and is also called “IMT-Advanced,” or “International Mobile Telecommunications Advanced.”

All 4G standards must conform to a set of specifications created by the International Telecommunications Union. For example, all 4G technologies are required to provide peak data transfer rates of at least 100 Mbps. While actual download and upload speeds may vary based on signal strength and wireless interference, 4G data transfer rates can actually surpass those of cable modem and DSL connections.

Like 3G, there is no single 4G standard. Instead, different cellular providers use different technologies that conform to the 4G requirements.

4G advantages:

- high spectral efficiency;
- very low latency;
- supports variable bandwidths;
- simple protocol architecture;
- compatibility and interworking with earlier 3GPP releases;
- Frequency division duplex (FDD) and time division duplex (TDD) within a single radio access technology;
- efficient multicast/broadcast.

3.9.3 5G

5G is the 5th generation mobile network. It is a new global wireless standard after 1G, 2G, 3G, and 4G networks. 5G enables a new kind of network that is designed to connect virtually everyone and everything together including machines, objects, and devices.

5G wireless technology is meant to deliver higher multi-Gbps peak data speeds, ultra low latency, more reliability, massive network capacity, increased availability, and a more uniform user experience to more users. Higher performance and improved efficiency empower new user experiences and connects new industries.

5G is based on OFDM (Orthogonal frequency-division multiplexing), a method of modulating a digital signal across several different channels to reduce interference. 5G uses 5G NR air interface alongside OFDM principles. 5G also uses wider bandwidth technologies such as sub-6 GHz and mmWave.

5G will bring wider bandwidths by expanding the usage of spectrum resources, from sub-3 GHz used in 4G to 100 GHz and beyond. 5G can operate in both lower bands (e. g. , sub-6 GHz) as well as mmWave (e. g. , 24 GHz and up), which will bring extreme capacity, multi-Gbps throughput, and low latency.

5G is designed to not only deliver faster, better mobile broadband services compared to 4G LTE, but can also expand into new service areas such as mission-critical communications and connecting the massive IoT. This is enabled by many new 5G NR air interface design techniques, such as a new self-contained TDD subframe design

Applications of 5G

5G is used across three main types of connected services, including:

- enhanced mobile broadband,
- mission-critical communications, and
- the massive IoT.

A defining capability of 5G is that it is designed for forward compatibility—the ability to flexibly support future services that are unknown today.

5G is designed to deliver peak data rates up to 20 Gbps based on IMT-2020 requirements.

5G technology has developed rapidly. The first real deployments went live in 2019, and further deployments soon followed. Although there were some teething issues, many noticed a significant increase in speed.

CHAPTER

4

MOBILE NETWORK LAYER

4.1 Mobile IP

Mobile IP is an IETF standard communications protocol that is designed to allow mobile device users to move from one network to another while maintaining a permanent IP address.

The Mobile IP allows for location-independent routing of IP datagrams on the Internet. Each mobile node is identified by its home address disregarding its current location on the Internet. Mobile IP specifies how a mobile node registers with its home agent and how the home agent routes datagrams to the mobile node through the tunnel.

4.1.1 Goals

- **Goal**: Use of mobile computer on the internet.
- **Problem**: As soon as mobile device leaves home network and reconnect at another place, it does not receive a single packet.
- **Reason**: Due to routing mechanisms used on the internet.

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 `27.36.36.69` shows that the receiver must be connected to the physical subnet with the network prefix `27.36.36`.

- Routers on 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.
- 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**.

4.1.2 Assumptions

This protocol assumes that mobile nodes will generally not change their point of attachment to the Internet more frequently than once per second.

This protocol assumes that IP unicast datagrams are routed based on the destination address in the datagram header (and not, for example, by source address).

Quick 'Solutions'

- Assign a new topologically correct IP address to the computer using Dynamic Host Configuration Protocol (DHCP).
- Using dynamic DNS an update of the mapping logical name — IP address is possible.
- Using DNS to dynamically adapting the IP address with regard to the current location.
- Creation of specific routes to the mobile node.

Problems

- Moving to a new location assigns new IP address which nobody knows and it is almost impossible to find a mobile host on the internet which has just changed its address.

- Dynamic Host Control Protocol (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.

There is a severe problem with higher layer protocols like Transmission Control Protocol (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.

4.1.3 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 on the internet. Several requirements accompanied the development of the standard:

4.1.3.1 Compatibility

- A new standard cannot introduce changes to existing network protocols.

- People do not want to change their applications just for mobility.

- Mobile has to work with current operating systems.

- Mobile IP must not require special media or Media Access Control (MAC)/LLC protocols.

- Mobile IP has to ensure that users can still access all the other servers and systems on the internet.

4.1.3.2 Transparency

- Mobility should remain *invisible* for many higher layer protocols and applications.

- Higher layers should continue to work even if the mobile computer has changed its point of attachment to the network.

4.1.3.3 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.
- It is crucial for a mobile IP to be scalable over a large number of participants in the whole internet, worldwide.

4.1.3.4 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.

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.*

4.1.4 Entities and Terminology

Mobile IP introduces the following entities and terms. Figure 4.1 illustrates an example scenario.

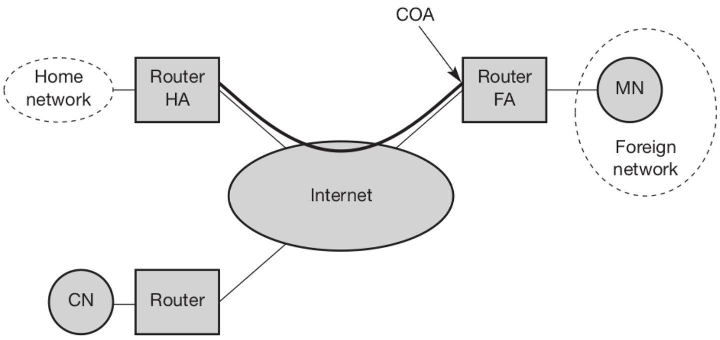


Figure 4.1: Mobile IP example network

Entities

4.1.4.1 Mobile Node (MN)

- 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 on the internet as long as link-layer connectivity is given.
- MNs are not necessarily small devices such as laptops with antennas or mobile phones; a router onboard an aircraft can be a powerful mobile node.

4.1.4.2 Home Agent (HA)

- The Home Agent (HA) provides several services for the MN and is located in 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 Care-of-address (COA).
- Three alternatives for the implementation of an HA exist:
  - On a router that is responsible for the home network.
  - On an arbitrary node in the subnet.
  - On the *router* but this time only acting as a manager for MNs belonging to a virtual home network.

4.1.4.3 Foreign Agent (FA)

- The Foreign Agent (FA) can provide several services to the MN during its visit to the foreign network.
- The FA can have the COA, 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.

Terminology

4.1.4.4 Correspondent Node (CN)

- A peer with which a mobile node is communicating.
- A (CN) may be either mobile or stationary.

4.1.4.5 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.

4.1.4.6 Foreign Network

Any network other than the MN's Home Network.

4.1.4.7 Home Address

- An IP address that is assigned for an extended period of time to an MN.
- It remains unchanged regardless of where the node is attached to the Internet.

4.1.4.8 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.
- 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 end-point and forwards packets to the MN.
    - **Co-located COA**: The COA is co-located if the MN temporarily acquired an additional IP address which acts as COA.

The example network in Figure 4.1 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.

4.1.5 IP Packet Delivery

Figure 4.2 illustrates packet delivery to and from the MN using the example network of Figure 4.1.

Step 1: A correspondent node CN wants to send an IP packet to 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. This means that CN sends an IP packet with MN as a destination address and CN as a source address.

Step 2: The HA now intercepts the packet. The packet is encapsulated and tunneled 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 3: 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 4: The MN sends the packet as usual with its own fixed IP address as source and CN's address as destination.

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 *Step 1*: through *Step 3*: would apply now in the other direction.

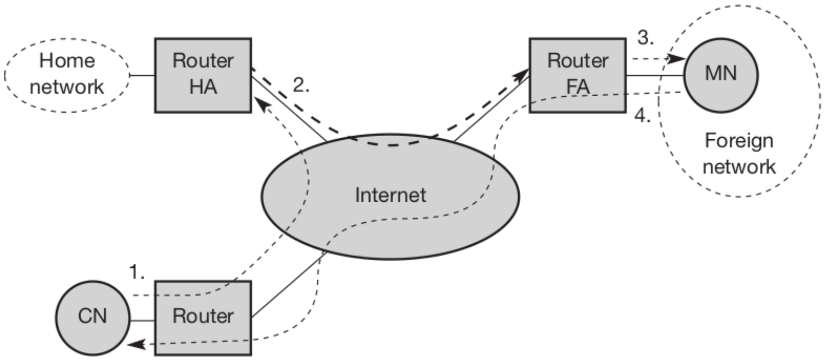


Figure 4.2: Packet delivery to and from the mobile node

Working of Mobile IP

Mobile IP has two addresses for a mobile host: *one home address* and *one care-of address.*

- The home address is permanent; the care-of addresses changes as the mobile host moves from one network to another.
- To make the change of address transparent to the rest of the Internet requires a home agent and a foreign agent.

- The specific function of an agent is performed in the application layer.
- When the mobile host and the foreign agent are the same, the care-of address is called a co-located care-of address.
- To communicate with a remote host, a mobile host goes through three phases:

    1. agent discovery,
    2. registration, and
    3. data transfer.

4.1.6 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.

4.1.6.1 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 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 with the extension for mobility is shown in Figure 4.3.

- The upper part represents the *ICMP packet*.
- 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

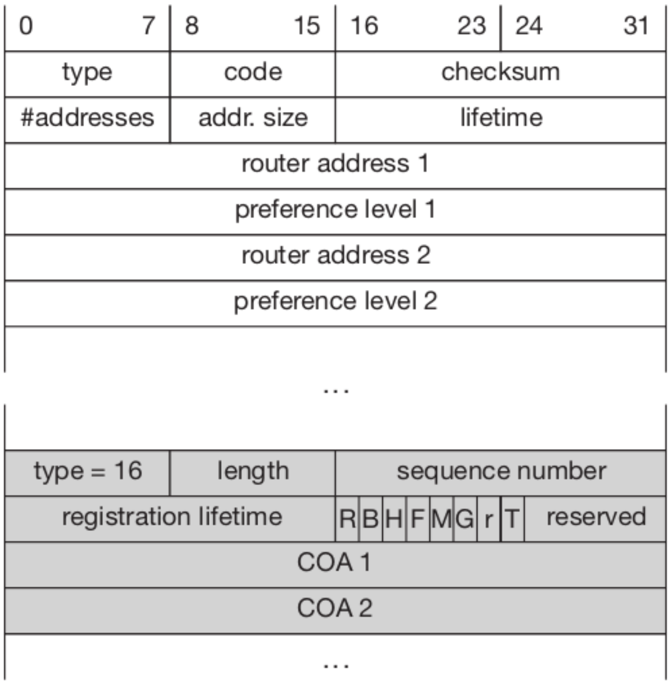


Figure 4.3: Agent advertisement packet (RFC 1256 + mobility extension)

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.

1. The **type** is set to 9.
2. 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.
3. **Lifetime** denotes the length of time this advertisement is valid.
4. Foreign agents are at least required to forward packets from the mobile node. The number of addresses advertised with this packet is in **#addresses**.
5. **Preference** levels for each address help a node to choose the router that is the most eager one to get a new node.

The difference compared with standard ICMP advertisements is what happens after the router addresses. This extension for mobility has the following fields defined:

1. **type** is set to 16.
2. **length** depends on the number of COAs provided with the message and equals $6 + 4 * (number of addresses)$.
3. An agent shows the total number of advertisements sent since initialization in the **sequence number**.
4. 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.

1. The **R** bit (registration) shows, if a registration with this agent is required even when using a co-located COA at the MN
2. If the agent is currently too busy to accept new registrations it can set the **B** bit
3. 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.
4. Bits **M** and **G** specify the method of encapsulation used for the tunnel.
5. **M** can specify minimal encapsulation and **G** generic routing encapsulation.
6. In the first version of mobile IP the **V** bit specified the use of header compression
7. Now the field **r** at the same bit position is set to zero and must be ignored.
8. 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.

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.

4.1.6.2 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, e. g. DHCP, the mobile node must send **agent 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.
- 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.
- Discovering a new agent can be done anytime, not just if the MN is not connected to one.
- 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.
- The next step for the MN is the registration with the HA if the MN is in a foreign network.

4.1.7 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.

4.1.7.1 COA at The FA

If the COA is at the FA, registration is done as illustrated in Figure 4.4 (left).

- The MN sends its registration request containing the COA (see Figure 4.5) 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 re-register 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.

4.1.7.2 COA is Co-located

If the COA is co-located, registration can be simpler, as shown in Figure 4.4 (right).

- The MN may send the request directly to the HA and vice versa.
- This 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.

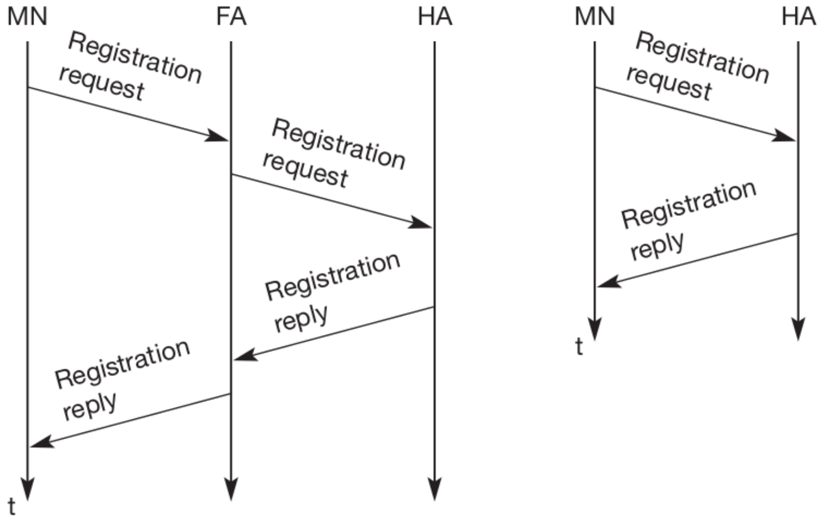


Figure 4.4: Registration of a mobile node via the FA or directly with the HA

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 (see Figure 4.6). The fields are defined as follows.

This allows for simultaneous bindings. The following bits denote the requested behavior for packet forwarding.

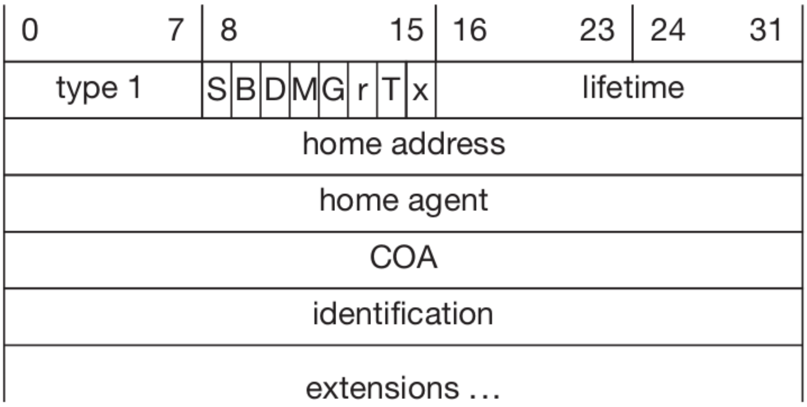


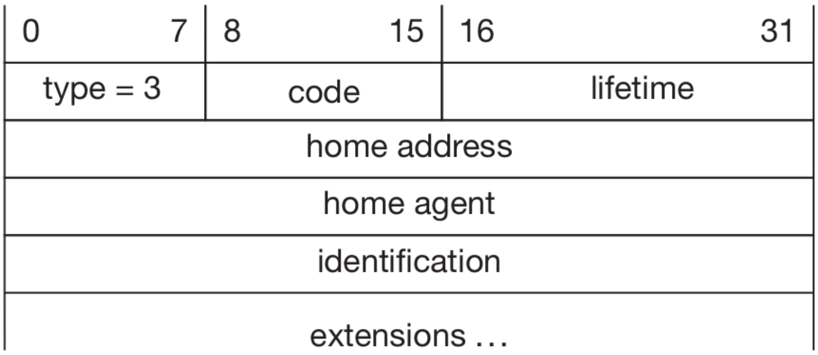
Figure 4.5: Registration request

1. The first field type is set to 1 for a registration request
2. With the **S** bit an MN can specify if it wants the HA to retain prior mobility bindings
3. 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
4. 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
5. **M** and **G** denote the use of minimal encapsulation or generic routing encapsulation, respectively
6. **T** indicates reverse tunneling
7. **r** and **x** are set to zero
8. **lifetime** denotes the validity of the registration in seconds
9. A value of zero indicates deregistration; all bits set indi-

cates infinity

10. The **home address** is the fixed IP address of the MN
11. **home agent** is the IP address of the HA, and COA represents the tunnel endpoint
12. 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.
13. The extensions must at least contain parameters for authentication

Registration Reply



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

Figure 4.6: Registration reply

A **registration reply**, which is conveyed in a UDP packet, contains:

1. a **type** field set to 3 and
2. a **code** indicating the result of the registration request.
3. **lifetime** field indicates how many seconds the registration is valid if it was successful.
4. **home address** and **home agent** are the addresses of the MN and the HA, respectively.
5. 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.
6. **extensions** must at least contain parameters for authentication.

4.1.8 Tunneling and Encapsulation

The following describes the mechanisms used for forwarding packets between the HA and the COA, as shown in Figure 4.2 *Step 2*: (**Figure 4.2 को Step 2: मा Tunneling र Encapsulation को कुरा बताइएको छ।**).

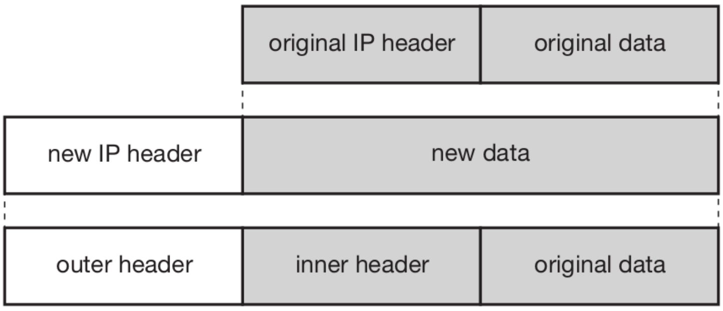


Figure 4.7: IP encapsulation

Tunnel

- 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

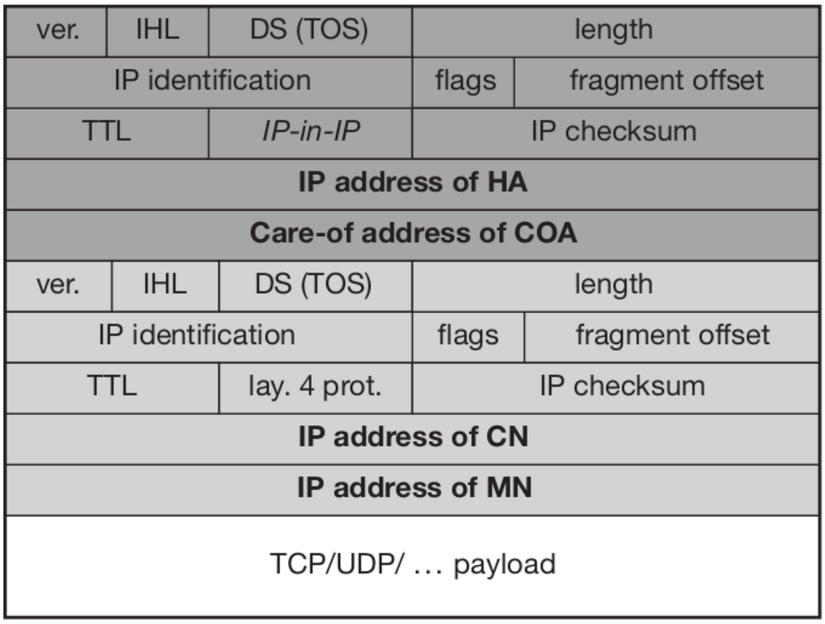
- **Encapsulation** is the mechanism of taking a packet consisting of packet header and data and putting it into the data part of a new packet.
- The reverse operation, taking a packet out of the data part of another packet, is called **decapsulation**.
- Encapsulation and decapsulation are the operations 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 4.7 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**.
- 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.

4.1.8.1 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**. Figure 4.8 shows a packet inside the tunnel.



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

Figure 4.8: IP-in-IP encapsulation

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.
- TTL have no special meaning for mobile IP.
- **TTL** must be high enough so the packet can reach the tunnel endpoint.
- field, **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 **IP address of the HA** is the tunnel entry as source address.
- **the COA** is the tunnel exit point as destination address

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.

4.1.8.2 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. **Minimal encapsulation** (shown in Figure 4.9) is an optional encapsulation method for mobile IP.

- 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.

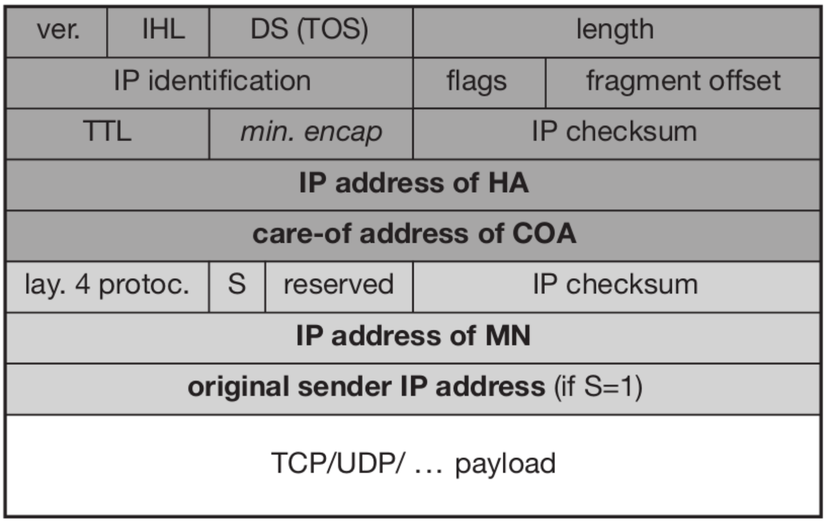


Figure 4.9: Minimal encapsulation

4.1.8.3 Generic Routing Encapsulation

While IP-in-IP encapsulation and minimal encapsulation work only for IP, the 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. Figure 4.10 shows this procedure.

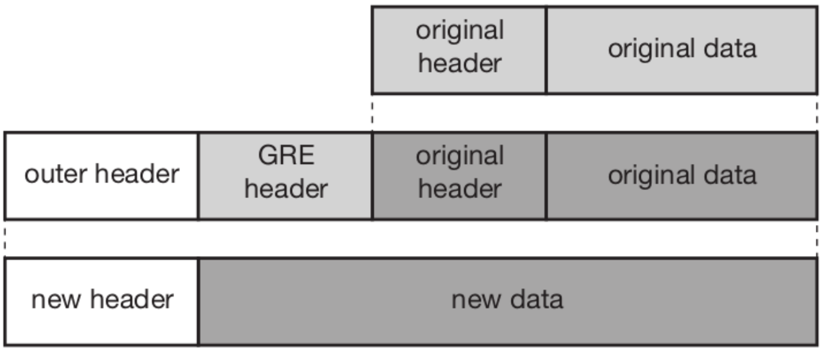
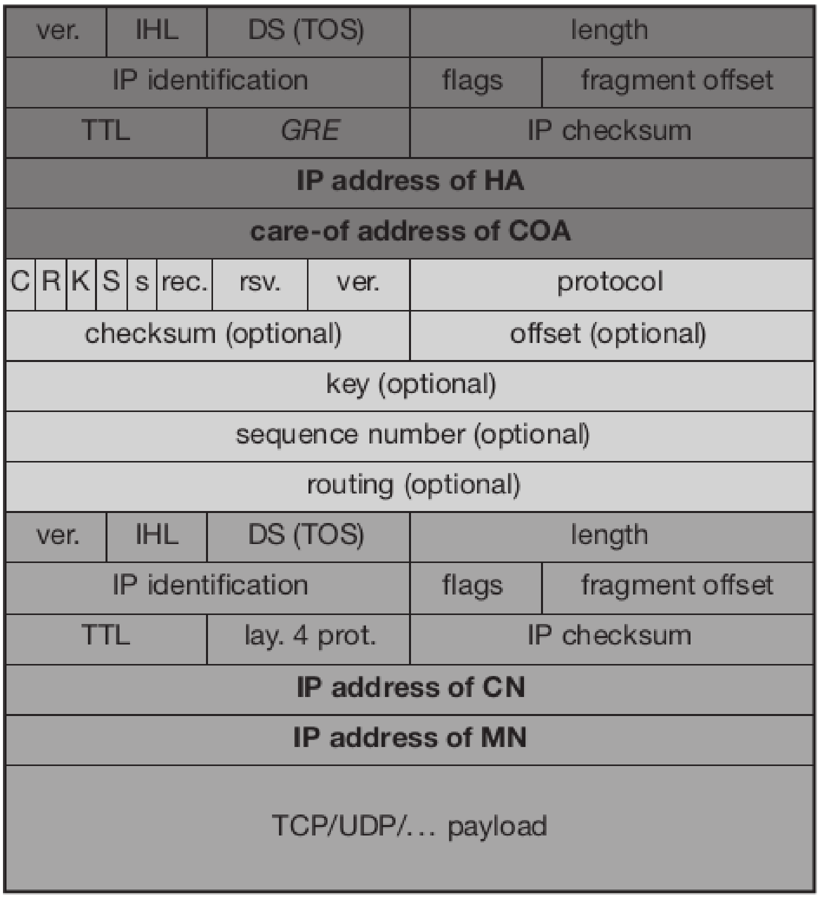


Figure 4.10: Generic routing encapsulation

- 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.

4.11 shows on the left side the fields of a packet inside the tunnel between home agent and COA using GRE as an encapsulation scheme. 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.



<table>
<tr><td colspan="3">ver.</td><td colspan="3">IHL</td><td colspan="2">DS (TOS)</td><td colspan="2">length</td></tr>
<tr><td colspan="8">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="6">TTL</td><td colspan="2"><i>GRE</i></td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="10"><b>IP address of HA</b></td></tr>
<tr><td colspan="10"><b>care-of address of COA</b></td></tr>
<tr><td>C</td><td>R</td><td>K</td><td>S</td><td>s</td><td>rec.</td><td>rsv.</td><td>ver.</td><td colspan="2">protocol</td></tr>
<tr><td colspan="8">checksum (optional)</td><td colspan="2">offset (optional)</td></tr>
<tr><td colspan="10">key (optional)</td></tr>
<tr><td colspan="10">sequence number (optional)</td></tr>
<tr><td colspan="10">routing (optional)</td></tr>
<tr><td colspan="3">ver.</td><td colspan="3">IHL</td><td colspan="2">DS (TOS)</td><td colspan="2">length</td></tr>
<tr><td colspan="8">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="6">TTL</td><td colspan="2">lay. 4 prot.</td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="10"><b>IP address of CN</b></td></tr>
<tr><td colspan="10"><b>IP address of MN</b></td></tr>
<tr><td colspan="10">TCP/UDP/... payload</td></tr>
</table>

Figure 4.11: Protocol fields for GRE

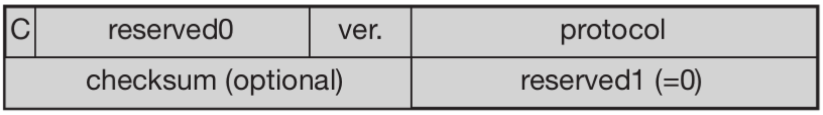
The GRE header starts with several flags indicating if certain fields are pre sent or not. A minimal GRE header uses only 4 bytes; nevertheless, GRE is flexible enough to include several mechanisms in its header. The C bit indicates if the checksum field is present and contains valid information.

- 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.
- **key** field may be used for authentication. If this field is present, the **K** bit is set.
- The **sequence** number bit **S** indicates if the sequence number field is present, if the s bit is set, strict source routing is used.
- The **recursion control** field (**rec.**) distinguishes GRE from IP-in-IP and minimal encapsulation.
- The **reserved** fields must be zero and are ignored on reception.
- The **version** field contains 0 for the GRE version.
- The 2 byte **protocol** field represents the protocol of the packet following the GRE header.

Figure 4.12 shows the simplified header of GRE, which is a more generalized version of GRE.

This version does not address mutual encapsulation and ignores several protocol-specific nuances on purpose.

- The field **C** indicates 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, defines the protocol of the payload.
- If the flag C is set, then **checksum** field and a field called **reserved1** follows.
- The reserved1 field is constant zero set to zero follow.



<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 4.12: Simplified header of GRE

4.1.9 Optimizations

Imagine the following scenario. A Nepalese and a Chinese meet at a conference on North Korea. Both want to use their laptops for exchanging data, both run mobile IP for mobility support. Now recall Figure 4.2 and think of the way the packets between both computers take.

- If the Nepalese sends a packet to the Chinese, his computer sends the data to the HA of the Chinese, i. e., from North Korea to China.
- The HA in China now encapsulates the packets and tunnels them to the COA of the Chinese laptop on North Korea.
- This means that although the computers might be only meters away, the packets have to travel around the world! This inefficient behavior of a non-optimized 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.

Triangle Routing Problem is considered as one of the problems facing the implementation of Mobile IP. When a CN sends traffics to MN, the following sequence must be done.

1. CN sends packets to the HA.
2. HA encapsulates these packets and tunnels them to the FA.
3. The FA de-tunnels the packets and delivers them to the Mobile Node.

This behavior is known as triangular routing. As shown in Figure 4.13, the route taken by these packets is triangle in nature, and the most extreme case of routing can be observed when the Correspondent Node and Mobile Node are in the same subnet.

- One way to optimize the route is to inform the CN of the current location of the MN.

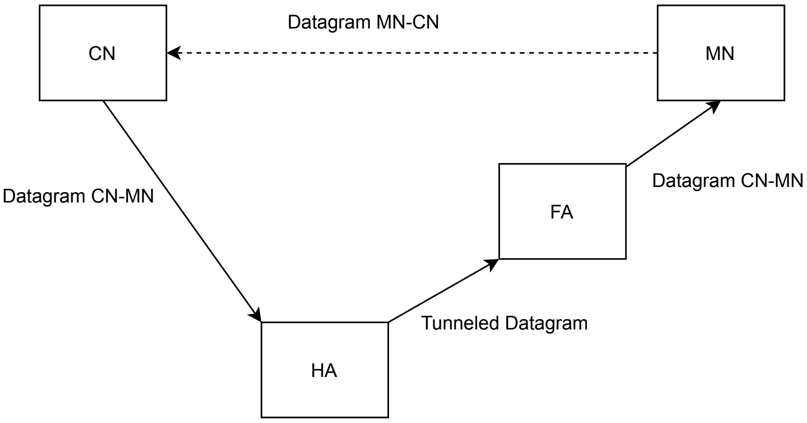


Figure 4.13: Triangular routing

- 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.

1. **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.
2. **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 acknowledgment.
3. **Binding acknowledgment**: If requested, a node returns this acknowledgment after receiving a binding update message.
4. **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 4.14 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}.

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.

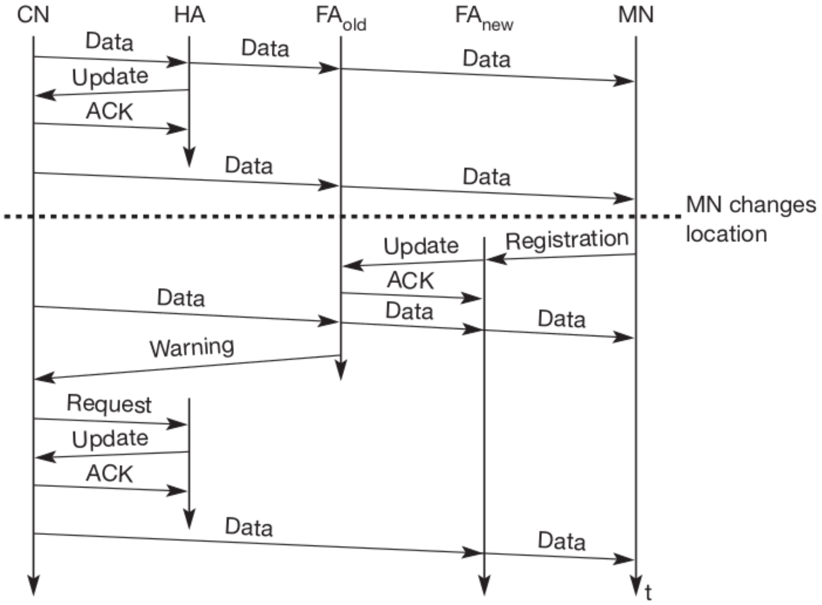


Figure 4.14: Change of the foreign agent with an optimized mobile IP

To tell CN that it has a stale binding cache:

1. FA_{old} sends, a binding warning message to CN.
2. CN then requests a binding update. (The warning could also be directly sent to the HA triggering an update).
3. The HA sends an update to inform the CN about the new location, which is acknowledged.
4. Now CN can send its packets directly to FA_{new}, again avoiding triangular routing.

Problems with this optimization

- Security problems such as *tunnel hijacking.*
- All users do not want to reveal their current location to others.

4.1.10 Dynamic Host Configuration Protocol (DHCP)

- DHCP is an automatic configuration protocol used on IP networks.
- DHCP allows a computer to join an IP-based network without having a pre-configured IP address.
- DHCP is a protocol that assigns unique IP addresses to devices, then releases and renews these addresses as devices leave and re-join the network.
- 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.

DHCP is based on a client/server model as shown in Figure 4.15. DHCP clients send a request to a server (DHCPDISCOVER in the example) 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.

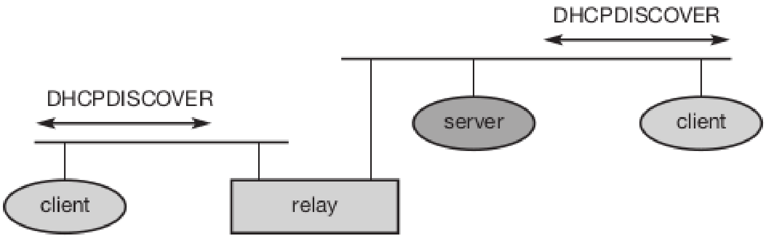


Figure 4.15: Basic DHCP configuration

A typical initialization of a DHCP client is shown in Figure 4.16. The figure shows one client and two servers.

- 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.

Advantages of DHCP

DHCP provides the following benefits.

Reliable IP Address Configuration DHCP minimizes configuration errors caused by manual IP address configuration, such as typographical errors, or address conflicts caused by the assignment of an IP address to more than one computer at the same time.

Reduced Network Administration DHCP includes the following features to reduce network administration:

- Centralized and automated TCP/IP configuration.
- The ability to define TCP/IP configurations from a central location.
- The ability to assign a full range of additional TCP/IP configuration values by means of DHCP options.

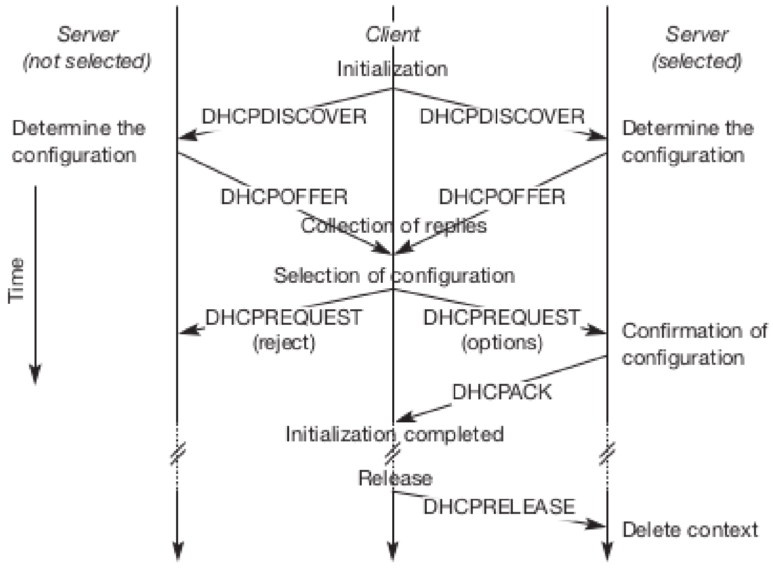


Figure 4.16: Client initialization via DHCP

- The efficient handling of IP address changes for clients that must be updated frequently, such as those for portable devices that move to different locations on a wireless network.
- The forwarding of initial DHCP messages by using a DHCP relay agent, which eliminates the need for a DHCP server on every subnet.

CHAPTER

5

MOBILE TRANSPORT LAYER

Supporting mobility only on lower layers up to the network layer is not enough to provide mobility support for applications. Most applications rely on a transport layer, such as TCP or User Datagram Protocol (UDP) in the case of the Internet.

Two functions of the transport layer in the internet are:

- checksumming over user data and
- multiplexing/demultiplexing of data from/to applications.

5.1 Traditional TCP

Several mechanisms of the TCP that influence the efficiency of TCP in a mobile environment are:

5.1.1 Congestion Control

- 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.

- Most of the hardware/software is not responsible for lost packets or bits flipping.
- The probable reason for a packet loss in a fixed network is a *state of congestion at a node*.
- Router drops packets when the packet buffers of a router are filled and router cannot forward the packets fast enough because sum of input rates of packets destined is higher than the capacity of the output.
- A dropped packet is lost for the transmission, and the receiver notices a gap in the packet stream.
- Receiver does not tell sender if packet is missing, bit continues to acknowledge packets in sequence up to the missing one.
- When sender notices missing acknowledgment, it assumes a packet loss due to congestion and re-transmits the missing packet at full speed. This increases congestion.
- To mitigate congestion, TCP slows down the transmission rate dramatically.

5.1.2 Slow Start

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 acknowledgment.
- If this acknowledgment arrives, the sender increases the congestion window by one, now sending two packets (congestion window = 2).
- This scheme doubles the congestion window every time the acknowledgments come back, which takes Round Trip Time (RTT).
- This is called the exponential growth of the congestion window in the slow start mechanism.

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 acknowledgments come back.

Linear increase continues until a time-out at the sender occurs due to a missing acknowledgment, or until the sender detects a gap in transmitted data because of continuous acknowledgments for the same packet. In either case the sender sets the congestion threshold to half of the current congestion window.

5.1.3 Fast Re-transmit/Fast Recovery

Two things lead to a reduction of the congestion threshold:

- One is a sender receiving continuous acknowledgments 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 acknowledgments only if it receives any packets from the sender. Receiving acknowledgments 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 re-transmit the missing packet(s) before the timer expires. This behavior is called *fast re-transmit*.
  - The receipt of acknowledgments 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.
- The other reason for activating slow start is a time-out due to a missing acknowledgment.

5.1.4 Implications on Mobility

- *Slow start* decreases the efficiency of TCP if used together with mobile receivers or senders. The reason being the use slow start under the wrong assumptions.
- Error rates on wireless links are orders of magnitude higher compared to fixed fiber or copper links.

- Mobility can cause packet loss. There are many situations where a soft handover from one access point to another is not possible for a mobile end system.
- Standard TCP reacts with slow start if acknowledgments 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.

5.2 Classical TCP Improvements

Together with the introduction of Wireless Local Area Network (WLAN)s in the mid-nineties several research projects were started with the goal to increase TCP's performance in wireless and mobile environments.

5.2.1 Indirect TCP (I-TCP)

Two competing insights led to the development of I-TCP:

- TCP performs poorly together with wireless links.
- TCP within the fixed network cannot be changed.

I-TCP segments a TCP connection into:

a) a fixed part and
b) a wireless part.

Figure 5.1 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.

- 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.

- 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 moves on.

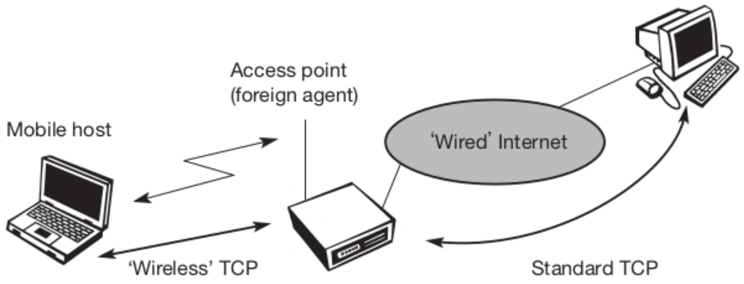


Figure 5.1: Indirect TCP segments a TCP connection into two parts

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 acknowledgment 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 re-transmit 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 re-transmit 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 (see Figure 5.2).
- In the example shown, the access point acts as a proxy buffering packets for re-transmission.
- After the handover, the old proxy must forward buffered data to the new proxy because it has already acknowledged the data.

- 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 connection, 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.

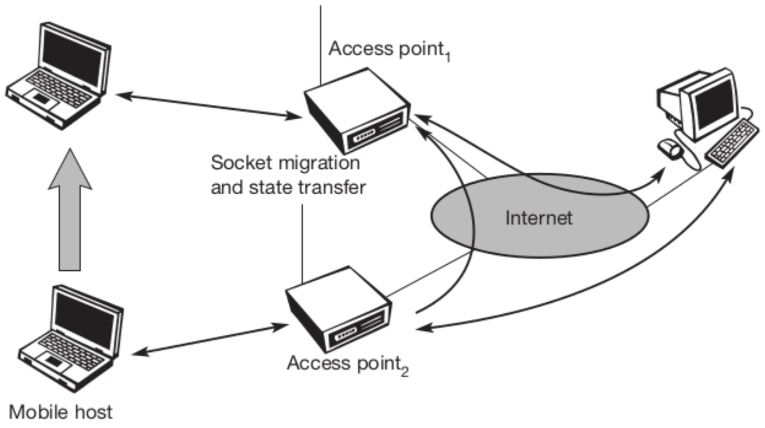


Figure 5.2: Socket and state migration after handover of a mobile host

5.2.1.1 Advantages With I-TCP

There are several advantages with I-TCP:

- I-TCP does not require any changes in the TCP protocol.
- All current optimizations for TCP still work between the foreign agent and the corresponding 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.
- Introduction to new mechanism into a huge network is always dangerous. I-TCP ensures Different solutions can be tested or used at the same time without jeopardizing the stability of the network.

- Optimizing new mechanism is simple because they only cover one single hop.
- An optimized TCP could use precise timeouts to guarantee retransmission as fast as possible.
- Partitioning into two connections also allows the use of a different transport layer protocol between the foreign agent and the mobile host.

5.2.1.2 I-TCP Disadvantages

Idea of segmentation in I-TCP also comes with some disadvantages:

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

Handover Latency 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.

Security Mechanism 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. the foreign agent has to be integrated into all security mechanisms.

5.2.2 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 re-transmission in case of packet loss.

- In this approach, the foreign agent buffers all packets with destination mobile host and additionally *snoops* the packet flow in both directions to recognize acknowledgments.
- The reason for buffering packets toward the mobile node is to enable the foreign agent to perform a local re-transmission in case of packet loss on the wireless link.

- The foreign agent buffers every packet until it receives an acknowledgment from the mobile host.
- If the foreign agent does not receive an acknowledgment from the mobile host within a certain amount of time, either the packet or the acknowledgment has been lost.
- Alternatively, the foreign agent could receive a duplicate ACK which also shows the loss of a packet.
- Now the foreign agent re-transmits the packet directly from the buffer, performing a much faster re-transmission compared to the correspondent host.
- The time out for acknowledgments can be much shorter, because it reflects only the delay of one hop plus processing time.

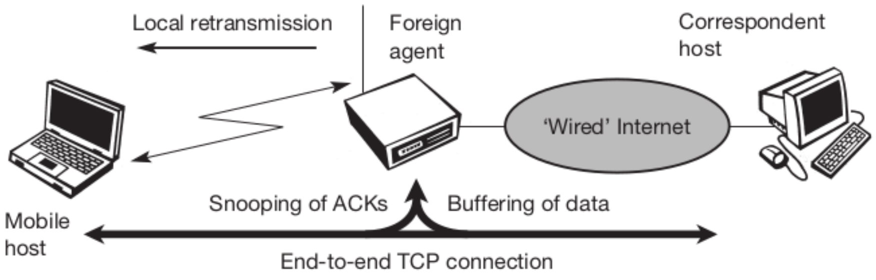


Figure 5.3: Snooping TCP as a transparent TCP extension

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 acknowledgment (NACK) to the mobile host.
- The mobile host can now re-transmit the missing packet immediately.
- Reordering of packets is done automatically at the correspondent host by TCP.

5.2.2.1 Advantages of Snooping TCP

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

Preserve End-to-End Semantics The end-to-end TCP semantic is preserved. No matter at what time the foreign agent crashes, 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.

Change in Correspondent Host Not Required The correspondent host does not need to be changed; most of the enhancements 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.

Immediate Handover of State Not Required 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 re-transmission of the packets, possibly already to the new care-of address.

Automatic Fallback to Standard Solution 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.

5.2.2.2 Disadvantages of Snooping TCP

The simplicity of the scheme also results in some disadvantages:

No Proper Wireless Link Isolation Snooping TCP does not isolate the behavior of the wireless link as well as I-TCP. 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.

Lack of Transparency Using negative acknowledgments 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.

Incompatible With Encryption 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.

5.2.3 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.

The **M-TCP (mobile TCP)**[1] approach has the same goals as I-TCP and snooping TCP:

- to prevent the sender window from shrinking if bit errors or disconnection but not congestion cause current problems.

M-TCP wants to improve overall throughput:

- to lower the delay,
- to maintain end-to-end semantics of TCP, and
- to provide a more efficient handover.

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

M-TCP splits the TCP connection into two parts as I-TCP does.

- An unmodified TCP is used on the standard host-supervisory host (SH) connection,
- while an optimized TCP is used on the SH-MH connection.

The supervisory host is responsible for exchanging data between both parts similar to the proxy in I-TCP (see Figure 5.1). The M-TCP approach assumes a relatively low bit error rate on the wireless link. Therefore, it does not perform caching/re-transmission of data via the SH. If a packet is lost on the wireless link, it has to be re-transmitted 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.

[1] Mobile TCP does not have the same status as mobile IP, which is an internet RFC.

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

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

5.2.3.1 Advantages of M-TCP

The advantages of M-TCP are the following:

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

Avoids Useless Re-transmissions If the MH is disconnected, it avoids useless re-transmissions, slow starts or breaking connections by simply shrinking the sender's window to 0.

Does Not Buffer Data Since it does not buffer data in the SH as I-TCP does, it is not necessary to forward buffers to a new SH. Lost packets will be automatically retransmitted to the new SH.

5.2.3.2 Disadvantages of M-TCP

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

Does Not Act as a Proxy 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.

Required Modification to Network Elements 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.

5.2.4 Fast Re-transmit/Fast Recovery

Moving to a new foreign agent can cause packet loss or time out at mobile hosts or corresponding hosts. TCP concludes congestion and goes into slow start, although there is no congestion.

The idea presented by 'Caceres' is to artificially force the fast re-transmit behavior on the mobile host and correspondent host side. As soon as the mobile host registers at a new foreign agent using mobile IP, it starts sending duplicated acknowledgments to correspondent hosts. The proposal is to send three duplicates. This forces the corresponding host to go into fast re-transmit 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 foreign agent, this approach additionally puts the mobile host into fast re-transmit. The mobile host re-transmits all unacknowledged packets using the current congestion window size without going into slow start.

5.2.4.1 Advantages

Simplicity 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.

5.2.4.2 Disadvantages

Insufficient Isolation of Packet Losses The main disadvantage of this scheme is the insufficient isolation of packet losses. Forcing fast re-transmission increases the efficiency, but re-transmitted 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 re-transmission.

5.2.5 Transmission/Time-out Freezing

- Quite often, the MAC layer has already noticed connection problems, before the connection is actually interrupted from a TCP point of view. Additionally, the MAC layer knows the real reason for the interruption and does not assume congestion.
- 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.

5.2.5.1 Advantages

The advantages of this approach:

Resumes TCP Connection It offers a way to resume TCP connections even after longer interruptions of the connection.

Works With Encrypted Data It is independent of any other TCP mechanism, such as acknowledgments or sequence numbers, so it can be used together with encrypted data.

5.2.5.2 Disadvantages

However, this scheme has some severe disadvantages.

Lots of Changes in Software Not only does the software on the mobile host have to be changed, to be more effective the correspondent host cannot remain unchanged.

Reliance on MAC Layer All mechanisms rely on the capability of the MAC layer to detect future interruptions.

Incompatible With Some Encryption Schemes Freezing the state of TCP does not help in case of some encryption schemes that use time-dependent random numbers.

Requires Re-synchronization These schemes need re-synchronization after interruption.

5.2.6 Selective Re-transmission

TCP acknowledgments 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 re-transmit everything starting from the lost packet (go-back-n re-transmission). 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 re-transmission 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 re-transmit it.

5.2.6.1 Advantages

The advantage(s) of this approach:

Lowers Bandwidth Requirements A sender re-transmits only the lost packets. This lowers bandwidth requirements and is extremely helpful in slow wireless links.

Not Restricted to Mobile Environments The gain in efficiency is not restricted to wireless links and mobile environments. Using selective re-transmission is also beneficial in all other networks.

5.2.6.2 Disadvantages

Minor Disadvantage(s):

More Complex Software Requires more complex software on the receiver side, because now more buffer is necessary to re-sequence data and to wait for gaps to be filled.

Cannot Extract Maximum Performance While memory sizes and CPU performance permanently increase, the bandwidth of the air interface remains almost the same.

5.2.7 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.

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.
- Connections with a long duration is not a problem with TCP.
- But in an example of only one data packet, TCP may need seven packets altogether.

Figure 5.4 shows an example for the overhead introduced by using TCP over General Packet Radio Service (GPRS) in a web scenario. Web services are based on HyperText Transfer Protocol (HTTP) which requires a reliable transport system. In the internet, TCP is used for this purpose. Before a 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 $500ms$ and more are quite common. The setup of a TCP connection already takes far more than a second.

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

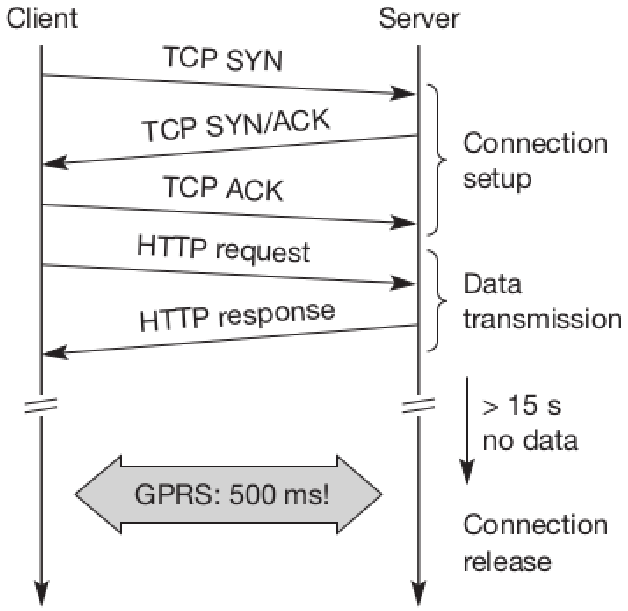


Figure 5.4: Example TCP connection setup overhead

- 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.

5.2.7.1 Advantages

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

5.2.7.2 Disadvantages

Requires Changes in MH and all CH 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.

Mobility No Longer Transparent This solution no longer hides mobility.

Exhibits Several Security Problems Furthermore, T/TCP exhibits several security problems.

5.3 Overview of Classical Enhancements to TCP for Mobility

Table 5.1: Overview of classical enhancements to TCP for mobility.

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 revcovery	Avoids slow-start after roaming	Simple and efficient	Mixed layers, not transparent

Continued on next page ...

Table 5.1 – *Continued from previous page*

Approach	Mechanism	Advantages	Disadvantages
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 re-transmission	Retransmits only lost data	Very effficient	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

5.4 TCP over 2.5G/3G Wireless Networks

The TCP over 2.5G/3G wireless networks describes a profile for optimizing TCP over wireless Wireless Local Area Network (WAN)s such as GSM, GPRS, Universal Mobile Telecommunication System (UMTS). The focus on 2.5G/3G for transport of internet data is important as already billions of people use mobile phones and it is obvious that the mobile phone systems will also be used to transport arbitrary internet data.

5.4.1 Characteristics to be Considered

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

5.4.1.1 Data Rates

Typically, data rates are asymmetric as it is expected that users will download more data compared to uploading. Uploading is limited by the limited battery power. In cellular networks, asymmetry does not exceed 3–6 times, however, considering broadcast systems as additional distribution media (digital radio, satellite systems), asymmetry may reach a factor of 1,000.

5.4.1.2 Latency

All wireless systems comprise elaborated algorithms for error correction and protection, such as forward error correction (FEC), checksumming, and interleaving. FEC and interleaving let the RTT grow to several hundred milliseconds up to some seconds.

5.4.1.3 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 10ms to several seconds.

5.4.1.4 Packet Loss

Packets might be lost during handovers or due to corruption. Due to link-level re-transmissions 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.

5.4.2 Configuration Parameters to Adapt TCP to Wireless Environments

Based on the characteristics (see Section 5.4.1), configuration parameters to adapt TCP to wireless environments are:

5.4.2.1 Large Windows

TCP should support large enough window sizes based on the bandwidth delay experienced in wireless systems. With the help of the windows scale option and larger buffer sizes this can be accomplished. A larger initial window segments may increase performance particularly for short transmissions.

5.4.2.2 Limited Transit

This mechanism is an extension of Fast Re-transmission/Fast Recovery and is particularly useful when small amounts of data are to be transmitted (standard for, e. g. web service requests).

5.4.2.3 Large MTU

The larger the MTU (Maximum Transfer Unit) the faster TCP increases the congestion window. Large MTUs may be used to increase performance. MTU path discovery should be used to employ larger segment sizes instead of assuming the small default MTU.

5.4.2.4 Selective Acknowledgment (SACK)

SACK allows the selective re-transmission of packets and is almost always beneficial compared to the standard cumulative scheme.

5.4.2.5 Explicit Congestion Notification (ECN)

ECN 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.

5.4.2.6 Timestamp

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.

5.4.2.7 No Header Compression

Header compression is not compatible with TCP options such as SACK or timestamps.

PURBANCHAL UNIVERSITY

2018/ २०७५

4 Years Bachelor of Computer Application (BCA/Eighth Semester/Final)

Time: 3.00 hrs. Full Marks: 60 / Pass Marks: 24

BCA454WN, Wireless Network and Mobile Computing (Elective-II)

Candidates are required to give their own answers in their own words as far as practicable. Figure in the margin indicate full marks.

Group A

Answer TWO questions. **$2 \times 12 = 24$**

1. Explain in detail UMTS architecture, features and handover with neat sketch.
2. Explain in detail about Ad-hoc networks and its routing strategies with neat sketch.
3. Draw and explain the architecture of mobile computing and also discuss on the mobility management.

Group B

Answer SIX questions. **$6 \times 6 = 36$**

4. Describe the phases of HIPERLAN1.
5. Explain the user scenarios of Bluetooth transmission.
6. Discuss on any one type of encapsulation techniques used in mobile IP schemes.
7. Discuss on the features and services provided by GSM along with its security issues.
8. Explain IPV4. What are the advantages of IPV6 over IPV4?
9. What is mobile computing? Explain the transaction oriented TCP with an example.
10. Write short notes on any TWO:
    (a) DHCP
    (b) Mobile Devices
    (c) Infrared Vs Radio Transmission

PURBANCHAL UNIVERSITY

2019/ २०७६

4 Years Bachelor of Computer Application (BCA/Eighth Semester/Final)

Time: 3.00 hrs. Full Marks: 60 / Pass Marks: 24

BCA454WN, Wireless Network and Mobile Computing (Elective-II)

Candidates are required to give their own answers in their own words as far as practicable. Figure in the margin indicate full marks.

Group A

Answer TWO questions. **$2 \times 12 = 24$**

1. Explain in detail about IEEE 802.11 protocol architecture, features and applications.
2. Draw the basic architecture of GSM and explain in detail about its subsystems along with its interfaces.
3. What is Mobile Computing? Draw and explain the architecture of mobile computing with various applications of it.

Group B

Answer SIX questions. **$6 \times 6 = 36$**

4. Describe the characteristics of an Ad-hoc networks..
5. What are the basic purposes of a DHCP? Explain.
6. Explain the features of tunneling and encapsulation.
7. What is mobile TCP? Explain. Also point out demerits of conventional TCP incorporated in wireless environment.
8. What is mobile IP? Describe the process of IP packet delivery with neat sketch.
9. Explain the various generations of wireless networks with examples.
10. Write short notes on any TWO: 3+3
    (a) LTE and Bluetooth
    (b) Mobility Management
    (c) Handoff Process in GSM

BIBLIOGRAPHY

Uprety, M. (2020). Wireless Networks and Mobile Communication: BCA-VIII. *Gomendra Multiple College.*

Schiller, J. H. (2003). *Mobile Communications* (Second Edition). Pearsn Education Limited.

Perkins, C. (1996). IP Mobility Support. Retrieved February 13, 2021, from https://tools.ietf.org/html/rfc2002

Dynamic Host Configuration Protocol (DHCP). (2020). Retrieved February 14, 2021, from https://docs.microsoft.com/en-us/windows-server/networking/technologies/dhcp/dhcp-top

Switching techniques. (n.d.). https://www.javatpoint.com/computer-network-switching-techniques

Kamel Hussein, S. H. A., Imane Aly Saroit Ismail. (n.d.). Triangle routing problem in mobile ip.

What is 5G | Everything You Need to Know About 5G | 5G FAQ. (2017). Retrieved February 18, 2021, from https://www.qualcomm.com/invention/5g/what-is-5g