1

Global System for Mobile Communications (GSM)

At the beginning of the 1990s, GSM, the Global System for Mobile Communications, triggered an unprecedented change in the way people communicate with each other. While earlier analog wireless systems were used only by a few, GSM is used worldwide by billions of people today. This has mostly been achieved by steady improvements in all areas of telecommunication technology and the resulting steady price reductions for both infrastructure equipment and mobile devices. This chapter discusses the architecture of this system, which also forms the basis for the packet-switched extension called General Packet Radio Service (GPRS), discussed in Chapter 2, for the Universal Mobile Telecommunications System (UMTS), which is described in Chapter 3 and Long-Term Evolution (LTE), which is discussed in Chapter 4.

Although the first standardization activities for GSM date back to the middle of the 1980s, GSM is still the most widely used wireless technology worldwide. In recent years, however, 4G LTE networks have become tremendously popular and a new service was standardized to support voice calls via the LTE radio network. This service is referred to as Voice over LTE (VoLTE) and is discussed separately in Chapter 5. Although efforts to roll out VoLTE are significant, a large percentage of mobile voice calls are still handled by GSM and UMTS networks to which devices without VoLTE support fall back for this service. In addition, even if a device and a network support VoLTE, a transfer to GSM or UMTS is still required when the user leaves the LTE-coverage area. As a consequence, knowledge of GSM is still required for a thorough understanding of how mobile networks are deployed and used in practice today.

1.1 Circuit-Switched Data Transmission

Initially, GSM was designed as a circuit-switched system that establishes a direct and exclusive connection between two users on every interface between all network nodes of the system. Section 1.1.1 gives a first overview of this traditional architecture. Over time, this physical circuit switching has been virtualized and many network nodes are connected over IP-based broadband connections today. The reasons for this and further details on virtual circuit switching can be found in Section 1.1.2.

From GSM to LTE-Advanced Pro and 5G: An Introduction to Mobile Networks and Mobile Broadband,
Third Edition. Martin Sauter.

1.1.1 Classic Circuit Switching

The GSM mobile telecommunication network has been designed as a circuit-switched network in a similar way to fixed-line phone networks. At the beginning of a call, the network establishes a direct connection between two parties, which is then used exclusively for this conversation. As shown in Figure 1.1, the switching center uses a switching matrix to connect any originating party to any destination party. Once the connection has been established, the conversation is then transparently transmitted via the switching matrix between the two parties. The switching center only becomes active again to clear the connection in the switching matrix if one of the parties wants to end the call. This approach is identical in both mobile and fixed-line networks. Early fixed-line telecommunication networks were designed only for voice communication, for which an analog connection between the parties was established. In the mid-1980s, analog technology was superseded by digital technology in the switching center. This meant that calls were no longer sent over an analog line from the originator to the terminator. Instead, the switching center digitized the analog signal that it received from the subscribers, which were directly attached to it, and forwarded the digitized signal to the terminating switching center. There, the digital signal was again converted back to an analog signal, which was then sent over the copper cable to the terminating party. In some countries, ISDN (Integrated Services Digital Network) lines were quite popular. With this system, the transmission became fully digital and the conversion back to an analog audio signal was done directly in the phone.

GSM reused much of the fixed-line technology that was already available at the time the standards were created. Thus, existing technologies such as switching centers and long-distance communication equipment were used. The main development for GSM, as shown in Figure 1.2, was the means to wirelessly connect the subscribers to the network. In fixed-line networks, subscriber connectivity is very simple as only two dedicated wires are necessary per user. In a GSM network, however, the subscribers are mobile and can change their location at any time. Thus, it is not possible to use the same input and output in the switching matrix for a user for each call as is the case in fixed-line networks.

As a mobile network consists of many switching centers, with each covering a certain geographical area, it is not even possible to predict in advance which switching center a

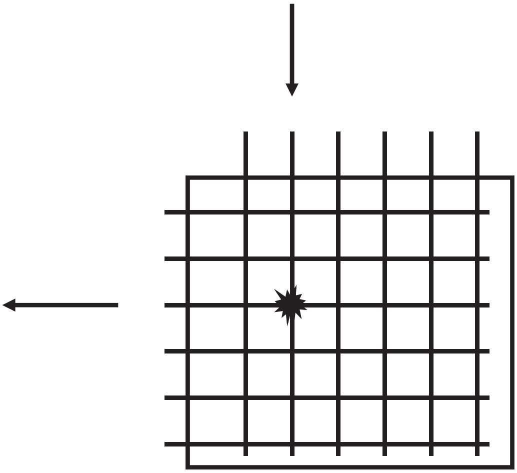


Figure 1.1 Switching matrix in a switching center.

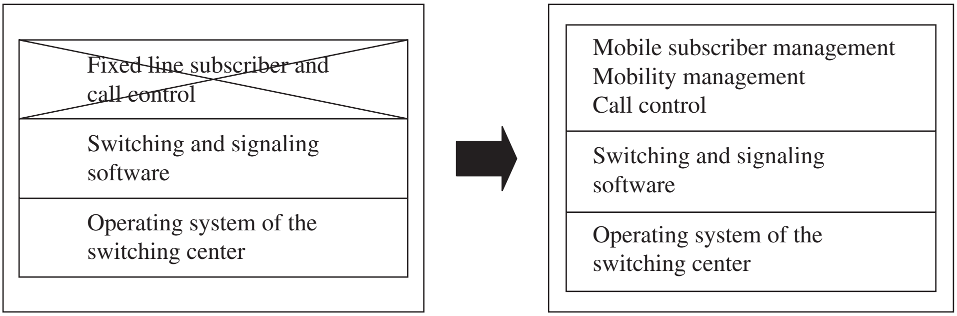

~~Fixed line subscriber and call control~~
Switching and signaling software
Operating system of the switching center
Mobile subscriber management
Mobility management
Call control
Switching and signaling software
Operating system of the switching center

Figure 1.2 Necessary software changes to adapt a fixed-line switching center for a wireless network.

call should be forwarded to for a certain subscriber. This means that the software for subscriber management and routing of calls of fixed-line networks cannot be used for GSM. Instead of a static call-routing mechanism, a flexible mobility management architecture in the core network became necessary, which needed to be aware of the current location of the subscriber and thus able to route calls to them at any time.

It was also necessary to be able to flexibly change the routing of an ongoing call as a subscriber can roam freely and thus might leave the coverage area of the radio transmitter of the network over which the call was established. While there was a big difference between the software of a fixed switching center and a Mobile Switching Center (MSC), the hardware as well as the lower layers of the software which are responsible, for example, for the handling of the switching matrix were mostly identical. Therefore, most telecommunication equipment vendors like Ericsson, Nokia Solutions and Networks, Huawei and Alcatel-Lucent offered their switching center hardware both for fixed-line and mobile networks. Only the software in the switching center determined whether the hardware was used in a fixed or mobile network (see Figure 1.2).

1.1.2 Virtual Circuit Switching over IP

While in the 1990s voice calls were the dominating form of communication, this has significantly changed today with the rise of the Internet. While voice calls remain important, other forms of communication such as e-mail, instant messaging (IM), social networks (e.g. Facebook), blogs, wikis and many more play an even bigger role. All these services share the Internet Protocol (IP) as a transport protocol and globally connect people via the Internet.

While circuit switching establishes an exclusive channel between two parties, the Internet is based on transferring individual data packets. A link with a high bandwidth is used to transfer the packets of many users. By using the destination address contained in each packet, each network node that the packet traverses decides over which outgoing link to forward the packet. Further details can be found in Chapter 2.

Owing to the rise of the Internet and IP-based applications, network operators thus had to maintain two separate networks: a circuit-switched network for voice calls and a packet-switched network for Internet-based services.

As the simultaneous operation of two different networks is very inefficient and costly, most network operators have, in the meantime, replaced the switching matrix in the

MSC with a device referred to as a media gateway. This allows them to virtualize circuit switching and to transfer voice calls over IP packets. The physical presence of a circuit-switched infrastructure is thus no longer necessary and the network operator can concentrate on maintaining and expanding a single IP-based network. This approach has been standardized under the name 'Bearer-Independent Core Network' (BICN).

The basic operation of GSM is not changed by this virtualization. The main differences can be found in the lower protocol levels for call signaling and voice call transmission. This will be looked at in more detail in the remainder of this chapter.

The trend toward IP-based communication can also be observed in the GSM radio network especially when a radio base station site supports GSM, UMTS and LTE simultaneously. Typically, connectivity is then established over a single IP-based link.

The air interface between the mobile devices and the network is not affected by the transition from circuit to packet switching. For mobile devices, whether the network uses classic or virtual circuit switching is therefore completely transparent.

1.2 Standards

As many network infrastructure manufacturers compete globally for orders from telecommunication network operators, standardization of interfaces and procedures is necessary. Without standards, which are defined by the International Telecommunication Union (ITU), it would not be possible to make phone calls internationally and network operators would be bound to the supplier they initially select for the delivery of their network components. One of the most important ITU standards, discussed in Section 1.4, is the Signaling System Number 7 (SS-7), which is used for call routing. Many ITU standards, however, only represent the lowest common denominator as most countries have specified their own national extensions. In practice, this incurs a high cost for software development for each country as a different set of extensions needs to be implemented in order for a vendor to be able to sell its equipment. Furthermore, the interconnection of networks of different countries is complicated by this.

GSM, for the first time, set a common standard for Europe for wireless networks. Due to its success it was later adopted around the globe. This is the main reason why subscribers can roam in GSM networks across the world that have roaming agreements with each other. The common standard also substantially reduces research and development costs as hardware and software can now be sold worldwide with only minor adaptations for the local market. The European Telecommunication Standards Institute (ETSI), which is also responsible for a number of other standards, was the main body responsible for the creation of the GSM standard. The ETSI GSM standards are composed of a substantial number of standards documents, each of which is called a technical specification (TS) and describes a particular part of the system. In the following chapters, many of these specifications are referenced and can thus be used for further information about a specific topic. Due to the global success of GSM, the 3rd Generation Partnership Project (3GPP) was later founded as a global organization and ETSI became one of the regional standardization bodies of the project. Today, 3GPP is responsible for maintaining and further developing the GSM, UMTS, LTE and 5G standards. All documents are freely available on the Internet at http://www.etsi.org [1] or at http://www.3gpp.org [2].

1.3 Transmission Speeds

The smallest transmission speed unit in a classic circuit-switched telecommunication network is the digital signal level 0 (DS0) channel. It has a fixed transmission speed of 64 kbit/s. Such a channel can be used to transfer voice or data, and thus it is usually not called a speech channel but simply referred to as a user data channel.

The reference unit of a telecommunication network is an E-1 connection in Europe and a T-1 connection in the United States, which use either a twisted pair or coaxial copper cable. The gross datarate is 2.048 Mbit/s for an E-1 connection and 1.544 Mbit/s for a T-1. An E-1 is divided into 32 timeslots of 64 kbit/s each, as shown in Figure 1.3 while a T-1 is divided into 24 timeslots of 64 kbit/s each. One of the timeslots is used for synchronization, which means that 31 timeslots for an E-1 or 23 timeslots for a T-1, respectively, can be used to transfer data. In practice, only 29 or 30 timeslots are used for user data transmission while the rest (usually one or two) are used for SS-7 signaling data (see Figure 1.3). More about SS-7 can be found in Section 1.4.

A single E-1 connection with 31 DS0s is not enough to connect two switching centers with each other. An alternative is an E-3 connection over twisted pair or coaxial cables. An E-3 connection is defined at a speed of 34.368 Mbit/s, which corresponds to 512 DS0s.

For higher transmission speeds and for long distances, optical systems that use the synchronous transfer mode (STM) standard are used. Table 1.1 shows some datarates and the number of 64 kbit/s DS0 channels that are transmitted per pair of fibers.

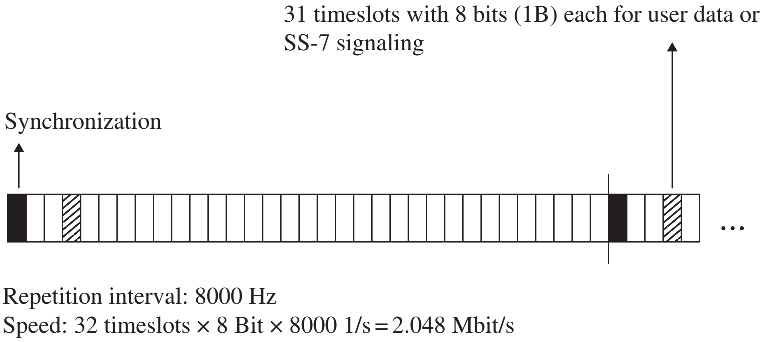

31 timeslots with 8 bits (1B) each for user data or SS-7 signaling
Synchronization
...
Repetition interval: 8000 Hz
Speed: 32 timeslots × 8 Bit × 8000 1/s = 2.048 Mbit/s

Figure 1.3 Timeslot architecture of an E-1 connection.

Table 1.1 STM transmission speeds and number of DS0s.

STM level	Speed (Mbit/s)	Approximate number of DS0 connections
STM-1	155.52	2300
STM-4	622.08	9500
STM-16	2488.32	37,000
STM-64	9953.28	148,279

For virtual circuit switching over IP, optical Ethernet links are often used between network nodes at the same location. Transmission speeds of 1 Gbit/s or more are used on these links. Unlike the circuit-switched technology described above, Ethernet is the de facto standard for IP-based communication over fiber and copper cables and is widely used. As a consequence, network equipment can be built much more cheaply.

1.4 The Signaling System Number 7

For establishing, maintaining and clearing a connection, signaling information needs to be exchanged between the end user and network devices. In the fixed-line network, analog phones signal their connection request when the receiver is lifted off the hook and a phone number dialed that is sent to the network either via pulses (pulse dialing) or via tone dialing, which is called dual tone multifrequency (DTMF) dialing. With fixed-line ISDN phones and GSM mobile phones, the signaling is done via a separate dedicated signaling channel, and information such as the destination phone number is sent as messages.

If several components in the network are involved in the call establishment, for example, if originating and terminating parties are not connected to the same switching center, it is also necessary that the different nodes in the network exchange information with each other. This signaling is transparent for the user, and a protocol called the SS-7 is used for this purpose. SS-7 is also used in GSM networks and the standard has been enhanced by ETSI to fulfill the special requirements of mobile networks, for example, subscriber mobility management.

The SS-7 standard defines three basic types of network nodes:

- Service Switching Points (SSPs) are switching centers that are more generally referred to as network elements and that are able to establish, transport or forward voice and data connections.
- Service Control Points (SCPs) are databases and application software that can influence the establishment of a connection. In a GSM network, SCPs can be used, for example, for storing the current location of a subscriber. During call establishment to a mobile subscriber, the switching centers query the database for the current location of the subscriber to be able to forward the call. More about this procedure can be found in Section 1.6.3 about the Home Location Register (HLR).
- Signaling Transfer Points (STPs) are responsible for the forwarding of signaling messages between SSPs and SCPs as not all network nodes have a dedicated link to all other nodes of the network. The principal functionality of an STP can be compared to an IP router in the Internet, which also forwards packets to different branches of the network. Unlike IP routers, however, STPs only forward signaling messages that are necessary for establishing, maintaining and clearing a call. The calls themselves are directly carried on dedicated links between the SSPs.

Figure 1.4 shows the general structure of an SS-7 circuit-switched telecommunication network and the way the nodes described above are interconnected with each other.

The SS-7 protocol stack is also used in virtual circuit-switched networks for communication between the network nodes. Instead of dedicated signaling timeslots on an

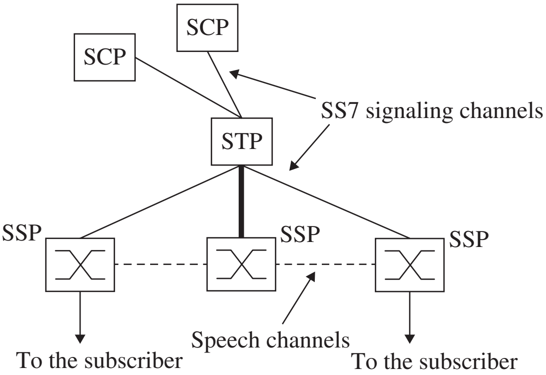


Figure 1.4 An SS-7 network with an STP, two SCP databases and three switching centers.

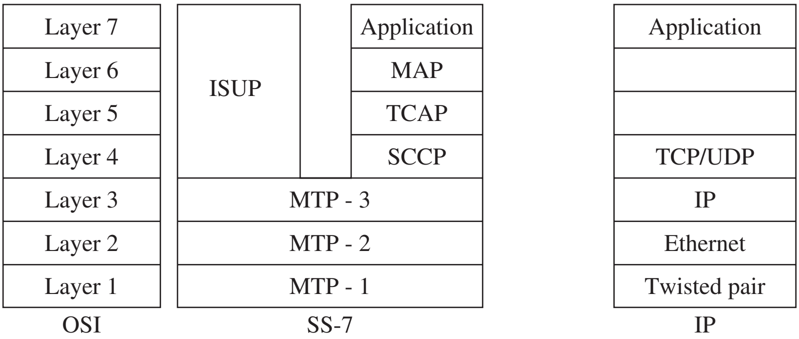
E-1 link, signaling messages are transported in IP packets. The following section describes the classic SS-7 protocol stack and afterward, the way SS-7 messages are transported over IP networks.

1.4.1 The Classic SS-7 Protocol Stack

SS-7 comprises a number of protocols and layers. A well-known model for describing telecommunication protocols and different layers is the Open System Interconnection (OSI) 7-layer model, which is used in Figure 1.5 to show the layers on which the different SS-7 protocols reside.

The Message Transfer Part 1 (MTP-1) protocol describes the physical properties of the transmission medium on layer 1 of the OSI model. Thus, this layer is also called the physical layer. Properties that are standardized in MTP-1 are, for example, the definition of the different kinds of cables that can be used to carry the signal, signal levels and transmission speeds.

On layer 2, the data link layer, messages are framed into packets and a start and stop identification at the beginning and end of each packet are inserted into the data stream so that the receiver is able to detect where a message ends and where a new message begins.



OSI	SS-7		IP
Layer 7	ISUP	Application	Application
Layer 6		MAP	
Layer 5		TCAP	
Layer 4		SCCP	TCP/UDP
Layer 3	MTP - 3		IP
Layer 2	MTP - 2		Ethernet
Layer 1	MTP - 1		Twisted pair

Figure 1.5 Comparison of the SS-7, OSI and TCP/IP protocol stacks.

Layer 3 of the OSI model, which is called the network layer, is responsible for packet routing. To enable network nodes to forward incoming packets to other nodes, each packet gets a source and destination address on this layer. This is done by the MTP-3 protocol of the SS-7 stack. For readers who are already familiar with the Transmission Control Protocol (TCP)/IP protocol stack, it may be noted at this point that the MTP-3 protocol fulfills the same tasks as the IP protocol. Instead of IP addresses, however, the MTP-3 protocol uses so-called point codes to identify the source and the destination of a message.

A number of different protocols are used on layers 4–7 depending on the application. If a message needs to be sent for the establishment or clearing of a call, the Integrated Services Digital Network User Part (ISUP) protocol is used. Figure 1.6 shows how a call is established between two parties by using ISUP messages. In the example, party A is a mobile subscriber while party B is a fixed-line subscriber. Thus, A is connected to the network via an MSC, while B is connected via a fixed-line switching center.

To call B, the phone number of B is sent by A to the MSC. The MSC then analyzes the national destination code (NDC) of the phone number, which usually comprises the first two to four digits of the number, and detects that the number belongs to a subscriber in the fixed-line network. In the example shown in Figure 1.6, the MSC and the fixed-line switching center are directly connected with each other. Therefore, the call can be directly forwarded to the terminating switching center. This is quite a realistic scenario as direct connections are often used if, for example, a mobile subscriber calls a fixed-line phone in the same city.

As B is a fixed-line subscriber, the next step for the MSC is to establish a voice channel to the fixed-line switching center. This is done by sending an ISUP Initial Address

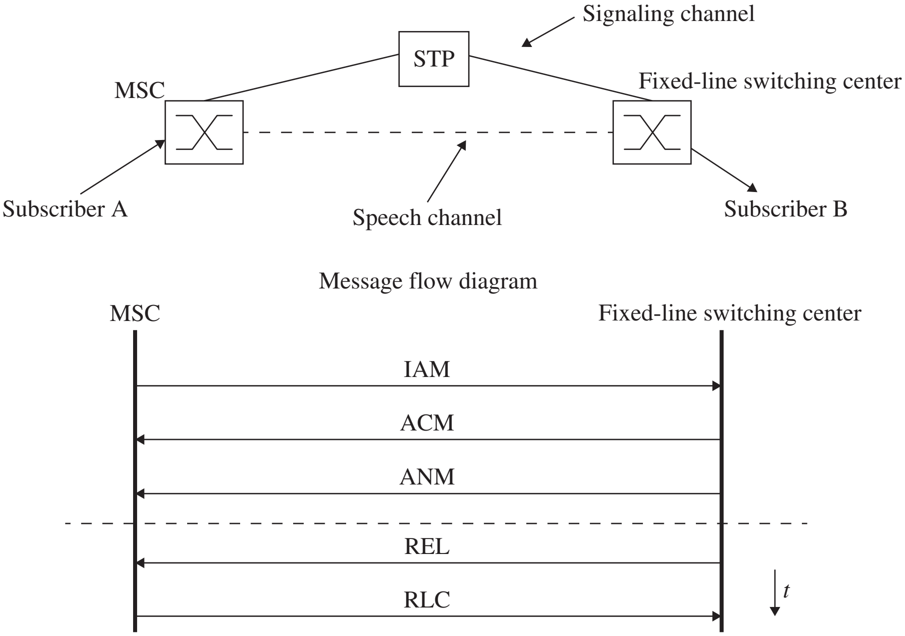


Figure 1.6 Establishment of a voice call between two switching centers.

Message (IAM). The message contains, among other data, the phone number of B and informs the fixed-line switching center of the channel that the MSC would like to use for the voice path. In the example, the IAM message is not sent directly to the fixed-line switching center. Instead, an STP is used to forward the message.

At the other end, the fixed-line switching center receives the message, analyzes the phone number and establishes a connection via its switching matrix to subscriber B. Once the connection is established via the switching matrix, the switch applies a periodic current to the line of the fixed-line subscriber so that the fixed-line phone can generate an alerting tone. To indicate to the originating subscriber that the phone number is complete and the destination party has been found, the fixed-line switch sends back an Address Complete Message (ACM). The MSC then knows that the number is complete and that the terminating party is being alerted about the incoming call.

If B answers the call, the fixed-line switching center sends an Answer Message (ANM) to the MSC and conversation can start.

When B ends the call, the fixed-line switching center resets the connection in the switching matrix and sends a Release (REL) message to the MSC. The MSC confirms the termination of the connection by sending back a Release Complete (RLC) message. If A had terminated the call, the messages would have been identical, with only the direction of the REL and RLC reversed.

For communication between the switching centers (SSPs) and the databases (SCPs), the Signaling Connection and Control Part (SCCP) is used on layer 4. SCCP is very similar to TCP and User Datagram Protocol (UDP) in the IP world. Protocols on layer 4 of the protocol stack enable the distinguishing of different applications on a single system. TCP and UDP use ports to do this. If a personal computer (PC), for example, is used as both a web server and a File Transfer Protocol (FTP) server at the same time, both applications would be accessed over the network via the same IP address. However, while the web server can be reached via port 80, the FTP server waits for incoming data on port 21. Therefore, it is quite easy for the network protocol stack to decide the application to which incoming data packets should be forwarded. In the SS-7 world, the task of forwarding incoming messages to the right application is done by SCCP. Instead of port numbers, SCCP uses Subsystem Numbers (SSNs).

For database access, the Transaction Capability Application Part (TCAP) protocol has been designed as part of the SS-7 family of protocols. TCAP defines a number of different modules and messages that can be used to query all kinds of different databases in a uniform way.

1.4.2 SS-7 Protocols for GSM

Apart from the fixed-line-network SS-7 protocols, the following additional protocols were defined to address the special needs of a GSM network.

- **The Mobile Application Part (MAP)**. This protocol has been standardized in 3GPP TS 29.002 [3] and is used for the communication between an MSC and the HLR, which maintains subscriber information. The HLR is queried, for example, if the MSC wants to establish a connection to a mobile subscriber. In this case, the HLR returns information about the current location of the subscriber. The MSC is then able to forward the call to the mobile subscriber's switching center establishing a voice channel between itself and the next hop by using the ISUP message flow that has been shown in Figure 1.6. MAP is also used between two MSCs if the subscriber moves into the coverage area of a different MSC while a call is ongoing. As shown in Figure 1.7, the MAP protocol uses the TCAP, SCCP and MTP protocols on lower layers.

- **The Base Station Subsystem Mobile Application Part (BSSMAP)**. This protocol is used for communication between the MSC and the radio network. Here, the additional protocol is necessary, for example, to establish a dedicated radio channel for a new connection to a mobile subscriber. As BSSMAP is not a database query language like the MAP protocol, it is based directly on SCCP instead of TCAP being used in between.
- **The Direct Transfer Application Part (DTAP)**. This protocol is used between the user's mobile device, which is also called mobile station (MS), and the MSC, to communicate transparently. To establish a voice call, the MS sends a Setup message to the MSC. As in the example in Section 1.4.1, this message contains among other things the phone number of the called subscriber. As it is only the MSC's task to forward calls, all network nodes between the MS and the MSC forward the message transparently and thus need not understand the DTAP protocol.

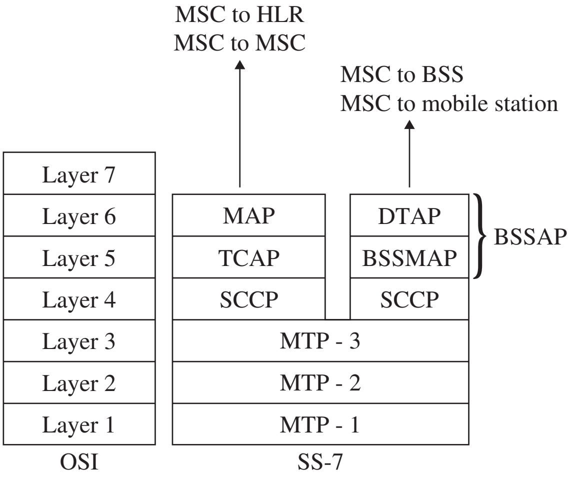


Figure 1.7 Enhancement of the SS-7 protocol stack for GSM.

1.4.3 IP-Based SS-7 Protocol Stack

When an IP network is used for the transmission of SS-7 signaling messages, the MTP-1 and MTP-2 protocols are replaced by the IP and the transport-medium-dependent lower-layer protocols (e.g. Ethernet). Figure 1.8 shows the difference between the IP stack and the classic stack presented in the previous section.

In the IP stack, layer-4 protocols are either UDP or TCP for most services. For the transmission of SS-7 messages, however, a new protocol has been specified, which is referred to as Stream Control Transmission Protocol (SCTP). When compared to TCP and UDP, it offers advantages when many signaling connections between two network nodes are active at the same time.

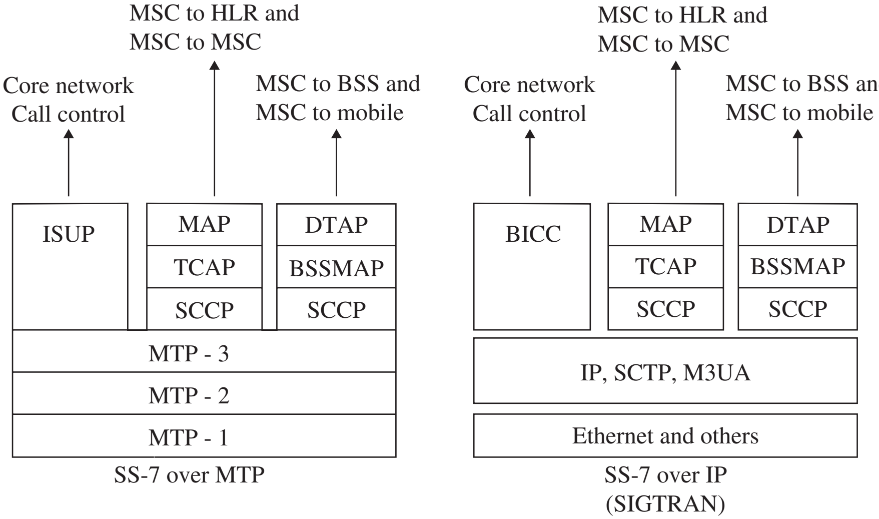


Figure 1.8 Comparison of the classic and IP-based SS-7 protocol stacks.

On the next protocol layer, SCTP is followed by the M3UA (MTP-3 User Adaptation Layer) protocol. As the name implies, the protocol is used to transfer information that is contained in the classic MTP-3 protocol. For higher protocol layers such as SCCP, M3UA simulates all functionalities of MTP-3. As a consequence, the use of an IP protocol stack is transparent to all higher-layer SS-7 protocols.

In the industry, the IP-based SS-7 protocol stack or the IP-based transmission of SS-7 messages is often referred to as SIGTRAN (signaling transmission). The abbreviation originated from the name of the IETF (Internet Engineering Task Force) working group that was created for the definition of these protocols.

As described in Section 1.1.1, the ISUP protocol is used for the establishment of voice calls between switching centers and the assignment of a 64 kbit/s timeslot. In an IP-based network, voice calls are transmitted in IP packets. As a consequence, the ISUP protocol has to be adapted as well. The resulting protocol is referred to as the Bearer-Independent Call Control (BICC) protocol, which largely resembles ISUP.

As IP links cannot be introduced on all interfaces in live networks at once, Signaling Gateways (SGWs) have been defined to bridge E-1-based and IP-based SS-7 communication. The SGWs adapt the lower layers of the protocol stack and thus make the differences transparent for both sides. This is necessary, for example, if the subscriber database has already been converted for IP interfaces while other components such as the switching centers are still using traditional signaling links.

To bridge voice calls between E-1-based and IP-based networks, Media Gateways (MGWs) are used. Connected to an MSC-Server, an MGW handles both IP-based and E-1-based voice calls transparently as it implements both the classic and IP-based signaling protocol stacks.

1.5 The GSM Subsystems

A GSM network is split into three subsystems which are described in more detail below:

- The Base Station Subsystem (BSS), which is also called 'radio network,' contains all nodes and functionalities that are necessary to wirelessly connect mobile subscribers over the radio interface to the network. The radio interface is usually also referred to as the 'air interface.'
- The Network Subsystem (NSS), which is also called 'core network,' contains all nodes and functionalities that are necessary for switching of calls, for subscriber management and mobility management.
- The Intelligent Network Subsystem (IN) comprises SCP databases that add optional functionality to the network. One of the most important optional IN functionalities of a mobile network is the prepaid service, which allows subscribers to first fund an account with a certain amount of money which can then be used for network services like phone calls, Short Messaging Service (SMS) messages and, of course, data services via GPRS and UMTS, as described in Chapters 2 and 3. When a prepaid subscriber uses a service of the network, the responsible IN node is contacted and the amount the network operator charges for a service is deducted from the account in real-time.

1.6 The Network Subsystem

The most important responsibilities of the NSS are call establishment, call control and routing of calls between different fixed and mobile switching centers and other networks. Other networks are, for example, the national fixed-line network, which is also called the Public Switched Telephone Network (PSTN), international fixed-line networks, other national and international mobile networks and Voice over Internet Protocol (VoIP) networks. Furthermore, the NSS is responsible for subscriber management. The nodes necessary for these tasks in a classic network architecture are shown in Figure 1.9. Figure 1.10 shows the nodes required in IP-based core networks. Both designs are further described in the following sections.

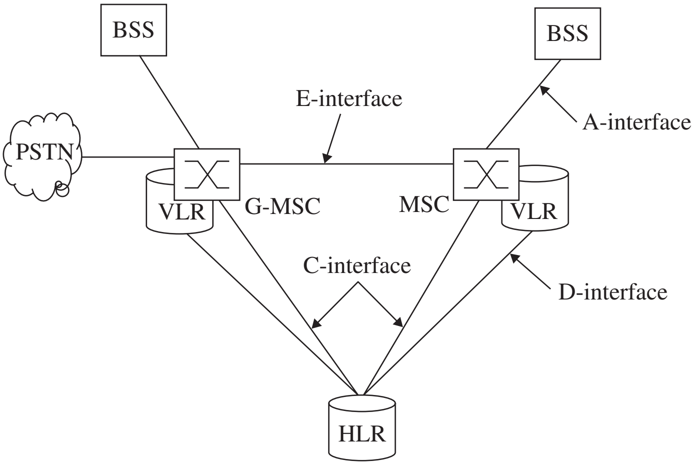


Figure 1.9 Interfaces and nodes in a classic NSS architecture.

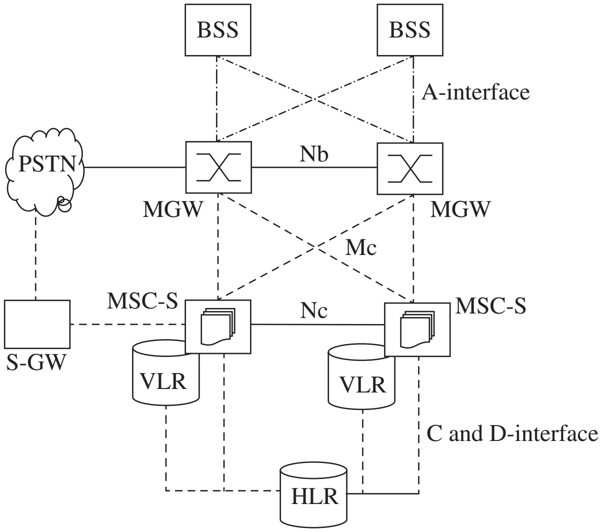


Figure 1.10 Interfaces and nodes in an IP-based NSS architecture.

1.6.1 The Mobile Switching Center (MSC), Server and Gateway

The MSC is the central element of a mobile telecommunication network, which is also called a Public Land Mobile Network (PLMN) in the standards. In a classic circuit-switched network, all connections between subscribers are managed by the MSC and are always routed over the switching matrix even if two subscribers who have established a connection communicate over the same radio cell.

The management activities to establish and maintain a connection are part of the Call Control (CC) protocol, which is generally responsible for the following tasks:

- Registration of mobile subscribers: When the mobile device, also referred to as MS, is switched on, it registers to the network and is then reachable by all other subscribers of the network.
- Call establishment and call routing between two subscribers.
- Forwarding of SMS messages.

As subscribers can roam freely in the network, the MSC is also responsible for the Mobility Management (MM) of subscribers. This activity comprises the following tasks:

- Authentication of subscribers at connection establishment is necessary because a subscriber cannot be identified as in the fixed network by the pair of copper cables over which the signal arrives. Authentication of subscribers and the authentication center (AuC) are further discussed in Section 1.6.4.
- If no active connection exists between the network and the mobile device, the MS has to report a change of location to the network to be reachable for incoming calls and SMS messages. This procedure is called location update and is further described in Section 1.8.1.
- If the subscriber changes their location while a connection is established with the network, the MSC is part of the process that ensures that the connection is not interrupted and is rerouted to the next cell. This procedure is called handover and is described in more detail in Section 1.8.3.

To enable the MSC to communicate with other nodes of the network, it is connected to them via standardized interfaces as shown in Figure 1.9. This allows network operators to acquire different components for the network from different network equipment vendors. The interfaces discussed below are either transmitted over timeslots in circuit-switched E-1 lines or over an IP-based network. As described earlier, only the lower protocol layers are affected by this. On the application layer, both variants are identical.

The BSS, which connects all subscribers to the core network, is connected to the MSCs via a number of 2 Mbit/s E-1 connections. This interface is called the A interface. As has been shown in Section 1.4, the BSSMAP and DTAP protocols are used over the A interface for communication between the MSC, the BSS and the mobile devices. As an E-1 connection can only carry 31 channels, many E-1 connections are necessary to connect an MSC to the BSS. In practice, this means that many E-1 s are bundled and sent over optical connections such as STM-1 to the BSS. Another reason to use an optical connection is that electrical signals can only be carried over long distances with great effort and it is not unusual that an MSC is over 100 km away from the next BSS node.

As an MSC has only a limited switching capacity and processing power, a PLMN is usually composed of dozens of independent MSCs. Each MSC thus covers only a certain area of the network. To ensure connectivity beyond the immediate coverage area of an MSC, E-1 s, which again are bundled into optical connections, are used to interconnect the different MSCs of a network. As a subscriber can roam into the area that is controlled by a different MSC while a connection is active, it is necessary to change the route of an active connection to the new MSC (handover). The necessary signaling connection is called the E interface. ISUP is used for the establishment of the speech path between different MSCs and the MAP protocol is used for the handover signaling between the MSCs. Further information on the handover process can be found in Section 1.8.3.

The C interface is used to connect the MSCs of a network with the HLR of the mobile network. While the A and E interfaces that were described previously always consist of signaling and speech path links, the C interface is a pure signaling link. Speech channels are not necessary for the C interface as the HLR is purely a database, which cannot accept or forward calls. Despite being only a signaling interface, E-1 connections are used for this interface. All timeslots are used for signaling purposes or are unused.

As has been shown in Section 1.3, a voice connection is carried over a 64 kbit/s E-1 timeslot in a classic circuit-switched fixed-line or mobile network. Before the voice signal can be forwarded, it needs to be digitized. For an analog fixed-line connection, this is done in the switching center, while an ISDN fixed-line phone or a GSM mobile phone digitizes the voice signal itself.

An analog voice signal is digitized in three steps, as shown in Figure 1.11: in the first step, the bandwidth of the input signal is limited to 300–3400 Hz to enable the signal with the limited bandwidth of a 64 kbit/s timeslot to be carried. Afterward, the signal is sampled at a rate of 8000 times per second. The next step in the processing is the quantization of the samples, which means that the analog samples are converted into 8-bit digital values that can each have a value from 0 to 255.

The higher the volume of the input signal, the higher the amplitude of the sampled value and its digital representation. To be able to also transmit low-volume conversations, the quantization is not linear over the whole input range but only in certain

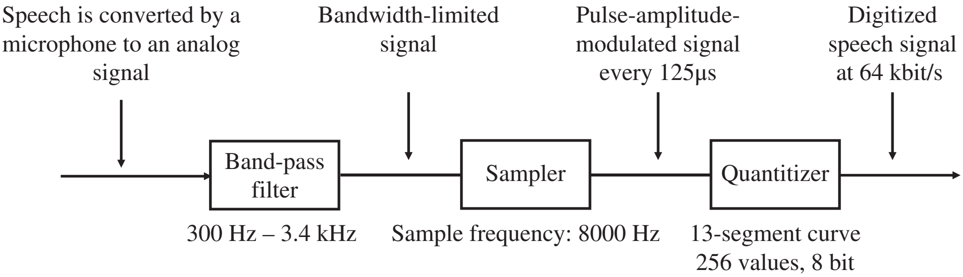

Speech is converted by a microphone to an analog signal
Bandwidth-limited signal
Pulse-amplitude-modulated signal every 125µs
Digitized speech signal at 64 kbit/s
Band-pass filter
Sampler
Quantitizer
300 Hz – 3.4 kHz
Sample frequency: 8000 Hz
13-segment curve 256 values, 8 bit

Figure 1.11 Digitization of an analog voice signal.

areas. For small input-signal amplitudes, a much higher range of digital values is used than for high-amplitude values. The resulting digital data stream is called a pulse code-modulated (PCM) signal. Which volume is represented by which digital 8-bit value is described in the A-law standard for European networks and in the µ-law standard in North America.

The use of different standards unfortunately complicates voice calls between networks using different standards. Therefore, it is necessary, for example, to convert a voice signal for a connection between France and the United States.

As the MSC controls all connections, it is also responsible for billing. This is done by creating a billing record for each call, which is later transferred to a billing server. The billing record contains information like the number of the caller and the calling party, cell ID of the cell from which the call originated, time of call origination, duration of the call, and so on. Calls for prepaid subscribers are treated differently as the charging is already done while the call is running. The prepaid billing service is usually implemented on an IN system and not on the MSC, as further described in Section 1.11.

MSC-Server and Media Gateway

In most of today's mobile voice networks, circuit-switched components have been replaced with IP-based devices. The MSC has been split into an MSC-Server (MSC-S) and an MGW. This is shown in Figure 1.10 and has been specified in 3GPP TS 23.205 [4]. The MSC-Ss are responsible for CC and MM (signaling), and the MGWs handle the transmission of virtual voice circuits (user data).

To establish a voice connection, MSC-Ss and MGWs communicate over the Mc interface. This interface does not exist in the classical model, as the MSC contained both components. 3GPP TS 29.232 [5] describes this interface on which the H.248/MEGACO (Media Gateway Control) protocol is used [6]. The protocol is used, for example, to establish voice channels to two parties and then to logically connect the two channels in the MGW. The protocol is also used to instruct the MGWs to play announcements to inform users of events, for example, where the called party is currently not available or is busy, and to establish conference calls between more than two subscribers. To add redundancy and for load-balancing reasons, several MSC-Ss and MGWs can be interconnected in a mesh. If an MSC-S fails, an MGW can thus still continue to operate, and is then controlled by another server. Thus, a single MSC-S is no longer solely responsible for a single geographical area as was the case in the traditional model.

On the radio network side, the A interface continues to be used to connect the radio network to the MSC-Ss and MGWs. The connection can be made without any changes in the radio network over the classic E-1-based A interface or over an IP-based A interface. In addition, the A interface has been made more flexible and can now be connected to several media gateways. This adds redundancy toward the radio network as well, as a geographical region can still be served even if a media gateway fails.

The Nc interface is used to transport voice calls within the core network, for example, to gateways, to other mobiles or to fixed networks. The protocol used on this interface is referred to as the BICC protocol and is very similar to the traditional ISUP protocol. This is specified in ITU Q.1901 [7] and 3GPP TS 29.205 [8]. By using an SGW as shown in Figure 1.10, the protocol can be converted into ISUP allowing the forwarding of calls to other core networks that are still based on the classic model. In practice, it can be observed that despite many networks having moved to an IP-based architecture, the gateways between them are still based on the classic architecture.

Virtual speech channels that have been negotiated over the Nc interface are transmitted between MGWs over the Nb interface. The combination of the Nb interface and Nc interface thus replaces the E interface of the classic network architecture. A voice channel is transmitted over IP connections either as PCM/G.711, Narrowband-AMR or Wideband-AMR, depending on the type of radio network, the configuration of the network and the capabilities of the mobile device. At the borders of the core network, for example, to and from the A interface to the GSM radio network or to and from a classic fixed-line PSTN network, MGWs can convert media streams, for example, between Narrowband-AMR over IP to G.711/PCM over E-1. This requires, however, that an MGW contain both Ethernet ports and E-1 ports.

Gateways between mobile networks are usually still based on ISUP and circuit-switched links, even though most networks are based on IP technology today. In the future, this is expected to change as advanced speech codecs such as Wideband-AMR can only be used over BICN and IP-based transport links.

Like in classic core networks, the C and D interfaces are used in a BICN network to communicate with the HLR. Instead of E-1 links, however, communication is based on IP links today.

1.6.2 The Visitor Location Register (VLR)

Each MSC has an associated Visitor Location Register (VLR), which holds the record of each subscriber that is currently served by the MSC (Figure 1.12). These records are only copies of the original records, which are stored in the HLR (see Section 1.6.3). The VLR is mainly used to reduce signaling between the MSC and the HLR. If a subscriber roams into the area of an MSC, the data are copied to the VLR of the MSC and are thus locally available for every connection establishment. Verification of the subscriber's record at every connection establishment is necessary as the record contains information about the services that are active and the services from which the subscriber is barred. Thus, it is possible, for example, to bar outgoing calls while allowing incoming calls, to prevent abuse of the system. While the standards allow implementation of the VLR as an independent hardware component, all vendors have implemented the VLR simply as a software component in the MSC. This is possible because MSC and VLR use different SCCP SSNs as shown in Figure 1.12 (see Section 1.4.1) and can thus run on a single physical node.

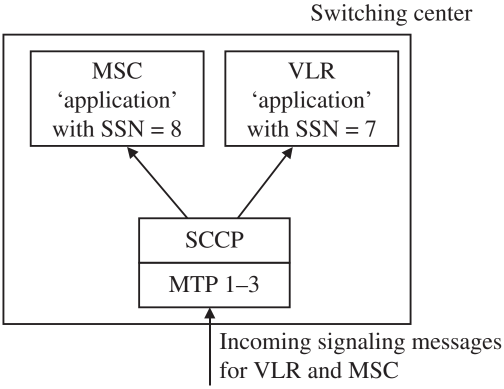

Switching center
MSC 'application' with SSN = 8
VLR 'application' with SSN = 7
SCCP
MTP 1–3
Incoming signaling messages for VLR and MSC

Figure 1.12 Mobile Switching Center (MSC) with integrated Visitor Location Register (VLR).

When a subscriber leaves the coverage area of an MSC, their record is copied from the HLR to the VLR of the new MSC, and is then removed from the VLR of the previous MSC. The communication with the HLR is standardized in the D interface specification, which is shown together with other MSC interfaces in Figure 1.9 and Figure 1.10.

1.6.3 The Home Location Register (HLR)

The HLR is the subscriber database of a GSM network. It contains a record for each subscriber, with information about the individually available services.

The International Mobile Subscriber Identity (IMSI) is an internationally unique number that identifies a subscriber and is used for most subscriber-related signaling in the network (Figure 1.13). The IMSI is stored in the subscriber's subscriber identity module (SIM) card and in the HLR and is thus the key to all information about the subscriber. The IMSI consists of the following parts:

- **The Mobile Country Code (MCC)**. The MCC identifies the subscriber's home country. Table 1.2 shows a number of MCC examples.
- **The Mobile Network Code (MNC)**. This part of the IMSI is the national part of a subscriber's home network identification. A national identification is necessary because there are usually several independent mobile networks in a single country. In the United Kingdom, for example, the following MNCs are used: 10 for O2, 15 for Vodafone, 30 for T-Mobile, 33 for Orange, 20 for Hutchison 3G, etc.
- **The Mobile Subscriber Identification Number (MSIN)**. The remaining digits of the IMSI form the MSIN, which uniquely identifies a subscriber within the home network.

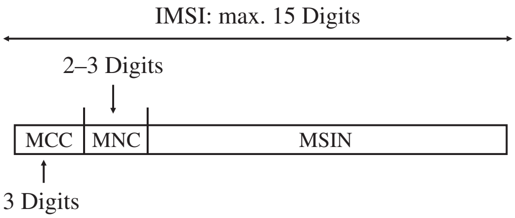


Figure 1.13 The International Mobile Subscriber Identity (IMSI).

Table 1.2 Mobile country codes.

MCC	Country
234	United Kingdom
310	United States
228	Switzerland
208	France
262	Germany
604	Morocco
505	Australia

As an IMSI is internationally unique, it enables a subscriber to use their phone abroad if a GSM network is available that has a roaming agreement with their home operator. When the mobile device is switched on, the IMSI is retrieved from the SIM card and sent to the MSC. There, the MCC and MNC of the IMSI are analyzed and the MSC is able to request the subscriber's record from the HLR of the subscriber's home network.

The phone number of the user, which is called the Mobile Subscriber Integrated Services Digital Network Number (MSISDN) in the GSM standards, has a length of up to 15 digits and consists of the following parts:

- The country code is the international code of the subscriber's home country. The country code has one to three digits such as +44 for the United Kingdom, +1 for the United States, and +353 for Ireland.
- The NDC usually represents the code with which the network operator can be reached. It is normally three digits in length. It should be noted that mobile networks in the United States use the same NDCs as fixed-line networks. Thus, it is not possible for a user to distinguish whether they are calling a fixed-line or a mobile phone. This impacts both billing and routing, as the originating network cannot deduct which tariff to apply from the NDC.
- The remainder of the MSISDN is the subscriber number, which is unique in the network.

There is usually a 1:1 or 1:N relationship in the HLR between the IMSI and the MSISDN. Furthermore, a mobile subscriber is normally assigned only a single MSISDN. However, as the IMSI is the unique identifier of a subscriber in the mobile network, it is also possible to assign several numbers to a single subscriber.

Another advantage of using the IMSI as the key to all subscriber information instead of the MSISDN is that the phone number of the subscriber can be changed without replacing the user's SIM card or changing any information on it. To change the MSISDN, only the HLR record of the subscriber needs to be changed. In effect, this means that the mobile device is not aware of its own phone number. This is not necessary because the MSC automatically adds the user's MSISDN to the message flow for a mobile-originated call establishment so that it can be presented to the called party.

Many countries have introduced functionality called mobile number portability (MNP), which allows a subscriber to retain their MSISDN even if they want to change

Table 1.3 Basic services of a GSM network.

Basic service	Description
Telephony	If this basic service is activated, a subscriber can use the voice telephony services of the network. This can be partly restricted by other supplementary services that are described below
Short messaging service (SMS)	If activated, a subscriber is allowed to use the SMS
Data service	Different circuit-switched data services can be activated for a subscriber with speeds of 2.4, 4.8, 9.6 and 14.4 kbit/s data calls
FAX	Allows or denies a subscriber the use of the FAX service, which can be used to exchange FAX messages with fixed-line or mobile devices

their mobile network operator. This is a great advantage for subscribers and for competition between mobile operators, but it also implies that it is no longer possible to discern the mobile network to which the call will be routed from the NDC. Furthermore, the introduction of MNP also increased the complexity of call routing and billing in both fixed-line and mobile networks, because it is no longer possible to use the NDC to decide which tariff to apply to a call. Instead of a simple call-routing scheme based on the NDC, the networks now have to query an MNP database for every call to a mobile subscriber to find out if the call can be routed inside the network or if it has to be forwarded to a different national mobile network.

Apart from the IMSI and MSISDN, the HLR contains a variety of information about each subscriber, such as which services they are allowed to use. Table 1.3 shows a number of 'basic services' that can be activated on a per subscriber basis.

In addition to the basic services described above, the GSM network offers a number of other services that can also be activated on a per subscriber basis. These services are called supplementary services and are shown in Table 1.4.

Most supplementary services can be activated by the network operator on a per subscriber basis and allow the operator to charge an additional monthly fee for some services if desired. Other services, like multiparty, can be charged on a per use basis. Although some network operators made use of this in the early years of GSM, most services are now included as part of the basic monthly fee.

Most services can be configured by the subscriber via a menu on the mobile device. The menu, however, is just a graphical front end for the user and the mobile device translates the user's commands into numerical strings which start with an '*' character. These strings are then sent to the network by use of an Unstructured Supplementary Service Data (USSD) message. The codes are standardized in 3GPP TS 22.030 [13] and are thus identical in all networks. As the menu is only a front end for the USSD service, the user can also input the USSD strings himself/herself via the keypad. After pressing the 'send' button, which is usually the button that is also used to start a phone call after typing in a phone number, the mobile device sends the string to the HLR via the MSC, where the string is analyzed and the requested operation is performed. For example, call forwarding to another phone (e.g. 0782 192 8355) while a user is already engaged in another call – call forward busy (CFB) is activated with the following string: **67* 07821928355# + call button.

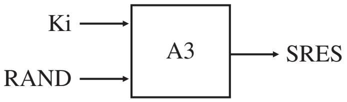
Table 1.4 Supplementary services of a GSM network.

Supplementary service	Description
Call forward unconditional (CFU)	If this service is activated, a number can be configured to which all incoming calls are forwarded immediately [9]. This means that the mobile device will not be notified of the incoming call even if it is switched on
Call forward busy (CFB)	This service allows a subscriber to define a number to which calls are forwarded if they are already engaged in a call when a second call comes in
Call forward no reply (CFNRY)	If this service is activated, it is possible to forward the call to a user-defined number if the subscriber does not answer the call within a certain time. The subscriber can change the number to which to forward the call as well as the timeout value (e.g. 25 seconds)
Call forward not reachable (CFNR)	This service forwards the call if the mobile device is attached to the network but is not reachable momentarily (e.g. temporary loss of network coverage)
Barring of all outgoing calls (BAOC)	This functionality can be activated by the network operator if, for example, the subscriber has not paid their monthly invoice in time. It is also possible for the network operator to allow the subscriber to change the state of this feature together with a PIN (personal identification number) so that the subscriber can lend the phone to another person for incoming calls only [10]
Barring of all incoming calls (BAIC)	Same functionality as provided by BAOC for incoming calls [10]
Call waiting (CW)	This feature allows signaling of an incoming call to a subscriber while they are already engaged in another call [11]. The first call can then be put on hold to allow the subscriber to accept the incoming call. The feature can be activated or barred by the operator and switched on or off by the subscriber
Call hold (HOLD)	This functionality is used to accept an incoming call during an already active call or to start a second call [11]
Calling line identification presentation (CLIP)	If activated by the operator for a subscriber, the functionality allows the switching center to forward the number of the caller
Calling line identification restriction (CLIR)	If allowed by the network, the caller can instruct the network not to show their phone number to the called party
Connected line presentation (COLP)	Shows the calling party the MSISDN to which a call is forwarded, if call forwarding is active at the called party side
Connected line presentation restriction (COLR)	If COLR is activated at the called party, the calling party will not be notified of the MSISDN to which the call is forwarded
Multiparty (MPTY)	Allows subscribers to establish conference bridges with up to six subscribers [12]

1.6.4 The Authentication Center

Another important part of the HLR is the AuC. The AuC contains an individual key per subscriber (Ki), which is a copy of the Ki on the SIM card of the subscriber. As the Ki is secret, it is stored in the AuC and especially on the SIM card in a way that prevents it from being read directly.

Figure 1.14 Creation of a signed response (SRES).



For many operations in the network, for instance, during the establishment of a call, the subscriber is identified by use of this key. Thus, it can be ensured that the subscriber's identity is not misused by a third party. Figure 1.15 shows how the authentication process is performed.

The authentication process, as shown in Figure 1.16, is initiated when a subscriber establishes a signaling connection with the network before the actual request (e.g. call establishment request) is sent. In the first step of the process, the MSC requests an authentication triplet from the HLR/AuC. The AuC retrieves the Ki of the subscriber and the authentication algorithm (A3 algorithm) based on the IMSI of the subscriber that is part of the message from the MSC. The Ki is then used together with the A3 algorithm and a random number to generate the authentication triplet, which contains the following values:

- **RAND**: A 128-bit random number.
- **SRES**: The signed response (SRES) is generated by using Ki, RAND and the A3 authentication algorithm, and has a length of 32 bits (see Figure 1.14).
- **Kc**: The ciphering key, Kc, is also generated by using Ki and RAND. It is used for the ciphering of the connection once the authentication has been performed successfully. Further information on this topic can be found in Section 1.7.7.

RAND, SRES and Kc are then returned to the MSC, which then performs authentication of the subscriber. It is important to note that the secret Ki key never leaves the AuC.

Extract of a decoded authentication request message
`SCCP MSG: Data Form 1`
`DEST. REF ID: 0B 02 00`
`DTAP MSG LENGTH: 19`
`PROTOCOL DISC.: Mobility Management`
`DTAP MM MSG: Auth. Request`
`Ciphering Key Seq.: 0`
`RAND in hex: 12 27 33 49 11 00 98 45 87 49 12 51 22 89 18 81 (16 B = 128 bit)`
Extract of a decoded authentication response message
`SCCP MSG: Data Form 1`
`DEST. REF ID: 00 25 FE`
`DTAP MSG LENGTH: 6`
`PROTOCOL DISC.: Mobility Management`
`DTAP MM MSG: Auth. Response`
`SRES in hex: 37 21 77 61 (4 B = 32 bit)`

Figure 1.15 Message flow during the authentication of a subscriber.

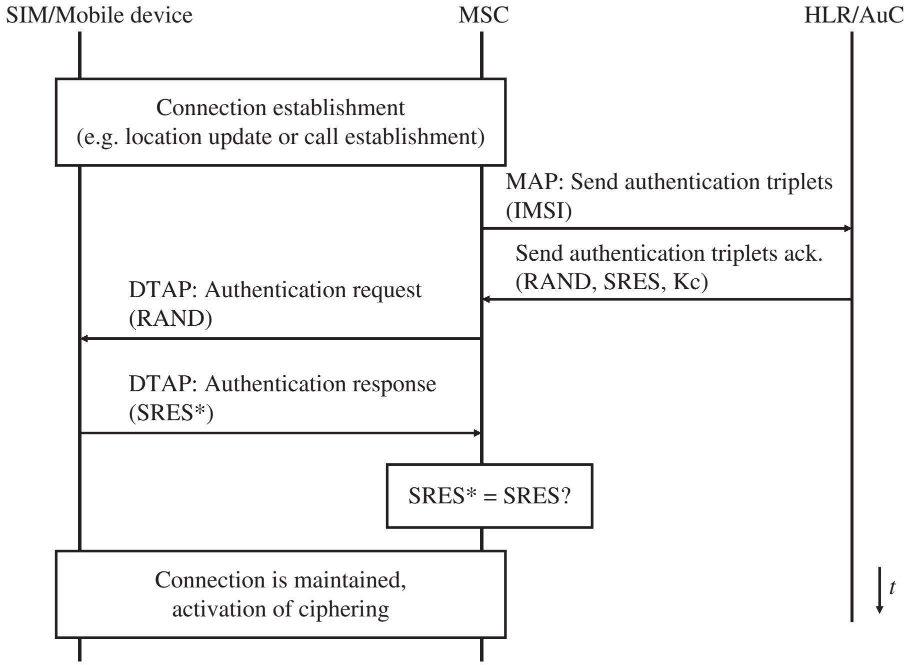

SIM/Mobile device
MSC
HLR/AuC
Connection establishment
(e.g. location update or call establishment)
MAP: Send authentication triplets
(IMSI)
Send authentication triplets ack.
(RAND, SRES, Kc)
DTAP: Authentication request
(RAND)
DTAP: Authentication response
(SRES*)
SRES* = SRES?
Connection is maintained,
activation of ciphering
t

Figure 1.16 Authentication between network and mobile device.

To speed up subsequent connection establishments, the AuC usually returns several authentication triplets per request. These are buffered by the MSC/VLR and are used during subsequent connection establishments.

In the next step, the MSC sends the RAND inside an Authentication Request message to the mobile device. The mobile device forwards the RAND to the SIM card, which then uses the Ki and the authentication A3 algorithm to generate a signed response (SRES*). The SRES* is returned to the mobile device and then sent back to the MSC inside an Authentication Response message. The MSC then compares SRES and SRES*, and if they are equal, the subscriber is authenticated and allowed to proceed with the communication.

As the secret key, Ki, is not transmitted over any interface that could be eavesdropped on, it is not possible for a third party to correctly calculate an SRES. As a fresh random number is used for the next authentication, it is also pointless to intercept the SRES* and use it for another authentication. A detailed description of the authentication procedure and many other procedures between the mobile device and the core network can be found in 3GPP TS 24.008 [14].

Figure 1.16 shows some parts of an authentication request and an Authentication Response message. Apart from the format of RAND and SRES, it is also interesting to note the different protocols that are used to encapsulate the message (see Section 1.4.2).

1.6.5 The Short Messaging Service Center (SMSC)

Another important network element is the Short Messaging Service Center (SMSC), which is used to store and forward short messages. The SMS was only introduced about four years after the first GSM networks went into operation, as an add-on, and has been specified in 3GPP TS 23.040 [15]. Most industry observers were quite skeptical at that time as the general opinion was that if it was necessary to convey some information, it would be done by calling someone rather than by the more cumbersome method of typing a text message on the small keypad. However, they were proved wrong and today most GSM operators (still) generate a significant amount of their revenue from the short message service, despite a trend towards replacing SMS messaging with other forms of mobile-Internet-based IM.

SMS can be used for person-to-person messaging as well as for providing notification of other events such as a missed call that was forwarded to the voice mail system. The transfer method for both cases is identical.

The sender of an SMS prepares the text for the message and then sends the SMS via a signaling channel to the MSC as shown in Figure 1.17. As a signaling channel is used, an SMS is just an ordinary DTAP SS-7 message and thus, apart from the content, very similar to other DTAP messages, such as a Location Update message or a Setup message to establish a voice call. Apart from the text, the SMS message also contains the MSISDN of the destination party and the address of the SMSC, which the mobile device has retrieved from the SIM card. When the MSC receives an SMS from a subscriber, it transparently forwards the SMS to the SMSC. As the message from the mobile device contains the address of the subscriber's SMSC, international roaming is possible and the foreign MSC can forward the SMS to the home SMSC without the need for an international SMSC database (see Figure 1.17).

To deliver a message, the SMSC analyzes the MSISDN of the recipient and retrieves its current location (the MSC concerned) from the HLR. The SMS is then forwarded to the MSC concerned. If the subscriber is currently attached, the MSC tries to contact the mobile device, and if an answer is received, the SMS is forwarded. Once the mobile device has confirmed the proper reception of the SMS, the MSC notifies the SMSC as well and the SMS is deleted from the SMSC's data storage.

If the subscriber is not reachable because the battery of the mobile device is empty, network coverage has been lost temporarily or the device is simply switched off, it is not possible to deliver the SMS. In this case, the message waiting flag is set in the VLR and the SMS is stored in the SMSC. Once the subscriber communicates with the MSC, the MSC notifies the SMSC to reattempt delivery.

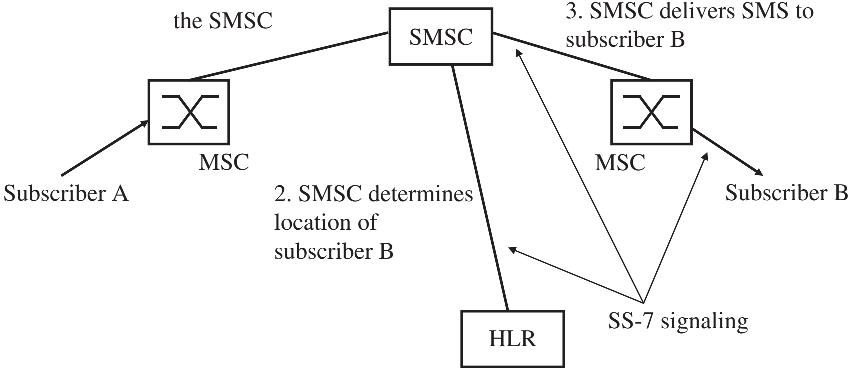

1. SMS is forwarded to the SMSC
SMSC
3. SMSC delivers SMS to subscriber B
MSC
MSC
Subscriber A
2. SMSC determines location of subscriber B
Subscriber B
HLR
SS-7 signaling

Figure 1.17 SMS delivery principle.

As the message waiting flag is also set in the HLR, the SMS also reaches a subscriber who has switched off the mobile device in London, for example, and switches it on again after a flight to Los Angeles. When the mobile device is switched on in Los Angeles, the visited MSC reports the location to the subscriber's home HLR (location update). The HLR then sends a copy of the user's subscription information to the MSC/VLR in Los Angeles including the message waiting flag and thus the SMSC can also be notified that the user is reachable again.

The SMS delivery mechanism does not include a delivery report for the sender of the SMS by default. The sender is only notified that the SMS has been correctly received by the SMSC. However, if supported by a device, it is also possible to request an end-to-end delivery notification from the SMSC. In practice, there are a number of different ways this is implemented in mobile devices. In some mobile operating systems, delivery reports can be activated in the SMS settings. Confirmations are then shown with a symbol next to the message or are displayed in the status bar. Other operating systems include a separate list of received or pending confirmations.

1.7 The Base Station Subsystem (BSS) and Voice Processing

While most functionality required in the NSS for GSM could be added via additional software, the BSS had to be developed from scratch. This was mainly necessary as earlier generation systems were based on analog transmission over the air interface and thus did not have much in common with the GSM BSS.

1.7.1 Frequency Bands

In Europe, GSM was initially specified only for operation in the 900 MHz band between 890 and 915 MHz in the uplink direction and between 935 and 960 MHz in the downlink direction, as shown in Figure 1.18. 'Uplink' refers to the transmission from the mobile device to the network and 'downlink' to the transmission from the network to the mobile device. The bandwidth of 25 MHz is split into 125 channels with a bandwidth of 200 kHz each.

It soon became apparent that the number of available channels was not sufficient to cope with the growing demand in many European countries. Therefore, the regulating bodies assigned an additional frequency range for GSM, which uses the frequency band from 1710 to 1785 MHz for the uplink and from 1805 to 1880 for the downlink. Instead of a total bandwidth of 25 MHz as in the 900 MHz range, the 1800 MHz band offers 75 MHz of bandwidth, which corresponds to 375 additional channels. The functionality

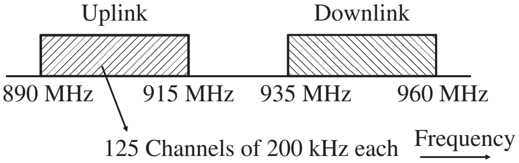


Figure 1.18 GSM uplink and downlink in the 900 MHz frequency band.

Table 1.5 GSM frequency bands.

Band	ARFCN	Uplink (MHz)	Downlink (MHz)
GSM 900 (primary)	0–124	890–915	935–960
GSM 900 (extended)	975–1023, 0–124	880–915	925–960
GSM 1800	512–885	1710–1785	1805–1880
GSM 1900 (North America)	512–810	1850–1910	1930–1990
GSM 850 (North America)	128–251	824–849	869–894
GSM-R	0–124, 955–1023	876–915	921–960

of GSM is identical on both frequency bands, with the channel numbers, also referred to as the Absolute Radio Frequency Channel Numbers (ARFCNs), being the only difference (see Table 1.5).

While GSM was originally intended only as a European standard, the system soon spread to countries in other parts of the globe. In North America, analog mobile networks continued to be used for some time before second-generation networks, which included the use of the GSM technology, were introduced. As the 900 MHz and the 1800 MHz bands were already in use by other systems the North American regulating body chose to open frequency bands for the new systems in the 1900 MHz band and later on in the 850 MHz band.

The GSM standard is also used by railway communication networks in Europe and other parts of the world. For this purpose, GSM was enhanced to support a number of private mobile radio and railway-specific functionalities and this version is known as GSM-R. The additional functionalities include the following:

- **The Voice Group Call Service (VGCS)**. This service offers a circuit-switched walkie-talkie functionality to allow subscribers who have registered to a VGCS group to communicate with all other subscribers in the area who have also subscribed to the group. To talk, the user has to press a ‘push to talk’ button. If no other subscriber holds the uplink, the network grants the request and blocks the uplink for all other subscribers while the push to talk button is pressed. The VGCS service is very efficient, especially if many subscribers participate in a group call, as all mobile devices that participate in the group call listen to the same timeslot in the downlink direction. Further information about this service can be found in 3GPP TS 43.068 [16].
- **The Voice Broadcast Service (VBS)**. It is similar to VGCS, with the restriction that only the originator of the call is allowed to speak. Further information about this service can be found in 3GPP TS 43.069 [17].
- **Enhanced Multi-Level Precedence and Preemption (EMLPP)**. This functionality, which is specified in 3GPP TS 23.067 [18], is used to attach a priority to a point-to-point, VBS or VGCS call. This enables the network and the mobile devices to automatically preempt ongoing calls for higher priority calls to ensure that emergency calls (e.g. a person has fallen on the track) are not blocked by lower priority calls and a lack of resources (e.g. because no timeslots are available).

As GSM-R networks are private networks, it has been decided to assign a private frequency band in Europe for this purpose, which is just below the public 900 MHz

GSM band. To use GSM-R, mobile phones need to be slightly modified to be able to send and receive in this frequency range. This requires only minor software and hardware modifications. To be also able to use the additional functionalities described above, further extensions of the mobile device software are necessary. More about GSM-R can be found at http://www.uic.org/gsm-r [19].

1.7.2 The Base Transceiver Station (BTS)

Base stations, which are also called base transceiver stations (BTSs), are the most visible network elements of a GSM system (Figure 1.19). Compared to fixed-line networks, the base stations replace the wired connection to the subscriber with a wireless connection, which is also referred to as the air interface. Base stations are also the most numerous components of a mobile network. In Germany, for example, Telefonica O2 has over 18,000 GSM base stations and the other three network operators are likely to have deployed similar numbers [20]. Figure 1.19 shows a typical base station antenna.

In theory, a base station can cover an area with a radius of up to 35 km. This area is also called a cell. As a base station can only serve a limited number of simultaneous users, cells are much smaller in practice, especially in dense urban environments. In these environments, cells cover areas within a radius from 1 to 2 km in residential and business areas, down to only several hundred meters with minimal transmission power in heavily frequented areas like shopping centers and downtown streets. Even in rural areas, a cell's coverage area is usually less than 15 km, with the transmission power of the mobile device of 1 or 2 W being the limiting factor in this case.

As the emissions of different base stations of the network must not interfere with each other, all neighboring cells have to send on different frequencies. As can be seen from Figure 1.20, a single base station usually has quite a number of neighboring

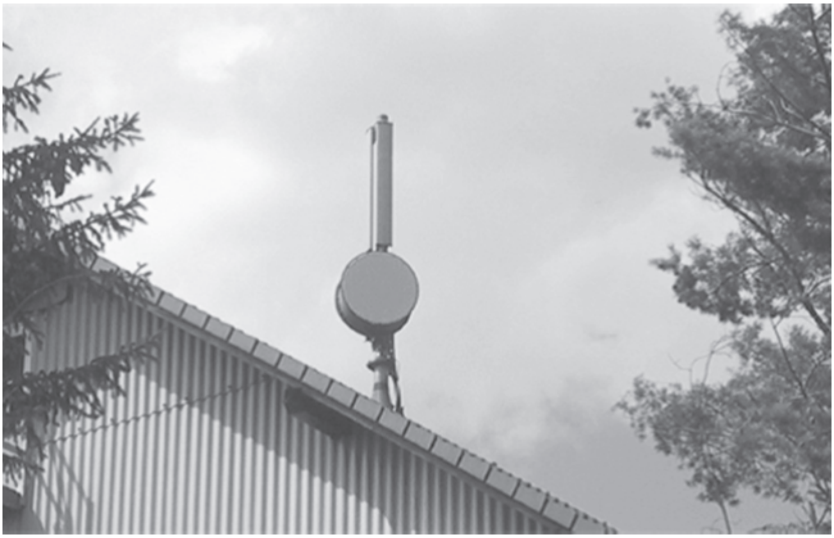


Figure 1.19 A typical antenna of a GSM base station. The optional microwave directional antenna (round antenna at the bottom of the mast) connects the base station with the GSM network. Source: Martin Sauter. Reproduced by permission of Martin Sauter.

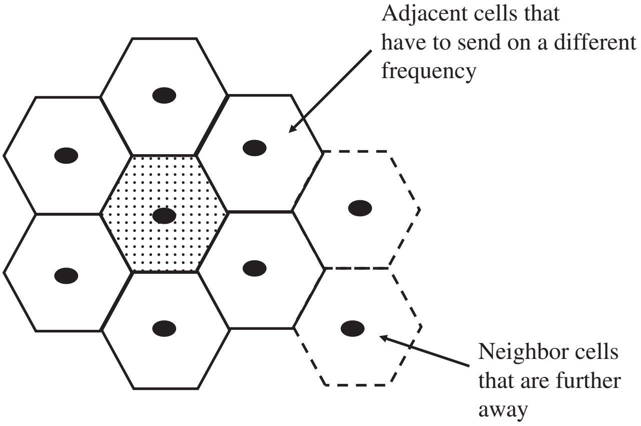
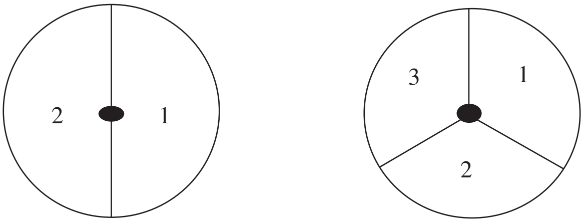

Adjacent cells that have to send on a different frequency
Neighbor cells that are further away

Figure 1.20 Cellular structure of a GSM network.

Figure 1.21 Sectorized cell configurations.



sites. Therefore, only a limited number of different frequencies can be used per base station to increase capacity.

To increase the capacity of a base station, the coverage area is usually split into two or three sectors, as shown in Figure 1.21, which are then covered on different frequencies by a dedicated transmitter. This allows a better reuse of frequencies in two-dimensional space than is the case where only a single frequency is used for the whole base station. Each sector of the base station, therefore, forms its own independent cell.

1.7.3 The GSM Air Interface

The transmission path between the BTS and the mobile device is referred to, in the GSM specifications, as the air interface or the Um interface. To allow the base station to communicate with several subscribers simultaneously, two methods are used. The first method is frequency division multiple access (FDMA), which means that users communicate with the base station on different frequencies. The second method used is time division multiple access (TDMA). GSM uses carrier frequencies with a bandwidth of 200 kHz over which up to eight subscribers can communicate with the base station simultaneously as shown in Figure 1.22.

Subscribers are time multiplexed by dividing the carrier into frames with durations of 4.615 milliseconds. Each frame contains eight physically independent timeslots, each for communication with a different subscriber. The time frame of a timeslot is called a burst and the burst duration is 577 microseconds. For example, if a mobile device is

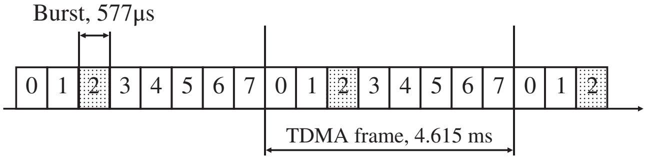


Figure 1.22 A GSM TDMA frame.

allocated timeslot number 2 for a voice call, then the mobile device will send and receive only during this burst. Afterward, it has to wait until the next frame before it is allowed to send again.

By combining the two multiple access schemes it is possible to approximately calculate the total capacity of a base station. For the following example, it is assumed that the base station is split into three sectors and each sector is covered by an independent cell. Each cell is typically equipped with three transmitters and receivers (transceivers). In each sector, 3 × 8 = 24 timeslots are thus available. Two timeslots are usually assigned for signaling purposes, which leaves 22 timeslots per sector for user channels. Let us further assume that four or more timeslots are used for the packet-switched GPRS service (see Chapter 2). Therefore, 18 timeslots are left for voice calls per sector, which amounts to 54 channels for all sectors of the base station. In other words, this means that 54 subscribers per base station can communicate simultaneously.

A single BTS, however, provides service to a much higher number of subscribers, as all of them do not communicate at the same time. Mobile operators, therefore, base their network dimensioning on a theoretical call profile model in which the number of minutes per hour that a subscriber statistically uses the system is one of the most important parameters. A commonly used value for the number of minutes per hour that a subscriber uses the system is 3. This means that a base station is able to provide service to 20 times the number of active subscribers. In this example, a base station with 54 channels is, therefore, able to provide service to about 1080 subscribers.

This number is quite realistic as the following calculation shows: Telefonica O2 Germany had a subscriber base of about 20 million in 2014 [20]. If this value is divided by the number of subscribers per cell, the total number of base stations required to serve such a large subscriber base can be determined. With our estimation above, the number of base stations required for the network would be about 18,500. This value is in line with the numbers published by the operator [20].

Each burst of a TDMA frame is divided into a number of different sections as shown in Figure 1.23. Each burst ends with a guard time in which no data is sent. This is necessary because the distance of the different subscribers from the base station can change while they are active. As airwaves propagate 'only' through space at the speed of light, the signal of a faraway subscriber takes a longer time to reach the base station compared to that of a subscriber who is closer to the base station. To prevent any overlap, guard times were introduced. These parts of the burst are very short, as the network actively controls the timing advance of the mobile device. More about this topic can be found below.

The training sequence in the middle of the burst always contains the same bit pattern. It is used to compensate for interference caused, for example, by reflection, absorption and multipath propagation. On the receiver side, these effects are countered by

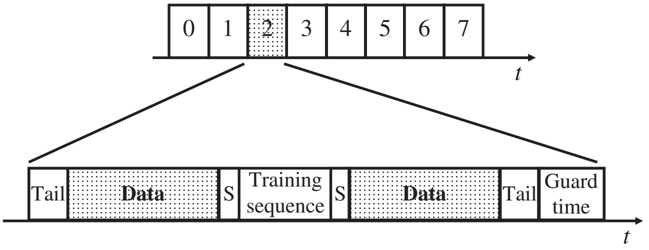


Figure 1.23 A GSM burst.

comparing the received signal with the training sequence and thus adapting the analog filter parameters for the signal. The filter parameters calculated this way can then be used to modify the rest of the signal and thus to better recreate the original signal.

At the beginning and end of each burst, another well-known bit pattern is sent to enable the receiver to detect the beginning and end of a burst correctly. These fields are called 'tails'. The actual user data of the burst, that is, the digitized voice signal, is sent in the two user data fields with a length of 57 bits each. This means that a 577-microsecond burst transports 114 bits of user data. Finally, each frame contains 2 bits to the left and right of the training sequence, which are called 'stealing bits'. These bits indicate whether the data fields contain user data or are used ('stolen') for urgent signaling information. User data from bursts that carry urgent signaling information are, however, lost. As shown below, the speech decoder is able to cope with short interruptions of the data stream quite well, and thus the interruptions are normally not audible to the user.

For the transmission of user or signaling data, the timeslots are arranged into logical channels. A user data channel for the transmission of digitized voice data, for example, is a logical channel. On the first carrier frequency of a cell, the first two timeslots are usually used for common logical signaling channels while the remaining six independent timeslots are used for user data channels or GPRS. As there are more logical channels than physical channels (timeslots) for signaling, 3GPP TS 45.002 [21] describes how 51 frames are grouped into a multiframe able to carry a number of different signaling channels over the same timeslot. In such a multiframe, which is infinitely repeated, which logical channels are transmitted in which bursts is specified on timeslots 0 and 1. For user data timeslots (e.g. voice), the same principle is used. Instead of 51 frames, these timeslots are grouped into a 26-multiframe pattern. For the visualization of this principle, a scheme is shown in Figure 1.24 which depicts how the eight timeslots of a frame are grouped into a two-dimensional table. In Figure 1.25, this principle is used to show how the logical channels are assigned to physical timeslots in the multiframe.

Logical channels are arranged into two groups. If data on a logical channel is dedicated to a single user, the channel is called a dedicated channel. If the channel is used for data that needs to be distributed to several users, the channel is called a common channel.

Let us take a look at the dedicated channels first:

- The traffic channel (TCH) is a user data channel. It can be used to transmit a digitized voice signal or circuit-switched data services of up to 14.4 kbit/s.
- The Fast Associated Control Channel (FACCH) is transmitted on the same timeslot as a TCH. It is used to send urgent signaling messages like a handover command. As

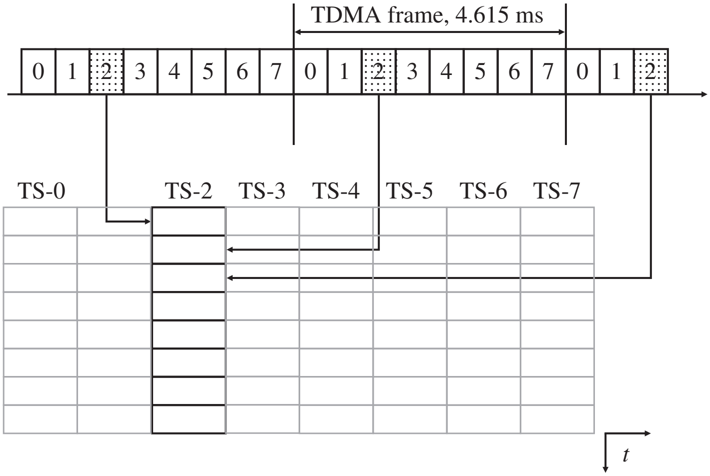


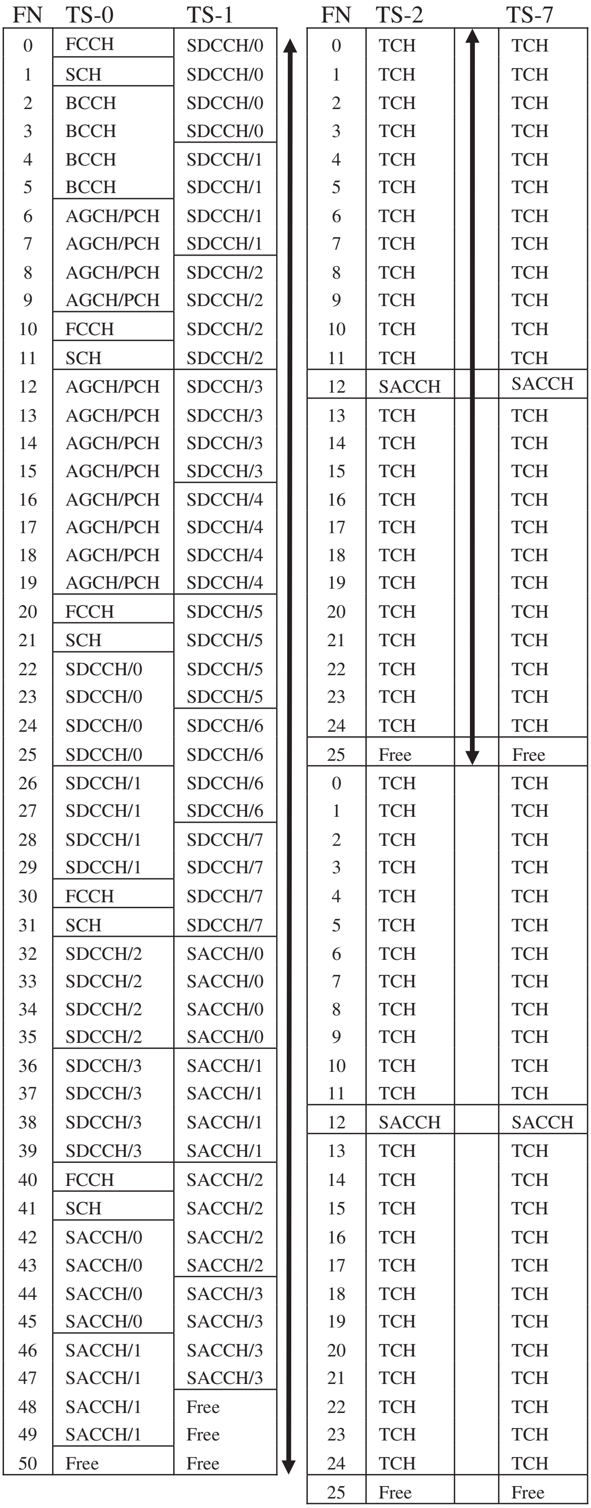
Figure 1.24 Arrangement of bursts of a frame for the visualization of logical channels in Figure 1.25.

these messages do not have to be sent very often, no dedicated physical bursts are allocated to the FACCH. Instead, user data is removed from a TCH burst. To inform the mobile device of this, the stealing bits to the left and right of the training sequence, as shown in Figure 1.23, are used. This is the reason why the FACCH is not shown in Figure 1.25.

- The Slow Associated Control Channel (SACCH) is also assigned to a dedicated connection. It is used in the uplink direction to report signal quality measurements of the serving cell and neighboring cells to the network. The network then uses these values for handover decisions and power control. In the downlink direction, the SACCH is used to send power control commands to the mobile device. Furthermore, the SACCH is used for timing advance control, which is described in Section 1.7.4 and Figure 1.26. As these messages are only of low priority and the necessary bandwidth is very small, only a few bursts are used on a 26-multiframe pattern at fixed intervals.
- The Standalone Dedicated Control Channel (SDCCH) is a pure signaling channel that is used during call establishment when a subscriber has not yet been assigned a TCH. Furthermore, the channel is used for signaling that is not related to call establishment, such as for the location update procedure or for sending or receiving a text message (SMS).

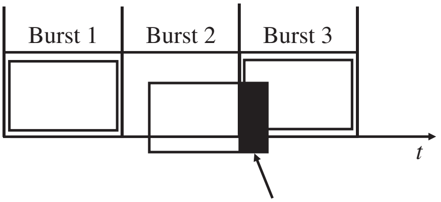
Besides the dedicated channels, which are always assigned to a single user, there are a number of common channels that are monitored by all subscribers in a cell:

- The Synchronization Channel (SCH) is used by mobile devices during network and cell searches.
- The Frequency Correction Channel (FCCH) is used by the mobile devices to calibrate their transceiver units, and is also used to detect the beginning of a multiframe.
- The Broadcast Common Control Channel (BCCH) is the main information channel of a cell and broadcasts SYS_INFO messages that contain a variety of information



FN	TS-0	TS-1	FN	TS-2	TS-7
0	FCCH	SDCCH/0	0	TCH	TCH
1	SCH	SDCCH/0	1	TCH	TCH
2	BCCH	SDCCH/0	2	TCH	TCH
3	BCCH	SDCCH/0	3	TCH	TCH
4	BCCH	SDCCH/1	4	TCH	TCH
5	BCCH	SDCCH/1	5	TCH	TCH
6	AGCH/PCH	SDCCH/1	6	TCH	TCH
7	AGCH/PCH	SDCCH/1	7	TCH	TCH
8	AGCH/PCH	SDCCH/2	8	TCH	TCH
9	AGCH/PCH	SDCCH/2	9	TCH	TCH
10	FCCH	SDCCH/2	10	TCH	TCH
11	SCH	SDCCH/2	11	TCH	TCH
12	AGCH/PCH	SDCCH/3	12	SACCH	SACCH
13	AGCH/PCH	SDCCH/3	13	TCH	TCH
14	AGCH/PCH	SDCCH/3	14	TCH	TCH
15	AGCH/PCH	SDCCH/3	15	TCH	TCH
16	AGCH/PCH	SDCCH/4	16	TCH	TCH
17	AGCH/PCH	SDCCH/4	17	TCH	TCH
18	AGCH/PCH	SDCCH/4	18	TCH	TCH
19	AGCH/PCH	SDCCH/4	19	TCH	TCH
20	FCCH	SDCCH/5	20	TCH	TCH
21	SCH	SDCCH/5	21	TCH	TCH
22	SDCCH/0	SDCCH/5	22	TCH	TCH
23	SDCCH/0	SDCCH/5	23	TCH	TCH
24	SDCCH/0	SDCCH/6	24	TCH	TCH
25	SDCCH/0	SDCCH/6	25	Free	Free
26	SDCCH/1	SDCCH/6	0	TCH	TCH
27	SDCCH/1	SDCCH/6	1	TCH	TCH
28	SDCCH/1	SDCCH/7	2	TCH	TCH
29	SDCCH/1	SDCCH/7	3	TCH	TCH
30	FCCH	SDCCH/7	4	TCH	TCH
31	SCH	SDCCH/7	5	TCH	TCH
32	SDCCH/2	SACCH/0	6	TCH	TCH
33	SDCCH/2	SACCH/0	7	TCH	TCH
34	SDCCH/2	SACCH/0	8	TCH	TCH
35	SDCCH/2	SACCH/0	9	TCH	TCH
36	SDCCH/3	SACCH/1	10	TCH	TCH
37	SDCCH/3	SACCH/1	11	TCH	TCH
38	SDCCH/3	SACCH/1	12	SACCH	SACCH
39	SDCCH/3	SACCH/1	13	TCH	TCH
40	FCCH	SACCH/2	14	TCH	TCH
41	SCH	SACCH/2	15	TCH	TCH
42	SACCH/0	SACCH/2	16	TCH	TCH
43	SACCH/0	SACCH/2	17	TCH	TCH
44	SACCH/0	SACCH/3	18	TCH	TCH
45	SACCH/0	SACCH/3	19	TCH	TCH
46	SACCH/1	SACCH/3	20	TCH	TCH
47	SACCH/1	SACCH/3	21	TCH	TCH
48	SACCH/1	Free	22	TCH	TCH
49	SACCH/1	Free	23	TCH	TCH
50	Free	Free	24	TCH	TCH
			25	Free	Free

Figure 1.25 Use of timeslots in the downlink direction per 3GPP TS 45.002 [21].



Without control, a burst arrives too late from subscribers at a far distance and overlaps with a burst of the next timeslot.

Figure 1.26 Time shift of bursts of distant subscribers without timing advance control.

about the network. The channel is monitored by all mobile devices which are switched on but currently not engaged in a call or signaling connection (idle mode), and broadcasts, among many other things, the following information:

- the MCC and MNC of the cell;
- the identification of the cell, which consists of the location area code (LAC) and the cell ID;
- to simplify the search for neighboring cells for a mobile device, the BCCH also contains information about the frequencies used by neighboring cells. Thus, the mobile device does not have to search the complete frequency band for neighboring cells.

- The Paging Channel (PCH) is used to inform idle subscribers of incoming calls or SMS messages. As the network alone is aware of the location area the subscriber is roaming in, the Paging message is broadcast in all cells belonging to the location area. The most important information element of the message is the IMSI of the subscriber or a temporary identification called the Temporary Mobile Subscriber Identity (TMSI). A TMSI is assigned to a mobile device during the network attach procedure and can be changed by the network every time the mobile device contacts the network once encryption has been activated. Thus, the subscriber has to be identified with the IMSI only once and is then addressed with a constantly changing temporary number when encryption is not yet activated for the communication. This increases anonymity in the network and prevents eavesdroppers from creating movement profiles of subscribers.
- The Random Access Channel (RACH) is the only common channel in the uplink direction. If the mobile device receives a message via the PCH that the network is requesting a connection establishment or if the user wants to establish a call or send an SMS, the RACH is used for the initial communication with the network. This is done by sending a Channel Request message. Requesting a channel has to be done via a ‘random’ channel because subscribers in a cell are not synchronized with each other. Thus, it cannot be ensured that two devices do not try to establish a connection at the same time. Only when a dedicated channel (SDCCH) has been assigned to the mobile device by the network can there no longer be any collision between different subscribers of a cell. If a collision occurs during the first network access, the colliding messages are lost and the mobile devices do not receive an answer from the network. Thus, they have to repeat their Channel Request messages after expiry of a timer that is set to an initial random value. This way, it is not very likely that the mobile devices will interfere with each other again during their next connection establishment attempts because they are performed at different times.

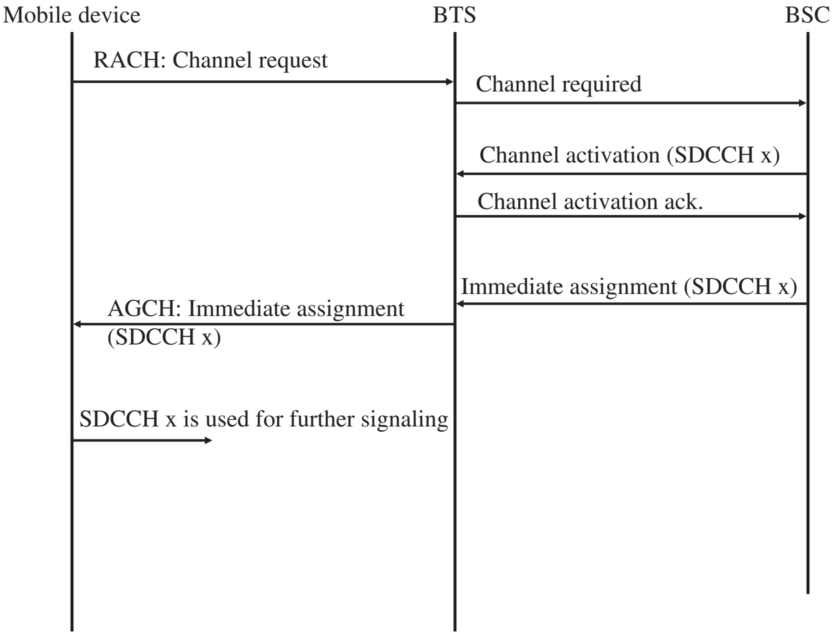

Mobile device
BTS
BSC
RACH: Channel request
Channel required
Channel activation (SDCCH x)
Channel activation ack.
Immediate assignment (SDCCH x)
AGCH: Immediate assignment (SDCCH x)
SDCCH x is used for further signaling

Figure 1.27 Establishment of a signaling connection.

- The Access Grant Channel (AGCH): If a subscriber sends a Channel Request message on the RACH, the network allocates an SDCCH or, in exceptional cases, a TCH, and notifies the subscriber on the AGCH via an Immediate Assignment message. The message contains information about which SDCCH or TCH the subscriber is allowed to use.

Figure 1.27 shows how PCH, AGCH and SDCCH are used during the establishment of a signaling link between the mobile device and the network. The base station controller (BSC), which is responsible for assigning SDCCH and TCH of a base station, is further described in Section 1.7.4.

As can also be seen from Figure 1.25, not all bursts on timeslots 2–7 are used for TCHs. Every 12th burst of a timeslot is used for the SACCH. Furthermore, the 25th burst is also not used for carrying user data. This gap is used to enable the mobile device to perform signal strength measurements of neighboring cells on other frequencies. This is necessary so that the network can redirect the connection to a different cell (handover) to maintain the call while the user is moving.

The GSM standard offers two possibilities to use the available frequencies. The simplest case, which has been described already, is the use of a constant carrier frequency (ARFCN) for each channel. To improve the transmission quality, it is also possible to use alternating frequencies for a single channel of a cell. This concept is known as frequency hopping, and it changes the carrier frequency for every burst during a transmission. This increases the probability that only few bits are lost if one carrier frequency experiences a lot of interference from other sources like neighboring cells. In the worst case, only a single burst is affected because the next burst is already sent on a different frequency. Up to 64 different frequencies can be used per base station for frequency hopping. To inform the mobile of the use of frequency hopping, the Immediate Assignment message used during the establishment of a signaling link contains all the information about the frequencies that are used and the hopping pattern that is applied to the connection.

For carriers that transport the SCH, FCCH and BCCH channels, frequency hopping must not be used. This restriction is necessary because it would be very difficult for mobile devices to find neighboring cells.

In practice, network operators use static frequencies as well as frequency hopping in their networks.

The interface which connects the base station to the network and which is used to carry the information for all logical channels is called the Abis interface. An E-1 connection is usually used for the Abis interface, and owing to its 64 kbit/s timeslot architecture the logical channels are transmitted in a way that differs from their transmission on the air interface. All common channels as well as the information sent and received on the SDCCH and SACCH channels are sent over one or more common 64 kbit/s E-1 timeslots. This is possible because these channels are only used for signaling data that are not time critical. On the Abis interface, these signaling messages are sent by using the Link Access Protocol (LAPD). This protocol was initially designed for the ISDN D-channel of fixed-line networks and has been reused for GSM with only minor modifications.

For TCHs that use a bandwidth of 13 kbit/s on the Abis interface, only one-quarter of an E-1 timeslot is used. This means that all eight timeslots of an air interface frame can be carried on only two timeslots of the E-1 interface. A base station composed of three sectors, which use two carriers each, thus requires 12 timeslots on the Abis interface plus an additional timeslot for the LAPD signaling. The remaining timeslots of the E-1 connection can be used for the communication between the network and other base stations as shown in Figure 1.28. For this purpose, several cells are usually daisy chained via a single E-1 connection (see Figure 1.28).

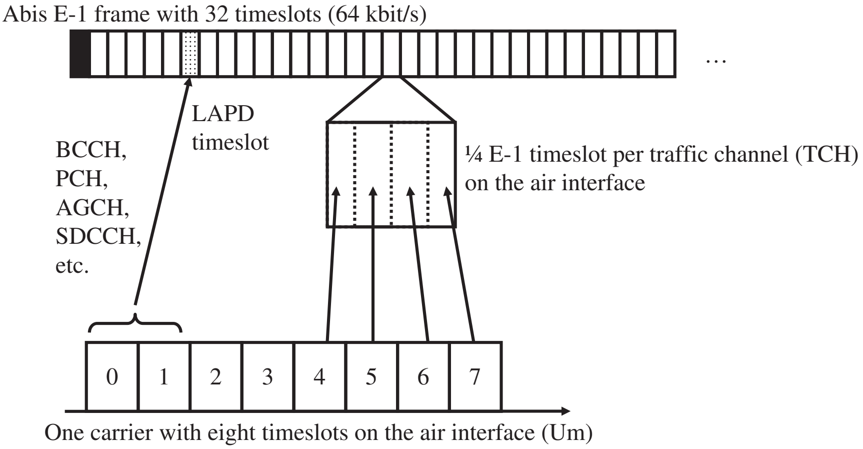

Abis E-1 frame with 32 timeslots (64 kbit/s)
...
LAPD timeslot
BCCH, PCH, AGCH, SDCCH, etc.
¼ E-1 timeslot per traffic channel (TCH) on the air interface
0
1
2
3
4
5
6
7
One carrier with eight timeslots on the air interface (Um)

Figure 1.28 Mapping of E-1 timeslots to air interface timeslots.

In practice, it can be observed today that physical E-1 links being replaced more and more with virtual connections over IP-based links. This is especially the case if a base station site is used for several radio technologies simultaneously (e.g. GSM, UMTS and LTE).

1.7.4 The Base Station Controller (BSC)

While the base station is the interface element that connects the mobile devices with the network, the BSC is responsible for the establishment, release and maintenance of all connections for cells that are connected to it.

If a subscriber wants to establish a voice call, send an SMS, and so on, the mobile device sends a Channel Request message to the BSC as shown in Figure 1.27. The BSC then checks if an SDCCH is available and activates the channel in the BTS. Afterward, the BSC sends an Immediate Assignment message to the mobile device on the AGCH that includes the number of the assigned SDCCH. The mobile device then uses the SDCCH to send DTAP messages that the BSC forwards to the MSC.

The BSC is also responsible for establishing signaling channels for incoming calls or SMS messages. In this case, the BSC receives a Paging message from the MSC, which contains the IMSI and TMSI of the subscriber as well as the location area ID in which the subscriber is currently located. The BSC in turn has a location area database that it uses to identify all cells in which the subscriber needs to be paged. When the mobile device receives the Paging message, it responds to the network in the same way as in the example above by sending a Channel Request message.

The establishment of a TCH for voice calls is always requested by the MSC for both mobile-originated and mobile-terminated calls. Once the mobile device and the MSC have exchanged all necessary information for the establishment of a voice call via an SDCCH, the MSC sends an assignment request for a voice channel to the BSC as shown in Figure 1.29.

The BSC then verifies if a TCH is available in the requested cell and, if so, activates the channel in the BTS. Afterward, the mobile device is informed via the SDCCH that a TCH is now available for the call. The mobile device then changes to the TCH and FACCH. To inform the BTS that it has switched to the new channel, the mobile device sends a message to the BTS on the FACCH, which is acknowledged by the BTS. In this way, the mobile also has a confirmation that its signal can be decoded correctly by the BTS. Finally, the mobile device sends an Assignment Complete message to the BSC, which in turn informs the MSC of the successful establishment of the TCH.

Apart from the establishment and release of a connection, another important task of the BSC is the maintenance of the connection. As subscribers can roam freely through the network while a call is ongoing, it can happen that the subscriber roams out of the coverage area of the cell in which the call was initially established. In this case, the BSC has to redirect the call to the appropriate cell. This procedure is called handover. To be able to perform a handover to another cell, the BSC requires signal quality measurements for the air interface. The results of the downlink signal quality measurements are reported to the BSC by the mobile device, which continuously performs signal quality measurements that it reports via the SACCH to the network. The uplink signal quality is constantly measured by the BTS and also reported to the BSC. Apart from the signal quality of the user's current cell, it is also important that the mobile device reports the

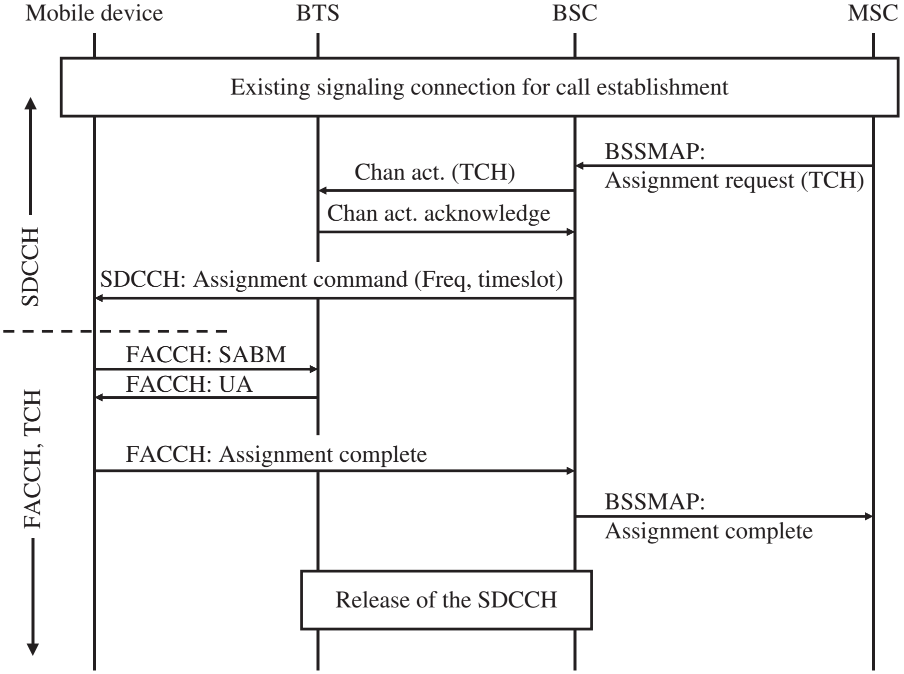


Figure 1.29 Establishment of a traffic channel (TCH).

quality of signals it receives from other cells. To enable the mobile device to perform these measurements, the network sends the frequencies of neighboring cells via the SACCH during an ongoing call. The mobile device then uses this information to perform the neighboring cell measurements while the network communicates with other subscribers and reports the result via measurement report messages in the uplink SACCH.

The network receives these measurement values and is thus able to periodically evaluate if a handover of an ongoing call to a different cell is necessary. Once the BSC decides to perform a handover, a TCH is activated in the new cell as shown in Figure 1.30. Afterward, the BSC informs the mobile device via the old cell with a Handover Command message that is sent over the FACCH. Important information elements of the message are the new frequency and timeslot number of the new TCH. The mobile device then changes its transmit and receive frequency, synchronizes to the new cell if necessary and sends a Handover Access message in four consecutive bursts. In the fifth burst, a Set Asynchronous Balanced Mode (SABM) message is sent, which is acknowledged by the BTS to signal to the mobile device that the signal can be received. At the same time, the BTS informs the BSC of the successful reception of the mobile device's signal with an Establish Indication message. The BSC then immediately redirects the speech path to the new cell.

From the mobile's point of view the handover is now finished. The BSC, however, has to release the TCH in the old cell and has to inform the MSC of the performed handover before the handover is finished from the network's point of view. The message to the MSC is only informative and has no impact on the continuation of the call.

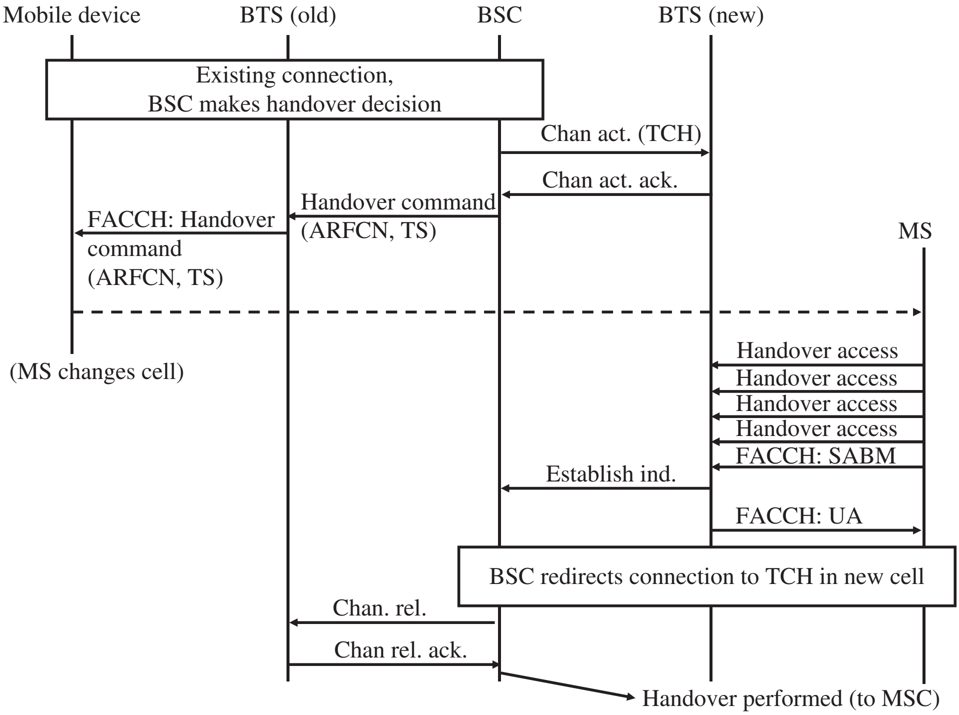


Figure 1.30 Message flow during a handover procedure.

To reduce interference, the BSC is also in charge of controlling the transmission power for every air interface connection. For the mobile device, an active power control has the advantage that the transmission power can be reduced under favorable reception conditions. Transmission power is controlled using the signal quality measurements of the BTS for the connection. If the mobile device's transmission power has to be increased or decreased, the BSC sends a Power Control message to the BTS. The BTS in turn forwards the message to the mobile device and repeats the message on the SACCH in every frame. In practice, it can be observed that power control and adaptation is performed every 1–2 seconds. During call establishment, the mobile device always uses the highest allowed power output level, which is then reduced or increased again by the network step by step. Table 1.6 gives an overview of the mobile device power levels. A distinction is made for the 900 MHz versus the 1800 MHz band. While mobile devices operating on the 900 MHz band are allowed to use up to 2 W, connections on the 1800 MHz band are limited to 1 W. For stationary devices or car phones with external antennas, power values of up to 8 W are allowed. The power values in the table represent the power output when the transmitter is active in the assigned timeslot. As the mobile device only sends on one of the eight timeslots of a frame, the average power output of the mobile device is only one-eighth of this value. The average power output of a mobile device that sends with a power output of 2 W is thus only 250 mW.

The BSC is also able to control the power output of the base station. This is done by evaluating the signal measurements of the mobile devices in the current cell. It is important to note that power control can only be performed for downlink carriers that do not

Table 1.6 GSM power levels and corresponding power output.

GSM 900 Power level	GSM 900 Power output	GSM 1800 Power level	GSM 1800 Power output
(0–2)	(8 W)	–	–
5	2 W	0	1 W
6	1.26 W	1	631 mW
7	794 mW	2	398 mW
8	501 mW	3	251 mW
9	316 mW	4	158 mW
10	200 mW	5	100 mW
11	126 mW	6	63 mW
12	79 mW	7	40 mW
13	50 mW	8	25 mW
14	32 mW	9	16 mW
15	20 mW	10	10 mW
16	13 mW	11	6.3 mW
17	8 mW	12	4 mW
18	5 mW	13	2.5 mW
19	3.2 mW	14	1.6 mW
–	–	15	1.0 mW

broadcast the common channels like frame control header (FCH), SCH and BCCH of a cell. On such carriers, the power output has to be constant to allow mobile devices which are currently located in other cells of the network to perform their neighboring cell measurements. This would not be possible if the signal amplitude varies over time as the mobile devices can only listen to the carrier signal of neighboring cells for a short time.

Owing to the limited speed of radio waves, a time shift of the arrival of the signal can be observed when a subscriber moves away from a base station during an ongoing call. If no countermeasures are taken, this would mean that at some point the signal of a subscriber would overlap with the next timeslot despite the guard time of each burst, which is shown in Figure 1.26. Thus, the signal of each subscriber has to be carefully monitored and the timing of the transmission of the subscriber has to be adapted. This procedure is called timing advance control (Figure 1.29).

The timing advance can be controlled in 64 steps (0–63) of 550 m. The maximum distance between a base station and a mobile subscriber is in theory 64 × 550 m = 35.2 km. In practice, such a distance is not reached very often as base stations usually cover a much smaller area for capacity reasons. Furthermore, the transmission power of the mobile device is also not sufficient to bridge such a distance under non-line-of-sight conditions to the base station. Therefore, one of the few scenarios where such a distance has to be overcome is in coastal areas, from ships at sea.

The control of the timing advance already starts with the first network access on the RACH with a Channel Request message. This message is encoded into a very short burst that can only transport a few bits in exchange for large guard periods at the beginning and end of the burst. This is necessary because the mobile device is unaware of the distance between itself and the base station when it attempts to contact the network. Thus, the mobile device is unable to select an appropriate timing advance value. When the base station receives the burst, it measures the delay and forwards the request, including a timing advance value required for this mobile device, to the BSC. As has been shown in Figure 1.27, the BSC reacts to the connection request by returning an Immediate Assignment message to the mobile device on the AGCH. Apart from the number of the assigned SDCCH, the message also contains a first timing advance value to be used for the subsequent communication on the SDCCH. Once the connection has been successfully established, the BTS continually monitors the delay experienced for this channel and reports any changes to the BSC. The BSC in turn instructs the mobile device to change its timing advance by sending a message on the SACCH.

For special applications, like coastal communication, the GSM standard offers an additional timeslot configuration to increase the maximum distance to the base station to up to 120 km. This is achieved by only using every second timeslot per carrier, which allows a burst to overlap onto the following (empty) timeslot. While this significantly increases the range of a cell, the number of available communication channels is cut in half. Another issue is that mobile devices that are limited to a transmission power of 1 W (1800 MHz band) or 2 W (900 MHz band) may be able to receive the BCCH of such a cell at a great distance but are unable to communicate with the cell in the uplink direction. Thus, such an extended-range configuration mostly makes sense with permanently installed mobile devices with external antennas that can transmit with a power level of up to 8 W.

1.7.5 The TRAU for Voice Encoding

For the transmission of voice data, a TCH is used in GSM as described in Section 1.7.3. A TCH uses all but two bursts of a 26-burst multiframe, with one being reserved for the SACCH, as shown in Figure 1.25, and the other remaining empty to allow the mobile device to perform neighboring cell measurements. As has been shown in the preceding section, a burst that is sent to or from the mobile every 4.615 milliseconds can carry exactly 114 bits of user data. When taking the two bursts which are not used for user data of a 26-burst multiframe into account, this results in a raw datarate of 22.8 kbit/s. As is shown in the remainder of this section, a substantial part of the bandwidth of a burst is required for error detection and correction bits. The resulting datarate for the actual user data is thus around 13 kbit/s.

The narrow bandwidth of a TCH stands in contrast to how a voice signal is transported in the core network. Here, the PCM algorithm is used (see Section 1.6.1) to digitize the voice signal, which makes full use of the available 64 kbit/s bandwidth of an E-1 timeslot to encode the voice signal (see Figure 1.31).

A simple solution for the air interface would have been to define air interface channels that can also carry 64 kbit/s PCM-coded voice channels. This has not been done because the scarce resources on the air interface have to be used as efficiently as possible. The decision to compress the speech signal was taken during the first standardization phase in the 1980s because it was foreseeable that advances in hardware and software-processing capabilities would allow compression of a voice data stream in real-time.

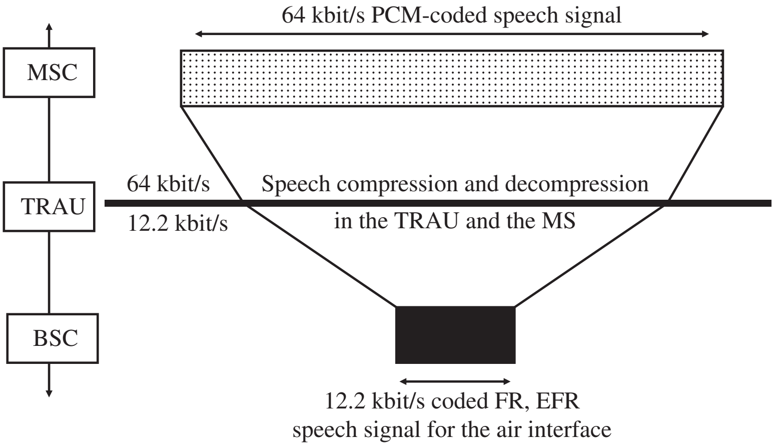

64 kbit/s PCM-coded speech signal
MSC
TRAU
BSC
64 kbit/s
12.2 kbit/s
Speech compression and decompression
in the TRAU and the MS
12.2 kbit/s coded FR, EFR
speech signal for the air interface

Figure 1.31 GSM speech compression.

In the mobile network, the compression and decompression of the voice data stream is performed in the Transcoding and Rate Adaptation Unit (TRAU), which is located between the MSC and a BSC and controlled by the BSC (see Figure 1.31). During an ongoing call, the MSC sends the 64 kbit/s PCM-encoded voice signal toward the radio network and the TRAU converts the voice stream in real-time into a 13 kbit/s compressed data stream, which is transmitted over the air interface. In the other direction, the BSC sends a continuous stream of compressed voice data toward the core network and the TRAU converts the stream into a 64 kbit/s coded PCM signal. In the mobile device, the same algorithms are implemented as in the TRAU to compress and decompress the speech signal (see Figure 1.32).

While the TRAU is a logical component of the BSS, it is most often installed next to an MSC in practice. This has the advantage that four compressed voice channels can be transmitted in a single E-1 timeslot. After compression, each voice channel uses a 16 kbit/s sub-timeslot. Thus, only one-quarter of the transmission capacity between an MSC and BSC is needed in comparison to an uncompressed transmission. As the BSCs of a network are usually located in the field and not close to an MSC, this helps to reduce transmission costs for the network operator substantially as shown in Figure 1.32.

The TRAU offers a number of different algorithms for speech compression. These algorithms are called speech codecs or simply codecs. The first codec that was standardized for GSM is the full-rate (FR) codec, which reduces the 64 kbit/s voice stream to about 13 kbit/s.

At the end of the 1990s, the enhanced full-rate (EFR) codec was introduced. The EFR codec not only compresses the speech signal to about 13 kbit/s but also offers superior voice quality compared to the FR codec. The disadvantage of the EFR codec is the higher complexity of the compression algorithm, which requires more processing power. However, the processing power available in mobile devices has increased significantly since the 1990s, and thus modern GSM phones easily cope with the additional complexity.

Besides those two codecs, a half-rate (HR) codec has been defined for GSM that only requires a bandwidth of 7 kbit/s. While there is almost no audible difference between

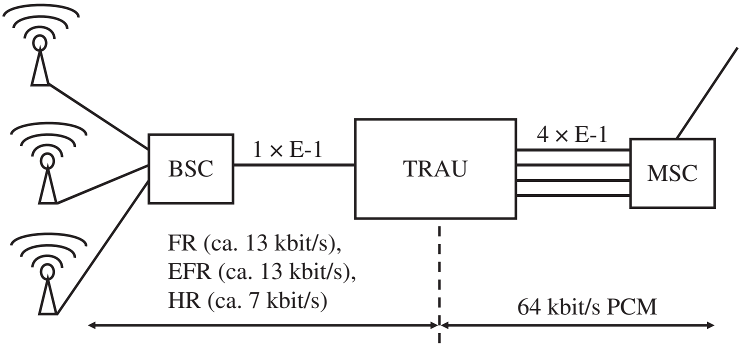


Figure 1.32 Speech compression with a 4:1 compression ratio in the TRAU.

the EFR codec and a PCM-coded speech signal, the voice quality of the HR codec is noticeably inferior. The advantage for the network operator of the HR codec is that the number of simultaneous voice connections per carrier can be doubled. With the HR codec, a single timeslot, which is used for a single EFR voice channel, can carry two (HR) TCHs.

Another speech codec development is the Adaptive Multirate (AMR) algorithm [22] that is used by most devices and networks today. Instead of using a single codec, which is selected at the beginning of the call, AMR allows a change of the codec during a call. The considerable advantage of this approach is the ability to switch to a speech codec with a higher compression rate during bad radio signal conditions to increase the number of error detection and correction bits. If signal conditions permit, a lower rate codec can be used, which only uses every second burst of a frame for the call. This in effect doubles the capacity of the cell as a single timeslot can be shared by two calls in a similar manner to the HR codec. Unlike the HR codec, however, the AMR codecs, which only use every second burst and which are thus called HR AMR codecs, still have a voice quality which is comparable to that of the EFR codec. While AMR is optional for GSM, it has been chosen for the UMTS system as a mandatory feature. In the United States, AMR is used by some network operators to increase the capacity of their network, especially in very dense traffic areas like New York, where it has become very difficult to increase the capacity of the network any further, with over half a dozen carrier frequencies per sector already used. Further information about AMR can also be found in Chapter 3.

The latest speech codec development used in practice is AMR-Wideband (AMR-WB) as specified in ITU G.722.2 [23] and 3GPP TS 26.190 [24]. The algorithm allows, as its name implies, digitization of a wider frequency spectrum than is possible with the PCM algorithm that was described earlier. Instead of an upper limit of 3400 Hz, AMR-WB digitizes a voice signal up to a frequency of 7000 Hz. As a consequence, the caller's voice sounds much clearer and more natural on the other end of a connection. A high compression rate is used in practice to reduce the datarate of a voice stream down to 12.65 kbit/s. This way, an AMR-WB data stream can be transmitted in a single GSM timeslot, and also requires no additional capacity in a UMTS network. As AMR-WB is not compatible with the PCM codec used between the BSC and MSC, it is sent transparently between the two nodes. This means that most of the bits in a 64 kbit/s PCM timeslot are unused, as the datarate required by the AMR-WB codec is only 12.65 kbit/s. In practice, AMR-WB is mostly used in UMTS networks today. Therefore, it is described in more detail in Chapter 3.

While the PCM algorithm digitizes analog volume levels by statically mapping them to digital values, GSM speech digitization is much more complex in order to reach the desired compression rate. In the case of the FR codec, which is specified in 3GPP TS 46.010 [25], the compression is achieved by emulating the human vocal system. This is done by using a source–filter model (Figure 1.33). In the human vocal system, speech is created in the larynx and by the vocal cords. This is emulated in the mathematical model in the signal creation part, while the filters represent the signal formation that occurs in the human throat and mouth.

On a mathematical level, speech formation is simulated by using two time-invariant filters. The period filter creates the periodic vibrations of the human voice while the vocal tract filter simulates the envelope. The filter parameters are generated from the human voice, which is the input signal into the system. To digitize and compress the human voice, the model is used in the reverse direction as shown in Figure 1.33. As time-variant filters are hard to model, the system is simplified by generating a pair of filter parameters for an interval of 20 milliseconds as shown in Figure 1.34. A speech signal that has previously been converted into an 8- or 13-bit PCM codec is used as an input to the algorithm. As the PCM algorithm delivers 8000 values per second, the FR codec requires 160 values for a 20-millisecond interval to calculate the filter parameters. As 8 bits are used per value, 8 bits × 160 values = 1280 input bits are used per 20-millisecond interval. For the period filter, the input bits are used to generate a filter parameter with a length of 36 bits. Afterward, the filter is applied to the original input signal. The resulting signal is then used to calculate another filter parameter with a length of 36 bits for the vocal tract filter. Afterward, the signal is again sent through the vocal tract filter with the filter parameter applied. The signal which is thus created is called the ‘rest signal’ and is coded into 188 bits (see Figure 1.34).

Once all parameters have been calculated, the two 36-bit filter parameters and the rest signal, which is coded into 188 bits, are sent to the receiver. Thus, the original information, which was coded in 1280 bits, has been reduced to 260 bits. In the receiver, the filter procedure is applied in reverse order on the rest signal and thus the original signal is recreated. As the procedure uses a lossy compression algorithm, the original signal and the recreated signal at the other end are no longer exactly identical. For the human ear, however, the differences are almost inaudible.

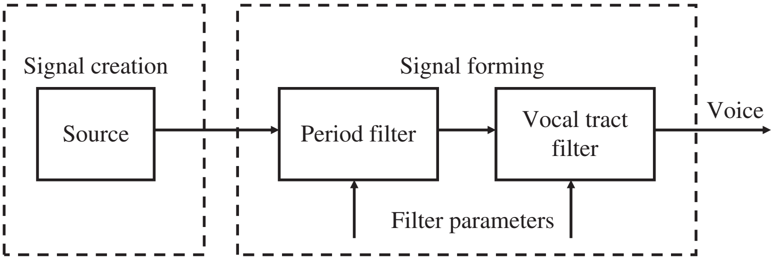


Figure 1.33 Source–filter model of the GSM FR codec.

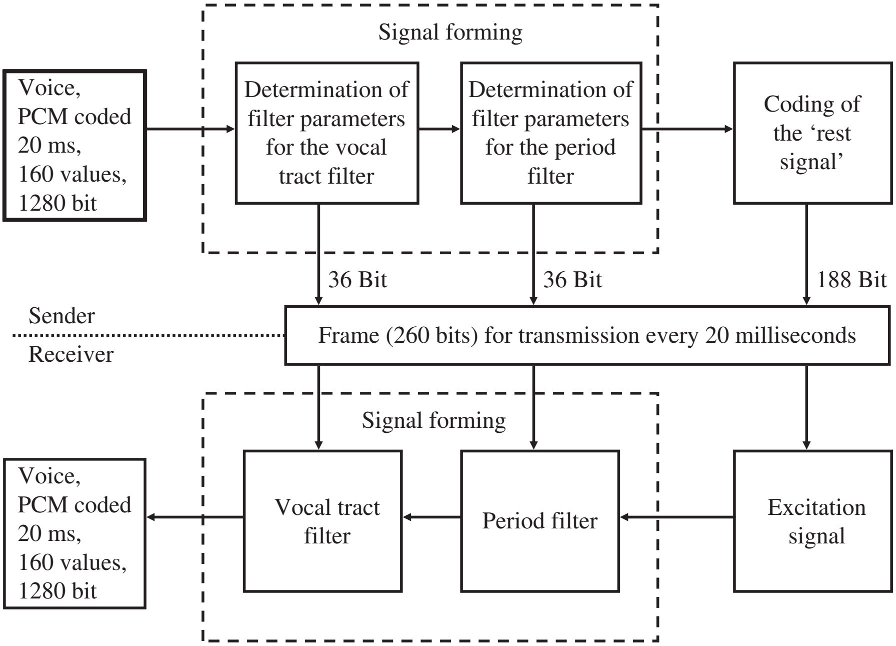

Signal forming
Voice,
PCM coded
20 ms,
160 values,
1280 bit
Determination of filter parameters for the vocal tract filter
Determination of filter parameters for the period filter
Coding of the 'rest signal'
36 Bit
36 Bit
188 Bit
Sender
Receiver
Frame (260 bits) for transmission every 20 milliseconds
Signal forming
Voice,
PCM coded
20 ms,
160 values,
1280 bit
Vocal tract filter
Period filter
Excitation signal

Figure 1.34 Complete transmission chain with the transmitter and receiver of the GSM FR codec.

1.7.6 Channel Coder and Interleaver in the BTS

When a 260-bit data frame from the TRAU arrives at the base station every 20 milliseconds, it is further processed before being sent over the air as shown in Figure 1.35. In the reverse direction, the tasks are performed in the mobile device.

In the first step, the voice frames are processed in the channel coder unit, which adds error detection and correction information to the data stream. This step is very important as the transmission over the air interface is prone to frequent transmission errors due to the constantly changing radio environment. Furthermore, the compressed voice information is very sensitive and even a few bits that might be changed while the frame is transmitted over the air interface create an audible distortion. To prevent this, the channel coder separates the 260 bits of a voice data frame into three different classes as shown in Figure 1.36.

Fifty of the 260 bits of a speech frame are class Ia bits and are extremely important for the overall reproduction of the voice signal at the receiver side. Such bits are, for example, the higher order bits of the filter parameters. To enable the receiver to verify the correct transmission of those bits, a three-bit cyclic redundancy check (CRC) checksum is calculated and added to the data stream. If the receiver cannot recreate the checksum with the received bits later on, the frame is discarded.

The other 132 bits of the frame are also quite important and are thus put into class Ib. However, no checksum is calculated for them. To generate the exact amount of bits that are necessary to fill a GSM burst, four filler bits are inserted. Afterward, the class Ia bits, checksum, class Ib bits and the four filler bits are treated by a convolutional coder that adds redundancy to the data stream. For each input bit, the convolutional decoder

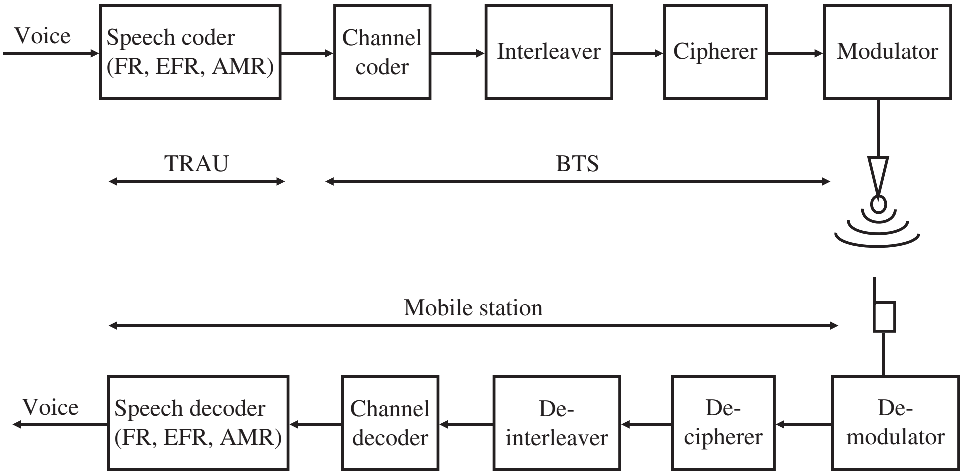


Figure 1.35 Transmission path in the downlink direction between the network and the mobile device.

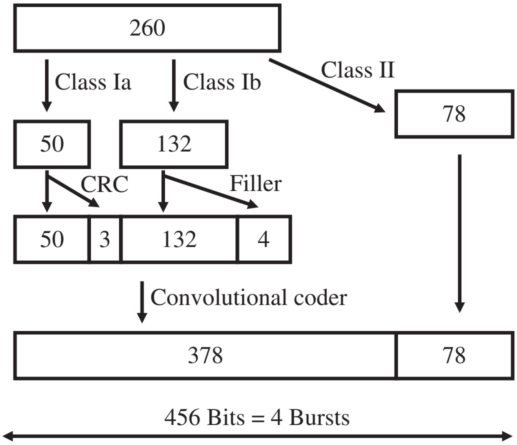


Figure 1.36 GSM channel coder for full-rate speech frames.

calculates two output bits. For the computation of the output bits, the coder uses not only the current bit but also the information about the values of the previous bits. For each input bit, two output bits are calculated. This mathematical algorithm is also called a HR convolutional coder.

The remaining 78 bits of the original 260-bit data frame belong to the third class, which is called class II. These are not protected by a checksum and no redundancy is added for them. Errors that occur during the transmission of these bits can neither be detected nor corrected.

As has been shown, the channel coder uses the 260-bit input frame to generate 456 bits on the output side. As a burst on the air interface can carry exactly 114 bits, four bursts are necessary to carry the frame. As the bursts of a TCH are transmitted every 4.6152 milliseconds, the time it takes to transmit the frame over the air interface is about 20 milliseconds. To get to exactly 20 milliseconds, the empty burst and the burst used for the SACCH per 26-burst multiframe has to be included in the calculation.

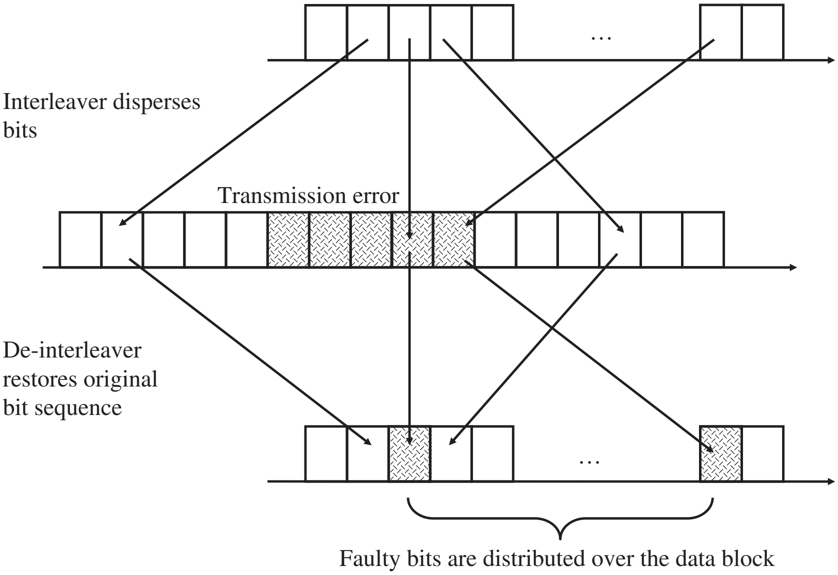

Interleaver disperses bits
Transmission error
De-interleaver restores original bit sequence
…
…
Faulty bits are distributed over the data block

Figure 1.37 Frame interleaving.

Owing to the redundancy added by the channel coder, it is possible to correct a high number of faulty bits per frame. The convolutional decoder, however, has one weak point: if several consecutive bits are changed during transmission over the air interface, the convolutional decoder on the receiver side is not able to correctly reconstruct the original frame. This effect is often observed as air interface disturbances usually affect several bits in a row.

To decrease this effect, the interleaver changes the bit order of a 456-bit data frame in a specified pattern over eight bursts, as shown in Figure 1.37. Consecutive frames are thus interlocked with each other. On the receiver side, the frames are put through the de-interleaver, which puts the bits again into the correct order. If several consecutive bits are changed because of air interface signal distortion, this operation disperses the faulty bits in the frame and the convolutional decoder can thus correctly restore the original bits. A disadvantage of the interleaver, however, is an increased delay in the voice signal. In addition to the delay of 20 milliseconds generated by the FR coder, the interleaver adds another 40 milliseconds, as a speech frame is spread over eight bursts instead of being transmitted consecutively in four bursts. Compared to a voice call in a fixed-line network, a mobile network thus introduces a delay of at least 60 milliseconds. If the call is established between two mobile devices, the delay is at least 120 milliseconds as the transmission chain is traversed twice.

1.7.7 Ciphering in the BTS and Security Aspects

The next module of the transmission chain is the cipherer (Figure 1.38), which encrypts the data frames it receives from the interleaver. GSM, like most communication systems, uses a stream cipher algorithm. To encrypt the data stream, a ciphering key (Kc) is calculated in the AuC and on the SIM card by using a random number (RAND) and the secret key (Ki) as input parameters for the A8 algorithm. Together with the GSM frame number, which is increased for every air interface frame, Kc is then used as input parameter for the A5 ciphering algorithm. The A5 algorithm computes a 114-bit sequence which is XOR combined with the bits of the original data stream. As the frame number is different for every burst, it is ensured that the 114-bit ciphering sequence also changes for every burst, which further enhances security.

To be as flexible as possible, a number of different ciphering algorithms have been specified for GSM. These are called A5/1, A5/2, A5/3 and so on. The intent of allowing several ciphering algorithms was to enable export of GSM network equipment to countries where export restrictions prevent the sale of some ciphering algorithms and technologies. Furthermore, it is possible to introduce new ciphering algorithms into already existing networks to react to security issues if a flaw is detected in one of the currently used algorithms. The selection of the ciphering algorithm also depends on the capabilities of the mobile device. During the establishment of a connection, the mobile device informs the network about the ciphering algorithms that it supports. The network can then choose an algorithm that is supported by the network and the mobile device.

When the mobile device establishes a new connection with the network, its identity is verified before it is allowed to proceed with the call setup. This procedure has already been described in Section 1.6.4. Once the mobile device and subscriber have been authenticated, the MSC usually starts encryption by sending a ciphering command to the mobile device. The ciphering command message contains, among other information elements, the ciphering key, Kc, which is used by the base station for the ciphering of the connection on the air interface. Before the BSC forwards the message to the mobile device, however, the ciphering key is removed from the message because this information must not be sent over the air interface. The mobile device does not need to receive the ciphering key from the network as the SIM card calculates the Kc on its own and forwards the key to the mobile device together with the SRES during the authentication

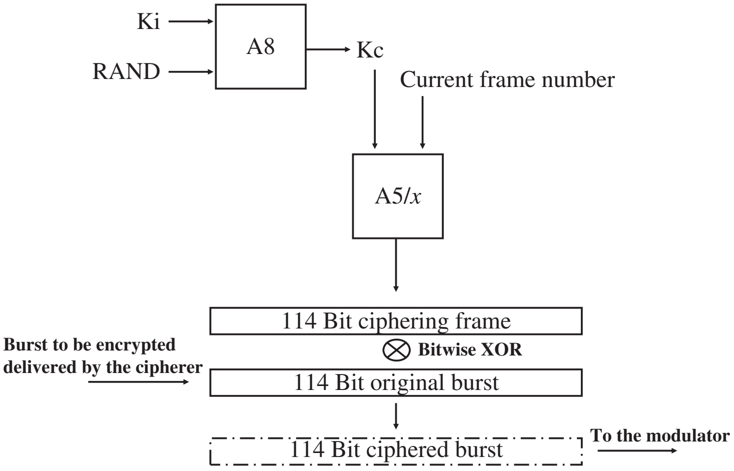


Figure 1.38 Ciphering of an air interface burst.

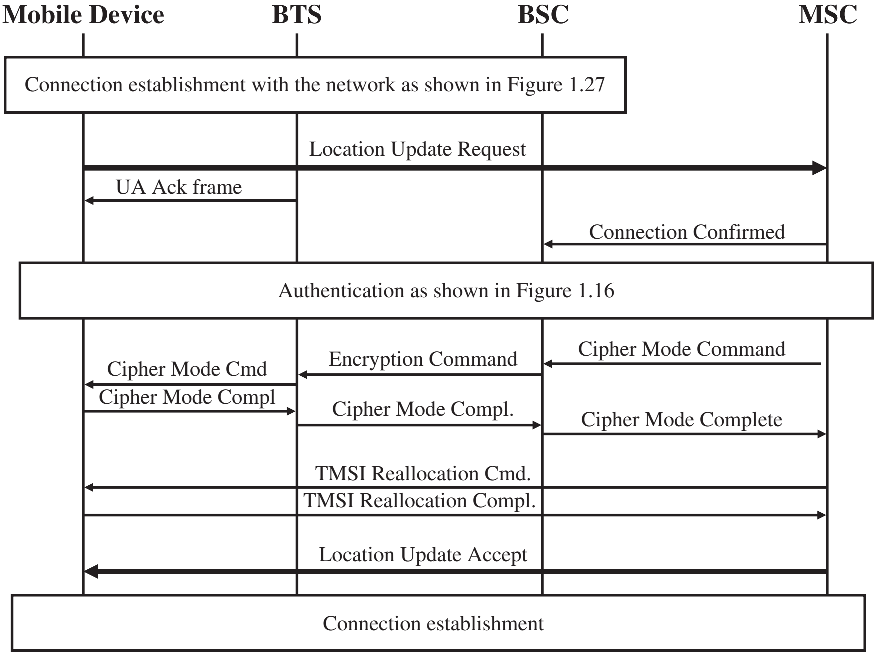


Figure 1.39 Message flow for a location update procedure.

procedure. Figure 1.39 further shows how ciphering is activated during a location update procedure.

With the rising popularity of GSM over the last 20 years, its authentication and encryption procedures have received a lot of scrutiny. From a user point of view, encryption and other security measures must prevent eavesdropping on any kind of communication such as voice conversations, SMS message transfers and signaling in general. Furthermore, it must prevent the theft and misuse of personal authentication data to ensure integrity of the system and to prevent false billing. Also, mobile devices must be protected from third-party attacks that attempt to steal or alter personal data from mobile devices that are directly based on the air interface.

At the time of writing, a number of security issues have been found in the GSM security architecture from a user point of view. In this regard, it is important to differentiate between several categories:

1) Theoretical security issues which, at the time of writing, cannot as yet be exploited.
2) Security issues for which practical exploits are likely to have been developed but which require sophisticated and expensive equipment which are not available to the general public.
3) The third group covers security issues which can be exploited with hardware and software available to the general public.

The following discussion gives an overview of a number of security issues from the second category, which are described in more detail in Barkan *et al.* [26], the 26C3 [27] and the 28C3 [28]:

- **No encryption on the Abis interface.** The communication link between the base station and the BSC is not ciphered today. Attackers with equipment that is able to intercept E1-based communication over a microwave or cable link can potentially intercept signaling messages and voice calls. In the future, this risk is likely to be reduced with the introduction of encrypted high-speed IP-based communication to multiradio access technology base stations as discussed in Chapters 3 and 4.
- **No mandatory air interface encryption.** Encryption is activated with a ciphering command message by the network. If not activated, all signaling and voice calls are transmitted without protection from eavesdropping. In practice, networks always activate air interface encryption today. Some phones indicate an unencrypted communication link with an open lock symbol.
- **No network authentication.** In practice, this allows attacks that are based on false base stations. By placing such a false base station close to a user and by using transmission power higher than that of any other base station from the network operator, the mobile device will automatically select the false base station and transmit its IMSI in the location update dialog that is further described in Section 1.8.1. The false base station can then use this information to intercept all incoming and outgoing communication by using the user's IMSI itself for communication with the network. By preventing the use of encryption, the need to get access to the shared secret Ki is dispensed with (cp. Section 1.6.4). Such devices are known as 'IMSI catchers' and further details can be found in [29] and Frick and Bott [30].

  Potential protection against such an attack would be to mandate authentication and encryption on the mobile side for every connection establishment. While it would still be possible to collect IMSIs, this would prevent the false base station from eavesdropping on SMS messages and voice calls. At the time of publication, however, such a protection is not implemented in mobile devices.
- **A5/2 Weaknesses.** This encryption algorithm was created to allow the export of GSM systems to countries for which export restrictions concerning security technologies exist. With the processing power of today's computers, it is possible to retrieve the ciphering key Kc within seconds with only little ciphering data collected. As A5/2 is not used in countries where no export restrictions apply, this in itself is not an issue.
- **A5/1 and A5/3 Active attacks.** The weakness of A5/2 can potentially be used for indirect attacks on communication encrypted with more secure A5 algorithms such as A5/1 and A5/3. This requires equipment that can not only intercept a data transfer but also act as a false base station as described above. In the first step, A5/1 or A5/3 encrypted data are recorded. In the second step, the secret ciphering key, Kc, that was used to encrypt the conversation is recovered by actively contacting the mobile device and instructing it to activate A5/2 ciphering without supplying new keying material. With subsequent frames now being encrypted with A5/2, its weaknesses can be exploited to calculate the secret ciphering key Kc. As no new ciphering material is supplied for this conversation, the recovered Kc is the same as that previously used for the recorded data that were encrypted using A5/1. To counter this attack, the GSM Association recommends that new mobile devices shall no longer support A5/2. This has been implemented in practice and today, only very old mobile devices are still vulnerable.

- **A5/1 Passive attacks.** Researchers have practically demonstrated that passive attacks on A5/1 are possible under the following conditions:
  - A correctly received data stream can be recorded.
  - Empty bits in GSM signaling frames (fillbits) are sent with a repeating bit pattern.
  - A precomputed decryption table with a size of around 4 TB is available.

While computing and storing the decryption table posed an insurmountable challenge even for specialized equipment at the time A5/1 was conceived, it has now become possible to compute the table in a reasonable amount of time and to store the result. The required hardware and open-source software are now easily available at low cost, and a practical real-time exploit has been demonstrated during the 28th CCC Congress in December 2011 [28].

This threat can be countered by using the A5/3 encryption algorithm for communication, which at the time of writing is considered to be secure. Today, A5/3 is supported by most new devices appearing in the market but only by a few networks. Further, the mobile device must not support A5/2, to deny an attacker the possibility of calculating the key later on as described above. Another method to protect communication against a passive A5/1 attack is to randomize the fillbits in GSM signaling frames in both the uplink and the downlink directions. This was standardized a number of years ago in 3GPP TS 44.008, Section 5.2. In practice, it can be observed that some devices and networks randomize the fillbits today, but widespread acceptance has still not been reached.

At this point, it is worth noting that the efforts described above were targeted at the ciphering key Kc. No practical methods are known that do not require physical access to the SIM card to break the authentication and key-generation algorithms A3 and A8 to get to the shared secret key Ki. This means that should an attacker get the ciphering key Kc of a user, they would still not be able to authenticate during the next network challenge. This means that if the network requires authentication and ciphering for each communication session, it is not possible for an attacker to impersonate another subscriber to receive calls or SMS messages in their place or to make outgoing calls.

1.7.8 Modulation

At the end of the transmission chain, the modulator maps the digital data onto an analog carrier, which uses a bandwidth of 200 kHz. This mapping is done by encoding the bits into changes of the carrier frequency. As the frequency change takes a finite amount of time, a method called Gaussian minimum shift keying (GMSK) is used, which smooths the flanks created by the frequency changes. GMSK has been selected for GSM as its modulation and demodulation properties are easy to handle and implement into hardware and as it interferes only slightly with neighboring channels.

1.7.9 Voice Activity Detection

To reduce the interference on the air interface and to increase the operating time of the mobile device, data bursts are only sent if a speech signal is detected. This method is called discontinuous transmission (DTX) and can be activated independently in the uplink and downlink directions (Figure 1.40). Since only one person speaks at a time during a conversation, one of the two speech channels can usually be deactivated. In the

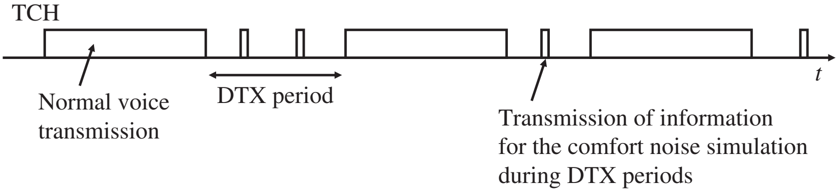

TCH
Normal voice transmission
DTX period
t
Transmission of information for the comfort noise simulation during DTX periods

Figure 1.40 Discontinuous transmission (DTX).

downlink direction, this is managed by the voice activity detection (VAD) algorithm in the TRAU, while in the uplink direction the VAD is implemented in the mobile device.

Simply deactivating a speech channel, however, creates a very undesirable side effect. As no speech signal is transmitted anymore, the receiver no longer hears the background noise on the other side. This can be very irritating, especially for high-volume background noise levels such as when a person is driving a car or sitting in a train. Therefore, it is necessary to generate artificial noise, called comfort noise, which simulates the background noise of the other party for the listener. As the background noise can change over time, the mobile device or the network, respectively, analyzes the background noise of the channel and calculates an approximation for the current situation. This approximation is then exchanged between the mobile device and the TRAU every 480 milliseconds. Additional benefits for the network and mobile device are the ability to perform periodic signal quality measurements of the channel and the ability to use these frames to get an estimation on the current signal timing to adapt the timing advance for the call if necessary. How well this method performs is clear from the audibility as this procedure is used in all mobile device calls today and the simulation of the background noise in most cases cannot be differentiated from the original signal.

Despite the use of sophisticated methods for error correction, it is still possible that parts of a frame are destroyed beyond repair during transmission on the air interface. In these cases, the complete 20-millisecond voice frame is discarded by the receiver and the previous data block is used instead to generate an output signal. Most errors that are repaired this way remain undetected by the listener. This trick, however, cannot be used indefinitely. If after 320 milliseconds a valid data block has still not been received, the channel is muted and the decoder keeps trying to decode the subsequent frames. If, during the following few seconds no valid data frame is received, the connection is terminated and the call drops.

Many of the previously mentioned procedures have specifically been developed for the transmission of voice frames. For example, for circuit-switched data connections that are used for fax transmissions or end-to-end encrypted voice calls, a number of modifications are necessary. While it is possible to tolerate a number of faulty bits for voice frames or to discard frames if a CRC error is detected, this is not possible for data calls. If even a single bit is faulty, a retransmission of at least a single frame has to be performed as most applications cannot tolerate a faulty data stream. To increase the likelihood of correctly reconstructing the initial data stream, the interleaver spreads the bits of a frame over a much larger number of bursts than the eight bursts used for voice frames. Furthermore, the channel coder, which separates the bits of a frame into different classes based on their importance, had to be adapted for data calls as well, as all bits are equally important. Thus, the convolutional decoder has to be used for all bits of a frame. Finally, it is also not possible to use a lossy data compression scheme for data calls. Therefore, the TRAU operates in a transparent mode for data calls. If the data stream can be compressed, this has to be performed by higher layers or by the data application itself.

With a radio receiver or an amplifier of a stereo set, the different states of a GSM connection can be made audible. This is possible as the activation and deactivation of the transmitter of the mobile device induce an audible sound in the amplifier part of audio devices. If the GSM mobile device is held close enough to an activated radio or an amplifier during the establishment of a call, the typical noise pattern can be heard, which is generated by the exchange of messages on the signaling channel (SDCCH). At some time during the signaling phase, a TCH is assigned to the mobile device at the point at which the noise pattern changes. As a TCH burst is transmitted every 4.615 milliseconds, the transmitter of the mobile device is switched on and off with a frequency of 217 Hz. If the background noise is low enough or the mute button of the telephone is pressed, the mobile device changes into DTX mode for the uplink part of the channel. This can be heard as well, as the constant 217 Hz hum is replaced by single short bursts every 0.5 seconds.

For incoming calls, this method can also be used to check that a mobile device has started communication with the network on the SDCCH one to two seconds before ringing. This delay is due to the fact that the mobile device first needs to go through the authentication phase and the activation of the ciphering for the channel. Only afterward can the network forward further information to the mobile device as to why the channel was established. This is also the reason why it takes a much longer time for the alerting tone to be heard when calling a mobile device compared to calling a fixed-line phone.

Some mobile devices possess a number of interesting network-monitoring functionalities, which are hidden in the mobile device software and are usually not directly accessible via the phone's menu. These network monitors allow visualization of many procedures and parameters that have been discussed in this chapter, such as the timing advance, channel allocation, power control, cell ID, neighboring cell information, handover and cell reselection. Various web pages can be found on the Internet that explain how these monitors can be activated, depending on the type and model of the phone. As the activation procedures are different for every phone, it is not possible to give a general recommendation. However, by using the manufacturer and model of the phone in combination with terms like 'GSM network monitor', 'GSM netmonitor' or 'GSM monitoring mode', it is relatively easy to discover if and how the monitoring mode can be activated for a specific phone.

1.8 Mobility Management and Call Control

As all components of a GSM mobile network have now been introduced, the following section gives an overview of the three processes that allow a subscriber to roam throughout the network.

1.8.1 Cell Reselection and Location Area Update

As the network needs to be able to forward an incoming call, the subscriber's location must be known. After the mobile device is switched on, its first action is to register with the network. Therefore, the network becomes aware of the current location of the user, which can change at any time because of the mobility of the user. If the user roams into the area of a new cell, it may need to inform the network of this change. To reduce the signaling load in the radio network, several cells are grouped into a location area. The network informs the mobile device via the BCCH of a cell not only of the cell ID but also of the LAC that the new cell belongs to. The mobile device thus only has to report its new location if the new cell belongs to a new location area. Grouping several cells into location areas not only reduces the signaling load in the network but also the power consumption of the mobile. A disadvantage of this method is that the network operator is only aware of the current location area of the subscriber but not of the exact cell. Therefore, the network has to search for the mobile device in all cells of a location area for an incoming call or SMS. This procedure is called paging. The size of a location area can be set by the operator depending on its particular needs. In operational networks, several dozen cells are usually grouped into a location area (Figure 1.41).

Figure 1.39 shows how a location area update procedure is performed. While idle, the mobile measures the signal strengths of the serving cell and of the neighboring cells. Neighboring cells can be found because their transmission frequency is announced on the broadcast channel (BCCH) of the serving cell. Typical values that a signal is received with are −100 dBm, which indicates that it is very far away from the base station, and −60 dBm, which indicates that it is very close to the base station. This value is also referred to as the received signal strength indication (RSSI). Once the signal of a neighboring cell becomes stronger than the signal of the current cell by a value that can be set by the network operator, the mobile reselects the new cell and reads the BCCH. If the LAC that is broadcast is different from that of the previous cell a location update procedure is started. After a signaling connection has been established, the mobile device sends a Location Update Request message to the MSC, which is transparently forwarded

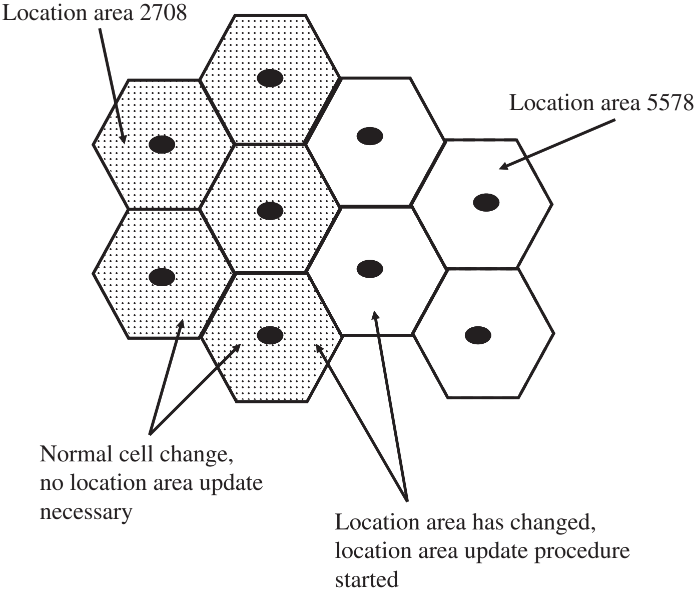

Location area 2708
Location area 5578
Normal cell change, no location area update necessary
Location area has changed, location area update procedure started

Figure 1.41 Cells in different location areas.

by the radio network. Before the message can be sent, however, the mobile device needs to authenticate itself and ciphering is usually activated as well.

Once the connection is secured against eavesdropping, the mobile device is usually assigned a new TMSI by the network, which it uses for the next connection establishment to identify itself instead of the IMSI. By the use of a constantly changing temporary ID, the identity of a subscriber is not revealed to listeners during the first phase of the call, which is not ciphered. Once TMSI reallocation has been performed, the location area update message is sent to the network, which acknowledges the correct reception. After receipt of the acknowledgment, the connection is terminated and the mobile device returns to idle state.

If the old and new location areas are under the administration of two different MSC/VLRs, a number of additional steps are necessary. In this case, the new MSC/VLR has to inform the HLR that the subscriber has roamed into its area of responsibility. The HLR then deletes the record of the subscriber in the old MSC/VLR. This procedure is called an inter-MSC location update. From the mobile point of view, however, there is no difference compared to a standard location update as the additional messages are only exchanged in the core network.

1.8.2 The Mobile-Terminated Call

An incoming call for a mobile subscriber is called a mobile-terminated call by the GSM standards. The main difference between a mobile network and a fixed-line PSTN network is that the telephone number of the mobile subscriber does not hold any information about where the subscriber is located. In the mobile network, it is thus necessary to query the HLR for the current location of the subscriber before the call can be forwarded to the correct switching center.

Figure 1.42 shows the first part of the message flow for a mobile-terminated call initiated from a fixed-line subscriber. From the fixed-line network, the Gateway-Mobile Switching Center (G-MSC) receives the telephone number (MSISDN) of the called party via an ISUP IAM message. The subsequent message flow on this interface is as shown in Figure 1.6 and the fixed-line network does not have to be aware that the called party is a mobile subscriber. The G-MSC in this example is simply a normal MSC with additional connections to other networks. When the G-MSC receives the IAM message, it sends a Send Routing Information (SRI) message to the HLR to locate the subscriber in the network. The MSC currently responsible for the subscriber is also called the subscriber's Visited Mobile Switching Center (V-MSC).

The HLR then determines the subscriber's IMSI by using the MSISDN to search through its database and thus is able to locate the subscriber's current V-MSC. The HLR then sends a Provide Roaming Number (PRN) message to the V-MSC/VLR to inform the switching center of the incoming call. In the V-MSC/VLR, the IMSI of the subscriber, which is part of the PRN message, is associated with a temporary Mobile Station Roaming Number (MSRN), which is returned to the HLR. The HLR then transparently returns the MSRN to the G-MSC.

The G-MSC uses the MSRN to forward the call to the V-MSC. This is possible as the MSRN not only temporarily identifies the subscriber in the V-MSC/VLR but also uniquely identifies the V-MSC to external switches. To forward the call from the G-MSC to the V-MSC, an IAM message is used again, which, instead of the MSISDN,

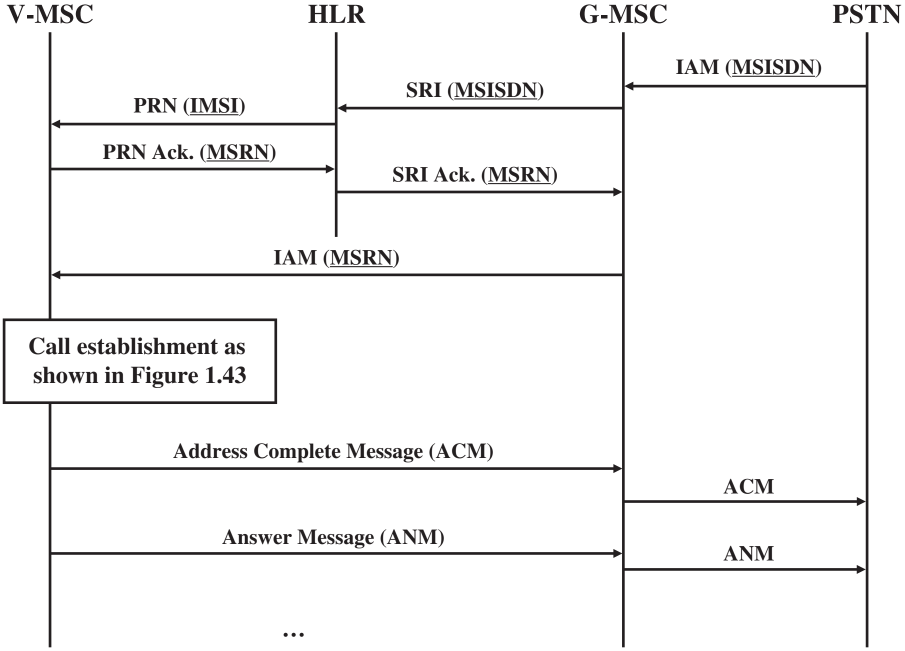


Figure 1.42 Mobile-terminated call establishment, part 1.

contains the MSRN to identify the subscriber. This has been done as it is possible, and even likely, that there are transit switching centers between the G-MSC and V-MSC, which are thus able to forward the call without querying the HLR themselves.

As the MSRN is internationally unique instead of only in the subscriber's home network, this procedure can still be used if the subscriber is roaming in a foreign network. The presented procedure, therefore, works for both national and international roaming. As the MSRN is saved in the billing record for the connection, it is also possible to invoice the terminating subscriber for forwarding the call to a foreign network and to transfer a certain amount of the revenue to the foreign network operator.

In the V-MSC/VLR, the MSRN is used to find the subscriber's IMSI and thus the complete subscriber record in the VLR. This is possible because the relationship between the IMSI and MSRN was saved when the HLR first requested the MSRN. After the subscriber's record has been found in the VLR database, the V-MSC continues the process and searches for the subscriber in the last reported location area, which was saved in the VLR record of the subscriber. The MSC then sends a Paging message to the responsible BSC. The BSC in turn sends a Paging message via each cell of the location area on the PCH. If no answer is received, then the message is repeated after a few seconds.

After the mobile device has answered the Paging message, an authentication and ciphering procedure has to be executed to secure the connection in a similar way as previously presented for a location update. Only then is the mobile device informed about the details of the incoming call with a Setup message. The Setup message contains, for example, the telephone number of the caller if the Calling Line Identification

Presentation (CLIP) supplementary service is active for this subscriber and not suppressed by the Calling Line Identification Restriction (CLIR) option that can be set by the caller (see Table 1.4).

If the mobile device confirms the incoming call with a call confirmed message, the MSC requests the establishment of a TCH for the voice path from the BSC (see Figure 1.43).

After successful establishment of the speech path, the mobile device returns an alerting message and thus informs the MSC that the subscriber is informed about the incoming call (the phone starts ringing). The V-MSC then forwards this information via the ACM to the G-MSC. The G-MSC then forwards the alerting indication to the fixed-line switch via its own ACM message.

Once the mobile subscriber accepts the call by pressing the answer button, the mobile device returns an Answer Message to the V-MSC. Here, an ISUP answer (ANM) message is generated and returned to the G-MSC. The G-MSC forwards this information again via an ANM message back to the fixed-line switching center.

While the conversation is ongoing, the network continues to exchange messages between different components to ensure that the connection is maintained. Most of the messages are measurement report messages, which are exchanged between the mobile device, the base station and the BSC. If necessary, the BSC can thus trigger a handover to a different cell. More details about the handover process can be found in Section 1.8.3.

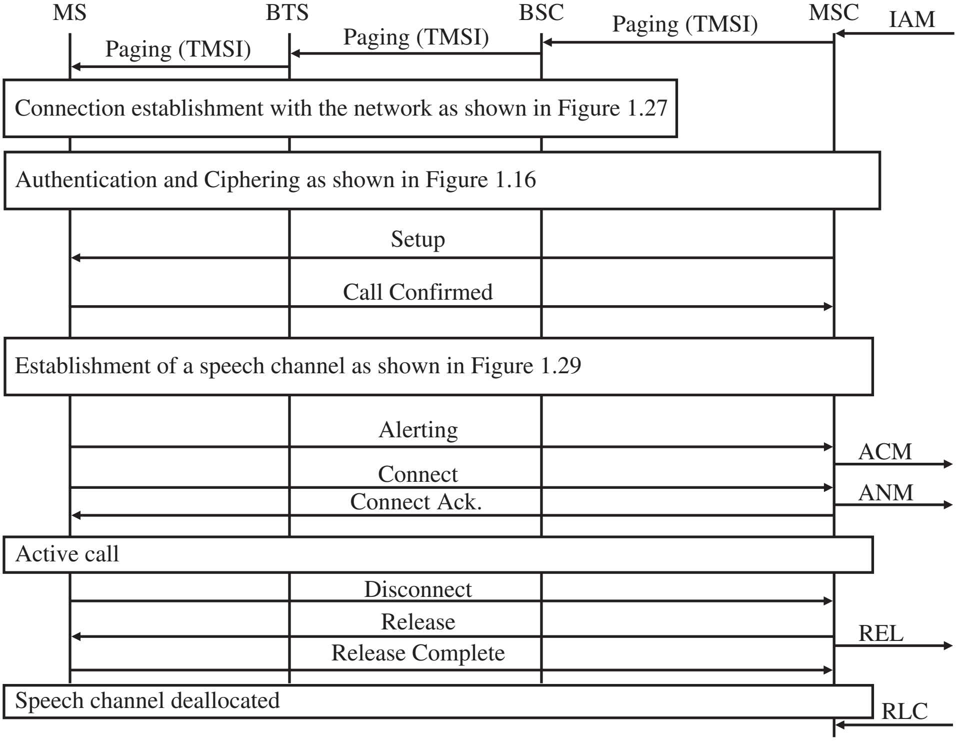

MS
BTS
BSC
MSC
IAM
Paging (TMSI)
Paging (TMSI)
Paging (TMSI)
Connection establishment with the network as shown in Figure 1.27
Authentication and Ciphering as shown in Figure 1.16
Setup
Call Confirmed
Establishment of a speech channel as shown in Figure 1.29
Alerting
ACM
Connect
ANM
Connect Ack.
Active call
Disconnect
Release
REL
Release Complete
Speech channel deallocated
RLC

Figure 1.43 Mobile-terminated call establishment, part 2.

If the mobile subscriber wants to end the call, the mobile device sends a disconnect message to the network. After the release of the TCH with the mobile device and the sending of an ISUP Release (REL) message to the other party, all resources in the network are freed and the call ends.

In this example, it has been assumed that the mobile subscriber is not in the area that is covered by the G-MSC. Such a scenario, however, is quite likely if a call is initiated by a fixed-line subscriber to a mobile subscriber who is currently roaming in the same region. As the fixed-line network usually forwards the call to the closest MSC to save costs, the G-MSC will, in many cases, also be the V-MSC for the connection. The G-MSC recognizes such a scenario if the MSRN returned by the HLR in the SRI acknowledge message contains a number which is from the MSRN pool of the G-MSC. In this case, the call is treated in the G-MSC right away and the ISUP signaling inside the mobile network (IAM, ACM and ANM) is left out. More details about call establishment procedures in GSM networks can be found in 3GPP TS 23.018 [31].

1.8.3 Handover Scenarios

If reception conditions deteriorate during a call because of a change in the location of the subscriber, the BSC has to initiate a handover procedure. The basic procedure and the necessary messages have already been shown in Figure 1.29. Depending on the parts of the network that are involved in the handover, one of the following handover scenarios described in 3GPP TS 23.009 [32] is used to ensure that the connection remains established:

- **Intra-BSC handover.** In this scenario, the current cell and the new cell are connected to the same BSC. This scenario is shown in Figure 1.30.
- **Inter-BSC handover.** If a handover has to be performed to a cell which is connected to a second BSC, the current BSC is not able to control the handover itself as no direct signaling connection exists between the BSCs of a network. Thus, the current BSC requests the MSC to initiate a handover to the other cell by sending a handover request message. Important parameters of the message are the cell ID and the LAC of the new cell. As the MSC administers a list of all LACs and cells under its control, it can find the correct BSC and request the establishment of a TCH for the handover in a subsequent step. Once the new BSC has prepared the speech channel (TCH) in the new cell, the MSC returns a handover command to the mobile device via the still existing connection over the current BSC. The mobile device then performs the handover to the new cell. Once the new cell and BSC have detected the successful handover, the MSC can switch over the speech path and inform the old BSC that the TCH for this connection can be released.
- **Inter-MSC handover.** If the current and new cells for a handover procedure are not connected to the same MSC, the handover procedure is even more complicated. As in the previous example, the BSC detects that the new cell is not in its area of responsibility and thus forwards the handover request to the MSC. The MSC also detects that the LAC of the new cell is not part of its coverage area. Therefore, the MSC looks into another table that lists all LACs of the neighboring MSCs. As the MSC in the next step contacts a second MSC, the following terminology is introduced to unambiguously identify the two MSCs: the MSC which has assigned an MSRN at the beginning of the call is called the Anchor-Mobile Switching Center (A-MSC) of the connection. The MSC that receives the call during a handover is called the Relay-Mobile Switching Center (R-MSC) (see Figure 1.44).

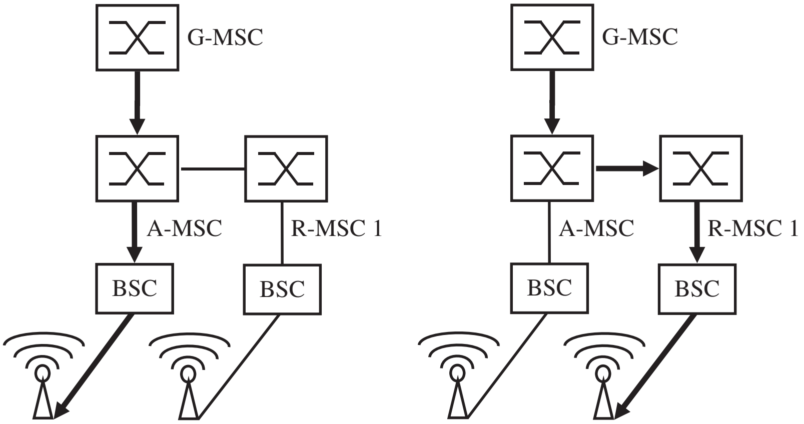


Figure 1.44 Inter-MSC handover.

To perform the handover, the A-MSC sends an MAP (see Section 1.4.2) handover message to the R-MSC. The R-MSC then asks the responsible BSC to establish a TCH in the requested cell and reports back to the A-MSC. The A-MSC then instructs the mobile device via the still existing connection over the current cell to perform the handover. Once the handover has been performed successfully, the R-MSC reports the successful handover to the A-MSC. The A-MSC can then switch the voice path toward the R-MSC. Afterward, the resources in the old BSC and cell are released.

If the subscriber changes again during the call to another cell controlled by yet another MSC, a subsequent inter-MSC handover has to be performed as shown in Figure 1.45.

For this scenario, the current Relay-MSC (R-MSC 1) reports to the A-MSC that a subsequent inter-MSC handover to R-MSC 2 is required to maintain the call. The A-MSC then instructs R-MSC 2 to establish a channel in the requested cell. Once the speech channel is ready in the new cell, the A-MSC sends the Handover Command message via R-MSC 1. The mobile device then performs the handover to R-MSC 2 and reports the successful execution to the A-MSC. The A-MSC can then redirect the speech path to R-MSC 2 and instruct R-MSC 1 to release the resources. By having the

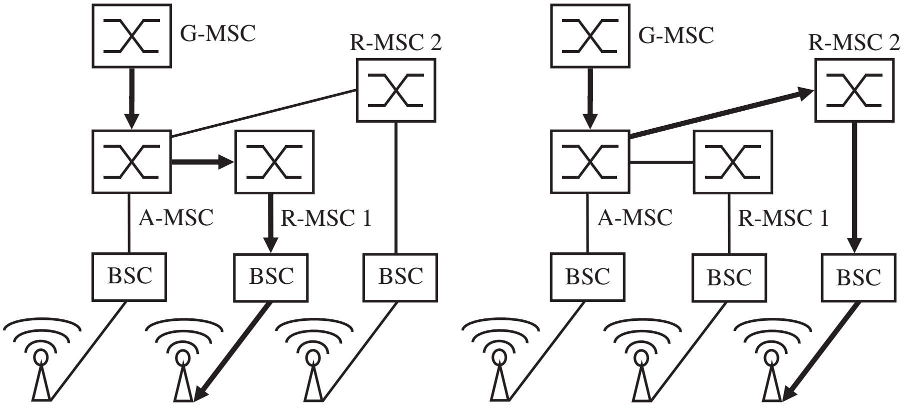


Figure 1.45 Subsequent inter-MSC handover.

A-MSC in command in all the different scenarios, it is ensured that during the lifetime of a call only the G-MSC, the A-MSC and at most one R-MSC are part of a call. In addition, tandem switches might be necessary to route the call through the network or to a roaming network. However, these switches purely forward the call and are thus transparent in this procedure.

Finally, there is also a handover case in which the subscriber who is served by an R-MSC returns to a cell which is connected to the A-MSC. Once this handover is performed, no R-MSC is part of the call. Therefore, this scenario is called a subsequent handback.

From the mobile device point of view, all handover variants are performed in the same way, as the handover messages are identical for all scenarios. To perform a handover as quickly as possible, however, GSM can send synchronization information for the new cell in the handover message. This allows the mobile device to immediately switch to the allocated timeslot instead of having to synchronize first. This can only be done, however, if the current and the new cells are synchronized with each other, which is not possible, for example, if they are controlled by different BSCs. As two cells that are controlled by the same BSC may not necessarily be synchronized, synchronization information is by no means an indication of what kind of handover is being performed in the radio and core network.

1.9 The Mobile Device

Owing to the progress of miniaturization of electronic components during the mid-1980s, it was possible to integrate all components of a mobile device into a single portable device. A few years later, mobile devices had shrunk to such a small size that the limiting factor in future miniaturization was no longer the size of the electronic components. Instead, the space required for user interface components like display and keypad limited a further reduction. Because of continuous improvement and miniaturization of electronic components, it is possible to integrate more and more functionalities into a mobile device and to improve the ease of use. While mobile devices were at first used only for voice calls, the trend today is toward feature-rich devices that also include telephony functions. This section, therefore, first describes the architecture of a, from today's perspective, simple voice-centric phone. Afterward, the architecture of a modern smartphone will be discussed.

1.9.1 Architecture of a Voice-Centric Mobile Device

Simple GSM devices for voice and SMS communication can be built with only a few parts today and can therefore be produced very cheaply. The simplest GSM mobile phones can thus be bought today for less than 20 euros.

Figure 1.46 shows the principle architecture of such a device. Its core is based on a baseband processor, which contains a reduced instruction set (RISC), a CPU and a digital signal processor (DSP).

The RISC processor is responsible for the following tasks:

- handling of information that is received via the different signaling channels (BCCH, PCH, AGCH, PCH and so on);
- call establishment (DTAP);

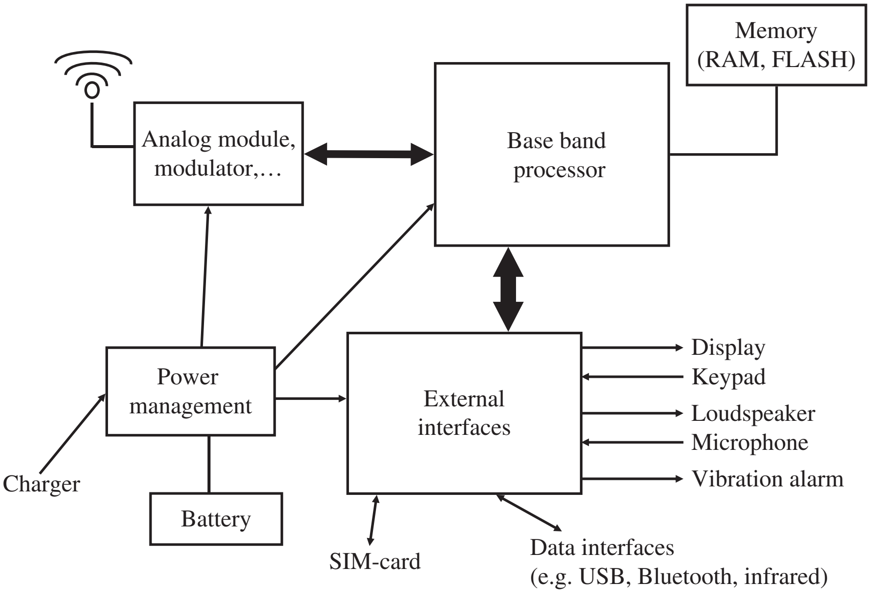


Figure 1.46 Basic architecture of a mobile phone.

- GPRS management and GPRS data flow;
- parts of the transmission chain, like channel coder, interleaver and cipherer (dedicated hardware component in some designs);
- mobility management (network search, cell reselection, location update, handover, timing advance, etc.);
- connections via external interfaces like Bluetooth, infrared and Universal Serial Bus (USB);
- user interface (keypad, display, graphical user interface).

As many of these tasks have to be performed in parallel, a multitasking embedded real-time operating system is used on the RISC processor. The real-time component of the operating system is especially important as the processor has to be able to provide data for transmission over the air interface according to the GSM frame structure and timing. All other tasks like keypad handling, display update and the graphical user interface, in general, have a lower priority.

In most devices, the baseband processor is based on the ARM Reduced Instruction Set (RISC) architecture that allows clock speeds of up to 2 GHz today. Such processors are built by several manufacturers that have obtained a license from ARM. For simple devices, ultra-low-power versions are used with a clock rate of only a few megahertz. Such chips are very power efficient and require only little energy while in sleep mode and while periodically observing the PCH. As a consequence, it is possible to reach standby times of well over a week. Also, the amount of RAM and ROM for such a device is very small by today's standards, usually in the order of only a few hundred kilobytes.

The DSP is another important component of a GSM chipset. Its main task is the decoding of the incoming signal and FR, EFR, HR or AMR speech compression. In GSM, signal decoding starts with the analysis of the training sequence of a burst (see

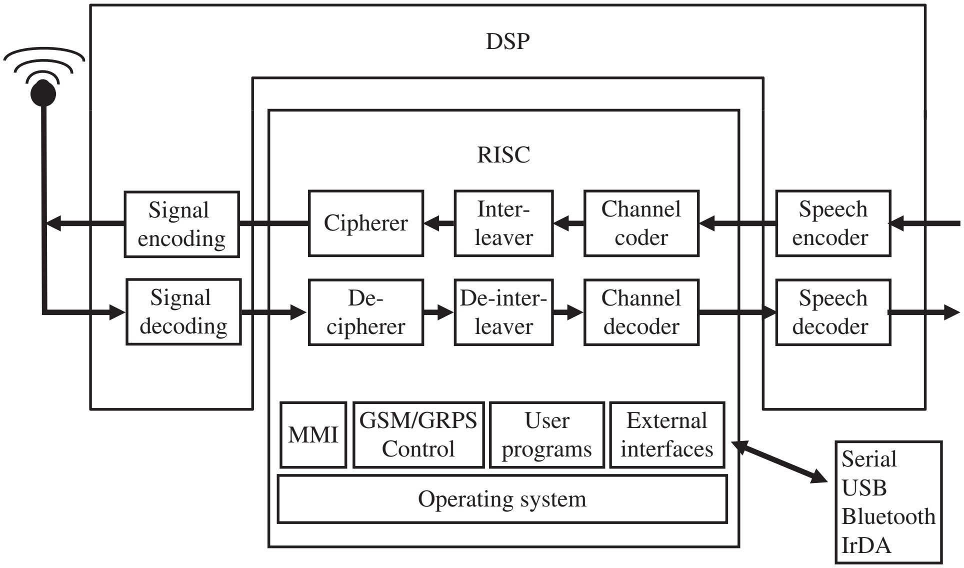


Figure 1.47 Overview of RISC and DSP functionalities.

Section 1.7.3). As the DSP is aware of the composition of the training sequence of a frame, the DSP can calculate a filter that is then used to decode the data part of the burst. This increases the probability that the data can be correctly reconstructed.

Figure 1.47 shows the respective tasks that are performed by the RISC processor and the DSP processor. If the transmission chain for a voice signal is compared between the mobile device and the network, it can be seen that the TRAU mostly performs the task of a DSP unit in the mobile device. All other tasks such as channel coding are performed by the BTS, which is thus the counterpart of the RISC CPU of the mobile device.

As hundreds of millions of mobile devices are sold every year, there is a great variety of chipsets available on the market. The chipset is in many cases not designed by the manufacturer of the mobile device. Chipset manufacturers that are not developing their own mobile devices include, for example, Qualcomm and Mediatek.

1.9.2 Architecture of a Smartphone

Simple GSM voice phones usually have one processor that handles both the modem functionality and the operating system for the user interface. In smartphones, these tasks are performed by independent processors. This has become necessary as each function has become much more complex over time, and the processor architecture required by each function has developed in different directions. In addition, smartphones include many new functionalities that require significant and specialized processing capabilities. Figure 1.48 gives an overview of the typical function blocks of a modern smartphone. Owing to increasing miniaturization, most or even all of the functions shown are included in a single chip. Such a combination is often also referred to as a System on a Chip (SoC).

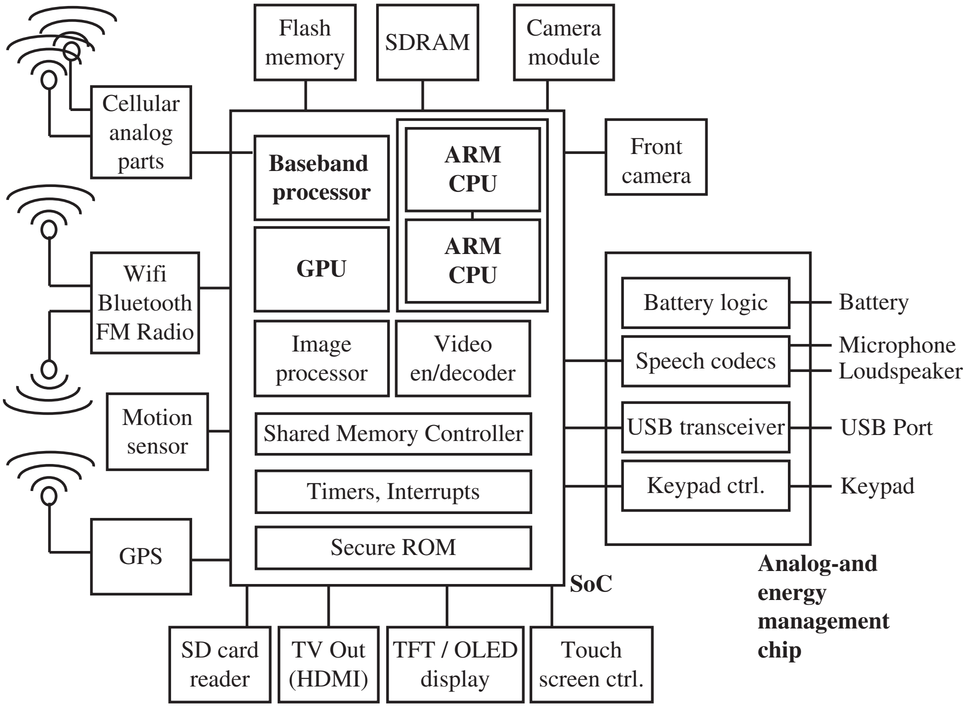


Figure 1.48 Architecture of a modern smartphone.

The baseband processor is responsible for communication with a mobile network and supports not only GSM but also UMTS and LTE. A few analog components cannot be included on the SoC because of their size and function and are thus shown separately in the figure.

The operating system for the user interface, such as Android or iOS, is executed on the application processor, which usually consists of several ARM processor cores. These ARM CPUs have an instruction set similar to that of the ARM processors of the voice-centric GSM phones discussed earlier. Owing to their higher complexity and higher clock rates, their power consumption is much higher. The baseband processor and application processor operate independently of each other and communicate with each other over a fast serial interface. This is also the case if both units are contained in a single SoC. Other important building blocks usually contained in an SoC are a dedicated Graphics Processing Unit (GPU) and a number of additional supporting functions for memory management, timers, interrupts and dedicated processing units allowing the external camera module to quickly process images taken by the user for the operating system. Memory chips are usually still physically separate from the SoC. And finally, additional external elements such as a camera module on the back, another one on the front, Wi-Fi, Bluetooth and FM radio chips, as well as GPS, SD card readers, TV-Out, display and touchscreen are also part of a modern smartphone.

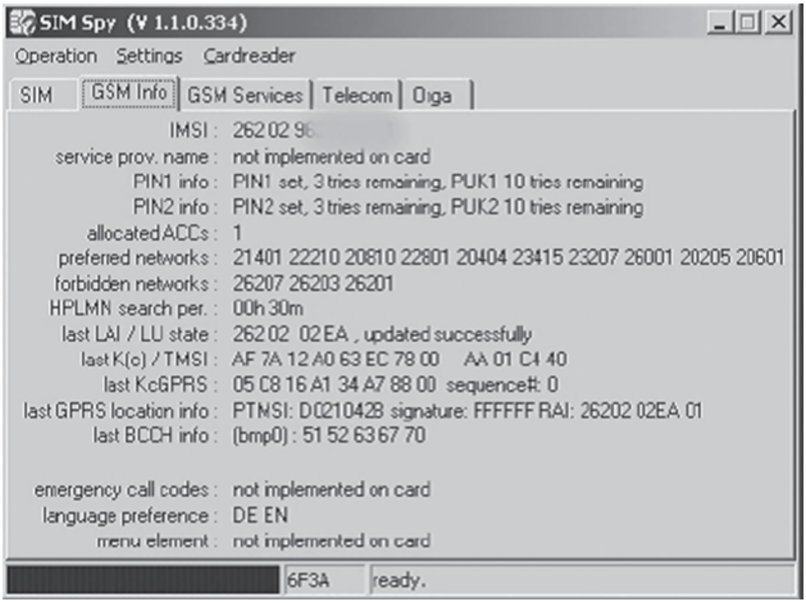
1.10 The SIM Card

Despite its small size, the SIM card, officially referred to as the Universal Integrated Circuit Card (UICC), is one of the most important parts of a GSM network because it contains all the subscription information of a subscriber. Since it is standardized, a subscriber can use any GSM or UMTS phone by simply inserting the SIM card. Exceptions are phones that contain a ‘SIM lock’ and thus only work with a single SIM card or only with the SIM card of a certain operator. However, this is not a GSM restriction. It was introduced by mobile network operators to ensure that a subsidized phone is used only with SIM cards of their network.

The most important parameters on the SIM card are the IMSI and the secret key (Ki), the latter of which is used for authentication and the generation of ciphering keys (Kc). With a number of tools, which are generally available on the Internet free of charge, it is possible to read out most parameters from the SIM card, except for sensitive parameters that are read protected. Figure 1.49 shows such a tool. Protected parameters can only be accessed with a special unlock code that is not available to the end user.

Astonishingly, a SIM card is much more than just a simple memory card as it contains a complete microcontroller system that can be used for a number of additional purposes. The typical properties of a SIM card are shown in Table 1.7.

As shown in Figure 1.50, the mobile device cannot access the information on the Electrically Erasable Programmable Read-Only Memory (EEPROM) directly, but has to request the information from the SIM’s CPU. Therefore, direct access to sensitive information is prohibited. The CPU is also used to generate the SRES during the network authentication procedure, based on the RAND which is supplied by the AuC (see Section 1.6.4). It is imperative that the calculation of the SRES is done on the SIM card



```
                 IMSI :  262 02 96[redacted]
    service prov. name :  not implemented on card
             PIN1 info :  PIN1 set, 3 tries remaining, PUK1 10 tries remaining
             PIN2 info :  PIN2 set, 3 tries remaining, PUK2 10 tries remaining
        allocated ACCs :  1
    prefered networks :  21401 22210 20810 22801 20404 23415 23207 26001 20205 20601
   forbidden networks :  26207 26203 26201
    HPLMN search per. :  00h 30m
    last LAI / LU state :  262 02  02 EA , updated successfully
      last K(c) / TMSI :  AF 7A 12 A0 63 EC 78 00     AA 01 C4 40
         last KcGPRS :  05 C8 16 A1 34 A7 88 00  sequence#: 0
last GPRS location info :  PTMSI: D0210428 signature: FFFFFF RAI: 26202 02EA 01
        last BCCH info :  (bmp0) : 51 52 63 67 70

  emergency call codes :  not implemented on card
   language preference :  DE EN
          menu element :  not implemented on card

6F3A    ready.
```

Figure 1.49 Example of a tool to visualize the data contained on a SIM card.

Table 1.7 SIM card properties.

CPU	8- or 16-bit CPU
ROM	40–100 kB
RAM	1–3 kB
EEPROM	16–64 kB
Clock rate	10 MHz, generated from clock supplied by mobile device
Operating voltage	1.8 V, 3 V and 5 V. Modern devices use 1.8 V but support SIM cards with higher voltage requirements as well

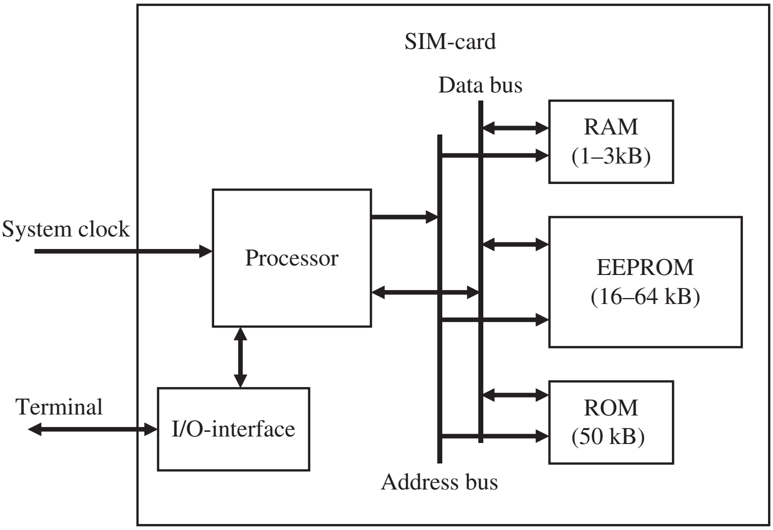


Figure 1.50 Block diagram of SIM card components.

itself and not in the mobile device to protect the secret Ki key. If the calculation was done in the mobile device itself, this would mean that the SIM card would have to hand over the Ki to the mobile device or any other device upon request. This would seriously undermine security, as tools like the one shown in Figure 1.49 would be able to read the Ki, which could then be used to make a copy of the SIM card.

Furthermore, the microcontroller system on the SIM can also execute programs that the network operator may have installed on the SIM card. This is done via the SIM application toolkit (SAT) interface, which is specified in 3GPP TS 31.111 [33]. With the SAT interface, programs on the SIM card can access functionalities of the mobile device such as waiting for user input after showing a text message or sending or receiving SMS messages without user intervention. While this functionality was used extensively by network operators in the past for value-added services, the SAT interface is now mainly used for background tasks such as sending a notification to the network when the SIM card detects that it has been inserted in a new device to trigger the transfer of welcome and configuration messages. Furthermore, the SAT interface still plays an important role in receiving 'silent' SMS messages from the network to update information on the SIM card such as the list of preferred roaming networks.

From a logical point of view, data are stored on a GSM SIM card in directories and files, in a manner similar to the storage on a PC's hard drive. The file and folder structures are specified in 3GPP TS 31.102 [34]. In the specification, the root directory is called the main file (MF), which is somewhat confusing at first. Subsequent directories are called dedicated files (DF), and normal files are called elementary files (EF). As there is only a very limited amount of memory on the SIM card, files are not identified via file and directory names. Instead, hexadecimal numbers with a length of four digits are used, which require only 2 B memory. The standard nevertheless assigns names to these numbers, which are, however, not stored on the SIM card. The root directory, for example, is identified by ID 0x3F00, the GSM directory is identified by ID 0x7F20 and the file containing the IMSI is identified by ID 0x6F07. To read the IMSI from the SIM card, the mobile device thus has to open the following path and file: 0x3F00 0x7F20 0x6F07.

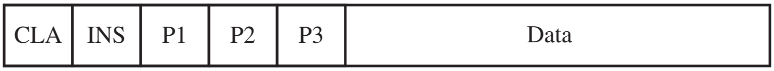
To simplify access to the data contained on the SIM card for the mobile device, a file can have one of the following three file formats:

- **Transparent.** The file is seen as a sequence of bytes. The file for the IMSI, for example, is of this format. How the mobile device has to interpret the content of the files is again specified in 3GPP TS 31.002 [34].
- **Linear fixed.** This file type contains records of a fixed length and is used, for example, for the file that contains the telephone book records. Each phone record uses one record of the linear fixed file.
- **Cyclic.** This file type is similar to the linear fixed file type but contains an additional pointer that points to the last modified record. Once the pointer reaches the last record of the file, it wraps over again to the first record of the file. This format is used, for example, for the file in which the phone numbers which have previously been called are stored.

A number of different access right attributes are used to protect the files on the SIM card. By using these attributes, the card manufacturer can control whether a file is read only or write only when accessed by the mobile device. A layered security concept also permits network operators to change files which are read only for the mobile device over the air by sending special provisioning SMS messages.

The mobile device can only access the SIM card if the user has typed in the PIN when the phone is started. The mobile device then uses the PIN to unlock the SIM card. SIM cards of some network operators, however, allow deactivation of the password protection and thus the user does not have to type in a PIN code when the mobile device is switched on. Despite unlocking the SIM card with the PIN, the mobile device is still restricted to only being able to read or write certain files. Thus, it is not possible, for example, to read or write to the file that contains the secret key Ki even after unlocking the SIM card with the PIN.

Details on how the mobile device and the SIM card communicate with each other have been specified in ETSI TS 102 221 [35]. For this interface, layer 2 command and response messages have been defined, which are called Application Protocol Data Units (APDUs). When a mobile device wants to exchange data with the SIM card, a command APDU is sent to the SIM card. The SIM card analyzes the command APDU, performs the requested operation and returns the result in a response APDU. The SIM card only has a passive role in this communication as it can only send response APDUs back to the mobile device.

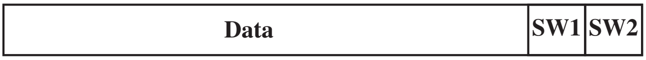


CLA	INS	P1	P2	P3	Data

Figure 1.51 Structure of a command APDU.

Table 1.8 Examples for APDU commands.

Command	ID	P1	P2	Length
Select (open file)	A4	00	00	02
Read binary (read file)	B0	Offset high	Offset low	Length
Update binary (write file)	D6	Offset high	Offset low	Length
Verify CHV (check PIN)	20	00	ID	08
Change CHV (change PIN)	24	00	ID	10
Run GSM algorithm (RAND, SRES, Kc, …)	88	00	00	10



Data	SW1	SW2

Figure 1.52 Response APDU.

If a file is to be read from the SIM card, the command APDU contains, among other information, the file ID and the number of bytes to read from the file. If the file is of cyclic or linear fixed type, the command also contains the record number. If access to the file is allowed, the SIM card then returns the requested information in one or more response APDUs.

If the mobile device wants to write some data into a file on the SIM card, the command APDUs contain the file ID and the data to be written into the file. In the response APDU, the SIM card then returns a response as to whether the data were successfully written to the file.

Figure 1.51 shows the format of a command APDU. The first field contains the class of instruction, which is always 0xA0 for GSM. The instruction (INS) field contains the ID of the command that has to be executed by the SIM card.

Table 1.8 shows some commands and their IDs. The fields P1 and P2 are used for additional parameters for the command. P3 contains the length of the following data field, which contains the data that the mobile device would like to write on the SIM card.

The format of a response APDU is shown in Figure 1.52. Apart from the data field, the response also contains two fields called SW1 and SW2. These are used by the SIM card to inform the mobile device whether the command was executed correctly.

For example, to open a file for reading or writing, the mobile device sends a SELECT command to the SIM card. The SELECT APDU is structured as shown in Figure 1.53.

As a response, the SIM card replies with a response APDU that contains a number of fields. Some of them are shown in Table 1.9.

For a complete list of information returned for the example, see [35]. In the next step, the READ BINARY or WRITE BINARY APDU can be used to read or modify the file.

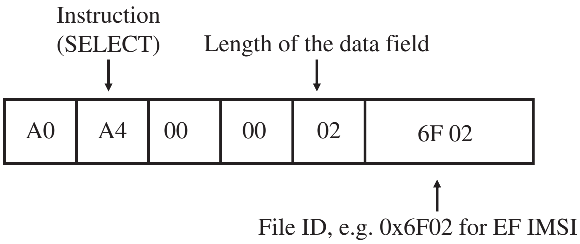


Figure 1.53 Structure of the SELECT command APDU.

Table 1.9 Some fields of the response APDU for a SELECT command.

Byte	Description	Length
3–4	File size	2
5–6	File ID	2
7	Type of file (transparent, linear fixed, cyclic)	1
9–11	Access rights	3
12	File status	1

To physically communicate with the SIM card, there are eight contact areas on the top side of the SIM card. Only five of those contacts are required:

- C1: power supply;
- C2: reset;
- C3: clock;
- C5: ground;
- C7: input/output.

As only a single line is used for the input and output of command and status APDUs, the data are transferred in half-duplex mode only. The clock speed for the transmission has been defined as C3/327. At a clock speed of 5 MHz on C3, the transmission speed is thus 13,440 bit/s.

1.11 The Intelligent Network Subsystem and CAMEL

All components that have been described in this chapter are mandatory elements for the operation of a mobile network in which billing records are collected and invoices sent once a month. To offer prepaid services for which subscribers have to be billed in real-time, additional logic and databases are necessary. These are referred to as the Intelligent Network (IN) and implemented on a Service Control Point (SCP) as described in Section 1.4. Prepaid services have become very popular in many countries since their introduction in the mid-1990s. Instead of receiving a bill once a month, a prepaid subscriber has an account with the network operator, which is funded in advance with a certain amount of money determined by the subscriber. The amount on the account can then be used for phone calls, SMS and data services. During every call or event, such as the user sending an SMS, the account is continually charged. If the account runs out of credit, the connection is interrupted and the transmission of further SMS messages is blocked. In the early years of GSM, the development of these services had been highly proprietary because of the lack of a common standard. The big disadvantage of such solutions was that they were customized to work only with very specific components of a single manufacturer. This meant that these services did not work abroad as foreign network operators used components of other network vendors. This was especially a problem for the prepaid service as prepaid subscribers were excluded from international roaming when the first services were launched.

To ensure the interoperability of intelligent network components between different vendors and in networks of different mobile operators, industry and operators standardized an IN protocol in 3GPP TS 22.078 [36], which is called Customized Applications for Mobile-Enhanced Logic, or CAMEL for short. While CAMEL also offers functionality for SMS and GPRS charging, the following discussion describes only the basic functionality necessary for circuit-switched connections.

CAMEL is not an application or a service, but forms the basis for creating services (customized applications) on an SCP which are compatible with network elements of other vendors and between networks. Thus, CAMEL can be compared with HTTP. HTTP is used for transferring web pages between a web server and a browser. HTTP ensures that any web server can communicate with any browser. Whether the content of the data transfer is a web page or a picture is of no concern to HTTP because this is managed on a higher layer directly by the web server and the web client. Transporting the analogy back to the GSM world, the CAMEL specification defines the protocol for communication between different network elements such as the MSC and the SCP, as well as a state model for call control.

The state model is called the Basic Call State Model (BCSM) in CAMEL. A circuit-switched call, for example, is divided into a number of different states. For the originator (O-BCSM), the following states, which are also shown in Figure 1.54, have been defined:

- call establishment;
- analysis of the called party's number;
- routing of the connection;
- notification of the called party (alerting);
- ongoing call (active);
- disconnection of the call;
- no answer from the called party;
- called party busy.

For a called subscriber, CAMEL also defines a state model which is called the Terminating Basic Call State Model (T-BCSM). T-BCSM can be used for prepaid subscribers who are currently roaming in a foreign network to control the call in the foreign network and to apply real-time charging.

For every state change in the state model, CAMEL defines a detection point (DP). If a DP is activated for a subscriber, the SCP is informed of the particular state change. Information contained in this message includes the IMSI of the subscriber, the current position (MCC, MNC, LAC and cell ID) and the number that was called. Whether a DP is activated is part of the subscriber's HLR entry. This allows the creation of specific services on a per subscriber basis. When the SCP is notified that the state model has triggered a DP, the SCP is able to influence the way the call should proceed. The SCP can take the call down, change the number that was called or return information to the MSC, which is put into the billing record of the call for later analysis on the billing system.

For the prepaid service, for example, the CAMEL protocol can be used between the MSC and the SCP as follows.

If a subscriber wants to establish a call, the MSC detects during the setup of the call that the 'authorize origination' DP is activated in the subscriber's HLR entry. Therefore, the MSC sends a message to the SCP and waits for a reply. As the message contains the IMSI of the subscriber as well as the CAMEL service number, the SCP recognizes that the request is for a prepaid subscriber. By using the destination number, the current time and other information, the SCP calculates the price per minute for the connection. If the subscriber's balance is sufficient, then the SCP allows the call to proceed and informs the MSC about the duration for which the authorization is valid. The MSC then continues and connects the call. At the end of the call, the MSC sends another message to the SCP to inform it of the total duration of the call. The SCP then modifies the subscriber's balance. If the time that the SCP initially granted for the call expires, the MSC has to contact the SCP again. The SCP then has the possibility of sending an additional authorization to the MSC which is again limited to a determined duration. Other options for the SCP to react are to send a reply in which the MSC is asked to terminate the call or to return a message in which the MSC is asked to play a tone as an indication to the user that the balance on the account is almost depleted.

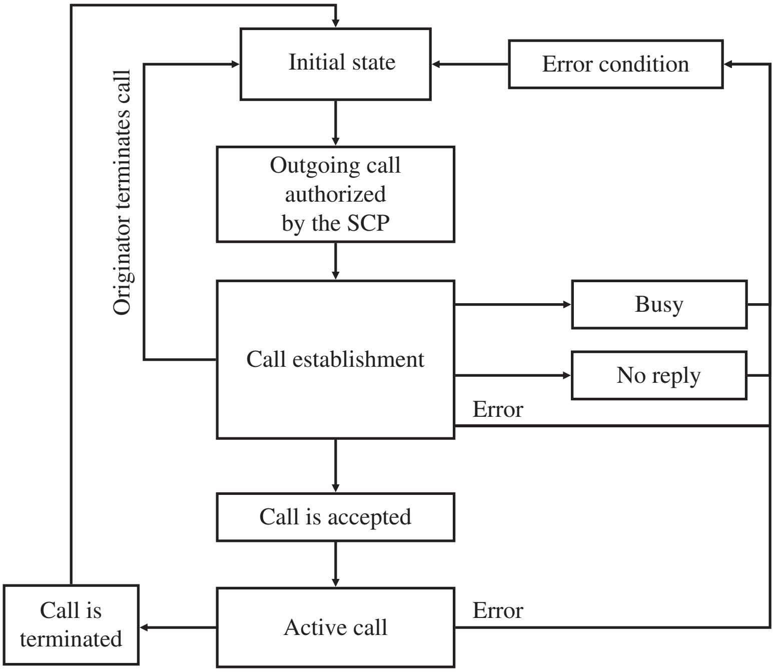


Figure 1.54 Simplified state model for an originator (O-BCSM) according to 3GPP TS 23.078 [37].

Questions

1. Which algorithm is used to digitize a voice signal for transmission in a digital circuit-switched network and at which datarate is the voice signal transmitted?
2. Name the most important components of the GSM NSS and their tasks.
3. Name the most important components of the GSM radio network (BSS) and their tasks.
4. How is a BTS able to communicate with several subscribers at the same time?
5. Which steps are necessary to digitize a speech signal in a mobile device before it can be sent over the GSM air interface?
6. What is a handover and which network components are involved?
7. How is the current location of a subscriber determined for a mobile-terminated call and how is the call forwarded through the network?
8. How is a subscriber authenticated in the GSM network? Why is an authentication necessary?
9. How is an SMS message exchanged between two subscribers?
10. Which tasks are performed by the RISC processor and which tasks are performed by the DSP in a mobile device?
11. How is data stored on the SIM card?
12. What is CAMEL and for which service is it typically used?

Answers to these questions can be found on the companion website for this book at http://www.wirelessmoves.com.

References

1 European Technical Standards Institute (ETSI), http://www.etsi.org.
2 The 3rd Generation Partnership Project, http://www.3gpp.org.
3 3GPP, Mobile Application Part (MAP) Specification, TS 29.002.
4 3GPP, Bearer-Independent Circuit-Switched Core Network – Stage 2, TS 23.205.
5 3GPP, Media Gateway Controller (MGC) – Media Gateway (MGW) Interface – Stage 3, TS 29.232.
6 ITU, H.248: Gateway Control Protocol, http://www.itu.int/rec/T-REC-H.248/.
7 ITU, Q.1901: Bearer Independent Call Control Protocol, http://www.itu.int/rec/T-REC-Q.1901.

8 3GPP, Application of Q.1900 Series to Bearer Independent Circuit Switched (CS) Core Network Architecture – Stage 3, TS 29.205.
9 3GPP, Call Forwarding (CF) Supplementary Services – Stage 1, TS 22.082.
10 3GPP, Call Barring (CB) Supplementary Services – Stage 1, TS 22.088.
11 3GPP, Call Waiting (CW) and Call Hold (HOLD) Supplementary Services – Stage 1, TS 22.083.
12 3GPP, Multi Party (MPTY) Supplementary Services – Stage 1, TS 22.084.
13 3GPP, Man–Machine Interface (MMI) of the User Equipment (UE), TS 22.030.
14 3GPP, Mobile Radio Interface Layer 3 Specification; Core Network Protocols – Stage 3, TS 24.008.
15 3GPP, Technical Realisation of Short Message Service (SMS), TS 23.040.
16 3GPP, Voice Group Call Service (VGCS) – Stage 2, TS 43.068.
17 3GPP, Voice Broadcast Service (VGS) – Stage 2, TS 43.069.
18 3GPP, Enhanced Multi-Level Precedence and Preemption Service (eMLPP) – Stage 2, TS 23.067.
19 Union Internationale des Chemins de Fer, GSM-R http://www.uic.org/gsm-r.
20 Telefonica O2 Germany, Zahlen und Fakten, January 2014.
21 3GPP, Multiplexing and Multiple Access on the Radio Path, TS 45.002.
22 3GPP, AMR Speech CODEC: General Description, TS 26.071.
23 ITU, G.722.2: Wideband Coding of Speech at Around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB), http://www.itu.int/rec/T-REC-G.722.2-200307-I/en.
24 3GPP, Speech Codec Speech Processing Functions; Adaptive Multi-Rate-Wideband (AMR-WB) Speech Codec; Transcoding Functions, TS 26.190.
25 3GPP, Full Speech Transcoding, TS 46.010.
26 E. Barkan, E. Biham and N. Keller, Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication, http://cryptome.org/gsm-crack-bbk.pdf, 2003.
27 26C3, GSM Related Activities and Presentations, https://events.ccc.de/congress/2009/, December 2009.
28 28C3, GSM Related Activities and Presentations, http://events.ccc.de/category/28c3/, December 2011.
29 Wikipedia – IMSI-Catcher, http://en.wikipedia.org/wiki/IMSI_catcher, 4 October 2016.
30 J. Frick and R. Bott, Method for Identifying a Mobile Phone User for Eavesdropping on Outgoing Calls, European Patent Office, EP1051053, http://v3.espacenet.com/publicationDetails/biblio?CC=DE&NR=19920222A1&KC = A1&FT = D&date = 20001109&DB = &locale=, November 2009.
31 3GPP, Basic Call Handling: Technical Realization, TS 23.018.
32 3GPP, Handover Procedures, TS 23.009.
33 3GPP, USIM Application Toolkit, TS 31.111.
34 3GPP, Characteristics of the USIM Application, TS 31.102.
35 ETSI, Smart Cards; UICC-Terminal Interface; Physical and Logical Characteristics, TS 102 221.
36 3GPP, Customised Applications for Mobile Network Enhanced Logic (CAMEL): Service Description – Stage 1, TS 22.078.
37 3GPP, Customised Applications for Mobile Network Enhanced Logic (CAMEL): Service Description – Stage 2, TS 23.078.