Mobile Communications for the working software engineer

A collection of notes from self-studying mobile communication.

Author: Kim Hammar

kimham@kth.se

2nd July 2017

"WHERE TO START!?"

If you're reading this, chances are that you have a background of a computer-engineering degree which covered everything from networking and software to logic and math but they didn't touch upon mobile communications. And now you have found yourself in a situation where you must learn the broad and daunting area of mobile communications, maybe because your job requires it, or maybe because you're simply curious. Where to even start?! look no further, this document serves as a fast-pasted introduction to a few of the most fundamental areas of mobile communication.

Kim Hammar, July 2017

Contents

1 Introduction

Mobile communication is known as communication where the devices are communicating without the use of any wired physical connection as well as being able to continue communication while moving over larger areas. It's not a wild guess that the future of computing will be moving more and more towards mobile and portable computing devices, if you could have your mobile/portable device do the same thing as your non-portable device and the added benefit of portability which one would you choose?

> *The vision of wireless communications supporting information exchange between people or devices is the communications frontier of the next century. This vision will allow people to operate a virtual office anywhere in the world using a small handheld device - with seamless telephone, modem, fax, and computer communications*[3]

Unfortunately with current technology, both the technology for communication (less bandwidth and higher delays and variation than wired technology) as well as the manufacturing of devices (space limitation of portable devices limits the power and user-experience) is holding this vision back and current mobile communication technologies cannot yet offer this ideal.

The ultimate portability in terms of mainstream computing devices today is the mobile phone system, where the system itself moves the device connection between different base stations (radio transmitters) as required, without the end-user having to think about it.

I mentioned that mobile communication is typically also wireless, but it's important not to mix these two properties. Wireless is a property denoting how a computing device is accessing the network (with wired technology or wireless (through electromagnetic waves)), while mobile is a more abstract property of the communication denoting that the user is not restricted to a certain area for communication and can move over larger areas whilst keeping the mobile connection.

> *While traditional communication paradigms deal with fixed networks, mobility raises a new set of questions, techniques, and solutions. For many countries, mobile communication is the only solution due to the lack of an appropriate fixed communication infrastructure.*[8]

As mentioned, there are some great benefits of mobile communication technology that is driving the development, but also some limitations. Below are some limitations currently holding the mobile technology back.

I **Low bandwidth** - Transmission rates for wireless devices is still a lot lower than for the wired counterpart.

II **Security vulnerability** - Wireless communication is known to be simpler for malicious users to attack.

III **Shared medium** - Radio access is always realised via a shared medium. As it is impossible to have a separate wire between a sender and each receiver, different competitors have to compete for the medium [8].

IV **Unreliability** - As a signal propagates through a wireless channel, it experiences random fluctuations in time if the transmitter or receiver is moving, due to changing reflections and attenuation. Thus, the characteristics of the channel appear to change randomly with time, which makes it difficult to design reliable systems with guaranteed performance[3].

2 Low-level: Wireless transmission

Just as traditional networking, mobile communication is broad and includes both the low-level transmission details which relate to electrical engineering as well as higher-level details more related to the field of computer science. This section will just give a brief overview of how the wireless transmission works on an abstract level, mainly comparing it to the wired counterpart.

Wireless transmission refers to systems which use some form of energy and signal, like radio waves, to transfer information without the use of wires. Examples of such systems are radio transmitters and receivers, remote controls, repeaters and alike.

2.1 Electrical Engineering

Wireless transmission In wireless transmission, space is the ultimate medium and communication is transferred through electromagnetic waves travelling through the air. Multiple variants of wireless technologies exists, e.g mobile phones, satellite signals, Wifi, Bluetooth, GPS etc. Wireless signals occupy a range of frequencies which denotes the rate at which a signal vibrates, which is measured in Hertz. Different types of wireless technologies occupy different frequency ranges. A device sending out wireless signals is called a transmitter, a device receiving wireless signals is called a receiver, a device both sending and receiving wireless signals is called a transceiver. A mobile phone is an example of a transceiver while an old-school radio-player typically only acts as a receiver and is never sending out any signals.

Signals and Frequency Radio waves are a type of electromagnetic radiation with wavelengths in the electromagnetic spectrum longer than infrared light. It is common to talk about frequencies of radio waves as that together with wavelength is what distinguishes waves. Frequency refer to the number of oscillations per a time unit. One typically say that radio waves covers a spectrum based on its frequency and wavelength.

Antennas Wireless transmission implies that the communication uses air as a medium, no wires are necessary, but for the communication to be useful it is necessary to convert the received signals into bits that our devices can use and vice verse to convert messages encoded in bits to radio waves ready for transmission. This is the task of *antennas*, which mobile devices are equipped with. One can think of the antennas as the necessary equipment for wireless transmission through the transmission and reception of electromagnetic waves, as opposed to physical wires as the required equipment for wired transmission.

Reaching the receiver, signal propagation A major difference of wireless transmission compared to wired is the diversity of the environment, in a wired transmission the signals travel through a homogeneous wire of some sort (copper, fiber, coax etc.), in wireless transmission the signals travel through a heterogeneous environment and has to deal with things such as large buildings, mountains, moving senders etc. For short distances such as in LAN's, travelling through different objects like for instance a brick-wall might not have a very big impact on the delay but for larger distances this can be a very serious concern for the ability to propagate the signal. Different type of radio signals have easier to travel through objects, e.g signals with lower frequency is known to better penetrate through objects.

Ultimately the speed of which radio signals can travel is limited to the speed of light, which is finite. It is also important to note that wireless signals can take different routes. For radio signals to reach its destination it typically needs to be intercepted and relayed multiple times. Further more a signal might only be transmitted a small part of the total distance as a radio-signal before it gets converted into a signal that can be transmitted through a wire for the rest of the distance. Different signals are transmitted at different frequencies and receivers pick out only the ones they are interested in.

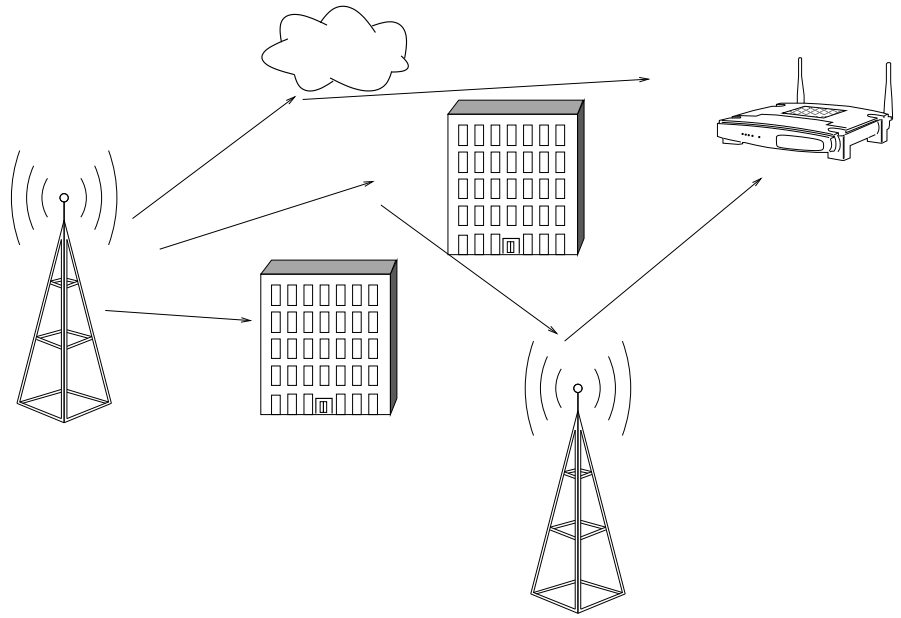

Figure 1: Wireless transmission of radio signals

Multiplexing Multiplexing is a way of combining multiple signals into one and transfer it over some shared medium, the obvious motivation of multiplexing is more efficient use of resources through sharing. For multiplexing it is important to minimise the interference the signals that share the medium has on each other. Multiplexing is ubiquitous in today's wireless communication.

Spread spectrum Spread spectrum is a technique to spread the bandwidth of a signal into a larger frequency domain, this can reduce the vulnerability of inferences.

Medium Access control (MAC) With cellular systems multiple users in the same cell will share the medium for communication with a base station. To accommodate multiple users, a mechanism for controlling the access to the medium is necessary. Medium Access Control (MAC) is part of the data link layer of the OSI model. You might be familiar with the concept of MAC-address which is a 48-bit unique identifier assigned to the network interface of networking devices, this identifier is used heavily when communicating at the data link layer, not least when using the Ethernet protocol for wired transmission. The purpose of the MAC layer and its associated algorithms is to regulate user access to the medium used. The most elaborated MAC scheme for wired transmission is called *carrier sense multiple access with collision detection (CSMA/CD)*. CSMA/CD controls the access to the medium by letting the transmitter decide when to transmit and when to stop by sensing the connection and trying to detect collisions.

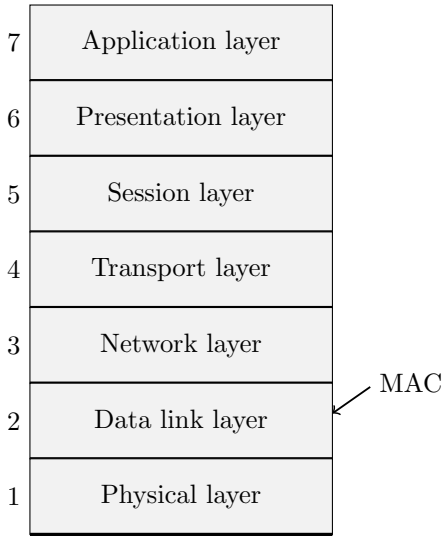

Figure 2: OSI model

CSMA/CD is more or less designed specifically for wired transmission and as such it is not surprising that CSMA/CD is not as effective for wireless transmission. The main problem with CSMA/CD for wireless transmission is that it assumes that the signal strength is close to the same over the whole wire which is convenient for detecting collisions in the transmission. This luxury is not present for wireless transmission where the strength of a signal varies greatly and decreases proportionally to the square of the distance to the sender[8]. How the MAC algorithms for wireless communication works is out of scope for this

document but it's good to be aware that it differs from medium access control for wired communication and that it presents certain challenges.

2.2 Cellular systems

With the increase in the number of wireless users in the same area, accommodating them within the limited available frequency spectrum becomes a major problem[6]. To resolve this problem, the concept of cellular communication was evolved. The present day cellular communication uses a basic unit called *cell*. Each cell consists of a small hexagonal area with a base station located at the center of the cell which communicates with the user[6].

The basic premise behind cellular system design is frequency reuse, which exploits path loss to reuse the same frequency spectrum at spatially-separated locations. Specifically, the coverage area of a cellular system is divided into non-overlapping cells where some set of channels is assigned to each cell[3].

So as mentioned, major rationales behind cellular systems are to enable radio coverage over larger geographic area without requiring single powerful radio base stations to transmit signals, but rather distribute the transmission over multiple radio base stations, each covering a cell in the network. These stations will require less power and be easier to maintain. When a mobile moves from the coverage area of one base station to the coverage area of another base station i.e., from one cell to another cell, then the signal strength of the initial base station may not be sufficient to continue the call in progress. So the call has to be transferred to the other base station. This is called hand-off[6].

Another important idea of the cellular concept is frequency reuse. Frequency reuse, or, frequency planning, is a technique of reusing frequencies and channels within a communication system to improve capacity and spectral efficiency[6]. Frequency reuse essentially means to arrange cells in a way such that cells that are close to each other use distinct frequencies, by doing this division we can avoid many users interfering each others signals and can provide higher quality of service for the users. Ideally we want cells that use the same frequencies to be far from each other.

Cellular systems implements a type of multiplexing called *Space division multiplexing* (SDM), which essentially means to transmit through multiple parallel channels. Each transmitter is typically called a base station, and covers a certain area that is referred to as a *cell*. The area that a single cell covers can vary from 10m to 10km, the shape of the cell depend on the enclosing environment that it covers, e.g the surrounding buildings, mountains etc. Cellular system is the typical approach used in mobile communication [8] where a mobile station located within a particular cell will communicate with the base station responsible for that cell and vice verse.

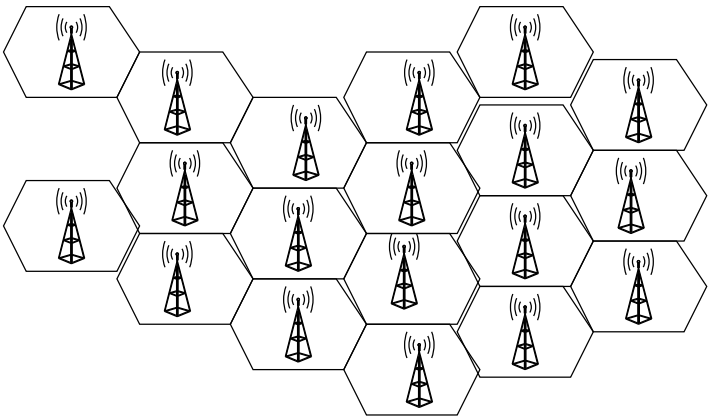

Figure 3: Cellular network

Small vs Large cells Most mobile network providers install many (thousands) base stations throughout a country to provide network access to their customers instead of using fewer more huge cells with more powerful transmission capabilities. Using smaller cells and more base stations is more expensive, but has the following advantages compared to larger cells:

- Higher capacity
- Requires less transmission power.
- Less interference.
- More robust towards failures.

Disadvantages of smaller cells is like mentioned, increased cost and also requires a more developed infrastructure. Another aspect that is important when considering cellular network with a large amount of cells is that the mobile stations connected to the network will have to move between cells more often depending on how large areas the mobile station is moving over and the size of the cells it is connected to. When a mobile station is moving from one cell to another we say that a *handover* happen.

Summary This section delved into the lower-level aspects of wireless and mobile communication. The main difference in wireless communication compared to the wired counterpart is its inherent unreliability and complexity due to its wireless nature. Signals and information is travelling through electromagnetic waves rather than digital signals through a wired medium. We briefly touched upon important aspects of wireless transmission such as antennas, signal propagation and spread spectrum as well as a common system design for mobile and wireless communication, called cellular systems. In summary cellular networks is a more convenient way for providers to distribute the connectivity to their users since the network can gradually be extended by new base stations. Cellular systems can better utilise the capacity due to frequency reuse and imposes less requirements on the individual base stations.

The electrical-engineering part of mobile communication is a very large research area on its own, this document is not targeted against professionals in this field but rather software engineers longing for a quick and graspable overview of the field of mobile communication.

3 Mobile Telecommunication Generations

1G,2G,3G... what does it all mean? What are the differences?

The "G" refer to *generation*, 1G, 2G, 3G, 4G, 5G are the current published generations of mobile-communications. This section will give you some overview of the different generations and how they differ and why it is necessary to distinguish newer mobile communication from previous generations.

The generations differ in many different aspects, architecture, technology, protocol stack etc. Naturally the generations also differ in performance, after all one would assume that the fifth generation of mobile technologies should have the upper hand compared to the first generation, and so is of course also the case. In fact the driving motivation for innovation and development of new generations is just that, to increase the data rates available to the end users (the customers). The grouping of the technologies into generations help to specify more precisely what is meant, just saying mobile technology is like referring to 50+ years of excessive research and development.

3.1 The Basics

Before we take a look at the different generations of mobile communication we'll quickly go over some basic notions that will be needed when comparing the generations.

Digital vs Analog When comparing different generations of mobile technology it is important to have grabbed the differences between analog and digital communication. To put it in mathematical terms, analog is continuous, whilst digital is discrete. You can think of an analog machine as a machine with a wave-formed signal while a digital machine would have an angular signal.

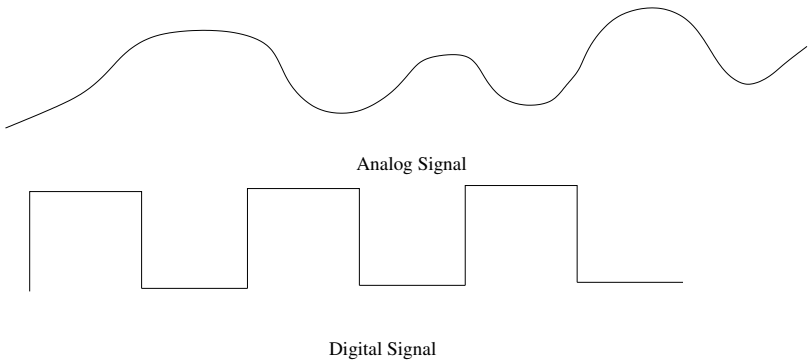

Figure 4: Analog vs Digital signals

Essentially digital means that you constrain the signal to only take values within certain bounds, this means you can loose some expressitivity possible with analog signals which are continuous, but it also makes digital signals a lot more precise. To generalise we can say that everything in the world is analog, digital signals simply means to quantify the signal and adding constraints to reproduce the analog signal with certain limits. In computing there is only digital signals, computers work with zeros and ones, which is a digital pattern.

Radio waves, voice and similar are examples of analog signals. Analog signals can suffer from signal degradation and noise. Transforming the analog signal to digital makes it more suited for transmission and easier to detect losses of signals. Phone systems have conventionally used analog technology but just as other type of analog signal systems, analog signals in phone systems are getting replaced by digital signals. As we mentioned roughly everything is analog, to make it digital the signals/data first run through a compression that transforms it into digital format. The digital format is then transmitted and on the receiver side the digital data is uncompressed into its analog form.

Circuit-Switched vs Packet-Switched Circuit-switched networks require dedicated point-to-point connections during calls. An example of a circuit-switched network is network built on technology from the first three generations of mobile technologies. The GSM standard which describes the second generation and which serves as a base also for the third generation is circuit-switched.

The canonical example of packet-switched network is the Internet and the Internet Protocol (IP). Packet-switched networks move data in separate, small blocks, packets, based on the destination address in each packet. When received, packets are reassembled in the proper sequence to make up the message. In the fourth generation of mobile communications circuit-switching is abandoned and solely packet switching is used.

Packet switching is very flexible - you only need to know which of your neighbours to send things to, and the route each packet takes can change without the endpoints knowing or noticing. On the downside, this means latency is unpredictable, and you're not guaranteed to have enough bandwidth even if the connection comes up.

The obvious comparison is with circuit switching. In a circuit switched network the full route is connected and reserved at the start of the connection, and then lasts until the connection is taken down. This is wasteful, since the connection reserves bandwidth even when nothing is being sent, and it's also not tolerant of changes. On the other hand, it's extremely predictable, since you know the latency will be stable and you'll never have to deal with congestion.

3.2 Mobile Communication Generations

1G The first generation was completely analog radio communication (think "walkie talkies"), and struggled with bad quality.

2G In the second generation, digital communication was introduced, still designed only for phone calls and was not for internet. In essence 2G was about

digital voice communication. 2G is described by a standard called GSM (Global System for Mobile Communications). 2G and GSM really took off and was implemented in big numbers over the world. 2G introduced the idea of using SIM Card to store user information and let phones be pre-programmed by the operators.

3G 3G added internet and was technically "CDMA" enhanced (remember CDMA stands for Code division multiple access (CDMA) and is a channel access method used by various radio communication technologies, as you remember CDMA is a kind of "spread spectrum" for cellular networks enabling many more wireless users to share airwaves than alternative technologies). The enhanced version of CDMA was called W-CDMA (Wideband-CDMA). With 3G operators could provide users with enough data rate for both voice and mobile internet. The standard describing 3G is called UMTS (Universal Mobile Telecommunications System). UMTS is based on the GSM standard describing the second generation.

4G 4G introduced mobile broadband. With 4G the operators can provide high enough data rates such that it can resemble wired broadband internet connections. With 4G everything is done over IP. 4G is described by the LTE (Long Term Evolution) standard.

5G 5G takes the data rates even further and promises to provide Gigabit Internet, 5G is not yet fully defined. The big thing with 5G apart from its speed is that it will be the first mobile technology generation explicitly designed for machines rather than for humans. 5G have been designed with Internet of Things and the increasing growth in small wireless devices in mind. For instance 5G will require minimal power and enable wireless devices to run on batteries for longer. A big shift in 5G will also be that it's possible for operators to give *guarantees*. As 5G is designed for machines it is important to be able to give minimum guarantees of the connectivity and data transfer rates.

The standards You can think of GSM as a widely deployed standard describing the protocols of the second generation (2G). Before GSM and 2G there was the first generation of mobile technology which was a generation consisting of analog cellular networks. GSM and 2G was a major uplift since it described a digital, circuit-switched network. And now in recent times the LTE standard have emerged, describing essentially 3G and 4G technologies. LTE is based on the old GSM technologies but have increased the capacity greatly. The upper layers of LTE are based upon TCP/IP, which result in an all-IP network similar to the current state of wired communications. LTE support mixed data, voice, video and messaging traffic.

3.3 Mobile Communication Concepts

Later in this document we'll look more in detail on the different standards that constitutes the generations outlined in this section (GSM, LTE, UMTS), but before doing that we'll briefly introduce a few related concepts that will make it easier to grasp the bigger picture.

Public Switched Telephone Network (PSTN) A short note on PSTN and the difference between today's cellular systems.

Traditionally the PSTN consisted solely of analog switches and communication. Today the description of the PSTN is rather blurry since it have evolved and now includes many technologies, including cellular networks.

The public switched telephone network (PSTN) is the aggregate of the world's circuit-switched telephone networks that are operated by national, regional, or local telephony operators, providing infrastructure and services for public telecommunication. The PSTN consists of telephone lines, fiber optic cables, microwave transmission links, cellular networks, communications satellites, and undersea telephone cables, all interconnected by switching centers, thus allowing most telephones to communicate with each other. Originally a network of fixed-line analog telephone systems, the PSTN is now almost entirely digital in its core network and includes mobile and other networks, as well as fixed telephones. [5]

The PSTN has been increasingly converted to digital carrier over the decades. But, the "last mile" is still analog, and entirely digital everywhere else, even transported by IP in many cases. It's only relation to "the internet" was by use of dial-up modems.

There was a time when PSTN was heavily used for DSL services (Digital Subscriber Line, a technology to transmit data over copper wires upon which the PSTN is built). Today the PSTN is rarely used as an access technology, more common is fiber or radio waves.

Mobile Telephone Switching Office (MTSO) The Mobile Telephone Switching Office (MTSO) is the mobile equivalent to a PSTN Central Office. The MTSO contains the switching equipment or Mobile Switching Center (MSC) for routing mobile phone calls. It also contains the equipment for controlling the cell sites that are connected to the MSC.

The systems in the MTSO are the heart of a cellular system. It is responsible for interconnecting calls with the local and long distance land-line telephone companies, compiling billing information (with the help of its CBM/SDM), etc. It also provides resources needed to efficiently serve a mobile subscriber such as registration, authentication, location updating and call routing.

All cellular systems have at least one MTSO which will contain at least one MSC. The MSC is responsible for switching calls to mobile units as well as to the local telephone system, recording billing data and processing data from the cell site controllers.

What happens when we make a call?

I When we switch on the mobile phone, it tries for an SID (System Identification Code, a unique digit number that is assigned to each carrier) on the Control channel. The Control channel is a special frequency that the phone and base station use to talk to one another. If the Mobile phone finds difficulty to get link with the control channel, it displays a "no service" message.

II If the Mobile phone gets the SID, it compares the SID with the SID programmed in the phone. If both SID match, the phone identifies that the

cell it is communicating with is the part of its home system.

III The phone also transmits a registration request along with the SID and the MTSO keeps track of your phone's location in a database. MTSO knows in which cell you are when it wants to ring the phone.

IV The MTSO then gets the signal, it tries to find the phone. The MTSO looks in its database to find the cell in which the phone is present. The MTSO then picks a frequency pair to take the call.

V The MTSO communicates with the Mobile phone over the control channel to tell it what frequencies to use. Once the Mobile phone and the tower switch on those frequencies, the call is connected.

VI When the Mobile phone move toward the edge of the cell, the cell's base station will note that the signal strength is diminishing. At the same time, the base station in the cell in which the phone is moving will be able to see the phone's signal strength increasing.

VII The two base stations coordinate themselves through the MTSO. At some point, the Mobile phone gets a signal on a control channel and directs it to change frequencies. This will switch the phone to the new cell.

[7]

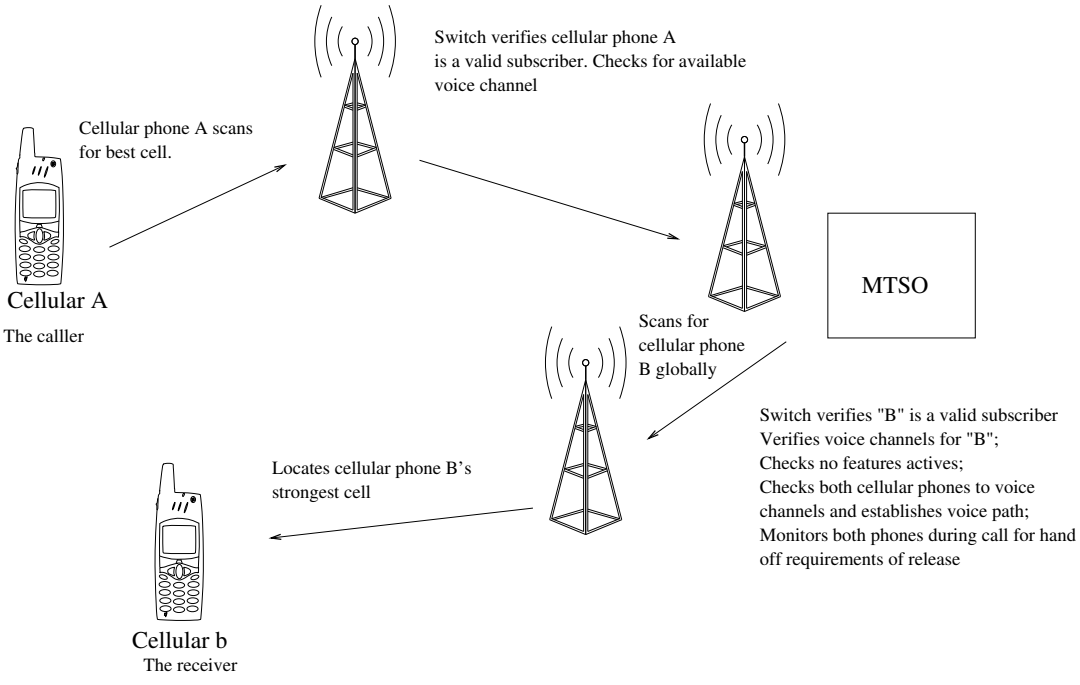

Figure 5: Call in a mobile network

Analog-to-Digital (ADC) Not surprisingly an analog-to-digital converter(ADC) is a system that converts an analog signal, such as a sound picked up by a microphone or light entering a digital camera, into a digital signal. The conversion involves quantization (quantization is the process of mapping a large set of input values to a (countable) smaller set. Rounding and truncation are typical examples of quantization processes) of the input, so it necessarily introduces a small amount of error. Furthermore, instead of continuously performing the conversion, an ADC does the conversion periodically, sampling the input. The result is a sequence of digital values that have been converted from a continuous-time and continuous-amplitude analog signal to a discrete-time and discrete-amplitude digital signal.

Digital-to-Analog (DAC) Digital-to-analog converter (DAC) is a device that converts a digital signal into an analog signal. An analog-to-digital converter (ADC) performs the reverse function. A DAC converts an abstract finite-precision number (usually a fixed-point binary number) into a physical quantity (e.g., a voltage or a pressure). In particular, DACs are often used to convert finite-precision time series data to a continually varying physical signal.

Digital Signal Processing (DSP) Digital Signal Processing (DSP) is e method of processing digital signals and often converting the signals to some specific format. DSP is used in many different scenarios where signals need to be processed, not least in context of wireless communication and processing of radio signals.

Digital Signal Processors (DSP) take real-world signals like voice, audio, video, temperature, pressure, or position that have been digitised and then mathematically manipulate them. A DSP is designed for performing mathematical functions like "add", "subtract", "multiply" and "divide" very quickly.

Signals need to be processed so that the information that they contain can be displayed, analysed, or converted to another type of signal that may be of use. In the real-world, analog products detect signals such as sound, light, temperature or pressure and manipulate them. Converters such as an Analog-to-Digital converter then take the real-world signal and turn it into the digital format of 1's and 0's. From here, the DSP takes over by capturing the digitised information and processing it. It then feeds the digitised information back for use in the real world. It does this in one of two ways, either digitally or in an analog format by going through a Digital-to-Analog converter. All of this occurs at very high speeds. [4]

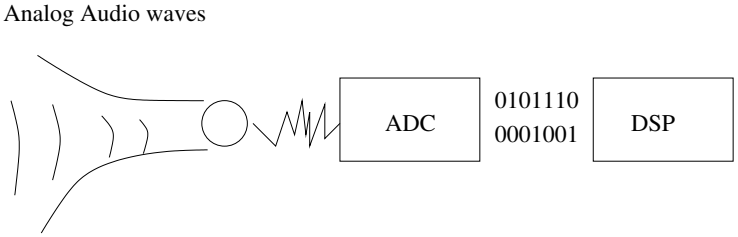

Figure 6: Digital Signal Processing

RAN A radio access network (RAN) is part of a mobile telecommunication system. It implements a radio access technology. Conceptually, it resides between a device such as a mobile phone, a computer, or any remotely controlled machine and provides connection with its core network (CN). Depending on the standard, mobile phones and other wireless connected devices are varyingly known as user equipment (UE), terminal equipment, mobile station (MS), etc. RAN functionality is typically provided by a silicon chip residing in both the core network as well as the user equipment. See the following diagram:

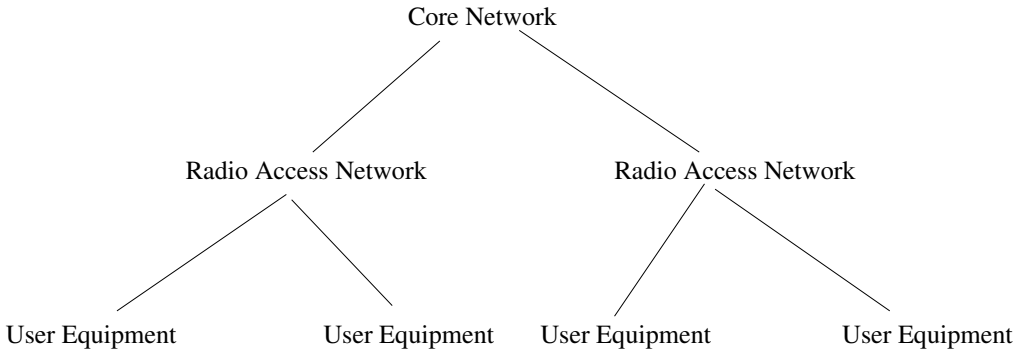

Figure 7: Radio Access Network

[13]

The idea, pioneered decades ago, is that a handset or other item can be wirelessly connected to a backbone or core network that transmits over the PSTN system or some other infrastructure. The radio access network gets the signal to and from the wireless end point, so it can travel with other traffic over networks built with a collective and deliberate purpose.

You can think of the RAN, as a middle-man in between the mobile devices and the actual network. It connects the two which enables communication.

RANs have different architectures/topologies depending on what standard if followed, today's RANs are commonly built of many base stations that are connected to provide a RAN of some topology.

The details of a RAN differs between different generations of mobile technology. Below is a summary of the basic RAN used in the second generation and which the later generations build upon.

RAN controls the radio interface to the mobile station. Basic components in RAN are:

- Base Transceiver Station (BTS)
- Base Station Controller (BSC)

Basic interfaces between each of these sub systems include:

- "Abis" interface between BSC and BTS (within the RAN)
- "Um" air interface between the RAN and the MS

RAN in 4G is called E-UTRAN and the Base stations (BTS) is called eNodeB.

Telephone Exchange and Switching A telephone exchange is a telecommunications system used in the public switched telephone network or in large enterprises. A telephone exchange traditionally was made manually by human operators that interconnected telephone subscriber lines or virtual circuits of digital systems to establish phone calls between subscribers. As you can imagine the digital switch was a big break through since it allowed better scalability and reliability than having human operators manually performing the switching.

A digital switch is a hardware device for handling digital signals. The main function of these switches is to manage digital signals generated or passed through a telephone exchange and then forward it to the telephone company's back-end network. The communication between the subscribers of a telephone company is established with the help of digital switching. Digital switches can be of different types based on the number of lines they handle and the included features. Ericsson's AXE telephone exchange is the most widely used digital switching platform in the world and was developed in Sweden in the 1970s.

Carrier Aggregation (CA) Carrier Aggregation (CA) ultimately is a method for achieving higher data rates that is used in for example the LTE standards. With carrier aggregation, it is possible to utilise more than one carrier and in this way increase the overall transmission bandwidth. Simplified you can say that CA means to merge frequency-bands to achieve double the space for data traffic and utilise more than one carrier. This means that data-traffic can be divided and optimised in a lot more effective way. A carrier can for example be a Base Station.

The simplest form of carrier aggregation, is where the carriers are contiguous and lie within the same frequency band. In this case it is feasible for a mobile device to handle the signals using a single transceiver, provided it is able to operate efficiently over the aggregate bandwidth.

Another option is intra-band non-contiguous carrier aggregation, in which the carriers lie within the same frequency band, but they are not adjacent. In this case it is necessary for the mobile device to use a separate transceiver for each carrier.

The final example of carrier aggregation is based on inter-band non-contiguous carriers. In this case the carriers fall in different parts of the radio spectrum, such as 900MHz and 1800MHz. The ability to combine such carriers is particularly useful for network operators with fragmented spectrum allocations

Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD) FDD and TDD are two different ways of sharing the same physical medium, both are used by different operators and mobile generations, which one to choose depends on the spectrum allocation.

In communication systems, a user needs to exchange data with one or more parties through a shared resource - a common channel. Depending on whether

the data is transmitted/received simultaneously, the following transmission techniques exist:

- **Simplex:** One party transmits data and the other party receives data. No simultaneous transmission is possible - the communication is one-way and only one frequency (channel) is used. Examples of simplex communication are traditional radio and TV (non interactive).
- **Half Duplex:** Each party can receive and transmit data, but not at same time. The communication is two-way and only one frequency (channel) is used. Walike talkie is a typical example.
- **Full Duplex:** Each party can transmit and receive data simultaneously. The communication is two-way and two frequencies are used. one for transmitting and one for receiving.

In the case of cellular networks, a limited shared resource (spectrum) needs to be shared with all users so full duplex communication is possible. The two main methods used are:

1. *Time Division Duplexing (TDD)* - The communication is done using one frequency, but the time for transmitting and receiving is different. This method emulates full duplex communication using a half duplex link.
2. *Frequency Division Duplexing (FDD)* - The communication is done using two frequencies and the transmitting and receiving of data is simultaneous.

LTE (The 4G standard) is defined to support both paired spectrum for Frequency Division Duplex (FDD) and unpaired spectrum for Time Division Duplex (TDD). LTE FDD uses paired spectrum that comes from a migration path of 3G network whereas TDD LTE uses unpaired spectrum that evolved from TD-SCDMA.

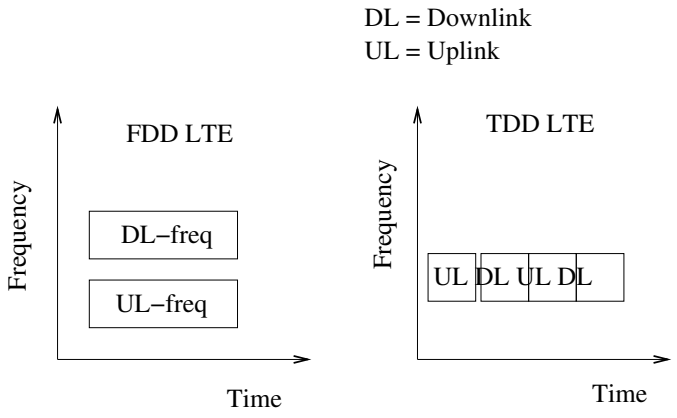

Figure 8: FDD vs TDD

Radio Link Control (RLC) Radio link control (RLC) is a layer 2 protocol used in UMTS (the 3G standard) and LTE on the Air interface.

The primary function of RLC is to transfer user data and signalling between the upper layers and the MAC layer. Data is transferred into the RLC in data blocks called Service Data Units (SDU). Data is transferred out of the RLC in data blocks called Protocol Data Units (PDU).

Integrated Services Digital Network (ISDN) Integrated Services Digital Network (ISDN) is a set of communication standards for simultaneous digital transmission of voice, video, data, and other network services over the traditional circuits of the public switched telephone network. Prior to ISDN, the telephone system was viewed as a way to transport voice, with some special services available for data. The key feature of ISDN is that it integrates speech and data on the same lines, adding features that were not available in the classic telephone system. ISDN is a circuit-switched telephone network system, which also provides access to packet switched networks, designed to allow digital transmission of voice and data over ordinary telephone copper wires, resulting in potentially better voice quality than an analog phone can provide. It offers circuit-switched connections (for either voice or data), and packet-switched connections (for data), in increments of 64 kilobit/s. In some countries, ISDN found major market application for Internet access, in which ISDN typically provides a maximum of 128 kbit/s bandwidth in both upstream and downstream directions.

The pro of ISDN is that it can reuse the old analog telephone network and can be built on the same copper wires as the PSTN already uses. It is only the phone stations themselves who need to be upgraded with new hardware to send digital signals over the copper wire, this can allow you to get internet access through your phone.[12]

Virtualization and Cloud-RAN Recall the base station which is ubiquitous in mobile communication infrastructure. The traditional architecture of base stations is that the base station itself is the interface between the RAN (Radio Access Network) and the core network. This is typically referred to as distributed baseband.

Today mobile communication is approaching a point where the current cellular infrastructure is reaching its limit and having problem to cope with the increasing demand of data-rates. A likely approach of 5G and future generations of mobile communication infrastructure is to shift from the distributed baseband to a centralised cloud-a-like baseband that exploit virtualization. Before we take a look at what a virtualized/cloud RAN looks like lets just repeat what is the purpose of the base station and what is its duties.

The area which a mobile network covers is divided into cells. Traditionally, in cellular networks, users communicate with a base station that serves the cell under coverage of which they are located. The main functions of a base station can be divided into deploying antennas for receiving radio signals, baseband processing and radio functions, this includes radio resource control, media access control, digital-to-analog-convertion, amplifying signals, analog-to-digital conversion and more [2]. The shortcoming of the traditional approach of having the base station do all this task is that it is not optimal utilisation of resources

to have each base station perform its own processing.

The main idea behind Cloud-RAN(C-RAN)/Virtualized-RAN(V-RAN) is to pool the Baseband Units (BBUs) from multiple base stations into a centralised BBU Pool which can serve many base stations simultaneously. You can view a BBU as a processing unit performing the processing of a base station.

By pooling the resources it is possibly to achieve cost savings in similar fashion as cloud computing is currently reducing the cost of higher level services. The cost savings comes in term of better utilisation of resources. Furthermore, C-RAN/V-RAN improves network capacity by performing load balancing and cooperative processing of signals originating from several base stations. C-RAN/V-RAN has also the potential to decrease the cost of network operation, because power and energy consumption are reduced compared to the traditional RAN architecture. New BBUs can be added and upgraded easily, thereby improving scalability and easing network maintenance.

With virtualization it is possible to have many virtual base stations on a single physical hardware system in the BBU pool effectively offering a isolated baseband for each base station and still capitalising on more effective resource-utilisation. Formally the virtualization part of V-RAN technology in the pooling of BBUs uses what is called software-defined networking (SDN) and Network function virtualization (NFV). A virtualized BBU Pool can be shared by many different network operators, allowing hem to rent Radio Access Network (RAN) as a cloud service. As BBUs from many sites are co-located in one pool, they can interact with lower delays.

Virtualization enables the creation of logically isolated networks over abstracted physical networks which can be shared in a flexible and dynamic way[2]. The network virtualization contains a group of virtual nodes and virtual links. Multiple virtual networks coexist on the same physical substrate. Deploying the virtual networks for the heterogeneous network architecture promotes flexible control, low cost, efficient resource usage, and diversified applications.

In the context of BBU pooling, network virtualization separates not only data storage but also applications, operating systems and management control. BBU Pool operates over a set of hardware platforms including CPU, memory, Network Interface Card (NIC) and so on. The virtualization solution is implemented via operating systems, i.e., Linux. The functions of a base station are realised as software instances, which are called the Virtual Base Stations (VBSs). Multiple VBSs share the common resources such as hardware and systems, which in turns offers the opportunity of efficient and flexible utilisation. Within the VBS Pool, several virtual operators share a common network environment, a common programming environment and IT platform.

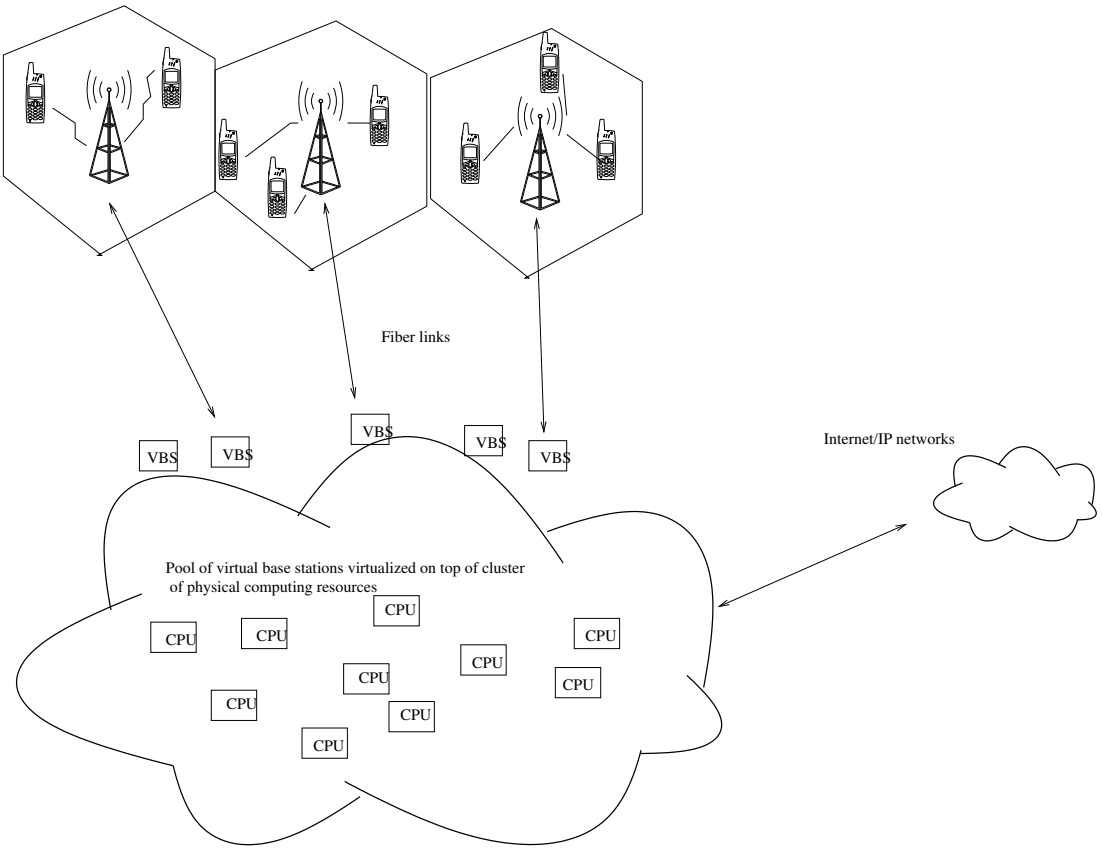

Figure 9: C-RAN/V-RAN

There are major distinctions, however, between cloud computing in the RAN compared to the core network and service layer. For example, the bulk of the cost of a mobile network lies in the large number of distributed base station and antenna sites, as well as in the last-mile transport-network links – not in central nodes and sites. Consequently, to calculate the return on investment benefit of implementing Cloud RAN, the costs associated both with the central parts of the network and its distributed elements and last-mile links must all be taken into account. A Cloud RAN architecture will therefore exploit a combination of virtualization, centralisation and coordination techniques, all of which interact with each other in a variety of ways within the network[1].

The terms virtualization and cloud are often used interchangeably. They do work well together in many cases, including in a RAN context. However, each concept brings different things to the table. In general, virtualization is a technique that can mean different things in different scenarios, and it is unlikely to mean the same thing in a RAN context as in, for example, a data server context. The reason for this is the substantial difference in real-time requirements imposed by the radio access protocol. Many of the synchronisation requirements that ensure the performance of the radio access protocol are on the microsecond

level and, in some cases, the nanosecond level. Thus, RAN functionality is not easily hosted by the so-called virtualized platform as a service (PaaS) model, as is possible with straightforward applications and server-type functions. On the other hand, there is no need to virtualize all RAN functionality to provide the benefits of Cloud RAN. Virtualization as an execution environment technique can be used to provide isolation, scalability and elasticity, among other things, for the Radio Resource Control (RRC) protocol layer. When applied in this manner, virtualization can be used to simplify the management and deployment of the RAN nodes, for example, by allowing the definition of arbitrarily-sized base stations (in terms of the number of cells) and for more flexible scaling of higher layer functionality separate from the scaling of other layers[1].

Virtualization can also be used to leverage a common execution environment for RAN, core and application functionality, providing the ultimate in execution proximity and ensuring maximum responsiveness of, for example, a certain service, or, as it is sometimes called, a certain type of network slice. The possibility to virtualize network functions in this way makes it feasible to place the functionality on a more generic and generally available execution platform together with cloud core applications and other latency-critical services, sometimes even in a PaaS environment[1].

Centralising base station processing with Cloud RAN simplifies network management and enables resource pooling and coordination of radio resources. Pooling, or statistical multiplexing, allows an execution platform to perform the same tasks with less hardware or capacity. This is of greatest interest for tasks that require a large number of computational resources. It also means that the most desirable pooling configuration is a fully centralised baseband approach[1].

Summary In this section we covered some history of mobile communication, it was pointed out how mobile communication have evolved and how different eras of mobile communication can be divided into generations. In addition you as a reader was brought up to speed in a lot of ground of sub-fields of mobile communication. Traditionally mobile communication was mostly analog but it is becoming more and more digital and in fact the 4G standard is entirely digital and packet-switched. The PSTN is a public telephone network whose architecture is a bit out of its time but is still around, the copper wires of PSTN have in some countries been used as an access technology (DSL services). We looked over what steps take place when a call is made and how the MTSO integrates mobile calls with the wired PSTN infrastructure. Finally the processing of base stations, a technique for utilising multiple carriers for increased data rates (CA), techniques for using a shared medium (FDD, TDD), ADC, DSP, RANs and the future of RANs, C-RAN and V-RAN was described.

The current generation of mobile communication is 4G but 3G is still around and 5G is in active development. In the next section we will go over two standards for mobile communication, GSM (2G/3G) and LTE (4G).

4 GSM and LTE

We will look a bit closer to two major standards of mobile technology, GSM and LTE. Both of these standards describe mobile technology systems based on cel-

lular networks, as mentioned earlier, GSM came along at the second generation of mobile technology while LTE is associated with 4G.

4.1 Global System for Mobile Communications (GSM)

Digital cellular networks are the segment of the market for mobile and wireless devices which are growing most rapidly and have been for quite some time. They are wireless extensions of traditional PSTN or ISDN networks and allow for seamless roaming with the same mobile phone nation or even worldwide. Traditionally, these systems are mainly used for voice traffic. However, data traffic is continuously growing.

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 ISDN and other PSTN systems. GSM permits the integration of different voice and data services and the interworking with existing networks.

A mobile station (MS) is connected to the GSM public land mobile network (PLMN). This network in turn is connected to transit networks, e.g ISDN or traditional PSTN.

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 (e.g fax). However as the main service is telephony, the primary goal of GSM was the provision of high-quality digital voice transmission, offering at least the typical bandwidth of 21 kHz of analog phone systems. A useful service for very simple message transfer is the short message service (SMS), which offers transmission of messages of up to 160 characters. SMS messages do not use the standard data channels of GSM but exploit unused capacity in the signalling channels. Sending and receiving of SMS is possible during data or voice transmission. SMS was in the GSM standard from the beginning; however, almost no one used it until millions of young people discovered this service in the mid-nineties as a fun service. [8]

The successor of SMS, enhanced message service (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 standardised way (some vendors offered similar proprietary features before. EMS never really took off as the multimedia message service (MMS) was available. MMS offers the transmission of larger pictures (GIF, JPG etc), short video clips and more.

GSM network architecture consists of different elements that all interact together to form the overall GSM system. These include elements like the base-station, controller, MSC, AuC, HLR, VLR, etc. GSM uses FDD for sharing a medium between multiple transceivers.

The GSM network architecture as defined in the GSM specifications can be grouped into four main areas:

- **Mobile station (MS)**
- **Base-Station Subsystem (BSS)**
- **Network and Switching Subsystem (NSS)**
- **Operation and Support Subsystem (OSS)**

The different elements of the GSM network operate together and the user is not aware of the different entities within the system.

Mobile stations (MS) are essentially the end-user equipment, they constitute the section of a GSM cellular network that the user sees and operates. There are a number of elements to MS, although the two main elements are the hardware and the SIM. The hardware contains the electronics used to generate the signal and to process the data received. The SIM or Subscriber Identity Module contains the information that provides the identity of the user to the network.

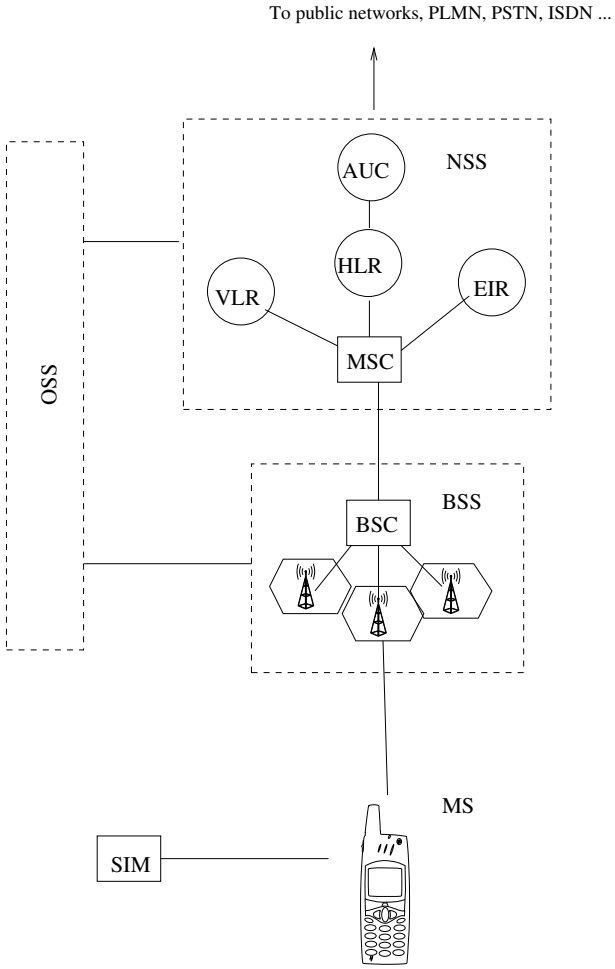

Figure 10: GSM architecture

The Base Station Subsystem (BSS) section of the GSM network architecture is fundamentally associated with communicating with the mobiles on the network. It consists of two elements:

- **Base Transceiver Station (BTS)**. This entity corresponds to the Base Station we have previously discussed in section 2, that is a defining part

of any cellular network. The BTS used in a GSM network comprises the radio transmitter receivers, and their associated antennas that transmit and receive to directly communicate with the mobiles. The BTS is the defining element for each cell. The BTS communicates with the mobiles and the interface between the two is known as the Um interface with its associated protocols.

- **Base Station Controller (BSC)**. This type of entity we haven't covered in our general discussion about cellular networks in section 2. The BSC forms the next stage back into the GSM network. It controls a group of BTSs, and is often co-located with one of the BTSs in its group. It manages the radio resources and controls items such as handover within the group of BTSs, allocates channels and the like. It communicates with the BTSs over what is termed the Abis interface.

Network Switching Subsystem (NSS) The Network Switching Subsystem (NSS) is often referred to as the core network, it provides the main control and interfacing for the whole mobile network. Is the really the "heart" of the GSM system . The NSS connects the wireless network (the RAN) with standard public networks, performs handovers between different BSSs, comprises functions for worldwide localisation of users and supports charging, accounting and roaming of users between different providers in different countries.

NSS includes:

- **The Mobile Switching Center (MSC)** performs the telephony switching functions for the network. It controls calls to and from other telephone and data communications networks, including the Public Switched Telephone Networks (PSTN), Integrated Services Digital Networks (ISDN), Public Land Mobile Networks (PLMN) and Public Data Networks.

  As you can imagine the MSC needs a bunch of information and data to be able to provide its customer with the specific service they have purchased and still avoid providing service to non-customers. Therefore the NSS also includes a set of subsystems providing just this information and also other information.

- **The Visitor Location Register (VLR)** database contains all temporary subscriber information needed by the MSC to serve visiting subscribers who are temporarily in the area of the MSC.

- **The Home Location Register (HLR)** database stores and manages user subscriptions. It contains all permanent subscriber information including their service profile, location information and activity status. This is the most important database in a GSM system.

- **Equipment Identity Register (EIR)**. The EIR is the entity that decides whether a given mobile equipment may be allowed onto the network. Each mobile equipment has a number known as the International Mobile Equipment Identity. This number, as mentioned above, is installed in the equipment and is checked by the network during registration.

- **Authentication Centre (AuC)**. The AuC is a protected database that contains the secret key also contained in the user's SIM card. It is used for authentication and for ciphering on the radio channel.
- **Gateway Mobile Switching Centre (GMSC)**. The GMSC is the point to which a ME terminating call is initially routed, without any knowledge of the MS's location. The GMSC is thus in charge of obtaining the MSRN (Mobile Station Roaming Number) from the HLR based on the MSISDN (Mobile Station ISDN number, the "directory number" of a MS) and routing the call to the correct visited MSC. The "MSC" part of the term GMSC is misleading, since the gateway operation does not require any linking to an MSC.
- **SMS Gateway (SMS-G)**. The SMS-G or SMS gateway is the term that is used to collectively describe the two Short Message Services Gateways defined in the GSM standards. The two gateways handle messages directed in different directions. The SMS-GMSC (Short Message Service Gateway Mobile Switching Centre) is for short messages being sent to an ME.

The final subsystem is the Operation and Support Subsystem (OSS). The OSS or operation support subsystem is an element within the overall GSM network architecture that is connected to components of the NSS and the BSC. It is used to control and monitor the overall GSM network and it is also used to control the traffic load of the BSS.

A rundown of GSM That was a lot of acronyms and subsystems/components but looking at figure 10 should give you the conceptual model of how a GSM network operates and how it is related to the general concept of cellular network as we described in chapter 2. I don't think you need to put an effort into memorising all the subsystems and their purpose, especially since GSM have been partly replaced by newer standards. What is important to take away from this section is the idea of what kind of functionality is desired to put together a cellular network that can provide services including phone calls and internet. It should be clear now that just putting out a bunch of base stations in cells and equipping users with antennas is not enough, some structure and utilities is required.

Localisation and calling One fundamental feature of the GSM system is the automatic, worldwide localisation 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 database always contains information about the current location (only the location area, not the precise geographical location), and the VLR database currently responsible for the mobile station (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. Typically people associate international roaming with the term roaming as it is this type of roaming that makes GSM very attractive: one device, over 190 countries! [8].

A call in GSM Below is the steps outlined which take place to perform a call from a device outside GSM to a device inside GSM:

1. A user dials the phone number of a GSM subscriber. The fixed network (PSTN) notices (looking at the destination code) that the number belongs to a user in the GSM network and forwards the call setup to the gateway MSC.
2. The GMSC identifiers the HLR for the subscriber and signals the call setup to the HLR
3. The HLR now checks whether the number exists and whether the user has subscribed to the requested services, and requests and MSRN from the current VLR
4. After receiving the MSRN the HLR can determine the MSC responsible for the mobile station (MS) and forwards this information to the GMSC
5. The GMSC can now forward the call setup request to the MSC indicated
6. From this point on, the MSC is responsible for all further steps. First it requests the current status of the MS from the VLR
7. If the MS is available, the MSC initiates paging in all cells it is responsible for as searching for the right cell would be too time consuming. The BTSs of all BSSs transmit this paging signal to the MS.
8. If the MS answers, the VLR has to perform security checks (set up encryption etc), the VLR then signals to the MSC to set up a connection to the MS.

A call originating from the GSM is easier:

1. The MS transmits a request for a new connection
2. The BSS forwards this request to the MSC
3. The MSC then checks if this user is allowed to set up a call with the requested service and checks the availability of resources through the GSM network and into the PSTN
4. If all resources are available the MSC sets up a connection between the MS and the fixed network.

[8]

Handover Cellular systems require **handover** procedures, as single cells do not cover the whole service area, but e.g. only up to 353km 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 60ms. [8]

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

- **Access control and authentication:** The first step includes the authentication of a valid user for the SIM. The user needs a secret PIN to access the SIM. The next step is the subscriber authentication. This step is based on a challenge-response scheme.
- **Confidentiality:** All user-related data is encrypted. After authentication, BTS and MS apply encryption to voice, data, and signalling. This confidentiality exist only between the mobile station and the base station, it does not exist end-to-end or within the whole fixed GSM/telephone network.
- **Anonymity:** To provide user anonymity, all data is encrypted before transmission, and user identifiers (which would reveal an identity) are not used over the air. Instead, GSM transmits a temporary identifier which is newly assigned by the VLR after each location update.

[8]

3G is in a sense built on top of GSM and the next standard we'll look at, Long Term Evolution (LTE) build upon the technology of 3G. 4G represents the future of mobile communications in the longer term. In 4G, the majority of the traffic is data and multimedia as opposed to voice only. Data rates in 4G systems will range from 20 to 100 Mbps.

4.2 Long-Term Evolution (LTE)

As opposed to earlier generations, a 4G system does not support traditional circuit-switched telephony service, but all-Internet Protocol (IP) based communication such as IP telephony. The IP Core network has much more stringent requirements and is developed further to support high data rates, provide advanced application services support and manage itself as well as the radio network more efficiently.

The architecture of the 4G network will more or less resemble the 3G architecture but there are some significant evolutionary changes. Circuit-switching capabilities are redundant in 4G and are thus removed. The MSC (Mobile Switching Centre) used previously to service legacy 2G voice traffic is discarded and all voice traffic is treated as packet data at the BS (Base Station). Backward

compatibility is maintained by segmenting voice data into packets and routing them through the IP backbone using VOIP (Voice Over IP) technology. A VOIP Gateway is used to connect to PSTN (Public Switched Telephone Network) or ISDN (Integrated Services Digital Network). Another major advancement that 4G makes is the integration of Wireless LANs into the total mobile network

LTE can be viewed as a way to build a network that can fulfil the speeds that 4G promises. The LTE technology enabled fast mobile internet connection. Actually, LTE is a path followed to achieve 4G speeds rather than the definition of 4G, which is a common misconception.

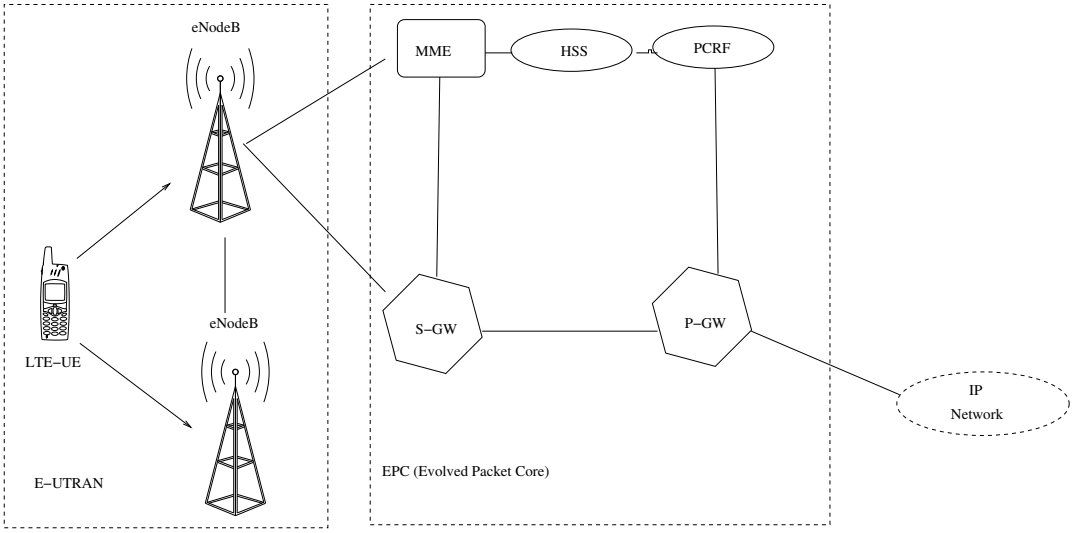

Figure 11: LTE architecture

Radio Access Network (EUTRAN) The radio access network is used for wireless radio connection between the mobile phones and antennas from the mobile operator. The radio access network is also called EUTRAN or Evolved Universal Mobile Telecommunications System Terrestrial Radio Access Network. EUTRAN can be also called just LTE (Long Term Evolution). Radio infrastructure is formed of the following nodes:

- **LTE Mobile Terminals:** LTE mobile terminals are the mobile phones and other devices that support the LTE standard.
- **Radio Interface:** Radio interface is a wireless connection between the LTE mobile terminals and eNodeB. It is wireless signals that form the mobile cells.
- **eNodeB:** E-UTRAN Node B or eNodeBs are situated all over the network of the mobile operator. They connect the LTE mobile terminal via radio interface to the core network. You can think of a eNodeB as a LTE-capable base station from the base-station we had earlier in the section about

cellular networks. The name node B was chosen during standardisation until a new and better name was found. However, no one came up with anything better so it remained. A node B connects to one or more antennas creating one or more cells respectively. The cells can either use FDD or TDD.

Core network (EPC) Core Network is the brain of the system. It is formed of telephony switches that enable the different services for the mobile users. Core network devices connect the mobile devices in the mobile network. They also connect the mobile network with the fixed telephony network and internet. The LTE core network is called EPC (Evolved Packet Core) or System Architecture Evolution (SAE). Just as for the GSM standard the core network (called EPC in the LTE standard and NSS in GMS standard) is the most involved part and is what differs the 2G cellular networks from the 4G cellular networks. You'll recognise the overall pattern of the core network from the GSM section. The core network of LTE is formed from the five nodes:

1. **MME:** Mobility Management Entity or MME is the central control node in the EPC network. It is responsible for mobility and security signalling, tracking and paging of mobile terminals.

2. **S-GW:** Serving Gateway or S-GW transports the user traffic between the mobile terminals and external networks. It also interconnects the radio access network with the EPC network. It is connected to the P-GW.

3. **P-GW:** PDN (Packet Data Network) Gateway connects the EPC network to the external networks. It routes traffic to and from PDN networks.

4. **HSS:** HSS (Home Subscriber Server) is the database of all mobile users that includes all subscriber data. It is also responsible for authentication and call and session setup.

5. **PCRF:** PCRF (Policy and Charging Rules Function) is node responsible for real-time policy rules and charging in EPC network.

[9]

A rundown of LTE To summarise, LTE is a technical standard for achieving 4G network speeds. LTE consists of some similar subsystems from earlier generations of cellular networks like base stations (called eNodeBs), user equipment (called mobile terminals), Databases containing customer data (called HSS), Gateways as intermediary points between base stations and outer IP networks. The main difference with LTE is that it is entirely packet-switched as opposed to GSM which is based on circuit-switching. LTE and 4G can also provide a lot higher traffic rates. To be backwards compatible with voice data LTE treats all voice data as packet data (essentially Voice over IP (VOIP)) and a special VOIP gateway is used to connect to PSTN, so essentially the data is carried as digital packets and only the very last part that is carried by PSTN is analog.

Summary This section went over a lot of technical details related to two major standards for mobile communication, GSM and LTE. LTE is the more modern of the two standards. GSM and LTE has a common overall idea and concept of the system architecture but has some defining differences that distinguishes the two. GSM is circuit-switched while LTE is packet-switched and LTE is able to provide higher data rates than GSM.

5 Satellite Systems

Satellite communication introduces another system supporting mobile communications. Satellites offer global coverage without wiring costs for base stations and are almost independent of varying population densities.

5.1 The Basics of Satellite Systems

History Satellite communication began after the Second World War. Scientists knew that it was possible to build rockets that would carry radio transmitters into space. In 1957, in the middle of the cold war, the first satellite SPUTNIK was launched by the Soviet Union. SPUTNIK is not at all comparable to a satellite today, it was basically a small sender transmitting a periodic 'beep'. [8]

Applications Traditionally satellites have been used in the following areas, among others.

- **Weather forecasting**: Several satellites deliver pictures of the earth using infra red or visible light.
- **Radio and TV broadcast satellites**: Hundreds of radio and TV programs are available via satellite. This technology competes with cable in many places as it is cheaper to install.
- **Satellites for navigation**: Even though it was only used for military purposes in the beginning, the global positioning system (GPS) is nowadays well-known and available for everyone. The system allows for precise localisation worldwide.

In the context of mobile communication, the capabilities of satellites to transmit data is of particular interest.

- **Global telephone backbone:** One of the first applications of satellites for communication was the establishment of international telephone backbones. Instead of using cables it was sometimes faster to launch a new satellite. However, while some applications still use them, these satellites are increasingly being replaced by fiber optical cables crossing the oceans. The main reason for this is the tremendous capacity of fiber optical links and the much lower delay compared to satellites.
- **Connections for remote or developing areas:** Due to their geographical location many places all over the world do not have direct wired connection to the telephone network or the internet. Satellites now offer a simple and quick connection to global networks.

- **Global mobile communication:** The latest trend for satellites is the support of global mobile data communication. Due to the high latency geostationary satellites are not ideal for this task, therefore satellites using lower orbits are needed. The basic purpose of satellites for mobile communication is not to replace the existing mobile phone network, but to extend the area of coverage. Cellular phone systems, such as GSM (and their successors) do not cover all parts of a country. Areas that are not covered usually have low population where it is too expensive to install a base station. With the integration of satellite communication, however, the mobile phone can switch to satellites offering worldwide connectivity to a customer.

  While in the beginning satellites were simple transponders, today's satellites rather resemble flying routers. Transponders basically receive a signal on one frequency, amplify the signal and transmit it on another frequency. Today's satellites provide many functions of higher communication layers, e.g inter-satellite routing, error-correction etc.

[8]

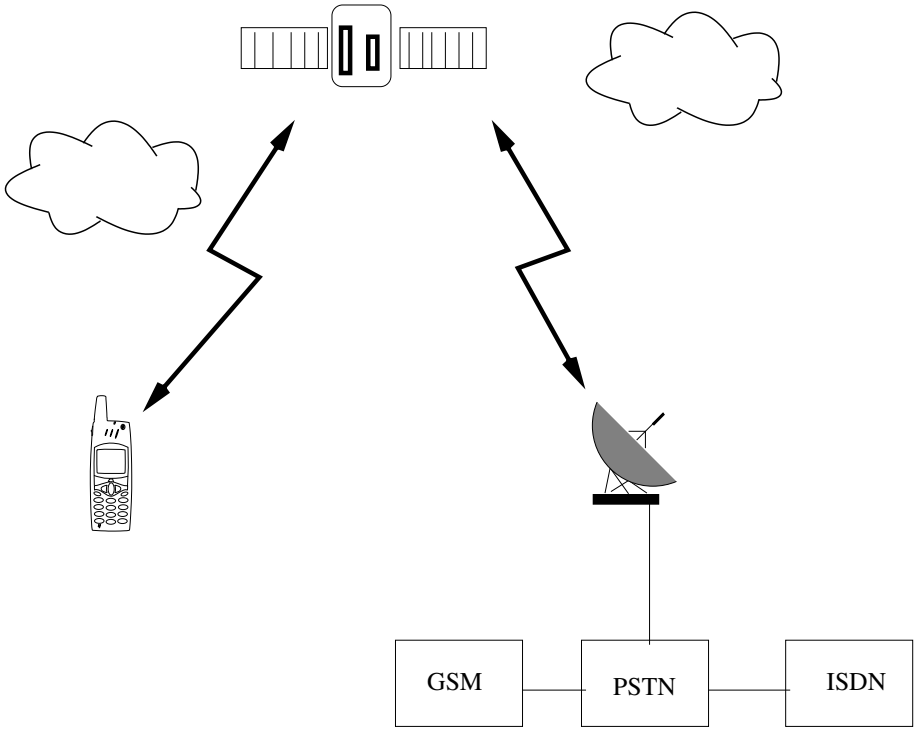

Figure 12: Satellites for extending the mobile network

Routing A satellite system together with gateways and fixed terrestrial networks has to route data transmissions from one user to another as any other

network does. Routing in the fixed segment (on earth) is achieved as usual, while two different solutions exists for the satellite network in space. If satellites offer ISLs, traffic can be routed between the satellites. If not, all traffic is relayed to earth, routed there and relayed back to the satellite. [8]

Localisation Localisation of users in satellite networks is similar to that of terrestrial cellular networks. One additional problem arises from the fact that now the "base stations" i.e the satellites, move as well. The gateways of a satellite network Maintains several registers. A home location register (HLR) stores all static information about a user as well as his or her current location. The last known location of a mobile user is stored in the visitor location register (VLR). Registration of a mobile station is achieved as follows. The mobile station initially sends a signal which one or several satellites can receive. Satellites receiving such a signal report this even to a gateway. The gateway can now determine the location of the user via the location of the satellites. User data is requested from the user's HLR and VLR. Calling a mobile station is again similar to GSM. The call is forwarded to a gateway which localises the mobile station using HLR and VLR.

5.2 Global Positioning System (GPS)

The Global Positioning System (GPS), originally Navstar GPS, is a space-based radio-navigation system owned by the United States government and operated by the United States Air Force. It is a global navigation satellite system that provides geo-location and time information to a GPS receiver anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites.

The GPS system does not require the user to transmit any data, and it operates independently of any telephonic or internet reception, though these technologies can enhance the usefulness of the GPS positioning information. The GPS system provides critical positioning capabilities to military, civil, and commercial users around the world. The United States government created the system, maintains it, and makes it freely accessible to anyone with a GPS receiver. However, the US government can selectively deny access to the system, as happened to the Indian military in 1999 during the Kargil War.

The Global Positioning System (GPS) is a network of about 30 satellites orbiting the Earth at an altitude of 20,000 km.

Wherever you are on the planet, at least four GPS satellites are 'visible' at any time. Each one transmits information about its position and the current time at regular intervals. These signals, travelling at the speed of light, are intercepted by your GPS receiver, which calculates how far away each satellite is based on how long it took for the messages to arrive.

Once it has information on how far away at least three satellites are, your GPS receiver can pinpoint your location. You do not need internet for GPS. [10]

Summary Satellite systems is not able to provide the increasing data rates that is required by users and is thus not going to be a replacement of cellular mobile networks. Satellite systems is a common way of extending the coverage of

a cellular network, it enables operator to provide coverage to customers located long distances from the nearest base station.

6 Wireless LAN

Wireless local area network technologies (WLAN) is a fast-growing market introducing the flexibility of wireless access into office, home or production environments. In contrast to the technologies described in earlier sections, WLANs are typically restricted by their diameter to buildings, a campus, single rooms etc. and are operated by individuals, not by large-scale network providers. The global goal of WLANs is to replace office cabling, to enable tether-less access to the Internet and, to introduce a higher flexibility for ad-hoc communication in, e.g., group meetings.

Some general advantages of WLANs are:

- **Flexibility:** Within radio coverage, nodes can communication without further restriction. Radio waves can penetrate walls, senders and receivers can be places anywhere.
- **Planning:** Only wireless ad-hoc networks allow for communication without previous planning, any wired network need wiring plans
- **Robustness:** Wireless networks can survive disasters, e.g earthquakes or users pulling a plug.

Some general disadvantages of WLANS:

- **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 error rates due to interference.
- **Restrictions:** All wireless products have to comply with national regulations. Several government and non-government institutions worldwide regulate the operation and restrict frequencies to minimise interference.

Infra red vs radio transmission Today, two different basic transmission technologies can be used to set up WLANs. One technology is based on the transmission of infra red light (e.g., at 900 nm wavelength), the other one, which is much more popular, uses radio transmission in the GHz range (e.g., 2.4 GHz in the license-free ISM band). Both technologies can be used to set up ad-hoc connections for work groups, to connect, e.g., a desktop with a printer without a wire, or to support mobility within a small area. Infra red technology uses diffuse light reflected at walls, furniture etc. or directed light if a line-of-sight (LOS) exists between sender and receiver. Senders can be simple light emitting diodes (LEDs) or laser diodes. Photo-diodes act as receivers.

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

Disadvantages of infra red transmission are its low bandwidth compared to other LAN technologies. However, their main disadvantage is that infra red is quite easily shielded. Infra red transmission cannot penetrate walls or other obstacles. Typically, for good transmission quality and high data rates a LOS, i.e., direct connection, is needed.

Most WLANs use radio waves for data transmission in a similar fashion as large-scale mobile networks like GSM, minus the big infrastructure.

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 infra red.

Again, the main advantage is also a big disadvantage of radio transmission. Shielding is not so simple. Radio transmission can interfere with other senders, or electrical devices can destroy data transmitted via radio. Additionally, radio transmission is only permitted in certain frequency bands. Very limited ranges of license-free bands are available worldwide and those that are available are not the same in all countries. However, a lot of harmonisation is going on due to market pressure. [8]

Infrastructure and ad-hoc networks 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, but not directly between the wireless nodes.

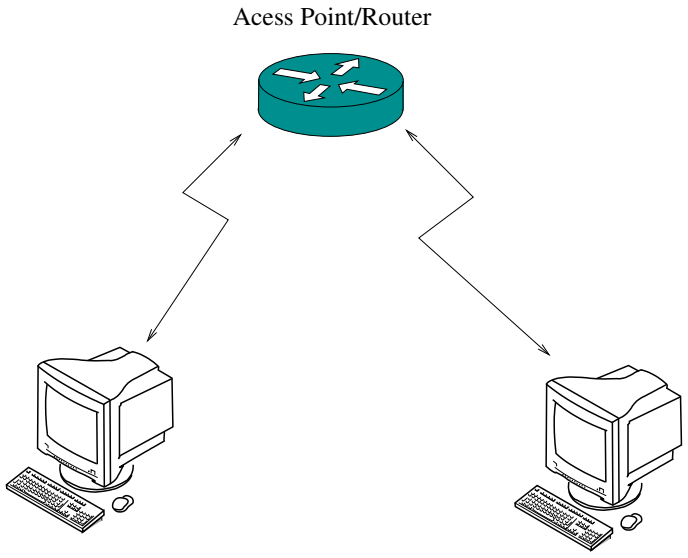

Figure 13: Infrastructure-based Wireless Local Area Network

Typically, the design of infrastructure-based wireless networks is simpler because most of the network functionality lies within the access point, the wireless clients can remain quite simple. This structure is reminiscent of switched Ethernet or other star-based networks, where a central element (e.g., a switch) controls network flow. This type of network can use different access schemes with or without collision. Collisions may occur if medium access of the wireless nodes and the access point is not coordinated. However, if only the access point controls medium access, no collisions are possible.

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.

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. Nodes within an ad-hoc network can only communicate if they can reach each other physically, i.e., if they are within each others range or if other nodes can forward the message.

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

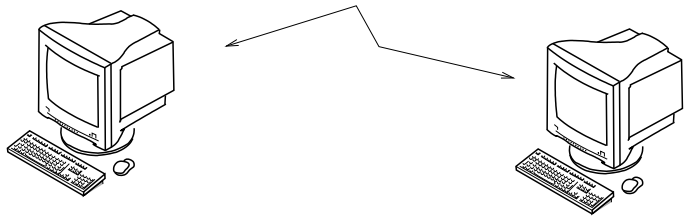

Figure 14: Ad-hoc Wireless Local Area Network

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

Protocol architecture As indicated by the standard number, IEEE 802.11 fits seamlessly into the other 802.x standards for wired LANs.

The IEEE 802.11 standard only covers the physical layer PHY and medium access layer MAC like the other 802.x LANs do. The physical layer is subdivided into the physical layer convergence protocol (PLCP) and the physical medium dependent sub-layer PMD.

Physical layer IEEE 802.11 supports three different physical layers: one layer based on infra red and two layers based on radio transmission.

Medium access control layer The MAC layer has to fulfil several tasks. First of all, it has to control medium access, but it can also offer support for roaming, authentication, and power conservation. The basic services provided by the MAC layer are the mandatory asynchronous data service and an optional time-bounded service.

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

- A station decides that the current link quality to its access point AP1 is too poor. The station then starts scanning for another access point.
- Scanning involves the active search for another BSS and can also be used for setting up a new BSS in case of ad-hoc networks.
- The station then selects the best access point for roaming based on, e.g., signal strength, and sends an association request to the selected access point AP2.
- The new access point AP2 answers with an association response. If the response is successful, the station has roamed to the new access point AP2. Otherwise, the station has to continue scanning for new access points.
- The access point accepting an association request indicates the new station in its BSS to the distribution system (DS). The DS then updates its database, which contains the current location of the wireless stations. This database is needed for forwarding frames between different BSSs, i.e. between the different access points controlling the BSSs, which combine to form an ESS. Additionally, the DS can inform the old access point AP1 that the station is no longer within its BSS.

Infrared Data Associated (IrDA) The Infrared Data Association (IrDA) is an industry-driven interest group that was founded in 1993 by around 50 companies. IrDA provides specifications for a complete set of protocols for wireless infrared communications, and the name "IrDA" also refers to that set of protocols. The main reason for using IrDA had been wireless data transfer over the "last one meter" using point-and-shoot principles. Thus, it has been implemented in portable devices such as mobile telephones, laptops, cameras, printers, and medical devices. Main characteristics of this kind of wireless optical communication is physically secure data transfer, line-of-sight (LOS) and very low bit error rate (BER) that makes it very efficient [11].

Bluetooth Compared to the WLAN technologies presented earlier, 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. A different type of network is needed to connect different small devices in close proximity (about 10 m) without expensive wiring or the need for a wireless infrastructure. The envisaged gross data rate is 1 Mbit/s, asynchronous (data) and synchronous (voice) services should be available. The necessary transceiver components should be cheap.

There are various problems with IrDA: its very limited range (typically 2 m for built-in interfaces), the need for a line-of-sight between the interfaces, and, it is usually limited to two participants, i.e., only point-to-point connections are supported. IrDA has no internet working functions, has no media access, or any other enhanced communication mechanisms. The big advantage of IrDA is its low cost, and it can be found in almost any mobile device (laptops, PDAs, mobile phones).

The history of Bluetooth starts in the tenth century, when Harald Gormsen, King of Denmark (son of Gorm), erected a rune stone in Jelling, Denmark, in memory of his parents. The stone has three sides with elaborate carvings. One side shows a picture of Christ, as Harald did not only unite Norway and but also brought Christianity to Scandinavia. Harald had the common epithet 'Blåtand', meaning that he had a rather dark complexion (not a blue tooth). It took a thousand years before the Swedish IT-company Ericsson initiated some studies in 1994 around a so-called multi-communicator link (Haartsen, 1998). The project was renamed (because a friend of the designers liked Vikings) and Bluetooth was born. In spring 1998 five companies (Ericsson, Intel, IBM, Nokia, Toshiba) founded the Bluetooth consortium with the goal of developing a single-chip, low-cost, radio-based wireless network technology.

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 can 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.

Summary This section covered wireless LAN technology (WLAN). WLAN is likely more familiar to the average reader than cellular networks as it is more exposed to the user. The user itself can configure Bluetooth or IrDA on its phone as well as setup a home wireless LAN through a router or access point.

LAN's based on Wi-Fi standard is the most common today but Bluetooth is still around. WLANs still battles with worse performance than its wired counterpart but shines in its flexibility as the user is free to move within certain areas without having to worry about wires.

7 Mobile IP

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

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 in the Internet. While away from its home network, a mobile node is associated with a care-of address which identifies its current location and its home address is associated with the local endpoint of a tunnel to its home agent. 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.

You are probably aware of Dynamic Host Configuration Protocol (DHCP) which is a protocol used to be able to assign IP addresses to computers dynamically and is very popular since it allows to take your laptop to any network without having to configure a particular IP-address.

The situation when you move your phone that is connected to an IP-network is similar. The UE (User Equipment) or "mobile phone" is assigned an address by a process similar to DHCP by the mobile operator in question. Every time the UE is restarted, or even when moving across an invisible network boundary that you're not even aware of, your IP address is likely to change. Furthermore, most operators would have to use NAT to translate a private address assigned to the UE to a public address in order to conserve public IP addresses and that is completely dynamic.

Summary In many applications (e.g., VPN, VoIP), sudden changes in network connectivity and IP address can cause problems. Mobile IP was designed to support seamless and continuous Internet connectivity.

Mobile IP is most often found in wired and wireless environments where users need to carry their mobile devices across multiple LAN subnets. Examples of use are in roaming between overlapping wireless systems.

Mobile IP is not required within cellular systems such as 3G, to provide transparency when Internet users migrate between cellular towers, since these systems provide their own data link layer handover and roaming mechanisms. However, it is often used in 3G systems to allow seamless IP mobility between different packet data serving node (PDSN) domains.

8 Conclusions

I hope this document sparked your interest in mobile communication and made you feel a bit more comfortable with the basic concepts and terminology. This document is a miscellany of different sources on mobile communication. Some sections are tightly coupled with established and thrust-worthy sources, some

sections are based on less trust-worthy sources e.g Wikipedia, and some sections are based entirely on my own understanding and simplification of things to make certain concepts graspable without requiring heavy background theory. You should not consider this document as a source of truth to cite but rather as a quick overview and a layman's sloppy description of the field.

References

[1] Ericsson. Cloud ran - ericsson white paper, 2015.

[2] Aleksandra Checko et al. Cloud ran for mobile networks—a technology overview, 2014.

[3] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005.

[4] Analog Devices Inc. A beginner's guide to digital signal processing (dsp). `http://www.analog.com/en/design-center/landing-pages/001/beginners-guide-to-dsp.html`, 2017. [Online; accessed 14-Jun-2017].

[5] Bruce Kushnick. What are the public switched telephone networks, 'pstn' and why you should care?, 2013.

[6] Abhijit Mitra. Lecture notes on mobile communication, 2009.

[7] D Mohankumar. Mobile phone communication. how it works? `http://www.electroschematics.com/5231/mobile-phone-how-it-works/`, 2017.

[8] J.H. Schiller. *Mobile Communications*. Addison-Wesley, 2003.

[9] Rakesh Kumar Singh and Ranjan Singh. 4g lte cellular technology: Network architecture and mobile standards, 2016.

[10] Wikipedia. Global positioning system. `https://en.wikipedia.org/wiki/Global_Positioning_System/`, 2017.

[11] Wikipedia. Infrared data association. `https://en.wikipedia.org/wiki/Infrared_Data_Association/`, 2017.

[12] Wikipedia. Integrated services digital network. `https://sv.wikipedia.org/wiki/Integrated_Services_Digital_Network/`, 2017.

[13] Wikipedia. Radio access networks. `https://en.wikipedia.org/wiki/Radio_access_network/`, 2017.