1988

NEAR EAST UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF
ELECTRICAL & ELECTRONIC ENGINEERING

EE 400

(GRADUATION PROJECT)

THE GSM SYSTEM FOR MOBILE COMMUNICATIONS

STUDENT'S NAME – SURNAME : MURAT AYDIN

STUDENT'S NUMBER : 940077

SUPERVISOR : PROF. DR. FAHRETTIN MAMEDOV

CONTENTS

3

TRANSMISSION 46

4

SIGNALING TRANSFER 68

5

RADIO RESOURCE MANAGEMENT 84

6

COMMUNICATION MANAGEMENT 102

1. SETTING THE SCENE

Mass-market mobile telecommunications is certainly one of the major breakthroughs of this end of millennium. The possibility to make and receive calls through a small wireless handset. Wherever you are, has an obvious appeal. The business opportunities are tremendous, since one can imagine that every person (and not only every home). Including children, could be equipped provided the service is cheap enough. Many people agree that the sociological consequences will be important, much more that for video communications. Wireline telephone allows us to reach a place, if someone is there to answer. Mobile telephony allows us to reach a particular person, wherever (almost) he or she is. This will greatly increase the accessibility to people, and increase the feeling of security. On the other hand, this increased accessibility can in many cases be a nuisance, and widespread social acceptability of mobile telephony requires that users have a high degree of control on the calls they receive (identification of the calling party, forwarding of calls to a third party. Message banks, etc.).

Mobile telecommunications is not a very recent technology, but it is a rapidly evolving one. Expensive vehicle mounted sets have been available for 30 or 40 years. A major step was made at the beginning of the 1980'S when analog cellular technology was introduced. We are now, at the beginning of the 1990's witnessing the emergence of the next step, which will enable this technology to reach a mass-market. Maybe approaching that of today's wireline telephony. The GSM system,. And its system was designed during the 1980's and entered operation in various European countries during 1992.

GSM was designed internationally, in standardisation committees, by the major European telecommunications operators and manufacturers. The understanding of the gain to be obtained by combining resources, and of the business opportunity offered by mass-market radiotelephony, resulted in a substantial-many-power and financial effort from the participants, thus making GSM a very dynamic project.

This first chapter will be quite general, and aims at setting the stage, the main lines of the plot and some of the main actors. The rest of the casting will be presented in the next chapter. The present chapter is divided into three parts. The first part will set the historical background of GSM. Some text is devoted to the description of how the standardisation committees worked; this give some insight of what happened behind the stage. The second part aims tom provide the basic technical foundations of public cellular radio telecommunications. Most of the addressed points are developed from different angles throughout the other chapters, recurring as leitmotis. The last part describes the services that GSM offers to the customers as well as to the system operators. This describes the objectives of the system and the rest of the book explains how they are accomplished.

1.1. The Genesis Of A Standard

From the start of the 80s. After NMT started operating successfully, it became obvious to several European countries that existing analogue systems had limitations. First, the potential demand for mobile services, even though systematically under-estimated in the early networks. Second, the different systems in operation offer no compatibility for mobile users; a TACS terminal cannot access an NMT network, and neither can an NMT terminal access a TACS network. What's more, the design of a new cellular system requires such a large investment that no European country on an individual basis can afford this investment if the only return expected is on its own national market, all these circumstances pointed toward the design of a new system, done in common between several countries.

The major prerequisite for a common radio system is a common radio bandwidth. This condition had already been met a few years before, in 1978, when it was decided to reserve a frequency band of twice 25 MHz at around 900 MHz mobile communication in Europe.

The need was clear and the major obstacle removed. It remained to organise the work. The world of telecommunication in Europe always was dominated by standardisation. The CEPT (Conference Europeennedes Postes et Telecommunications) is a standardisation arena which-in the early 80's-included the European Administrations of Posts and Telecommunications of more that 20 countries. All these factors, both circumstantial and market-driven, led to the creation in 1982 of a new standardisation body within CEPT, whose task was to specify a unique radiocommunication system for Europe, at 900 MHz. The new-born "Groupe Special Mobile" (GSM) held its first meeting in December 1982 in Stockholm, under the chairmanship of Thomas Haug, from the Swedish Administration. 31 persons from 11 countries were present at this first meeting.

Date	**Achievement**
1982	"Groupe Special Mobile" Is Created Within CEPT
1986	A Permanent Nucleus Is Set Up
1987	Main Radio Transmission Techniques Are Chosen, Based On Prototype Evaluation (1986)
1989	GSM Becomes An ETSI Technical Committee
	The Phase 1 GSM900[1] specifications (drafted 1987-1990) are frozen
1990	DCS 1800 adaptation starts
1991	First systems are running (Telecom 91 exhibition,...)
	DCS 1900 specifications are frozen
1992	All major European GSM1900 operators begin commercial operations

Table 1.1.1. GSM project milestones

The span of GSM work from the very start to commercial service extended over some 10 years, but the actual specification work did not start until 1987

In 1990. By request of the United Kingdom, the specification of a version of GSM adapted to the 1800 MHz frequency bad was added to the scope of the standardisation group, with a frequency allocation of twice 75 MHz. This variant, referred to as DCS1800 (Digital Cellular System 1800) is aimed at reaching higher capacities in urban areas for example for the type of mass-market approach known as PCN (Personal Communications Network).

The elaboration of the GSM standard took almost a decade. Major milestones are shown in table 1.2. and the corresponding stages are described in more detail in the following.

1.1.1. Organization Of The Work

The first two years of the GSM were dedicated to discussion of the fundamental principles. The frequency of meetings and the number of participants increased steadily. At the beginning of meetings and the number of participants increased steadily. At the beginning of 1984, three "working parties" were created: the meetings were split to allow a more in-depth technical work. The amount of contributions and problems to solve increased steadily, and at the end of 1985 it became obvious that the number of meeting was insufficient. It was then decided (for the first time in CEPT) that the working parties would meet independently, reporting to the GSM plenary meetings which still endorsed the decisions. The role of the different working parties had already been clearly established for some time:

- WP1 (Working party 1), for the definition of services:
- WP2 (Working party 2), for the specification of radio transmission: this subject stayed dominant until 1987;
- WP3 (Working party 3), for all other issues, i.e., mainly network architecture, specification of the signalling protocols and of the open interfaces between network entities.

Later on, a fourth working party (WP4) was created to deal specifically with the implementation of data services.

During 1985, a detailed list of Recommendations to be output by the group,. Following the model technical Recommendations of the CCITT (Committee Consultatif International Telegraphique et Telephonique), was discussed and settled. From 1986 onwards, work became centred around a major objective: drafting these recommendations. The list includes more that 100z recommendations, sorted in 12 series. A detailed action plan was generated to follow the progress of this huge task, and was updated at each meeting. In 1991, the list includes 130 recommendations,. With a total of more than 5000 pages. These recommendations cover the fullmk specification of the radio interface '(i.e.. the interface between the mobile stations and the infrastructure) as well as fairly detailed specifications of infrastructure architecture and of some of the interfaces and signalling protocols between network entities.

In order to co-ordinate the work between the working parties and to manage the edition and the updating of the recommendations, a "permanent nucleus" (PN) was created at the beginning o2f 1986. It consisted of a small team of full time members and was located in Paris. In 1988. The European Telecommunications Standard institute (ETSI) was created and most of the CEPT technical standardisation activities were transferred tom this new body, including GSM. Contrary to CEPT, ETSI is not restricted to Administrations, but includes members from industry and user groups as well as operators. GSM had already anticipated this need by officially allowing industry representatives to participate directly to the working parties on an ad hoc basis since 1987

With the transfer to ETSI, most GSM Recommendations were due to become I-ETSs (Interim European Telecommunications Standards). Then ETSs. Becoming an ETS requires several stages of approval. Including public enquiries and voting, and this process takes several months. In the meantime. GSM Recommendations are called "GSM technical Specifications"(GSM TS). In this book we will refer to the set of the GSM Technical Specifications as the Specifications. When integrated in the ETSI structure, the Working Parties became Technical Sub-Committees (STCs. Or "Sub Technical Committees" as the official ETSI terminology stands: GSM itself is a Technical Committee. TC. of ETSI) and are named GSM1, GSM2, GSM3 and ĞSMA4, the Permanent Nucleus became the Project Team n^{0}12 (PT12) of ETSI . FİGURE 1.2 shows the ETSI entities involved in the GSM project in 1990-1991. The date at which the first GSM Technical Specifications were published.

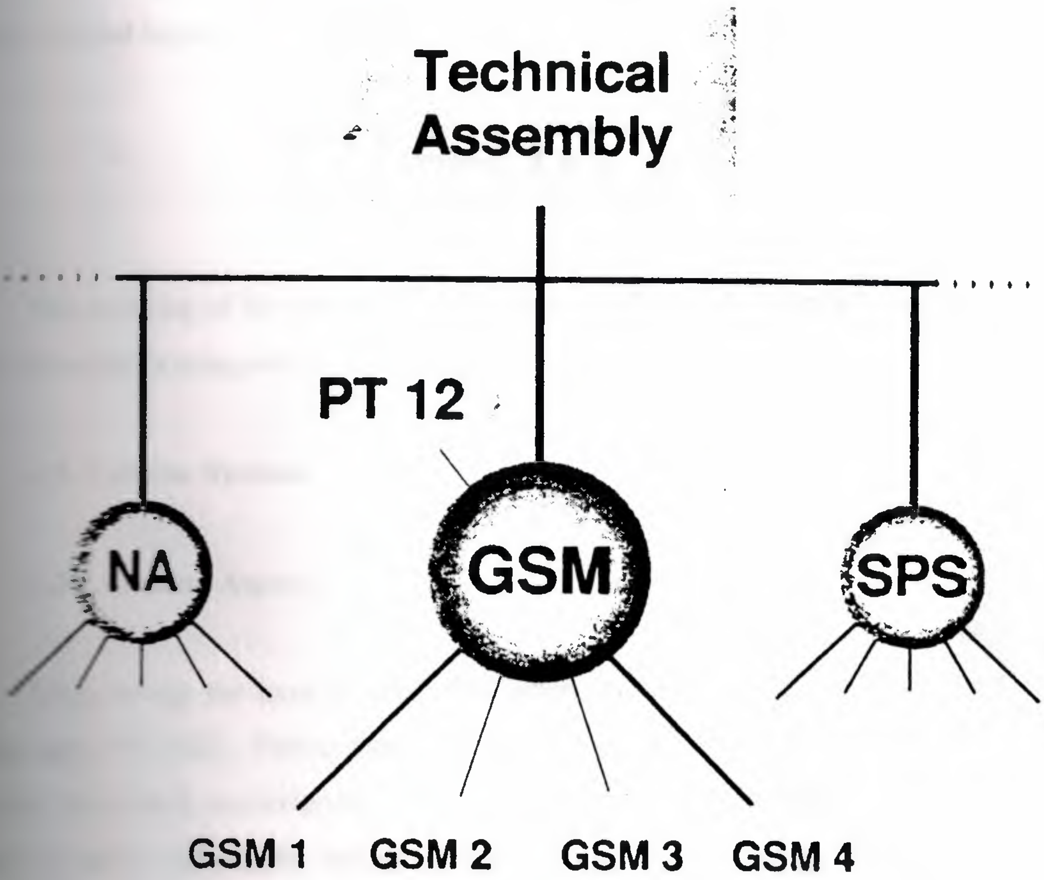


Figure 1.1.1. Partial view of the ETSI organization relevant to GSM

ETSI Technical committes such as GSM, NA (Network Aspects) and SPS (Signaling, Protocols and Switching) report to the Technical Assembly (TA), ETSI Project Team 12 supports the work of Technical Committe GSM)

At the end of 1991, activities concerning the pos-GSM generation of mobile communications were added to the scope of the GSM Technical Committee, which was renamed SMG ("Special Mobile Group"). With Technical Sub-Committee SMG1 to 4 being the same as the previous GSM1 to 4, and SMG5 dedicated to the post-GSM generation, the UMTS ("Universal Mobile Telecommunication System"). A sub-group first established under the responsibility OF THE Permanent nucleus to draft specifications in the area of Operation and Maintenance GSM to SMG to distinguish it from the 900 MHz system: the term GSM is still kept to refer tom the standard and to the corresponding trademark for the European 900 MHz system, meaning in that context "Global System for Mobile communications", with the corresponding logo:

The meaning of the dots in this logo. If any. Is unknown tom the authors and is left to the reader's imagination.

1.2. Cellular Systems

1.2.1. General Aspects

Even though the term of network is often used when speaking about the GSM architecture ("PLMN", Public Land Mobile Network), it would be more proper to consider the system. as designed, as an access toexisting telecommunication networks, aimed at users with GSM mobile stations. A GSM PLMN, as presented by the specifications, is indeed not able to establish autonomously calls other than local ones between mobile subscribers. For all other call configurations, the PLMN relies on fixed networks to route the calls. Most of the time the service provided by a GSM switch hierarchy is defined, and transport of information between GSM machines is based on the usual 64 kbit/s units multiplexed onto 2 Mbit/s links, or higher rate multiplexes. Ho Ever, nothing in the technical choices really limits GSM to be an access network. One can imagine in the future, in some countries without extensive wireline-access infrastructure, that GSM, complemented with a suitable switch hierarchy, would be used as the basic telecommunication network.

Considering GSM as another "local-loop" leads us tom compare its characteristics with wireline telecommunication access, highlighting two major differences;

- First, the wide-area mobility of subscribers leads them to change their point of access to the network: this poses a serious problem for the routing of calls toward subscribers: this is the realm of mobility management:
- Second, the link between the subscriber terminal and the fixed infrastructure is not permanent and is subject to fluctuating transmission requirements: this is the realm of radio resource management.

The consequences of both aspects will now be described, after some preliminary considerations on the cellular coverage.

1.2.2. Cellular Coverage

The major problems with radio distribution arise from electromagnetic wave propagation. With a decreasing weight and an increasing autonomy, the mobile terminals have a limited transmission range. Every telecommunication engineer will remember that the power of radio wives decreases with the inverse of the squared distance (d^{-2}): however, it must be remembered that this applies only in empty space. Between two stations close to the ground, interference, creating reflections from the ground can not be neglected, mand it is very likely that obstacles intervene on the direct path between them. As a consequence, propagation at ground level in an urban environment is more difficult, and the received varies typically with d^{-4}[1]

A second problem is spectrum scarcity: the number of simultaneous radio communications supported by a single fixed station (a base station) is therefore limited.

Cellular coverage allows a high traffic density in a wide area despite both problems at the expense of infrastructure cost and of complexity. Because of the limited transmission range of the terminals, cellular systems are based on a large number of reception and transmission devices on the infrastructure side (the base stations).

Scattered over the area to cover and each one covering a fairly small geographical zone called a cell. The underlying image is the one of a stations allows low-power mobile stations to access the system anywhere within a widearea: they are never very far from a base station. Cells are often represented by hexagons, in order to model the system by paving the plane with a single geometrical figure (see figure 1.2.1).

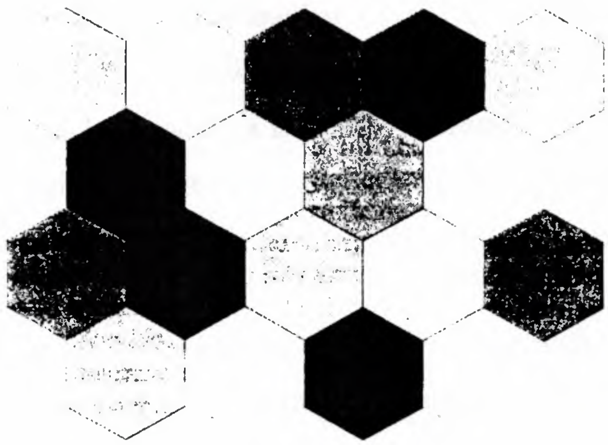


Figure 1.2.1. Cellular coverage representation

Hexagons nicely pave the plane without overlapping and are commonly used for calculating theoretical frequency reuse in cellular systems

Spectrum scarcity is circumvented by the reuse of radio resources. Frequencies used in a given cell are reused a few cells away, at a distance sufficient so the unavoidable interference created by the close use of the same spectrum has fallen to an acceptable level, which depends in particular on the transmission method. This concept of frequency reuse is the key to high capacity. As an example, if the same frequency may be reused in every ninth cell, a spectrum allocation of N frequencies allows N/9 carriers to be used simultaneously in any given cell. The total system throughput, often expressed in number of simultaneous calls Per km^2 Per MHz. Can therefore be increased by reducing the cell size, has some all cell sizes from 0 to infinity. In the of GSM, the design was aimed at the beginning at medium-sized cells, of a diameter expressed in kilometres or tens of kilometres. Yet, the lower boundary is difficult tom determine; cells of more than one kilometre radius should be no problem. Whereas the system may not be fully suitable to cells with a radius below, say. 300meters. One source of limitation is more economical than due to physical laws. The efficiency of the system decreases when cell size is reduced and then the ratio between the expenditure and the traffic increases, and eventually reaches a point where economical considerations call for a halt. Another important point is the capacity of the system to move a communication from one cell to another rapidly, and GSM requires too long a time to prepare such a transfer to cope with fast moving users in very small cells. The cell size upper bound is more obvious: a first, non-absolute, limitation in GSM is a range of 35 kilometres. Cells of bigger sizes are possible but require specially designed cell-site equipment and incur some loss in terms of maximum capacity.

расход, трата
потребление

потерпеть убытки, потери

The number of sites to cover a given area with a given high traffic density, and hence the cost of the infrastructure, is determined directly by the reuse factor and the number of traffic channels that can be extracted from the available spectrum. These two factors are compounded in what is called the spectral efficiency of the system. Not all systems allow the same performance in this domain: they depend in particular on the robustness of the radio transmission scheme against interference, but also on the use of a number of technical tricks such as reducing transmission together with the constraints on the cell size, determines also the possible compromises between the capacity and the cost of the infrastructure. All this explains the importance given to spectral efficiency.

Many technical tricks to improve spectral efficiency were conceived during the systems design and have been introduced in GSM. They increase the complexity, but this is balanced by the economical advantages of a better efficiency. The major points are the following.

- The control of the transmitted power on the radio path aims at minimising the average power broadcast by mobile stations as well as by base stations. Whilst keeping transmission quality above a given threshold. This reduces the level of interface caused to other communications;

- Frequency hopping improves transmission quality at slow speeds through frequency diversity, and improves spectral efficiency through interfere diversity:
- Discontinuous transmission. Whereby transmission is suppressed when possible, allows reduction in the interference level of other communications. Depending on the type of user information transmitted, it is possible to derive the need for effective transmission. In the case of speech, the mechanism called (VAD (Voice Activity Detection) allows transmission requirements to be reduced as an important factor (typically, reduced by half);
- The mobile assisted handover, whereby the mobile station provides measurements concerning neighbouring cells, enables efficient handover decision algorithms aimed at minimising the interference generated by the call (whilst keepeng the transmission quality above some threshold).

1.3. GSM Functionalities

In this section we will describe the functions of GSM as seen by the users, that is to say the services that are provided to users, abstracting the details of how it is done.

1.3.1. GSM: A Multiservice System For The User.

GSM is a multiservice system, allowing communications of various types, depending on the nature of the transmitted information as perceived by the end users. By tradition. One distinguishes speech services from data services: in speech services, the information is voice. Whereas the term "data services" groups everything else, such as text. Images, facsimile, computer files, messages, and so on. GSM provides a large palette of the services offered to fixed telecommunication users, as will be described further on in this section. GSM provides in addition a non-traditional set of services. The "short message services". Closer to the paging services (one-way radio messaging services) than to any service provided in fixed networks. These services will be described separately from the other data services.

More generally, the definition of a telecommunication service includes more than just the nature of transported information. Other characteristics of the communications are also relevant, such as the transmission configuration (point-to-point or point-to-multipoint, half or full-duplex) and the potential partners. But also the possible "supplementary services", which refer to the possible control of varies aspects of the communications by the user, and more administrative nations such as for example the charging aspects.

Service provision depends on three independent factor;

- the contents of the subscription held bay the subscriber, in terms of services as well as in terms of geographical areas. The subscription packages offered by each operator will vary, and well as the corresponding subscription rates, and each subscriber will perform a choice between these packages:
- The capabilities of the network from which the user is getting service. All networks will not offer exactly the same range of services at a given date, and therefore the user might find some restrictions on the available services depending on the network in which he is currently roaming:
- The capabilities of the equipment held by the user. For instance, it is obvious that faxes cannot be sent or received on a speech only mobile station if it is not connected to a proper fax machine.

The basic services provided by a telecommunication network. Which consist in transmission media and in the means to set up calls, are distinguished from a number of supplementary features enabling users to better control the provision of these basic services, or to simplifying the daily use of telecommunications. These features will represent an important contribution to the user comport, and include the possibility for the user to forward calls, to visualise the calling number etc. Some of these features are performed locally by the terminal, and are not properly speaking provided by the network (through the limit is difficult to tell for the user). Other features are provided by the network, and are called the "supplementary services". Several reasons warrant the distinction between basic and supplementary services , including the fact that supplementary services apply generally to several basic service (and hence a separate presentation is better than repetition): and also because in many existing networks these supplementary services must be asked and paid for in addition to the basic, services in the future it is quite likely that some ofl these supplementary features will be automatically pat of a service package. This statement holds in particular for those services which improve network efficiency (such as the forwarding of calls which cannot be completed).

The description of the services provided by GSM following . We will first address the speech services, the data services and the short message services. Then the "supplementary services" will be considered. Afterwards, a small section will deal with the "local services". i.e., the facilities offered by the GSM terminals on their own. Then an important particularity of the GSM mobile terminal will be described: the Subscriber Identity Module (SIM), Finally, the security features of GSM. Another complement of service for the users; will be addressed.

1.5.1.1. Speech Services

The most important service provide by GSM is telephony. This service enables bi-directional speech calls to be placed between GSM users and any telephone network. Fixed telephone subscribers world-wide can be reached as well as mobile network subscribers or subscribers of specific networks connected to a public telephone network. With the rise of ISDN telephony has been somewhat eclipsed by data services, which are often presented as the future of telecommunications. However, speech remains and will remain the most important service for mobile systems. The absence of a fixed wire and the arrival of handheld terminals (less than 400 g. And fit to be carried in a pocket9 make cellular telephony the foremost technique for inter-personal communications. The true Tele-communications that human beings have used for maybe several hundred millennia.

Following the GSM official terminology, emergency calling is a distinct service, derived from telephony. It allows the user of a mobile station to reach a nearby emergency service (such as police or the fire brigade) through a simple and united [illegible] by dialling 112 (the number which h has been agreed as the standard emergency number throughout Europe).

Another service derived from telephony is voice messaging. The Specifications do not identify this service as a separate one. But many an operator will offer it as a basic feature. It enables a voice message to be stored for later retrievel by the mobile recipient. Either because he was not reachable at the time of the call, or even because the calling party chose to access directly the voice mailbox of the GSM subscriber.

1.3.1.2. Data Services

As an heir of ISDN, GSM has been designed from the start to offer many data services. Basically, most services which are provided to fixed telephony users and to ISDN users (thus covering access to other more specialised networks such as Packet Switched Public Data Networks) Have been included. As far as the limitations due to radio transmission allow. This corresponds to a very large number of different cases (the Specifications list some 35 services). To cope with all the variants stemming from the history of telecommunications.

Data services are distinguished mainly by the potential correspondents (users of the telephone network, of ISDN, or of specialised networks), by the nature of the end-to-end information flow drive (raw data, facsimile, videotext, teletex.....) by the nature of the terminal, and so on.

For practical reasons, data services shall not be presented here with the full descriptive methodology used in the Specifications (i.e.. as tables of attribute values). They shall be grouped according primarily to the type of potential correspondents, that is firstly according to the type of network to which the other correspondent is a subscriber and a summary of all data services as specified in the Specifications will then be given.

1.3.1.3. Short Message Services

The different data services listed in the previous section are not really adapted to the mobile environment. They are simply extensions to GSM subscribers of the services available to fixed subscribers. One of the problems is that these data services normally require rather bulky terminals, compared to the size of a handheld . They are then suited to semi-fixed usage, such as temporary installations. Or for use with vehicle mounted terminals (e.g..fax in the car). But none of the data services is totally fit for an integrated implementation in the handheld case (they require a complete computer keyboard and a comfortable-sized screen).

Still some data services are of interest for persons on the move and not desiring to be encumbered with a bulky terminal. An example of such services is the paging service, where simple messages. A few tens of octets long. Can be received on a very small terminal. Since its designers recognised this need. GSM was designed to support such a service, in order to spare GSM subscribers the trouble of carrying two terminals one for speech, the other for paging.

GSM enables the transmission of point-to-points short message, and distinguishes between the “Mobile Terminating Short Message Service, Point to Point” (SMS-MT/PP), for the reception of short message and the “Mobile Originating Short Message Service, Point to Point” (SMS-MO/PP), enables for instance a GSM user to send such a message to another GSM user. Another short message service is the “Cell Broadcast Short Message Service” (SMS-CB), enabling short messages of a general nature to be broadcast at regular intervals to all subscribers in a given geographical area. Let us examine in more details what is provided, by taking the examine in more details what is provided, by taking the example of a GSM user called Kevin.

1.3.1.4. Point-to-Point Short Message

These services enable alphanumerical messages of several tens of message is usually displayed on Kevin’s GSM terminals upon reception. The service provided by

SMS is akin to paging, but with many enhancements, which make use of the other capabilities of GSM, in particular the possibility to have a bi-directional dialogue between the mobile station and the network (paging services are based on unidirectional links. For instance, the network is informed whether the message has been received or not by the mobile station; and messages can therefore be kept in the network in cases of delivery failure, and repeated once Kevin is known to be reachable. Another improvement is that the originator of the message can be advised of the outcome, which was not the case in standard purging systems. On the mobile station side, the last received messages can be stored in a non-volatile memory.

The way the message is sent from original sender to the PLMN is operator dependent. This service is not an extension of some fixed network service, and as a result does not correspond to any standardisation prior to GSM. Generating methods as widespread as keying on a dual-tone multi-frenquency phone allow many PSTN subscribers to access the service; however, the message contents are in that cases restricted to numerical characters. Other access possibilities include simple videotext terminals such as the French Minitel. Of course, any computer accessible through a packet data network would do, but access to the service would not be so popular! Delivery through a human operator could also be envisaged. But is not very cost effective for the PLMN operator. In the same field, voice recognition machines may be a futuristic solution to obtain a widespread access through the PSTN.

The way the message is sent from the original sender to the PLMN is operator dependent. This service is not an extension of some fixed network service. And as a result does not correspond to any standardisation prior to GSM. Generating methods as widespread as keying on a dual-tone service: however, the message contents are in that case restricted to numerical characters. Other access possibilities include simple videotext terminals such as the French Minitel. Of course, any computer accessible through a packet data network would do. But any computer access to the service would not be so popular! Delivery through a human operator could also be envisaged. But is not very cost effective for the PLMN operator. In the same field. Voice recognition machines may be a futuristic solution to obtain a widespread access through the PSTN.

GSM subscribers are luckier than paging subscribers, since the Mobile Originating Point-to-Point short message service enables them to send short alphanumerical messages to other parties. For the time being. These other parties are not precisely defined. They include certainly the other GSM users (which will receive them according to the SMS-MT/PP service). But these other parties could also be subscribers in a paging system, an electronic mailbox or alternatively one could foresee gateway devices transforming a short message into, e.g.. a facsimile or any kind of format suitable for the equipment of the recipient. Interwoking between GSM PLMNs will certainly be possible.

2. ARCHITECTURE

GSM as a modern telecommunication system is a complex object. To the multi-service aspect it shares with ISDN. It adds all the difficulties coming from cellular networks. As such, its specification. Its implementation and its operation are no simple tasks. Neither is its description. It the course of the specification of GSM, much effort was expended to sort out this complexity, In this book. We have built upon this foundation so as to fallow a structured approach for the presentation of the system.

The purpose of the chapter is first to present this structure, both of the system and of the book. GSM will be analysed in terms of subsystems and the main machines and functional domains will be identified. This is the first step of the system description. Of which the subsequent chapters can be considered as refinements.

Another aim of this chapter is to present the structuring method. Which is of interest in itself. The question of how to deal with complex systems has been encountered in many different domains such as electronics, biology, ecology, economics, and of course telecommunications. Where the Open System Interconnection (OSI) model for data networks represents one step towards a structured approach of complex meshed networks. While the authors do not claim mastery of this field. They think that the study of the structure of a concrete system like GSM and the general study of systems can be mutually helpful. Hence this chapter will introduce concepts such as functional planes, interfaces and protocols.

A last, but important, aspect of this chapter is to introduce many terms of the GSM jargon. These terms and abbreviations are used heavily in the specifications. In the GSM literature in general, and in the rest of this book. Though we try to limit the use of the technical jargon, it is unavoidable: when a concept or an object has to be referred to every third line, it is best to give it a name.

2.8.1. The Three Description Axes

From one point of view a telecommunications system is a collection of electronic boards transferring analogue electrical signals than this "reductions" approach, equating the system with electromagnetic field values or transistor states. The opposite viewpoint would consist in looking at the system as a black box, been only through its interfaces with the external world. Though not satisfactory either, this approach is helpful in raising two fundamental questions:

- What is part of the system (here, GSM) and what is not?
- What does GSM interface with?

These two questions will be answered in turn. But it would be somewhat frustrating to stop there, only looking at GSM from the outside. A middle way has been sought in this book between this black box approach and sheer reductionism. On this route, the system will be divided into pieces; each of these pieces will be described as a black box in itself, and the overall operation of the system will be presented as the interactions between the pieces. This approach results in a description of GSM as a set of interconnected co-operating sub-systems.

But, as already expressed. GSM is more than a concatenation of sub-systems; some areas involve many pieces of equipment and cannot be described satisfactorily by looking at each sub-system independently There fore, we must additionally look at how GSM operates from two different viewpoint: a static one and a dynamic one.

The static viewpoint enables us to identify and describe several functions which are fulfilled through the co-operation of several machines. The term machine is used here to refer to an assembly of interconnected system components, physically close to each other, working together to perform identifiable task. The term function is of the used in the technical literature (and in the Specification) to refer to some abstract machine. The use here is closer to the basic meaning of the word. A function is something to fulfil, an activity, in a sound architecture, a machine corresponds to some function. However other structures should.

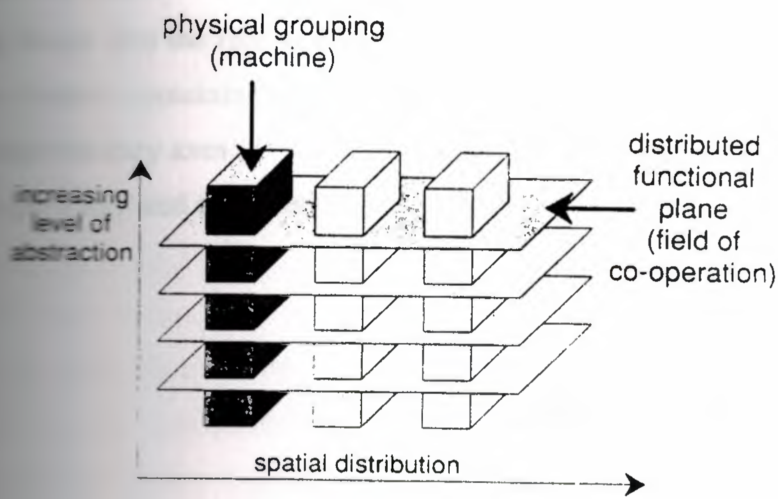


Figure 2.0.1. Two-dimensional views of a network

Physical groupings (machines or entities) are represented by vertical blocks, whereas co-operating functions are grouped in horizontal layers each one corresponding to a different functional domain

also be identified, which group similar functions in a system. These functional groupings will put together elementary functions, possibly from different machines, which fulfil by co-operation the same goal. The usual representation, such as used in the layering method of the Open System Interconnection model, consists in showing functional groupings over several machines as "horizontal" layers, a machine being a "vertical" structure in this representation, as shown in figure 2.0.1.

The dynamic view consists in looking at the events affecting the system. Events happen at many scales, from microseconds for transmission aspect. To years when one considers the deployment of a network. The description of these events, their organisation and of the way they trigger other events inturn is very important in understanding how the system performs its functions.

The GSM architecture, which is the subject of this chapter, will first be described in terms of machines, then through a functional layer view. The subsequent chapters (i.e. the bulk of the book) will be devoted to the study functional planes in detail, going deeper into the role of each machine within each plane, with a substantial part of these chapters containing a description of event sequences. Having thus covered the three complementary axes (as shown in figure 2.0.2.), the description should enable the reader to get a full- and consistent- picture of GSM

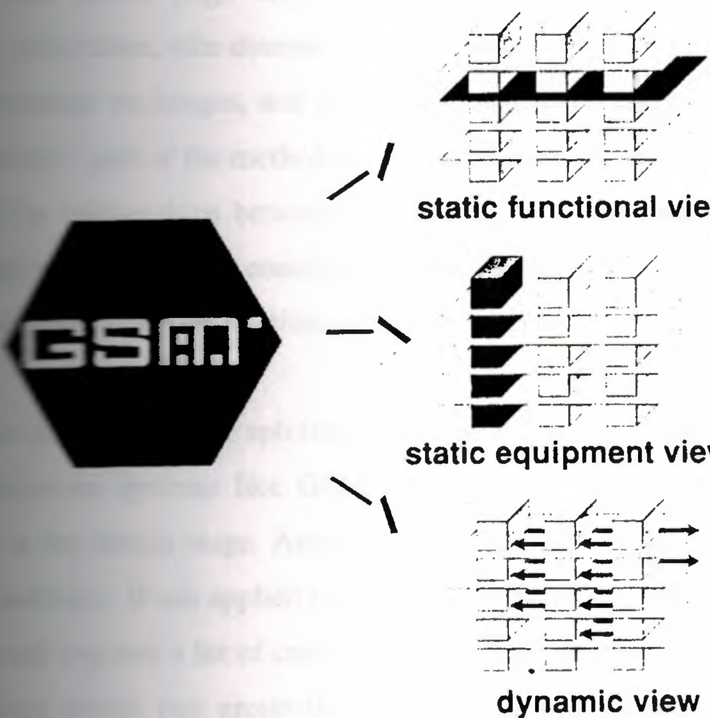


Figure 2.0.2. The three axes of the description

GSM functions can be described along several axes each one from a different and complementary viewpoint

Now, although this three-axes concept helps in structuring the description of the system, it is not sufficient to tackle the overall complexity. The number of possible functional groupings (that is to say machines or abstract portions of machines) , or of event sequences, is still rather big. The second trick we will use is a recursive description, or top-down approach. Taking the horizontal axis as a first example, this approach consists of describing the system in a few (less than 5 when possible) sub-

[illegible] analysing the interactions at this level whilst taking each subsystem as a black [illegible] applying this method for each subsystem. Similarly, functional planes will [illegible] of bus-palens, themselves composed of sub-sub-planes, and so on. [illegible] for the temporal sequences, which can be analysed from the large scale down [illegible] small scale. This allows us to identify first general, compound events, of [illegible] duration (e.g.. system deployment, a site installation, a communication, [illegible] into finer events (e.g.. the call set-up, its release), then stepping up the [illegible] (e.g. procedures, like channel allocation, start of ciphering), and so on down [illegible] message exchanges, and finally to transmission and bit modulation. This [illegible] "reductionism" part of the methodology, which must be balanced at each step by [illegible] of the relationships between the sub-parts, in order not to lose the large-[illegible]. This will be helped considerably by the three-axes approach, each axis [illegible] relationships between entities along the other axes.

[illegible] this book, this recursive splitting is used as a descriptive approach. In the case [illegible] telecommunications systems like GSM this approach is also used, more or less [illegible], at the design stage. Another application, which we will see in the last [illegible] the software. When applied to description, but even more so when applied to [illegible] the method requires a lot of care for the location of each split. The that there is [illegible] interworking within one group than between different groups. When correctly [illegible]. This method leads to a sound implementation. Some effort has indeed been [illegible] in the GSM standardisation bodies to obtain a clear-cut architectural model, [illegible] will be used in this book.

The architecture of GSM described in the Specifications will be referred to as [illegible] "canonical" architecture. This architecture does not describe the actual machines [illegible] will be found in a GSM implementation. First because freedom has been left to [illegible] in terms of physical grouping choices, second because the Specifications [illegible] only a small part of the specifications of an actual machine. The canonical [illegible] can be seen as the description of a network model, serving as a template for [illegible] implementation. However, the interfaces, that is to say what happens [illegible] canonical machines, are described and specified without abstraction. The manufacturers consists most often in the possibility to group canonical machine (in which case canonical interfaces disappear), or to split an distinct, possibly distant, machines (thus creating "proprietary" – specified – interfaces). Throughout this book. We will describe the system of the canonical architecture; additional information will be given on the choices left to manufacturers

groupings and their borders are two sides of the same coin. In fact, machines are extremely important in a system such as GSM, since the specify in fact mainly the behaviour of the system as seen on interfaces and not the internal working of these machines (though this can be extent from the external behaviour9. It is then an important function model to define the system interfaces, and the last section of this devoted to this aspect. In this book, an interface represen8t the frontier machines which are in contact via a transmission medium. It should be Specifications use in some places a wider meaning for this term, as for some places a wider meaning for this term , as for instance when MSC to HLR interface (where the "interface" may well include a full network). In the following, both the machines and the interfaces stage will be indicated, with a particular emphasis on the interfaces for exchange of information is specified in the Specifications.

between functional layers can also be described in terms of these interactions happen inside machines. They are not to be a detailed and formal description of primitives between layers). In description takes up a substantial portion of the Specifications, and is quest for as little ambiguous specification as possible. It should be model, which does not constrain implementation in any way, though it guideline. In this book we will not refer to this formalism any more.

time to apply these fine principles, and we will start by the first first along he vertical axis (machines), then across the horizontal axis

This analysis is then refined in the rest of the book. Whereas the main subject of the signalling chapters.

Frontiers Of The System: Where Are The Borders Of GSM?

The first task is to identify the extent of GSM itself as a whole, thereby also external interfaces, i.e. its interfaces with the rest of the world.

Figure 2.0.3. External interfaces of GSM

a black box, GSM shares borders with three major realms; users of the service through their mobile station other telecommunication networks calls transit and the operating company controlling the GSM domain

at from a distance. GSM is part of the Global Telecommunications a part of the human organisation. As such. GSM is in direct contact services through which are being provided with other network (e.g.. the global telephony network) and with the operating companies. These are indeed the three main external GSM , as shown in figure 2.0.2.

Other interfaces with the external world exist, such as the contact of machines with ground and power supplies(which we may term environmental aspects) as well with other systems using the radio spectrum (electromagnetic compatibility - EMC – aspects). These pragmatic aspects, which are far from negligible for manufacturers and operators, if not for users, are not directly related to the provision of telecommunication services and will be dealt with here. Let us now look at the three main border lines, with respectively users other telecommunication networks and operators

On the user side the limit lies somewhere between the user himself, who can be excluded from GSM, and the radio interface which represents the principal part of the specifications. But is everything between the user and the radio interface part of GSM? The mobile station are only partly specified by the Specifications: an example is that terminal equipment functional entities similar to those defined in ISDN are not defined specifically for GSM; another example is the manmachine interface of mobile stations, which is in no way specified in a binding manner in the Specifications, and could include functions which have nothing to do with GSM. The point of contact between GSM and the user lies therefore somewhere inside the mobile station. We will come back to this point when addressing the internal mobile station architecture.

Now for the interface with other telecommunications networks. GSM is specified mainly as an access network, enabling the setting up of calls between GSM subscribers and subscribers to other telecommunication networks. For practical reasons, machines belonging to GSM are most often kept separate from machines belonging to [illegible] this comes from the division, which is now the general rule between [illegible] and the fixed PSTN/ISDN operator(s) even in companies of marked [illegible] origin such as FRANCE TELECOM or DEUTSCHE BUNDES[illegible]. Other choices exist: one could imagine telecommunication switches [illegible] GSM functions as well as managing PSTN/ISDN subscribers. This is not [illegible] the Specifications, but the canonical GSM architecture does not consider [illegible], and the interface between GSM and other telecommunication networks [illegible] defined.

The interface includes three aspects. The first and major one concerns the point where communications transit between GSM machine (a switching exchange) at the corresponding contact point is referred to as a "Mobile services Switching Centre" (MSC). The second aspect of interfacing with fixed networks concerns the provision of some telecommunications transport between GSM machines. In many instances, regulations do not allow GSM operators to operate the terrestrial links between their machines, often leaving them no other choice but leasing lines from a fixed network of operator. This second point will be ignored, since in practice it has no effect on the functioning of GSM. A third aspect of interconnection with fixed networks, which the cases of using external networks as support for GSM functions concerns the routing of non-call related signalling between different GSM networks. For instance the management of location data for GSM users roaming abroad, requires an exchange of data with no direct relation with given calls to be transmitted through an international signalling network. There again, GSM operators of many countries are forbidden to access the international signalling network directly, and must therefore interface with a fixed network operator in their country to transmit this signalling information between GSM entities (MSCs, location registers) of different networks.

Although the point of contact (at least for the transit of calls) is well defined, the interfaces between GSM and external networks are not specified in the Specifications. CCITT recommendations, as well as ETSI standards based on them, include such specifications, as well as ETSI standards based on them, include such specifications, but usually national variants exist. Fixed network operators have their networks must be customised to meet the exact interface requirements specific to each network. In this book, as in the Specifications, the basis for description will be the CCITT Recommendations related to ISDN or telephony.

Now comes the border between GSM and operator personnel. As in the case of the user interface, the border is basically between machines and the human employees of operators. The set of machines intervening between this boundary and the machines handling the telecommunications traffic (as well as some parts of these machines) are generally referred to as the Operation and Maintenance Sub-system (OSS). It includes various entities such as workstations (or terminals) handling the man-machine interface with operator personnel, and dedicated computers managing a number of tasks required for operation and maintenance of the system, as well as parts of the software of traffic handling machines themselves.

Most of the OSS aspects are not specified by the Specifications. Only a small part of the interfaces between traffic handling entities defined in the GSM architecture is related to OSS functions, and interfaces between these entities and OSS machines are only partially specified. However these entities and OSS area must be only partially specified. However, the whole OSS area must be considered part of GSM –and this would indeed be the opinion of many an operator - since what would a complex system such as GSM be without means to drive and maintain ?

2.1 Sub-Systems

This section will deal with each subsystem in turn. The main purpose is to introduce a number of terms, not to exhaust the subject. The level of detail here is just sufficient to get a general idea of the functional splits. More complete architectural considerations will be found in each of the chapters of the book, where a more thorough description of the required functions will be found.

2.1.1. The Mobile Station (MS)

The mobile station usually represents the only equipment the user ever sees from the whole system. Examples taken from the first types of GSM mobile stations to be on the market are shown in figures 2.1.1 and 2.1.2. Mobile station types include not only vehicle-mounted and portable equipment, but also handheld stations, which will probably make up most of the market.

Figure 2.1.1. A GSM portable mobile station

The first GSM mobile stations on the market were portable mobile stations weighing around 2 kilograms, and also capable of being installed in a vehicle

The first GSM mobile stations on the market were portable mobile stations. Weighing around 2 kilograms, and also capable of being installed in a vehicle.

But what does a mobile station involve? Beside generic radio and processing functions to access the network through the radio interface, a mobile station must offer either an interface to the human user (such as a microphone, loudspeaker, display and keyboard for the management of speech calls), or an interface to some other terminal equipment (such as an interface towards a personal computer or a facsimile machine), or [illegible] them. An effort has been made to allow off-the-shelf terminal facsimile machines designed for connection to the telephone network), and specific terminal adaptation functions have been specified for his purpose. However, all implementation variants are possible and left open to manufacturers, enabling fully integrated compact mobile stations to coexist with mobile stations featuring standard interfaces.

Figure 2.1.2. A GSM handheld mobile station

With their ever-decreasing weight and volume
handheld mobile stations represent a very attractive product for the user

This leads to the identification of three main functions, as shown in figure 2.1.3.:

- The terminal equipment, carrying out functions specific to the service, without any GSM –specific functions: e.g., a fax machine:
- *The mobile termination, carrying out, among others, all functions related to transmission on the radio interface;*
- Possibly a terminal adapter, which acts as a gateway between the terminal and the mobile termination. A terminal adapter is introduced when the

external interface of the mobile termination fallows the ISDN standard for a terminal installation, and the terminal equipment has a terminal-to-modem interface.

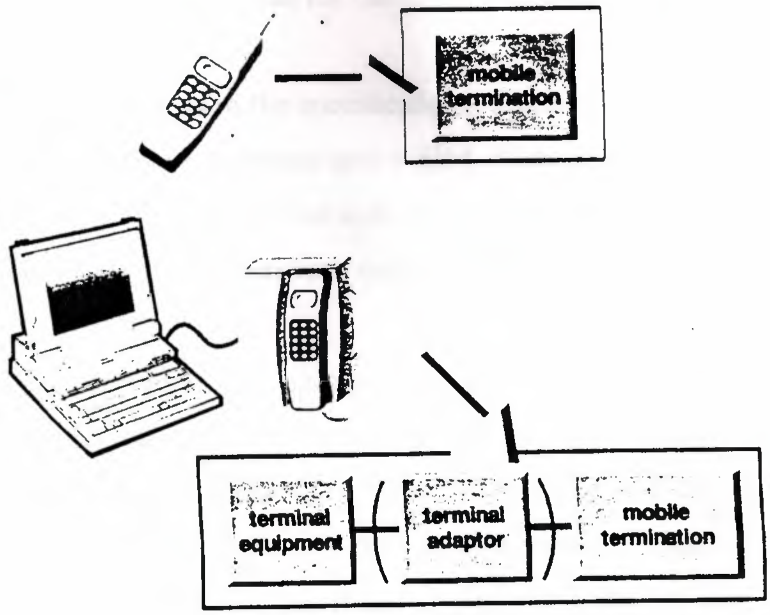


Figure 2.1.3. Mobile station functional architecture

The mobile station may be a standalone equipment for certain services or support the connection of external terminals
Either directly or through relevant adaptation functions

The function split between split between mobile termination, terminal adapter and terminal equipment is very much related to the transmission needs of each service, and will be detailed in Chapter3.

GSM another more significant, architectural aspect of the mobile station relates to the concept of subscriber module, or SIM (Subscriber Identity Module, a slightly restrictive name, as more than identity is involved). As described in Chapter 1. The SIM is basically a smart card (or a cut-out thereof), following ISO standards. Containing all the subscriber-related information stared on the user's side of the radio interface. Its [illegible] besides this information storage capability, relate also to the [illegible] area. The rest of the mobile station contains all the generic transmission [illegible] means to access the network. The interface between the and the rest of [illegible] is fully specified in the specifications, and is simply referred to as the [illegible] INTERFACE (ME stands for "mobile equipment").

[illegible] book as well as in the specifications. The term mobile station (MS) shall [illegible] a Mobile Equipment and a SIM. Although a rare case exist where a [illegible] could be opared without a SIM (i.e.., reduced to the mobile equipment) [illegible] of anonymous emergency calls when so parmitted by the network.

[illegible] of a removable storage device for subscriber data has far-reaching [illegible]. In previous cellular systems, except for the German C-network which [illegible] card concept at the time when it was making its way in GSM [illegible]. The personalisation of the mobile station required a non-trivial [illegible]. Only possible for technical specialists and not for the operator's [illegible]. This situation lead to several drawbacks. A mobile station could [illegible] specialist dealers, able not only to install the equipment in a vehicle. But [illegible] intermediary between the user and the service provider to personalise [illegible]. Should the mobile station fail (unfortunately not such a rareevent). It [illegible] provide the user with a replacement during the repair period, and almost [illegible] allow the user to keep the same directory number during this time.

[illegible] removable SIM simplifies these issues, and also brings other benefit. A [illegible] of course buy a mobile equipment. But the may also lease or borrow [illegible] of time, and change it as he wishes without a lot of administration. All [illegible] his own SIM. Obtained through an operator or a service provider, [illegible] of any equipment choice. The last steps of the SIM personalisation can [illegible] through a small computer and a simple adapter. Mobile equipment will [illegible] much larger-scale than ever before, since their acquisition will not [illegible] intervention of an operator or a service provider. Car phones will still require installed in the vehicle, but portables or handholds will encourage users to buy their mobile equipment from any store.

Other advantages can be envisaged. For instance, rented cars could be equipped with mobile equipment usable with any SIM , whether userowned or also rented. The separation may also bring benefit to subscribers, if not to operators; a subscriber can change his serving operator without replacing his ME. But most of all as explained in Chapter 1. This personal chip secured in its plastic case and called SIM is the first step in the building of a personal communication system enabling wide-ranging roaming between different telecommunications networks.

2.2.2. The Base Station Sub-System (BSS)

Largely speaking. The Base Station Subsystem groups. The infrastructure equipment which are specific tom the radio cellular aspects of GSM . The BSS is in direct contact with mobile stations through the radio interface. as such it includes the equipment in charge of transmission and reception on the radio path, and the management thereof. On the other side. The BSS is in contact with the switches of the NSS. The role of the BSS has to be controlled and is thus also in contact with the OSS, The external interfaces of the BSS are summarised in figure 2.1.4.

According to the canonical GSM architecture, the BSS includes two types of equipment, the BTS (Base Transceiver Station). In contact with the mobile stations through the radio interface, and the BSC (Base Station Controller) the latter being in contact with the switches of the NSS. The functional split is basically between a transmission equipment, the BTS, and a managing equipment, the BSC.

A BTS comprises radio transmission and reception devices, up to and including the antennas, and also all the signal processing specific to the radio interface. BTSs can be considered as complex radio modems, and have little other function. A typical first-generation BTS consists of a cabinet (2m high and 80 cm wide) containing all electronic devices necessary for the transmission functions, as shown in figure 2.1.5 for a GSM900 BTS and figure 2.10 for a DCS1800 BTS . The antennas are usually a few tens of figure 2.10 for a DCI1800 BTS. The antennas are usually a few tens of meters away, on a mast, and the racks are connected to it through a feeder cable. A one-racak first-generation BTS is typically able to handle three to five radio carriers. Carrying between 20 and 40 simultaneous communications. Reducing the BTS volume is important to keep down the cost of the cell sites, and progress can be expected in this area.

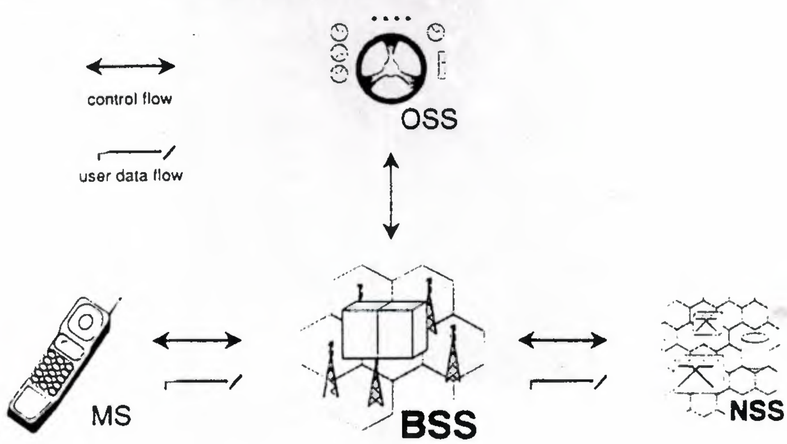


Figure 2.1.4. The external environment of the BSS

The BSS bridges the space between the mobile stations on one side (through the radio interface), and the switching functions on the other.
It is controlled by the operator through the OSS

An important component of the BSS which is considered in the canonical GSM architecture as a part of the BTS. Is the TRAU or Transcoder/ Rate Adapter Unit. The TRAU is equipment in which the GSM-specific speech encoding and decoding is carried out, as well as the rate adaptation in case of data. Although the Specifications consider the

TRAU as a sub-part of the BTS, it can be sited away from the BTS, and even more so since in many cases it is actually between the BSC and the MSC. Its remote position allows the advantage of more compressed transmission between BTS and TRAU, and its impact will be discussed in detail in Chapter 3.

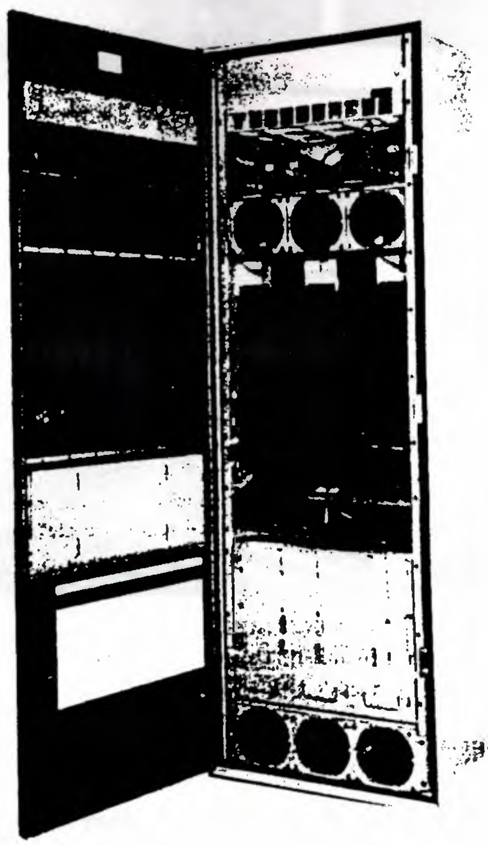


Figure 2.1.5. A GSM BTS

A one-rack BTS such as the one shown, is typically able to handle up to 5 carriers. The picture shows the rack equipped for 3 carriers.

BSC the internal structure of the BSS is represented in figure 2.1.6. On top of the BTS, t shown the second "canonical" component of the BSS, the BSC, the BSC is in charge of all the radio interface management through the remote command of the BTS and the MS. Mainly the allocation and release of radio channels and the handover management, the BSC is connected, on one side, to several BTSs and on the other side, to the NSS(more exactly to an MSC).

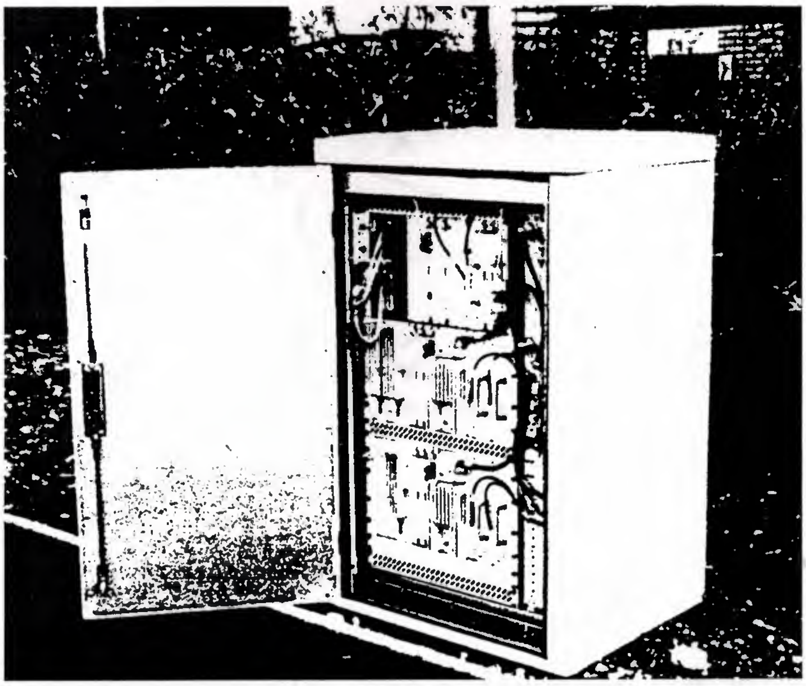


Figure 2.1.6. A DCS 1800 BTS

The small rack shown is to be used outside
Typically below the antenna mast

A BSC is in fact a small switch with a substantial computational capability. Its main roles are the management of the channels on the radio interface, and of the handovers. A typical BSC consists of one or two racks, as shown in figure 2.1.7. and can manage up to some tens of BTSs depending on their traffic capacity.

The concept of the interface between BSC and MSC called the A interface. Was *introduced fairly early in the GSM standard elaboration process. Only later was it* [illegible] standardise the interface between BTS and BSC, AND this interface [illegible] the (not any more meaningful than A!) name of "Abis" interface.

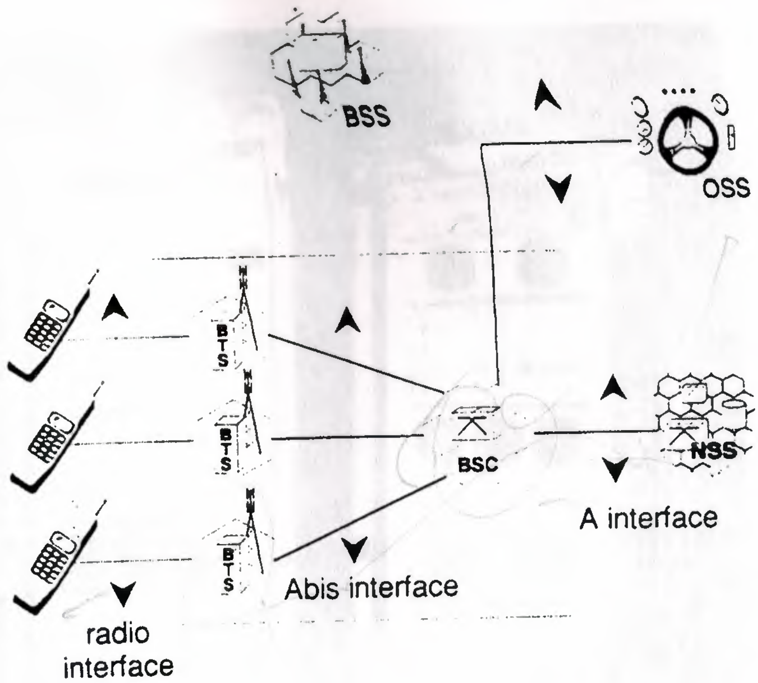


Figure 2.1.7. BSS components and interfaces

The base Sub-system consists of BTSs, situated on the antenna sites, and of BSCs, each one in control of several BTSs

In the GSM vocabulary, a BSS means the set of one BSC and all the BSTs under its control, not to be confused with the BSS as the subsystem including all the BSCs an BTSs.

2.1.3. The Network And Switching Sub-System (NSS)

The Network and Switching SubSystem or Nss, includes the main switching functions of GSM, as well as the data bases needed for subscriber data and mobility management. It is also sometimes called the switching sub-system. Which is indeed more appropriate since a GSM network includes the BSS as well as the NSS. The main role of the NSS is to manage the communications between the GSM users and other telecommunications network users.

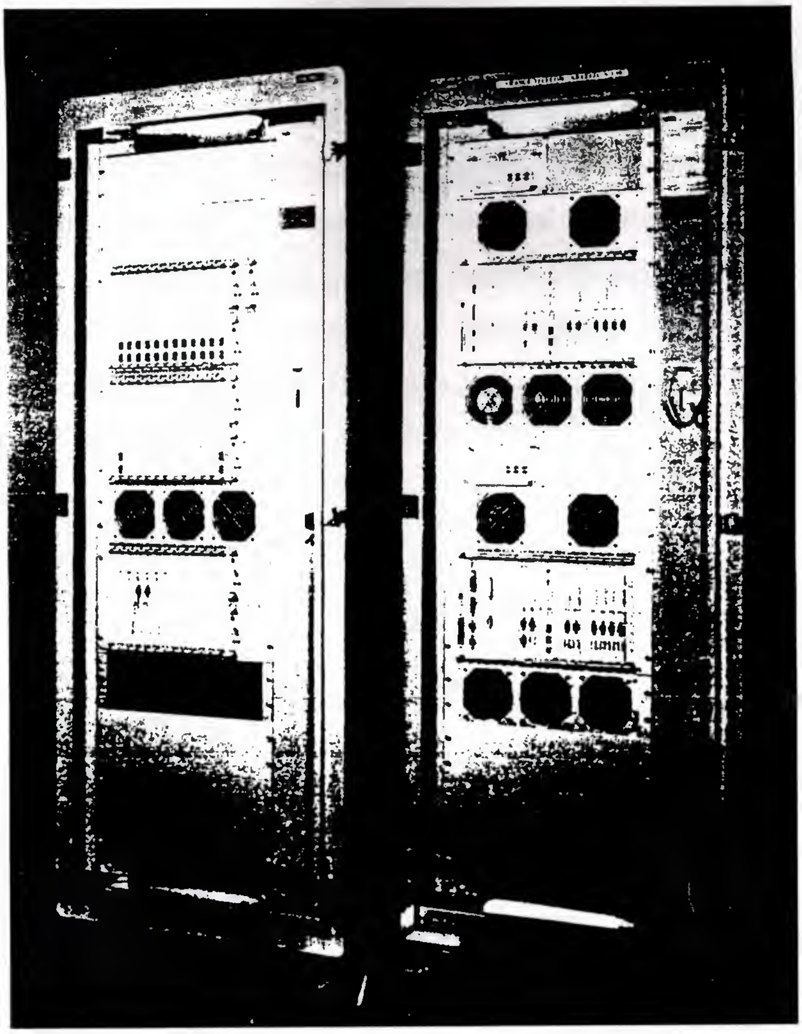


Figure 2.1.8. A GSM BSC

The BSC shown here consistes of two cabinets
A control cabinet holding the duplicated central control and switching functions, and a cabinet handling the interfaces

Within the NSS, the basic switching function is performed by the MSC (Mobile Services Switching Centre), whose main function is to co-ordinate the setting-up of calls to and from GSM users. The MSC has interfaces with the BSS on one side (through which it is in contact with GSM users), and with external networks on the other. The interface with external networks for communication with users outside GSM may require a gateway for

Interworking Functions, or IWF). The role of which may be more or less important depending on the type of user data and the network it interfaces with. The NSS also needs to interface external networks to make use of their capability to transport user data or signalling between GSM entities. In particular, the NSS makes use of a signalling support network at least partly external to GSM following the CCITT Signalling System no 7 protocols (and therefore usually referred to as the SS7 network): this signalling network enables co-operative interworking between NSS machines within one or several GSM networks. The external interfaces of the NSS are represented schematically in figure 2.1.8.

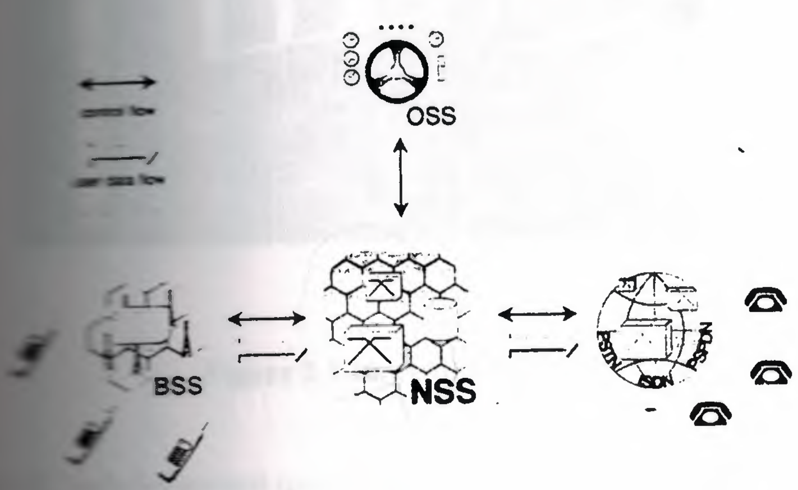


Figure 2.1.9. The external environment of the NSS

The NSS, through its MSC, is in contact with the BSS and external networks.

Like the BSS, it is controlled by the operator through the OSS

NSS entities are also in contact with NSS entities of other GSM networks for the exchange of data through SS7 signalling networks

In terms of equipment, an MSC controls a few BSCs and is usually a rather big switching machine. With a medium population penetration percentage , a typical MSC at the time of writing is suitable for covering a regional capital and its surroundings, totalling say 1 million inhabitants. Such an MSC includes about half a dozen racks. Figure 2.1.9. shows a GSM MSC.

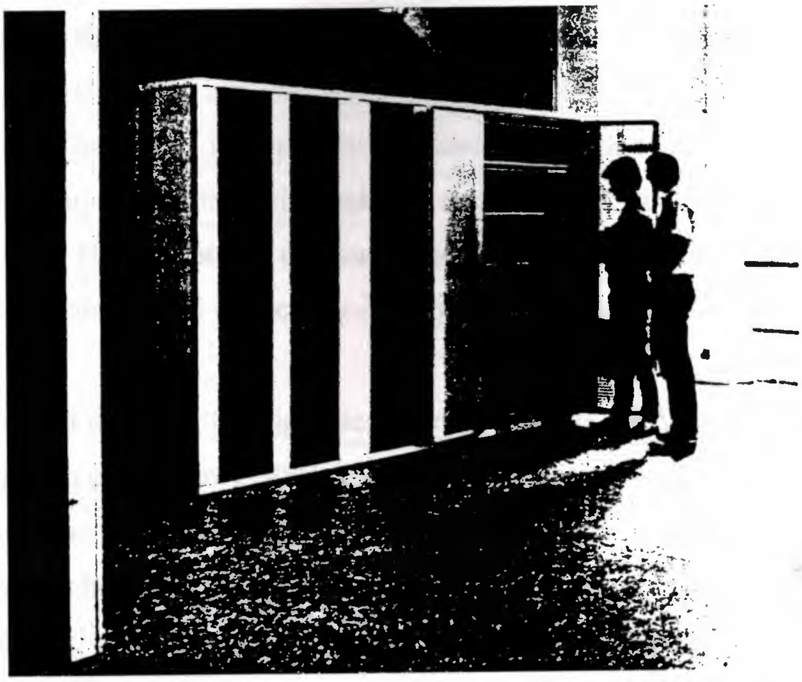


Figure 2.1.10. A GSM MSC

An MSC includes several rows of cabinets, on of which is shown here

The interconnection of the MSC with certain networks requires the adaptation of the GSM transmission peculiarities to those of the partner network. These adaptations are the Interworking Functions (IWF). This term refers by extension to the functional entity in charge of them. It basically consists of a transmission and protocol adaptation equipment. It enables interconnection with networks such as PSDNs (Packet-Switched Public Data Networks) or CSPDNs (Circuit-Switched Public Data Networks), but it also exists when the partner network is simply the PSTN or the ISDN. Interworking functions may be implemented together with the MSC function, or they may be performed by a separate equipment. In the second case, the interface between MSC and IWF is left open by the specifications.

Besides MSCs the NSS includes data bases. Subscriber information relevant to the provision of telecommunications services is held on the infrastructure side in the HLR (Home Location Register), independently of the actual location of the subscriber. The HLR also includes some information related to the current location of the subscriber. The HLR is typically a standalone computer, without switching capacities, and able to handle hundreds of thousands of subscribers. A functional subdivision of the HLR identifies the Authentication Centre, or AuC, the role of which is limited to the management of security data for the authentication of subscribers.

The second database function identified in GSM is the VLR (Visitors Location Register), linked to one or more MSCs and in charge of temporarily storing subscription data for those subscribers currently situated in the service area of the corresponding MSC(s) as well as holding data on their location at a more precise level than the HLR, In current practice, as will be explained in more detail in Chapter 7, a VLR functions always integrated with each MSC.

But the NSS contains more than MSCs, VLRs and HLRs, in order to set up a call towards a GSM user, this call is first routed to a gateway switch, referred to as GMSC, without any knowledge, of the whereabouts of the subscriber. The gateway switches are in charge of fetching the location information and of routing the call towards the MSC through which the subscriber can obtain service at this instant (the Visited MSC). To do this, they must first find the right HLR, knowing only the directory number of the GSM subscriber, and interrogate it. The gateway switch has an interface with external networks for which it provides gatewaying as well as with the SS7 signalling network to interwork with other NSS entities. The term GSMC is somewhat misleading, because the GMSC function is not by technical necessity linked to an MSC. It could be thought of as an independent equipment, or as a function integrated in a digital telephony switch. However, charging considerations are such that gateway functions will not for some time be set outside GSM networks, and economic considerations make it undesirable to have standalone machines for this function. This results in the widespread implementation of the GSMC function in the same machines as the MSC function itself.

Having seen the pieces, let us look at the glue. Depending upon national regulations. A GSM operator may or may not be allowed to operate the full SS7 network between NSS machines, if the GSM operator has the full control of this signalling network, then Signalling Transfer Points (STPs) will probably be part of the NSS functions, and could be implemented either as stand-alone nodes or in the same machines as the MSCs in order to optimise the cost of the signalling transport between NSS entities (MSC/VLRs. GSMCs, HLRs....)

Similarly depending upon the terms of its license, a GSM operator may have the right to implement its own network for routing calls between GMSC and MSC or even for routing outgoing calls as near as possible to the destination point before using the fixed network. In this case, Transit Exchanges (TE, not to be confused with "Terminal

Equipment" as used for the mobile station architecture) may be part of the GSM network as well, and there again may be implemented as standalone nodes or together with some MSCs.

As a summary, figure 2.1.11 shows the main components of a GSM NSS and the interconnections between them.

2.2. Transmission

Some of the GSM machines are concerned with transmission only, An obvious example is the transcoder and rate adapter unit (TRAU), which is only concerned in adapting speech or data representations. Another example is provided by a transit exchange, whose role is limited to the routing of signalling exchanges between distant NSS entities. But most other GSM machines also play a more or less complex role in transmission. The mobile station obviously does so, and so does the BSC. , the MSC and the interworking function (IWF), which may all be along the transmission path between two users. Conversely, some of the machines have no relation to transmission, except for the minimum needs concerning signalling with the other machines. These include the data bases (HLR,VLR,EIR) and the OSS in general.

As already mentioned, the transmission plane includes two more or less independent functions. The first one is to provide the means to carry user information (whether speech or data) on all segments along the path followed by a communication. The second one is to provide the means to carry signalling messages between entities. The transport of signalling is needed between adjacent machine (i.e., MS to BTS, BTS to BSC, BSC to MSC), but also through networks such as the SS7 network used between NSS machines.

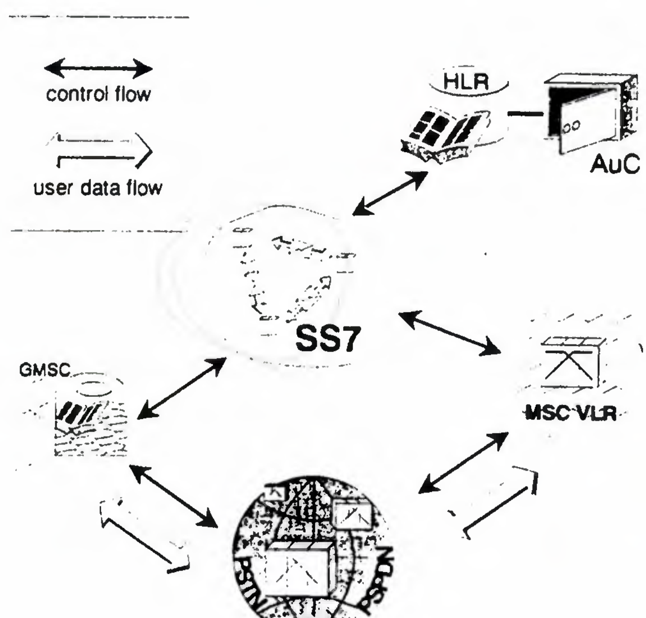


Figure 2.1.11. Internal Structure of the NSS

Includes in this plane are aspects indeed traditionally called transmission, such as modulation, coding, multiplexing, but also other aspect such as low level protocols to format data, to ensure proper sequencing, to correct errors through repetitions and to route information throughout networks. Three chapters in this book are dedicated to the transmission plane (Chapters 3.4. and 5) Chapters 3 and 4 deal with the "traditional" transmission aspects, also referred to as "layer 1" or "physical layer" in the specifications. In accordance with the Open System Interconnection model. Chapter 4 is entirely devoted to the radio interface, because of its prime importance in GSM. Chapter 5 deals specifically with the additional functions provided by "link layers" or "network layers" (according to ISO terminology) in order to transport user data and signalling messages between communicating entities.

2.2.1. Radio Resource Management (RR)

The role of the radio resource management layer is to establish and release stable connections between mobile and an MSC for the duration of a call for instance, and to maintain them despite user movements. It must cope with a limited radio resource (and the corresponding terrestrial resources) and share it dynamically between all needs. The functions of the RR layer are mainly performed by the MS and the BSC. In addition, since the responsibility for the handover process lies entirely within the RR layer part of the functions implemented in the MSC are within the RR domain, in particular the ones related to inter-MSC handovers.

The detailed study of the RR layer functions and protocols is to be found in Chapter 6. Which is devoted to this subject.

2.2.2. Mobility Management (MM)

The machines concerned with mobility management are mainly the mobile station (and more precisely the SIM inside the mobile station) the HLR and the MSC/VLR. The management of the security functions are done by the same machines, and more particularly by the AuC inside the HLR, the BSS is not concerned with the MM plane

2.2.3. Communication Management (CM)

The functions of the communication management layer, or CM layer, consist in setting up calls between users at their, request as f course maintaining these calls and releasing them. It includes the means for the user to have some control over the management of the calls he originates or receives, through the "Supplementary Services". The variety of the Communication Management functions makes it easier to describe as three sub-domains.

2.2.4. Call Control

The MSC/VLRs. GMSCs. IWFs and HLRs. Through basic call management functions, are able to manage most of the circuit oriented services provided to GSM users, including speech and circuit data. This functional core represent a sub-part of the CM layer and is called CC (Call Control) in the specifications.

An important aspect of communication management, beside establishing, maintaining and releasing calls, is the routing function, i.e.., the choice of transmission segments liking distant users and their concatenation through switching entities. GSM mostly relies on external networks to perform this task, interfacing these networks through MSCs z and GMSCs. The IWF may also have a switching function for communications to and from the networks it interfaces with. Call management requires access to the subscription data. In order to check the profile of the subscriber, and therefore the HLR also intervenes in the CM layer.

2.2.5. Supplementary Services Management

A second aspect of the CM layer concerns the management of the so-called supplementary services. as explained through examples in Chapter 1. Users in GSM have some control on the way their calls are handled by the network. This potentiality is described as "supplementary services" each one of them corresponding to some specific variation of the way the basic service is rendered to the user. The impact of supplementary services on calls is mainly a CC function. However the management of the supplementary service itself. i.e.. the modification or checking of their actual configuration, can be done through GSM independently of the calls. It is the object of a separate subpart of the CM plane; the SS management part.

The entities involved in SS management are very few; the mobile station and HLR are the only entities involved.

2.2.6. Short Message Services

The last aspect of the CM layer is related to the point-to-point short message services (SMS-PP). For the purpose of these services. GSM is in contact with a Short Message Service Centre (SM-SC). A service centre may be connected to several GSM networks. In each of these, one or several functional entities are in charge of interfacing the SME-CM . They are basically gateway functions, with the general definition we gave previously. However, the Specifications have special terms for the gateway functions when applied to Short Messages. They define two types of such entities; the SMS-GMSC for Mobile Terminating Short Messages (SMS?MT?PP=) and the SMS?IWMSC for Mobile Originating Short Messages (SMS-MO/PP). The role of the SMS-GMSC is identical to the role of the GMSC for incoming speech or data calls. The role of the SMS-IWMSC is much less obvious and adds little value to the service except providing a fixed GSM point of interconnection for an SM-SC, rather than enforcing its connection to the SS' network which would enable information transfer with any MSC.

3. TRANSMISSION

The first goal of a telecommunication network is to provide a means of transmission between end users. What exactly is to be provided, and how , varies, according to the different kinds of information to be transported, and to the specific constraints of the different interfaces to be crossed.

Both points are relevant to GSM. The multi-service nature of GSM requires that it interconnects with various kinds of external networks each with their own transmission requirements. As far as internal interfaces are concerned, the radio interface is usually the focal point in a cellular network. GSM is no exception, and the specifications of its radio interface include more original features than any other public radio interface yet developed. Though transmission on fixed links is more constrained by existing standards, some new features have been introduced on the restrain links between GSM infrastructure machines.

Despite this variety, the whole system must provide consistent end-to-end transmission paths, taking into account different optimisation schemes on the successive segments along the way. This calls for translation functions between some of the transmission segments, and is a source of complexity.

The purpose of this and the next two chapters is to give an overview of transmission, with a focus on the specific aspects of GSM. A top-down approach will be followed. This chapter will start with a presentation of how end-to-end transmission paths are structured. The way in which user data is transformed along the way will be addressed: this includes the GSM digital speech coding, and the various rate adaptation schemes for data services. Then transmission between the GSM infrastructure machines will be detailed. It was felt that the radio interface deserved a chapter of its own which follows this one. Both chapters will focus on the circuit transmission mode which covers almost all of the services offered by GSM. This leaves aside the packet mode transmission, which is used in GSM for the exchange of control information between machines, and for the Short Message services, which will be described in Chapter 5. The third and last chapter dealing with the transmission functions.

3.1. AN END-TO-END VIEW OF TRANSMISSION

In this section we will address first the transmission of speech between a GSM user and another user in the GSM or in some telecommunications network accessible through the PSTN or the ISDN. Then the other types of transmission will be addressed, that is to say the transmission of non-speech data between GSM users and users in networks such as the PSTN, GSM, ISDN, packet or circuit data networks.

3.1.1. Speech

Telephony is by far the most popular service offered by public networks, including fixed networks and mobile cellular networks. After a general presentation of how voice is transported from mouth to ear through these networks .we will attempt to describe the environment with which GSM must interwork .

3.2.1.1 General Overview

Let s take the example of a speech call between a GSM user called Bernard and a subscriber of a fixed telephony network .called Fred .The function of the transmission plan is to transfer speech signals from Bernad's mouth toFred's ear .as well as from Fred's mouth to Bernard's ear .The transmission path goes through GSM equipment from Bernard to some interworking point ,and from there to some interworking point, and from there tom Fred through the PSTN.

Starting with Bernard, the first item encountered is the microphone of his mobile station. Inside this station, the analog voice signal is transformed into a digital information stream at a rate of 13 kbit/s, which represents the speech signal. Other process in turn change this digital bit streal into a high frequency analog signal transmitted over the air.

After being detected by a base station antenna, this radio signal is processed to recover the digital signal representing speech, which is transported over coaxial cables towards a speech transcoder. From this incoming 13 kbit/s input, the speech transcoder derives another digital representation of the speech signal, at a rate of 64 kbit/s (the standard used in fixed network transmission). It is routed through the mobile services switching centre (MSC) and various links and switches in the PSTN until it reaches the local switch to which Fred's telephone is connected. There (or possibly elsewhere inside the PSTN), the analogy speech signal is reconstructed from the digital 64 kbit/s flow, and transported on Fred's subscriber line until it reaches his telephone, where the loudspeaker enables. Fred to hear an acoustic signal which should be recognised as Bernard's voice.

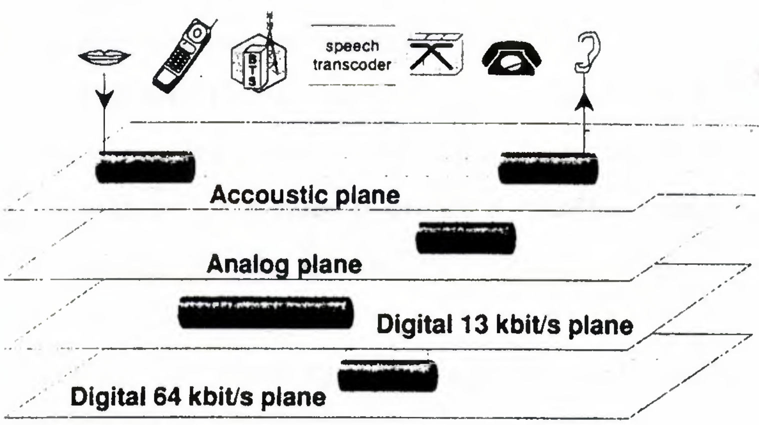


Figure 3.1.1. Speech representations

Transmission between two-end user takes place on several transmission planes

Each one corresponding to a different representation of the speech signal

This elementary description of a telephony call involves several transmission modes for speech, and therefore several transmission planes.

Starting with the GSM subscriber's end, these transmission planes unfold as follows (see figure 3.1.1): acoustic transmission, analogy transmission, digital transmission at 13 kbit/s (this transmission being performed in two speech transcoder), and finally another digital transmission mode in which speech is represented by a 64 kbit/s signal.

Closest to the end user is the acoustic transmission plane; digital transmission at 13 kbit/s and 64 kbit/s as well as analog transmission are found further down the transmission path. The transmission means vary from one interface to another, even within the same transmission plane.

3.1.2 Non Speech Services

Non-speech services, or data services, cover the exchange of a lot of different types of information. Data transmission encompasses the exchange of text of drawing, of computer files, of animated images, of messages and soon. An important part of the information processing is done the two extremities in a machine most often outside the scope of the specifications. We call such a machine "terminal equipment" or more simply "terminal" though in some cases it can be a complex installation, such as a videotext server or a message handing system.

The main function performed by a data terminal in the realm of end-to-end information are the following.

- Source coding which transforms back and forth text, images, sound, etc. In the international which are the binary digits:
- End-to-end protocol, dealing with the organisation of the communication, juggling with such concepts as pages, sessions. Languages;
- And most important the presentation of the information to the user, by display, sound generation, printing, and so on.

In most cases it is possible to confine such processes to the ends of the transmission chain. This enables the reduction of the number of different cases which need to be taken into account by eliminating the need to study the intermedia transmission devices. The relevant characteristics distinguishing the different cases are few, and include the bit rate, the acceptable transmission delay (fixed or variable) and the maximum acceptable degradation due to transmission errors. The concept of hearer capability is used to describe and to refer to what is provided by the intervening equipment, i.e,. the transportation of information between two user-network interfaces. A bearer capability is then characterised mainly by the attributes listed above.

Conversely, the concept of teleservice corresponds to the full chain, Including end-to-end processing of specific information. The detailed functions in the data terminals are outside the scope of this book, just as they are outside the scope of the specifications. This is natural, GSM being basically a distribution network. In most case, only one of the Specifications. The terminal functions, and the corresponding end-to-end protocols between the terminals are by necessity those that are used in the networks GSM between two fax machines is the one used between PSTN users, because one of the two machines may be in the PSTN, A terminal equipment can be for instance a facsimile device, a personal computer, a computer terminal or a videotext terminal.

We will concentrate on the bearer capabilities used between the terminals. Looking closely at the structure of the transmission path we immediately detect an important point; the boundary between GSM and the external network. GSM can be connected to a variety of external networks, since we are not yet in the promised land of broadband ISDN where a single international long-distance network supports all possible telecommunication services. Examples of networks include the good old PSTN (this ubiquitous telephone network which is stell the principal carrier of data transmission) Packet Switched Public Data Networks (PSDNs), such as TRANSPAC in France. or Circuit Switched Public data Networks (CPDNs).

The existence of an external network divides the transmission path into two segments. The segment between the GSM user terminal and the boundary point is entirely within GSM. But the other segment, form the boundary point to the other terminal, is entirely outside the control of GSM , and follows transmission rules that are specific to the external networks.

To reduce the number of cases dealt with by transmission equipment within GSM despite the variety of interworking cess, two generic functions are inserted on each side of the GSM segment, as shown in figure 3.1.2. these functions enable GSM to deal with a small amount of internal transmission modes, and still accommodate the various interworking needs. The adaptation function at the boundary between GSM and the external network is called the network interworking function, most often reduced to the last two terms, and often further reduced to "IWF", On the GSM user side, the functional part of the mobile station which performs the adaptation between a specific terminal equipment (TE) and the generic radio transmission part is called the Terminal Adaptation Function, or TAF.

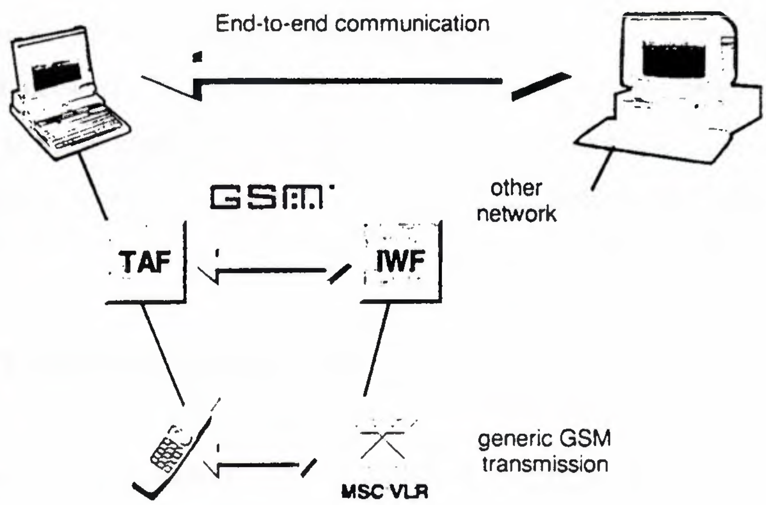


Figure 3.1.2. Data transmission planes

Data transmission can be studied on three levels, the end-to-end transmission plane between terminal equipment

The TAF-IWF plane inside GSM

And the generic GSM transmission plane

The TAF on one side and the IWF on the other sideact as entry points into the GSM world. Their functions depend on the type of end-to-end service. Conversely, GSM entities in-between TAF and IWF are not concerned with the end-to-end service, but solely with the bearer capabilities required to transport the corresponding data flow.

Figure 3.3ğ shows how data transmission can be looked at on three different levels the end-to-end level, the TAF-IWF level and the generic GSM transmission level.

As already noted, most of the end-to-end domain is out of the scope of the Specifications. But the adaptation functions (in TAF and IWF) are of direct concern to GSM For most of the data services, the tasks fulfilled by the adaptation functions can be inferred from the single knowledge of the bearer capability. There is however an exception of importance, which is the facsimile teleservice. In this case, the adaptation includes additional functions dedicated to facsimile, and will be further discussed later.

So, with the sole exception of facsimile, the adaptation functions and the general configuration of the transmission paths depend mainly on the bearer capability and on the external network, It appears that the key factor, at the origin of most of the differences, is the external network. As a consequence, we chose to present the different cases not service preservice, but external network Per external network starting with the PSTN.

3.2. Transmission Inside GSM

In the previous sections we have looked mainly at the aspects of transmission outside, or at the boundary, of GSM it is time to tackle the innards of the system. The inner part of the GSM transmission system extends from a point somewhere in the mobile station (inside the TAF for data services, and where speech is an acoustic signal for the speech cace). Between these two points lie several machines and several interfaces. Our first tass will be present them.

3.2.1. Architecture

Let us stard by looking a little bit more in detail at the functions situated at the borders of GSM (the IWF on one side, the TAF on the other) before describing the more internal parts of GSM.

The IWF is a set of functions fulfilling the adaptations necessary between GSM and external networks. As will be seen. It can be rather limited for speech toward the PSTN and for basic data when interfacing with ISDN. But in other cases, such as facsimile, interworking Functions can be quite extensive. The IWF as a function lies somewhere between the MSC and the external network it interfaces with a first implementation approach is simply to put the IWF in the MSC and this is the usual approach for simple cases such as speech. For the complex cases, it can also be imagined to have special machines devoted only to the interworking functions, and linked to several MSCs. This centralised approach is sensible if the traffic through the IWF is but a small proportion of the overall traffic. This implementation is not precluded by the Specifications, but there is no standard specification of an MSC/IWF interface , and any such interfaces will be proprietary.

Let us turn to the mobile estimation side. The canonical GSM architecture identifies on one side the terminal (TE), in direct contact with the user, and on the other the core functionalities of a mobile station which are common to all services. in between lie the Terminal Adaptation Functions, and in addition for facsimile the Fax Adapter. The piece of mobile station equipment which contains the functions common to all services is called in the Specifications the Mobile Termination (MT):

If we turn now to concrete implementations of this functional model, we find a number of possibilities, differing by the grouping of the functions in specific machines, and on the interface specifications between these machines. The simplest case is when everything is integrated, generic functions, terminal equipment and adaptation functions if applicable. Such a machine is called MT0 (Mobile Termination type 0) in the Specifications. Integrated mobile stations for other service will certainly appear sooner or later, for instance for facsimile.

The next simplest case, which will be our basis of work in the following for data applications, is when the TAF is totally integrated with the genetic functions and interfaces with the generic functions, and interfaces with the terminal through a classic modem to terminal interfaces with the terminal through a classic modem to terminal interface this integrated device is called MT2 (Mobile Termination type2).

Another identified possibility is when the external interface of the Mobile Termination is the ISDN "S" interface to which off-the-shelf ISDN terminal equipment can be directly connected. In this case, the machine is called. MTI (Mobile Termination type 1) A terminal using a modem to terminal interface can still be connected to an MTI, provided an ISDN Terminal Adapter (TAF) are spread between the MTI (where a synchronous adaptation is performed) and the TA (Where a synchronous /synchronous adaptation is performed) The different mobile station configurations are illustrated in figure 3.2.1. In the following. We will not distinguish these different physical implementations. We will not distinguish these different physical implementations. We will refer to the Terminal Adaptation Functions in general, whatever their implementations; similarly, we will refer to the TE-TAF interface, which can be according to the case the interface between the TE and the MT or the interface between the TE and the TA.

Let us now look at what exists between the mobile station and the IWF. Along the transmission path the canonical architecture of GSM distinguishes the BTS (Base Transceiver Station), the BSC (Base Station Controller) and the MSC. Between the mobile station and the BTS is a clear reference point, the radio interface, where the information crosses the space riding the 900 or 1800 MHz electromagnetic waves. The BTS/BSC/MSC split is adequate for the study of the signalling aspects. But the MSC and BSC have little role to play in the transmission chain. Historically, the BTS and IWF were the main actors in the transmission scene, and only basic transmission functions were found between them. Then another piece of equipment was introduced : the TRAU (Transcoder / Rate Adapter Unit), which is definitely transmission equipment, and which was conceived to be distinct from the BSC or the MSC. The TRAU will take the starring role in this section.

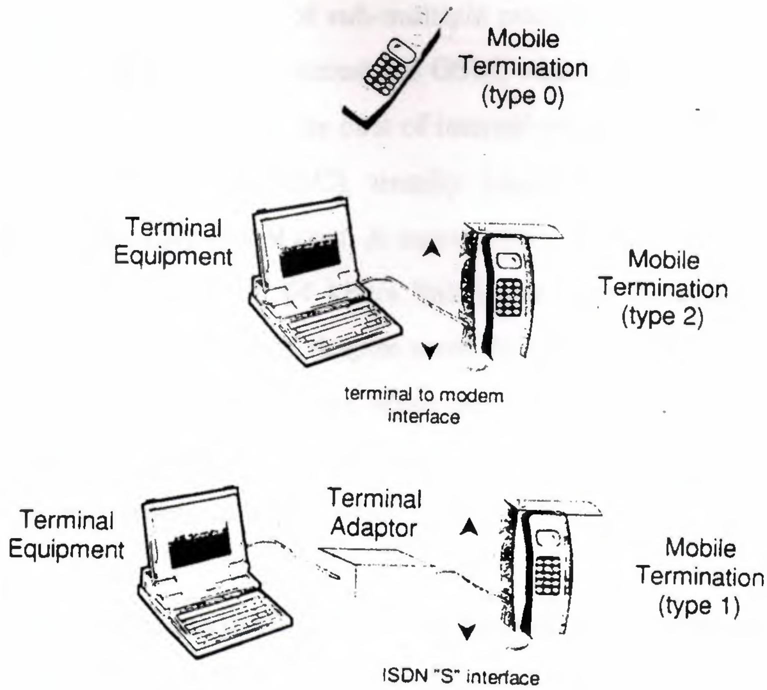


Figure 3.2.1. Mobile stations configurations

Mobile stations can be either fully integrated or include a separate terminal equipment connected to a Mobile Termination through a Terminal Adaptation Function which can be either integrated or kept as an independent piece of equipment.

The rationale behind the existence of the TRAU, distinct from the MSC and BSC, consists of several points. The implicit assumption during the elaboration of the concept of MSC was that it would be implemented more or less as a modified ISDN switch. As a consequence, the transmission at the level of the MSC is very close to that of the ISDN specifications. In particular only 64 kbit/s circuits are switched. As a consequence, the A interface must comform to the lower layers of the ISDN specification. Indeed, the 2Mbits/s standard multiplex structure used on the A interface (and also on the Abis interface) is not specific to GSM, but follows the CCITT G.703 standard. Their basic usage is to carry 64 kbit/s circuits compliant with the needs of ISDN. The multiplexing is based on a 125 μs cycle, each cycle transporting one octet Per circuit. This structure is aimed at the transport and switching of 4 kbit/s circuits, but, in addition, enables the transport of sub-multiple rates such as 32, 16 or even 8 kbit/s. This possibility is effectively of interest for GSM, which does not require connections of more than 16 kbit/s, and where the cost of internal terrestrial links (between BTS and BSC, and between BSC and MSC), usually leased by the operator, represents a substantial part of the operational cost. A transmission method using only 16 kbit/s for user data (signalling is kept on 64 kbit/s links) was then devised, to allow this cost reduction which seems compelling despite some drawbacks. First, this introduces some extra delay for the transmission, and hence lowers the overall speech transmission quality. Second, it introduces a gateway function at a border between 16 and 64 kbit/s. Which is really the purpose of the TRAU.

The late introduction of the TRAU, and the will to keep the switching capability of the MSC strictly similar to the one of an ISDN switch, is the source of its eccentric architectural location. The TRAU may be allocated in different places along the transmission chain, between the BTS, to which it belongs functionally, and (but not including) the MSC. One may then deduce that the only site possible when not on the MTS site is the BSC site. This is however not quite so: the implementations of many manufacturers include a remote transcoder situated on the MSC site. The BSC, as a functional unit, is then "spread" over its own site and the MSC site, and includes the link between these two sites. Conversely, the BSC – MSC interface (or A interface) is situated on the MSc site, over a very short distance. Figure 3.2.2. shows the positions of the TRAU relatively to the other BSS machines.

This somewhat artificial definition stems from historical reasons (i.e.. reasons which were meaningful when the decision was taken, even through this meaning have been lost through later evolution...). The definition was chosen in order to avoid the introduction of an option on the A interface, between transporting user data 16 kbit/s or at 64 kbit/s. since operators have always been keen on limiting the number of options on the A interface to help multi-vedort inter-operability. Manufacturers also might have found an advantage in the decision of having a single rate of 64 kbit/s on the A interface, by avoiding the implementation of switching matrices at 16 kbit/s when the TRAU is put on the BTS side of the BSC matrix.

As a consequence, the Specifications strictly speaking do not allow the placing of the transcoder inside the MSC. Every call between two GSM users must then undergo two transformations (from 16 kbit/s to 64 kbit/s and vice versa, entailing for speech two transcoding operations between the 13 kbit/s and the 64 kbit/s representations), even if the two users are connected to the same BTS. One may imagine that such a restriction could be removed in future phases of GSM...

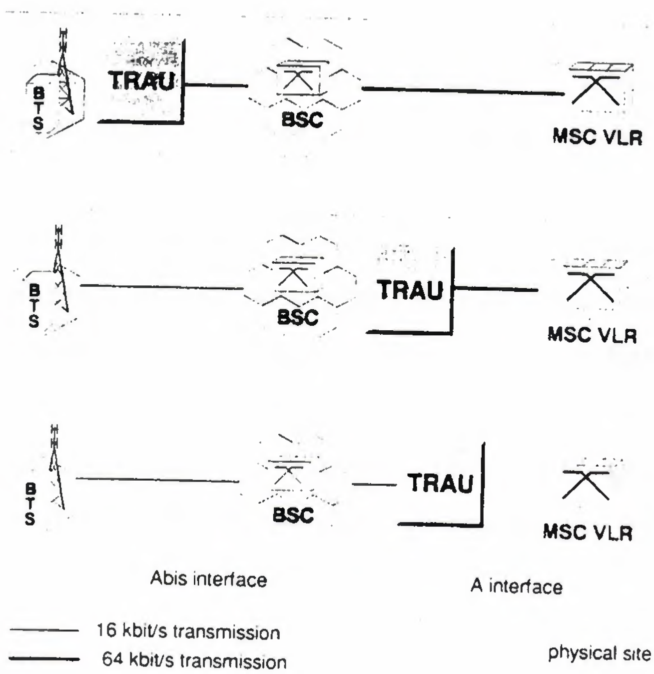


Figure 3.2.2. Positions of the TRAU

The TRAU, which is functionally part of the BTS. can be installed in a remote location (up to the MSC site), to save link costs between the BTS and the TRAU, thanks to the increase in transmission capacity from 16 to 64 kbit/s.

Because the TRAU is the true intermediate equipment for transmission, and because of its architectural predicament, we will not use the notion of A and Abis interfaces in this section, but instead the BTS/TRAU and the TRAU/IWF (or TRAU/MSC) interfaces. The BSC (and the MSC in the case of data) is simply ignored, as it has no special role as transmission equipment (but some as a switching equipment!). In the following paragraphs we will describe the transmission of speech, then data inside GSM, but we will exclude the details of the radio interface, since they will be studied in the next chapter. The relevant interfaces are in each case the following:

- the radio interface ;
- the BTS – TRAU interface (which can be non-existent if the transcoders are situated at the BTS);
- and the TRAU – IWF interface, or more generally the interface between the transcoder and the point of interconnect with other networks.

Across those interfaces the transmission system has the task of carrying speech or binary flows for a variety of data services. this skill represents quite a few different cases, and additional treatment is performed to obtain a higher uniformity. In the rest of the chapter we will tackle these aspects and the detailed transmission schemes inside the infrastructure, first for the speech and then for data. The study of speech will take us to the very specific area of digital speech encoding, whereas for data we will discover the subtleties of the numerous rate adaptation schemes.

3.2.2. Speech

Digital speech transmission over a radio interface in a mobile environment is quite a challenge. As already mentioned, a special digital speech coding algorithm is used in GSM. chosen for its low bit rate (13 kbit/s) and its resistance to high error rate conditions. The description of this algorithm will be first topic of this section. This description is somewhere between the view of laymen (which the authors basically are in this field) and that of a specialist. Some emphasis will be given to some features of voice transmission, such as vocal activity detection and discontinuous transmission, which are important for the spectral efficiency of GSM. the rest of the section is devoted to the rate adaptation which enables speech encoded with this algorithm to be carried not only over the radio interface, for which it was originally designed, but also over fixed digital links, between the BTS and the TRAU.

The GSM transmission path for speech can be divided into the following segments :

a) the mobile station;
b) from the mobile station to the base station : the radio path;
c) from the base station to the (remote) voice transcoder;
d) from the voice transcoder to the MSC.

The junction points separating the segments a) to d) described above correspond to places where a speech representation is changed to another one. These transcoding points are of major importance here, since the description of a transmission scheme is intimately related to the description of the corresponding transcoding functions. Transcoding points are the "elevators" between the different floors of figure 3.2. The following transcoding points are identified inside the GSM domain :

- Acoustic to Analogue Electric transcoding, implemented in the microphone, and the reverse Analogue Electric to Acoustic transcoding, implemented in the loudspeaker; this type of transcoding is not however specific to GSM;
- Analogue to 13 kbit/s Digital transcoding (and the reverse operation), implemented in the mobile station.
- 13 kbit/s Digital to 64 kbit/s Digital transcoding (and the reverse operation), implemented in the voice transcoder, either in the BTS or in the TRAU.

This does not mean that the signal is transported exactly in the same way on all links between two transcoding points. The signal representation is adapted to the transmission medium in intermediate processing points. The two main adaptations are:

- adaptation of the 13 kbit/s digital representation for transmission on the radio path;

- adaptation of the same 13 kbit/s digital representation to transmission on fixed links between the BTS and the voice transcoder in the TRAU.

3.2.3. Data

We have seen in the first part of this chapter the general aspects of the connection between users when circuit data services are provided. The grand result was that between the mobile user terminal and some point inside the IWF, the GSM segment was an over-stretched DTE (data terminal equipment, or terminal) to DCE (data communication equipment, usually modem) junction. This statement is true even when the IWF does not effectively include a modem, e.g. when direct digital interworking is possible. Seen like that the different cases for data transmission correspond to different cases of "modems", with various data rates and other properties. We will devote the rest of this chapter to study how the synchronous GSM radio interface has been made suitable to become part of the various types of terminal to modem connections.

The design of the GSM data transmission was led by two main issues. The first one is simply how to transport a multi-wire interface such as a standard terminal to modem interface, possibly using the start/stop transmission format, over basically a single wire and synchronous medium. A similar issue exists in ISDN, and technical answers are specified in the ISDN recommendations, such as V.110. The second problem is the radio interface itself, with its high raw error rate.

Through an important design goal for a transmission system is to reduce the number of different ways in which the data has to be transported, several connection types have been defined in GSM, they reflect two widely different approaches of how to adapt a terminal to modem interface, and warrant a detour to explain the different types of connections which have been designed between the mobile station and the IWF.

3.2.4 The GSM T Connections

The transmission path between the TAF on the mobile user side and the IWF is functionally totally equivalent to what appeals between the terminal to "modem" interface and the 64 kbit/s circuit in the case of an ISDN connection using V.110. So the RA0. RA1 and RA2 functions will appear somewhere between the TAF (included) and the IWF (excluded). However the transmission over the radio interface must be introduced somewhere in the picture.

Data transmission on the radio interface is not done at 64 kbit/s and kbit/s and V.110 obviously can not be used in its pure form. A first idea could be to keep V.110 as it is with the exception of the RA2 function, which is very simple, and has clearly to do only with transportation over 64 kbit/s circuits. Between RA1 functions, the transmission is done at an intermediate rate, 16 kbit/s or 8 kbit/s, which could have been fitted onto the transmission over the radio interface. Yet, the problem on the radio interface is to limit as much as possible the information to be transmitted, so that the maximum part of the raw throughput can be devoted to optimised redundancy, in order to maximise the transmission quality.

When the V.110 bit stream at the intermediate rate is looked at, it becomes apparent that an important part of the exchanged bits can be removed in GSM. the first of these are the synchronisation bits. Out of the 80 bits of a V.110 frame 17 are used for synchronisation. GSM radio transmission is based on a complex synchronisation scheme, and there is no difficulty in deriving the V.110 frame boundaries from the GSM synchronisation (thanks in particular to the choice of 20 ms as a fundamental GSM synchronisation period, which is a multiple of 5 and 10 ms). In fact, another important aspect of synchronisation comes from the forward error correcting scheme used over the radio interface. With such schemes, residual errors are grouped into bursts, corresponding to an ill-fated radio coding block. There are some advantages to map precisely the V.110 frames and the coding blocks. This is possible because yje coding block recurrence has been chosen to be 20 ms. The rule is then simple : a radio coding block corresponds exactly to 2 or 4 V.110 frames.

But 63 bits still remain per V.110 frame. Out of those, three are not transmitted over the radio interface, because they can be reconstructed by the receiver. These are the bits E1, E2 and E3, which indicate the true end-to-end rate. This does not correspond to new information between the mobile station and the infrastructure, since the rate is transmitted separately by signalling meanjs for setting-up purposes, and thus can be dispensed with. What remains consists finally of 60 bit frames, which can be seen as a simple subset of the original V.110 frame.

The resulting "intermediate rate" for GSM is then 12 kbit/s (derived from the 16 kbit/s) or 6 kbit/s (derived from the 8 kbit/s). in fact, a third and lower rate has been introduced for user bit rates below 2400 bit/s, once again to optimise the redundancy. We have seen that in ISDN, rates below 4800 bit/s are rate-adapted to 4800 bit/s by simple bir repetition. In GSM, this simple repetition is done only up to 2400 bit/s. Because the same amount of auxiliary information is kept, the intermediate rate corresponding to user data rates of 2400 bit/s or less is then 3.6 kbit/s (2.4 + 1.2). The "V.110 like" frame in this case is not 60 bits long any more, but 36 bits long. The transformation from ISDN frames at 4800 bit/s to GSM frames is done simply by taking every other user bit. The reverse transformation consists in duplication each user bit.

The RA0 function is performed on the mobile side, as well as the rate adaptation inspired by the ISDN RA1, called RA1'. This includes in the synchronous cases the network independent clocking control as defined in V.110.

On the infrastructure side the RA1'/RA1 function performs the translation between the radio interface format and the ISDN format, and an RA2 function completes the ISDN adaptation, so that the data flow reaching the IWF is in a full ISDN format.

The difference between a V.110 frame and a radio rate adaptation frame is simple, and the translation between the two is easy. It is just a matter of adding (respectively removing) the synchronisation bits, synchronisation the V.110 frames with coding blocks; and adding (respectively removing) bits E1, E2, E3, whose contents are known thanks to a signalling inside the GSM infrastructure.

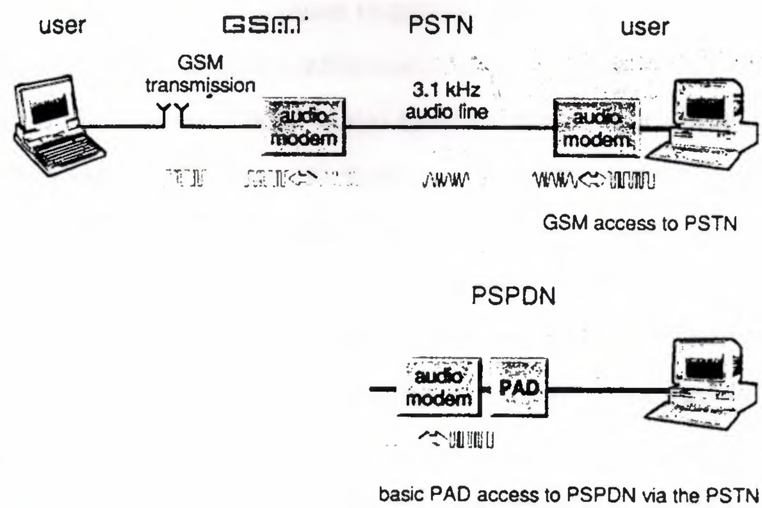


Figure 3.2.3. Rate Adaptation in GSM

Adaptation functions RA0 (for asynchronous data only) and part of RA1 (called RA1') are performed in the TAF (inside the mobile station), whereas the complement of RA1 and RA2 are performed in the BTS/TRAU.

3.2.5. The GSM NT Connections

Since in any case the IWF has a lot to do for an NT connection, there is no reason why GSM NT connections need to strictly follow ISDN specifications as in the T case. The needs are basically to transport a flow of 240 bit frames between the TAF and the IWF, using a maximum total rate of 12 kbit/s. The adaptation to ISDN is done, if need be, at the ends of the connection (TAF and IWF).

However there are spome advantages in using as close as possible transmission methods for different modes, and NT transmission has been designed to have a common core with the 9.6 kbit/s T connections. Indeed, the data rate effectively carried between the TAF and the IWF for 9600 kbit/s T connections is also 12 kbit/s, as explained above. In addition, the distinction between user data and auxiliary information is irrelevant for the RA1' function or for the transmission over the radio interface. Let us see how these common aspects are exploited.

Let us first tart with a difference requiring some explanation. A simple solution for the transmission of the auxiliary information on NT connections would have been to do the same as in the T cases. There would have been no major obstacle to this choice, and the result would have been elegant. A different choice was made. For T connections, auxiliary information adds 12 bits for each period of 5 or 10 ms. This would have resulted respectively 48 or 24 bits Per RLP frame. This was felt to be a high overhead, of which at least a third is useless (bits E4 to E7), since network independent clocking signals are not used in NT connections. Moreover the rest of the bits, which correspond to side signals to and from the terminal, rarely vary. A more complex approach has been chosen to reduce the load incurred by the auxiliary information in most cases. The key act is that at most, three side signals from the terminal are sampled in each direction. The idea was to transmit the values of these signals once Per RLP frame, plus to indicate the transitions, if any, during the period corresponding to the user bits in the frame. The formatting is such that a minimum of 8 bits is consumed Per frame, plus a further 8 bits to indicate a transition. So, if the signals are stable, which is the case when the connection is operational, only one octet per RLP frame is used for auxiliary information may use much more than 6 octets Per frame, but such cases cannot occur when effective transmission of user data takes place.

Another important point for NT connections is the need for frame delimitation. As usual, frame delimitation can be easily obtained on the radio path as a side product of the comprehensive synchronisation arrangement. Then, despite the difference of frame length (RLP frames are four times as long as V.110 frames for the 9.6 kbit/s T connection). It was possible to use exactly the same transmission scheme for the T and NT cases over the radio interface, and so it was done. (Well! This is somewhat a rewriting of history, In reality, the NT mode was contrived by starting from the T case, wondering how the latter could be used to obtain a better transmission quality!).

Unfortunately, the radio path is not the only segment in the way, and the RLP frame delimitation must also be transported between the BTS and the IWF. The V.110 frame delimitation is available, but is not sufficient, the RLP frames being larger. And there like a devil out of its box we find the three E bits: they are free for any use in the case of an NT connection. They are used to convey the frame delimitation.

For T connections, the three E bits are dealt with in the RA1'/RA1 translation. NT connections use a slightly modified RA1'/RA1 translation, which manages the correspondence between the RLP frame delimitation over the radio interface and the one given by the V.110 frame delimitation and the E bits. In the NT case, these bits are used to indicate the position of a V.110 frame in a group of four constituting an RLP block. In the downlink direction, the RA1'/RA1 uses this information to put the four frames of same RLP block in a same radio interface block. The size of the RLP block has been chosen according to the radio block size, such that errors affecting one coding block affect only one RLP frame. Conversely, the E bits are set in the uplink direction according to the position of the contents of the V.110 frame in the corresponding radio block.

The next point to look at is the protocol conversion. It has already been mentioned that the RLP replaces the start/stop protocol or the packet protocol used by the terminal. The RLP between the TAF and the IWF provides the same functionality as the original protocol, but adapted to the GSM transmission. The conversion is done by relay functions in the TAF and the IWF. These functions depend of course on the terminal protocols. These are the three cases : the start/stop protocol. LAPB (X.25-2) and the protocol used for fax. The fax protocol is in essence the same as X.25-2 but with additional signalling so that transmission is basically identical and the three cases are effectively only two. The specifications distinguish two relay protocols, namely that L2R-COP (Layer 2 Relay Character Oriented Protocol) and the L2R-BOP (Layer 2 relay Bit Oriented Protocol).

In the asynchronous case, all the functionalities of the start/stop protocol are fulfilled by the NT mode functions. The frame synchronisation enables bits inside the frame to be constructed into octets, thus removing the need for start and stop signals. Only the 7 or 8 bits of user information are transported (1 fill bit is added in the case of 7-bit characters, so that in all cases there are 8 bits Per character in the frame). At the receiver's end, the start/stop button signals can be reinserted at their correct position. Note that the duration of the stop signal is lost, and as a consequence the relative timing of the characters is lost, but their order is kept. This just makes asynchronous transmission even more asynchronous, and no application is known which suffers from this. Flow control is fulfilled in the start/stop protocol either by using special characters, or by toggling modem control signals. The RLP provides its own flow control, in particular because it needs to regulate the flow when for instance too many repetitions are needed at one moment, which can happen in the case of bad luck. So the start/stop flow control protocol can be relayed by the similar functionality of RLP, once the type flow of control in use is known. A last interesting function of the start/stop protocol is the 'break' signal, which is basically a violation of the start/stop rules (a break signal is a start signal longer than a character, i.e.. such that the stop signal arrives too late). The break signal is used basically as a reset mechanism at the disposal of the user, to be used when things are going strangely. A special method is provided in the relay protocol to convey an indication of the reception of a break signal, which is generated by the receiver.

A last point for this presentation of the start/stop relay protocol concerns the efficiency of the method. If we take the worst case, which is a 9600 bit/s asynchronous connection with 7-bit characters and 1-bit long start and stop signals, we have a maximum throughput of 9600/9=1067 characters per second, that is 21 and one third characters per RLP frame period. A frame contains as a whole 240 bits, 40 of which being used for error detection, for frame numbering and for the acknowledgement and flow control protocol. A minimum of 8 is used for the modem control signals. There remains 192 bits, that is to say 24 octets (which corresponds to a data rate of 9.6 kbit/s exactly). A benefit 2 2/3 octets is then obtained, allowing on average, one frame out of 8 to be a repeated frame. For user rates of 4800 bit/s or less, things are obviously much better.

The relay method for protocols which deal with frames, such as HDLC, is rather different. Because the RLP replaces the protocol, the digital stream coming from the external world can be stripped from the overhead introduced by the replaced protocol. In the case of HDCL, this corresponds to the link layer header (2 octets per HDCL frame), the synchronisation overhead (a minimum of 1 octet) and the error detection overhead (2 octets). The remaining data consists of chunks of variable length, which is in general do not fin into the fixed length RLP frames. Frame delimitation is then needed, and this is done using a special status octet after the last octet of the content of an HDCL frame. The final gain is then 4 octets Per HDCL frame. The relative gain depends on the HDCL frame length. The frame rate in RLP frames being exactly equal to the original rate, the breathing space obtained for repetition is never null, but is greater when the HDCL frames are smaller.

4. SIGNALING TRANSFER

In the two pervious chapters the focus has been put on the transmission functions close to the physical medium, and we mainly studied the means to transport user information. But user information is not the only thing to be transported in a complex network such as GSM. most functions performed by such a network are distributed over several distant machines, and information exchanges are needed to co-ordinate what is done in these machines. We will address these exchanges from two points of view. One is simply to present what these exchanges simplify, why messages are sent and what reactions they induce. The other point of view is simply how the messages are transported from one point in the network to another. These transmission aspects call for a number of additional functions in comparison to those presented in the two previous chapters, and will be the topic of this chapter.

In all cases, the information needed for the co-operation of distinct entities. In other words, the signalling information, is organised into messages. The sending of a message is triggered by some event, and its reception triggers a chain of events. A typical elementary message consists of some "message type", which indicates what reaction the message will trigger in the recipient, and some qualifying information, under the form of mandatory or optional parameters.

One of the tasks of the transmission protocols is to provide message delimitation out of bit streams. Another is to guarantee a very low level of undetected errors, since such errors can have important consequences, e.g., changing the meaning of a message into another one. These functions are part of the link layer functions. We will address in this chapter a variety of link layer protocols, akin to the famous HDCL. including the link protocol adapted to the GSM radio interface.

Another aspect of signalling transfer is the organisation of the message flows, and their routing. If the exchanges between machines were to be done by speech, a network such as GSM would appear as a very noisy crowd, a source of an immense cacophony. Everybody would be speaking to almost everybody else at the same time using other people as intermediaries to overcome the problem of distance. There is then some need for organisation, independent from the actual significance of the exchanged information. Two main aspects can be, identified. The first is the routing problem i.e. how messages are passed from one point to another until they reach their final destination. The second is how to use references to handle several dialogues in parallel. These are the main aspects of the network layer, which will be the topic of the second part of this chapter. There we will see how messages are carried between the mobile station and the KSC .We will also visit the realm of the Signalling System Number 7. A packet data network designed for signalling exchanges.

In fact our subject-matter is not limited exclusively to signalling information. Although most user information in GSM is of a circuit nature, there are a few user services which are basically of a non-circuit nature, such as short messages. Such services are very much akin to signalling messages in so far as they both require the same kind of transport mechanism. They will then be studied here. Another special case is the RLP, the Radio Link Protocol which was presented at the end of Chapter 3. Because of its relationship with link layer protocols, it will be presented in a bit more detail here.

Through the link and network protocols from the backbone of any system composed of co-operating communicating machines, their detailed understanding is not a pre-requisite for the understanding of the next chapters, which deal with the co-operation itself.

4.1. The Needs

Both signalling messages and short messages require a packet switched system to transfer them. Packets can be stored, combined, segmented, multiplexed, and tortured in many ways along their route. They can be carried over interfaces in a variety of ways, and this is actually what happens in GSM, where virtually each segment along the transmission path bears its own stack of protocols.

One may then wonder why there should be so many different methods when all of them basically fulfil similar purposes? There is no simple answer to this question, but some sensible reasons do exist. One of them is the search for optimization, which applies in particular for the radio interface, on which traffic load is critical. Another reason is the reuse of the existing standards. For instance the CCITT Signalling System number 7 (SS7) forms the basis of signalling exchanges between the machines of the Network and Switching Sub-system (NSS) since it would have been very uneconomical to choose anything else in a switch environment but beside these arguments the historical background must also be recalled. Each interface was effectively designed by a different standardization group or sub-group, with different antecedents, and this situation certainly contributed to the variety of protocols one can find in GSM.

When trying to understand the needs for signalling transportation, let us first remark that each entity in the system may communicate many others. There are cases when the two communicating entities are contiguous, i.e. directly linked through a physical interface, but this is not the general rule. We will start by listing the communication needs between contiguous entities, before proceeding to the relaying cases.

4.2. Linking

The link protocols used in GSM are not the same on all interfaces. The ones which will not be described are summarised in table 4.2.1. They represent the major protocols used by GSM signalling in the linking area. In the mobile station itself, the interfaces between SIM and Mobile Equipment, as well as between Terminal Equipment and Mobile Termination should for completion be cited as well. In order not to be drawn into too many details, they will not be studied here.

These three protocols (LAPDm, LAPD and MTP 2) have very similar functionality. The presentation will therefore be done on a functional basis, along with explanations of the differences between the cases. The reader will therefore be able to gain a better knowledge of their functions, which should help in the understanding of the individual protocol's behaviour. For a detailed interface-per-interface specification, the reader is referred to the specifications, as well as the relevant ETSI and CCITT recommendations. In the case of the Message Transfer Part (MTP) in particular. GSM has not introduced any new functionalities, but has used the protocol as defined in the ETSI specifications, themselves referring extensively to the CCITT recommendations of the Q series. The examples in this chapter will mostly be taken from the LAPDm protocol – and to some extent from the Link Access Protocol for the ISDN "D" channel (LAPD) protocol from which it derives – which is, among the examples shown, the link protocol with the most original features.

In all cases except for the radio interface, signalling messages are sent over plain 64 kbit/s circuits. But between MS and BTS, the physical medium is very specific to GSM. On this interface, the transport of point-to-point messages can be done in two ways, as explained in the previous chapter ;

Interface	Link Protocol
MS – BTS	LAPDm (GSM specific)
BTS – BSC	LAPD (adapted from ISDN)
BSC – MSC	MTP level 2 (SS7 protocol)
MSC/VLR/HLR – SS7 network	

Table 4.2.1. – Link protocols on GSM interfaces

Despite their apparent variety, the link factors used on the GSM interfaces have similar functionality

- using the main channel, if necessary by preempting the resource (Fast Associated Signalling). This stealing method typically consists of not transmitting one block of user data (representing 20 ms), and using the corresponding resource to send a signalling message istead;
- using the Slow Associated Control Channel (SACCH).

Let us now consider how the data is structured on the link layer to be transported on these physical media.

4.2.1. Structuring in Frames

The prime functionality of a link layer is to structure the information to be transmitted on the channel in units bigger than a single bit. The resulting atomic units are the basic structure on which all link layer functions work. In the signalling world, such a unit is called a frame. The whole issue consists of including sufficient information in the bit stream so that the receiver is able to find out the beginning and end of each frame. Both LAPD and MTP 2 are the heirs of HDCL in this area, whereas LAPDm makes use of the synchronisation scheme of the radio interface to convey information on frame limits.

In HDLC frames start and end with an eight-bit long pattern called a flag (see Fig. 4.2.). to prevent false starts or ends, a mechanism ('0' bit insertion) is introduced to disguise the flag pattern when it occurs inside the data. The resulting bit stream only contains the "legal" flags.

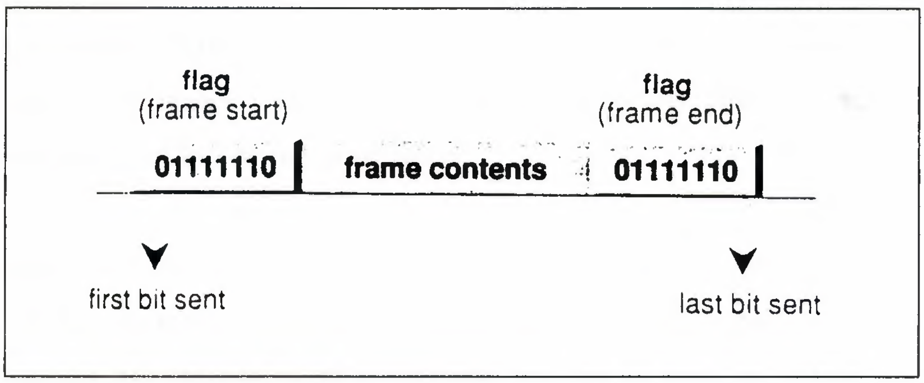


Figure 4.2.1. HDLC frame flags

Frames start and stop with a defined pattern called a flag.

The figure applies both for LAPD and MTP2.

Between transmitter and receiver, a `0` 's added to each sequence of 5 consecutive "I"s in the data, in order to disguise the flag pattern when it appears inside the frame.

4.3. Networking

The link protocols described above enable the exchange of frames between two entities which are directly interconnected through some physical medium. Now, there are a number of application protocols which involve two entities not directly interconnected. Additional transmission functions, described in this section are needed to provide these application protocols with end-to-end connections for the transfer of the corresponding messages. These connections make use of successive elementary links, as described in the previous section, a long a route between the two extremities. Conversely an elementary link is used for a number of network connections in parallel with possibly different origins and ends. For example, call control messages originating from the mobile station must be routed to the MSC, whereas radio resource management messages generated by the same mobile station must terminate their course in the BSC, even though both flows of messages use the same signalling link on the radio path and on the Abis interface.

One of the networking functions is the routing, i.e. the choice of the successive segments composing the route. In this area two broad techniques are used and we will give examples of both of them. With datagrams, each message is routed following analysis on its arrival. With virtual circuits, a route is established for sometyime, for the use of complete dialogues : the route is established by the first message, and the following message just follow the same route.

Another function we will see here is closely related, and consists in the possibility of having several independent connections existing in parallel between two entities. This can be used between contiguous entities as well as distant ones. The connection correspond to independent application dialogues, for instance to the management of different user communications.

A concept common to all the aspects of networking is the address, which will be a leitmotiv of this section. The network protocols add tax to the messages to discrimşinate between the different follows. These tax can be addresses identifying the

4.3. Networking

The link protocols described above enable the exchange of frames between two entities which are directly interconnected through some physical medium. Now, there are a number of application protocols which involve two entities not directly interconnected. Additional transmission functions, described in this section are needed to provide these application protocols with end-to-end connections for the transfer of the corresponding messages. These connections make use of successive elementary links, as described in the previous section, a long a route between the two extremities. Conversely an elementary link is used for a number of network connections in parallel with possibly different origins and ends. For example, call control messages originating from the mobile station must be routed to the MSC, whereas radio resource management messages generated by the same mobile station must terminate their course in the BSC, even though both flows of messages use the same signalling link on the radio path and on the Abis interface.

One of the networking functions is the routing, i.e. the choice of the successive segments composing the route. In this area two broad techniques are used and we will give examples of both of them. With datagrams, each message is routed following analysis on its arrival. With virtual circuits, a route is established for sometyime, for the use of complete dialogues : the route is established by the first message, and the following message just follow the same route.

Another function we will see here is closely related, and consists in the possibility of having several independent connections existing in parallel between two entities. This can be used between contiguous entities as well as distant ones. The connection correspond to independent application dialogues, for instance to the management of different user communications.

A concept common to all the aspects of networking is the address, which will be a leitmotiv of this section. The network protocols add tax to the messages to discrimşinate between the different follows. These tax can be addresses identifying the origin or the destination, or connection references, or route references. They are used to choose a route, i.e. to forward the message onto the next appropriate segment, or to distribute it to the right application software.

Let us now consider how this issue is tackled in the different subsystems, starting with the BSS.

4.3.1. Networking In The BSS

4.3.1.1. The Mobile Station Point of View

From the point of view of a mobile station, the origin or destination of the messages depends on the application protocol. The mobile station addresses different network functional entities, and these addresses are then used by the network to actually route the message to the appropriate equipment (BSC, MSC or HLR). Moreover, several protocols are handled between the mobile station and MSC. In addition several user communications may have to be managed in parallel between the mobile station and the MSC, as we have seen in chapter 1, e.g. for the indication of a new incoming call when another already exists.

The identification of the link on the radio interface gives the ability to distinguish signalling messages from short messages for any single mobile station, but this distinction is not enough to determinate the application protocol a message pertains to. It must be complimented by a networking address. This function is fulfilled by the Protocol Discriminator (or PD). Several PDs are defined (see table 4.3.1.), which were originally introduced just as a functional partition of the messages. As a consequence of the GSM functional architecture, this partition also corresponds to an entity or the infrastructure side, and this is why we can consider the protocol discriminator as an address. The case of supplementary services (SS Protocol Discriminator) raises specific problems which will be dealt with in a dedicated section.

The BTS does not appear in the list of partners on the infrastructure side (third column of the table 4.3.1). This reflects the fact that the mobile station does not have a dialogue with the BTS for reason other than for link management. There is however a single exception to this rule : during handover, one message is sent directly from the BTS to mobile station, for the sake of speed.

The protocol discriminator is specified as part of the application protocols (in TS GSM 04.08), though it belongs to a sublayer common to several protocols, the PD is inserted by the originator. In the mobile station to infrastructure direction it is used by the BSC to decide whether it is the destination (RR) or if the message has to be forwarded to the MSC (all other cases). The PD of received messages is used by the mobile station and the MSC to distribute them to the right software module.

PD	function	Origin/destination
CC, SS	Call control management and supplementary services management	MS from/to MSC (and HLR)
MM	Location management Security management	MS from/in MSC/VLR
RR	Radio resource management	MS from/to BSC

Table 4.3.1. Protocol discriminations on the radio interface

Three protocol discriminations are defined
Which correspond to an origin/destination on the infrastructure side

Now, this is not sufficient to discriminate between CC and SS messages pertaining to different user communications. The term transaction is used in this context, each transaction corresponding to a communication. In fact, transactions also exists which are used for supplementary service management. In all cases, distinct transactions can exist in parallel. Messages pertaining to different transactions are distinguished by a Transaction Identifier (TI). The TI is inserted by the originating entity (MS or MSC), and is used by the other to relate the message to the right context. The specification related to the TI are part of specifications of the RR protocol which appears in this case as below the MM and CC protocols.

4.3.2. Networking For Supplementary Services Management

One case of signalling message transmission has been left aside so far : the transmission of supplementary service management messages between the mobile station and HLR. Let us first recall that, even though a supplementary service is in *general fulfilled by the MSC/VLR, its management (EG activation / deactivation of the* service, interrogation of its status, etc.) is performed by the HLR of the subscriber. It is the HLR which is in charge of eventually modifying the context in which MSC / VLR to keep it in line with the service state in the HLR. This choice of having a single point of control (HLR) ensures the consistency of data throughout the network.

We must consider two legs for the dialogue between the mobile station and the HLR : one from MS to MSC, which makes use of the mechanism described in the BSS section, and another in the NSS, using MTP, SCCP, TCAP as described in the previous sections. The disparity between the two worlds made things difficult, at the conceptual level as well as for the intervening MSC / VLR. The conceptual approach in this domain is somewhat fuzzy, and the vision presented here reflects the view of the authors.

On the radio interface, supplementary services management information can be carried either in standalone messages, or be part of some call control messages. Here we find the first fuzzy conceptual point. In our view, the call control protocol can be considered as a carrier of supplementary services messages, in the same way as e.g. the messages belonging to MM protocol between MS and MSC are carried on the Abis interface. The comparison between the Abis interface is actually interesting, because some Abis messages from the BSC are also “mixed blood” messages, in so far as they contain an order for the BTS as well as a message to be carried forward.

But the radio interface is not the only area in which supplementary services raise architectural questions. The MSC / FLR is another fuzzy area. It acts as a gateway between MS – MSC / VLR protocol, and the MAP protocol towards the HLR. As such, ι must perform some analysis of the messages, be it only to know whether messages received from the HLR must be transmitted towards the mobile station or not. The corresponding information can typically be found in the message type, which is for instance the only information which provides the ability to distinguish between messages belonging to different MAP / X protocols. In addition to this analysis for routing, the specification is such that the MSC / VLR checks most of the syntax of the messages, and may reject them if they do not comply with the specifications. For these reasons it is difficult to see the MSC / VLR as a pure transit mode for the supplementary service management messages. But it can not be considered either that the MSC / VLR makes use of the semantics of these messages : it only deals with them as objects to be transported or rejected. Taking al these points into consideration, we prefer here to present the MSC / VLR only as a relay on the path between the mobile station and HLR as far as supplementary services management is concerned, and consider the application protocol (MAP/1) as an MS / HLR protocol, eventhough it is not carried by SS7 all the way. This approach puts the stress more on the functional aspects (in particule on the fact that the dialogue is between the mobile station and HLR), rather than on marginal issues such as syntax analysis.

Figure 4.3.1. shows the overall picture : the MAP/1 protocol is an application protocol between the mobile station and the HLR. Between the mobile station and the MSC, its messages are carried as encapsulated messages either insidecall control messages (protocol discriminator : CC) or inside messages using the SS protocol discriminator. Between the MSC and the HLR, it makes use of all the SS7 protocol stack, and its messages are distinguished from other MAP messages simply by the message type. Messages pertaining to different MSs are distinguished on the MAP leg by the use of TCAP transactions, on the A-interface by the use of SCCP connections, and finally between the MSC and the mobile station by the radio channel. Several SS management transactions can exist in parallel. They are distinguished on the BSS leg by

different Tis, and on the MAP leg as different TCAP transactions. It falls on the MSC / VLR to make all the needed translations!

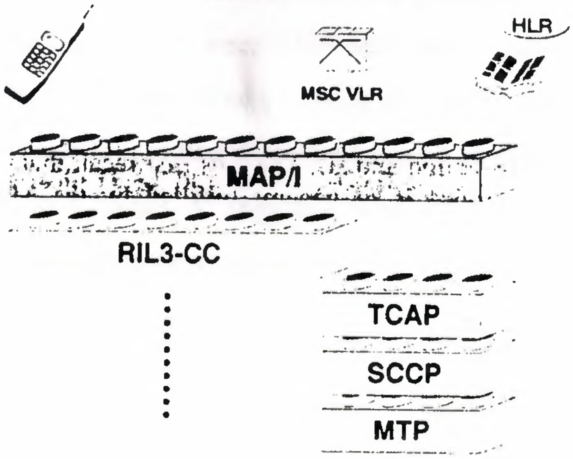


Figure 4.3.1. Protocol stack for supplementary service management

The actual application protocol between MS and HLR relies on two different transportation protocol sets,
One between MS and MSC, the other one inside the NSS

4.3.3. Networking for Point-to-Point Short Messages

The short message services is the second domain where networking encompasses both the BSS and the NSS worlds. In fact it encompasses even more as GSM interworks for these functions with external networks.

As explained in chapter 2, the role of GSM for short messages is to transport them between the mobile station and the Short Message Service Center (SM-SC), the latter being out of the scope of the Specifications, and possibly outside the control of GSm network operators. The SM-SC is connected to one or more MSCs which act as gateways between the GSM world and the SM-SC. The corresponding functions are called SMS-GMSC in the case of Mobile Terminating short messages and SMS-IWMSC (InterWorking MSC) in the case of Mobile Originating short messages. We will call them both the SMS-gateway for short. The stack of protocols between such gateways and the SM-SC is left open. What the Specifications specify is a set of transportation means for the conveying the short messages between the mobile station and the SMS-gateway. The transportation involves potentially two domains : the MS to MSC segment, and the segment between MSC and SMS – gateway. The SM-TP ("Short Message Transport Protocol") specified between the mobile station and the SM-SC is in fact an end-to-end protocol including some features of an application protocol, and will therefore dealt with in Chapter 5, together with other short message management functions. We will concentrate here on the relying of the shot message between GSM entities.

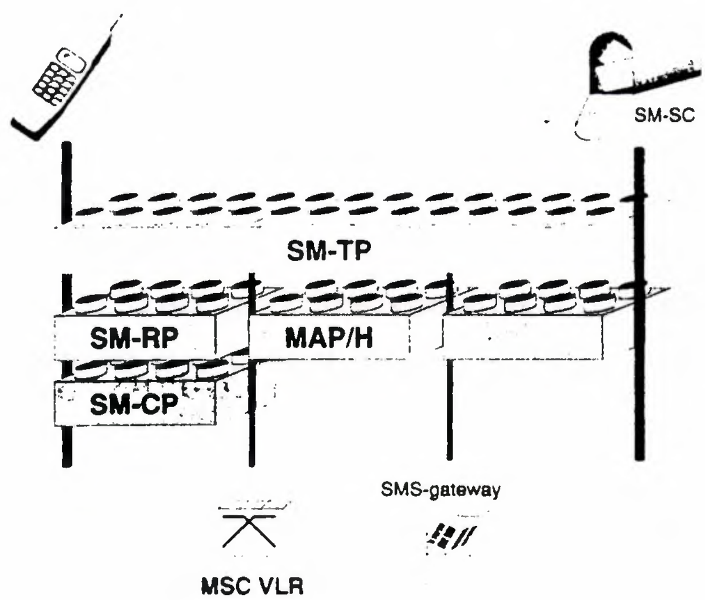


Figure 4.3.2. Protocol architecture for the transport of short messages in GSM

The relay protocols co-operate to convey a short message between the MS and the point of interconnection of the SM-SC in GSM
MAP/H interconnects with SM-RP itself relying on SM-CP
for transport between MS and MSC

The stack of protocols between MSC and SMS – gateway includes MAP/H and the usual stack on the lying Map protocols, i.e. (from the bottom up) NTP, SSCP SCCP and TCAP. This interface is the most greedy in terms of message overhead, and the maximum length of the short messages in GSM (140 Octets, sufficient coding for 160 7-bit ASCII characters) is then a direct consequence of the MTP maximum frame length (see page 270).

We will now consider the BSS leg (from mobile station to MSC) and the NSS leg (from MSC to SMS gateway) in turn. Since on both legs the Mobile Originating and Mobile Terminating short messages use similar transport mechanism but in opposite directions, only one of the directions will be described in detail (the Mobile Originating short message). As a last point in the description of the each leg we will identify the differences for the Mobile Terminating case.

4.3.3.1. The BSS Leg

The lower layers used to convey the messages from the Mobile station to the MSC have been described in the section. On the radio path, they include and acknowledge-mode SAPI 3 link on a TCH/8 or on an SACCH. This link is then relayed up to the MSC using the relay protocol on the Abis interface and DTAP on the A interface.

We now consider the short message "Control Protocol" (SM-CP), which has very little added value (if any?) compared to the service offered by the underlying layers. It is a very simple protocol, consisting of a command/answer procedure with three message types, as shown in figure 4.3.3., and without even a reference to correlate them since its mode of operation is basically of an alternate nature ("send-and-wait"). The SM-CP CP-DATA message is the only one carrying upper layer information, which does not necessarily include the short message itself, but could be some upper layer acknowledgement or error report. Each message of the SM-CP protocol includes a protocol discriminator specific for short messages, and a transaction identifier. This last field could enable a mobile station to manage several parallel short message transactions at the same time. This would then be the only real added value of the protocol. But the Specifications do not indicate clearly whether it can be used as such.

The next protocol on our list is the SM-RP (Short Message Relay protocol), whose functions include the management of references and addressing. This protocol for the networking functions between MS and SM-SC. The SM-RP protocol interworks with the MAP/H protocol in the MSC/VLR.

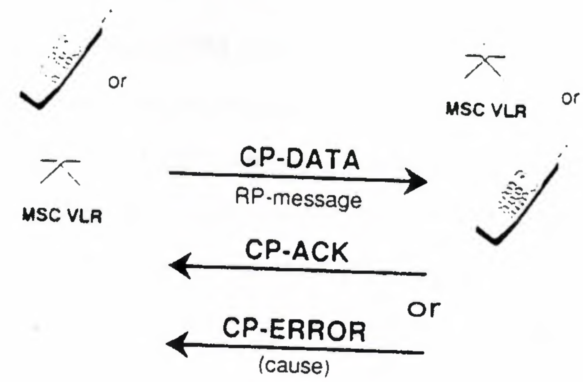


Figure 4.3.3. The SM-CP protocol between MS and MSC

The SM-CP protocol forces and alternate mode of operation (sending of the message, then waiting for the acknowledgement) with the possibility of the message being repeated once if a timer expires

Three messages are defined in the SM-RP protocol, one to carry the message (RP-DATA),one for transporting the acknowledgement (RP-AACK) and on for indicating an error (RP-ERROR). Though the message list is similar to the one of SM-CP, the functions fulfilled are quite different. The messages are correlated together through a one-octet message reference generated by this protocol, enabling the sending or receiving of different messages in parallel at this level. Addressing is performed by including in the SM-RP RP-DATA message the destination address (for an MO message), or the origin address (for a MT message), i.e. in all cases the address of the SM-SC.

There are a few differences between the messages for Mobile Originating messages, and for Mobile Terminating messages. One of them is the inclusion in the mobile terminating SM-RP RP-DATA message of a priority indicator. However, since this priority indicator is not transported in the MAP, it is quite useless the MSC tampers with upper layers to find the relevant priority information.

4.3.3.2. The NSS Leg

Between the MSC and the GMSC, the transport of short messages is performed by the same means as signalling messages, i.e. using the SS7 stack of protocols supporting MAP (MTP, SCCP, TCAP), on top of which is added one of the MMAP protocols, the MAP/h. Taking into account the underlying SS7 layers, the MAP/h provides the same functions as SM-RP in the BSS leg. It contains three messages, which can be mapped onto the three SM-RP messages, as indicated in figure 3.3.4.

4.3.3.3. The Relay in the MSC/VLR

As was the case for the protocol supporting supplementary service management, the MSC/VLR is the translator between two worlds. The key points of the translation are summarised in figure 4.18.As far as upper layers are concerned, the combination of the SM-RP protocol on the BSS leg, the MAP/H protocol on the NSS leg, and the relay in the MSC/VLR, can all be considered as a single network protocol, providing the routing between MS and GMSC (and from then on to the Short Message Service Centre), as well as the possibility to deal with several messages in parallel. This will be how these protocols will be considered in the description of the upper layer protocols for the short message services, in Chapter 6.

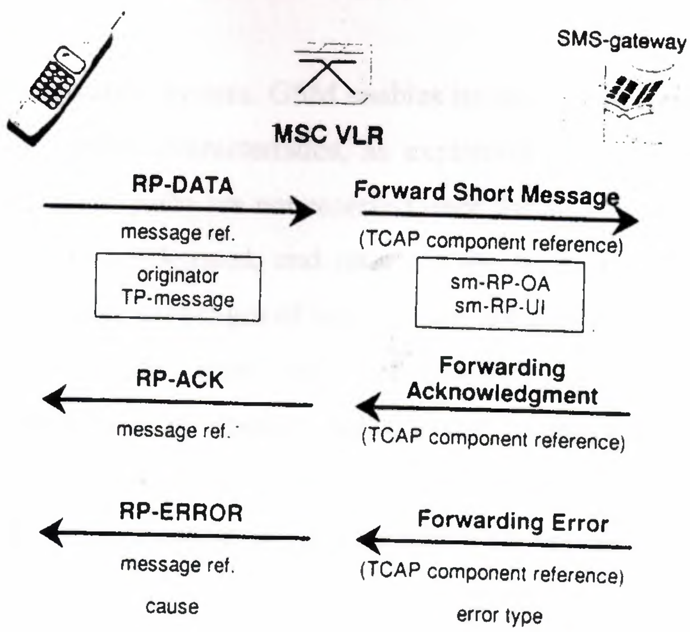


Figure 4.3.4. Short message relying with SM-RP and MAPH

Messages from SM-RP and MAP/H map onto one another, relaying both the short message and its acknowledgement between the MS and the MSC to which the SM-SC is connected

5. RADIO RESOURCE MANAGEMENT

As a telecommunication system. GSM enables its users to communicate through transmission paths of various characteristics, as explained in the previous chapters. However, these transmission paths are not reserved once for all between any two pairs of users. They are set up on demand, and only for the time necessary for a given communication. This requires exchanges of information, not only between the users and the network they are in direct contact with, but also between machines within the network. This and the two next chapters are devoted to the description of these information exchanges, which enable distant participants, users and machines, to act together to provide the communication services for which the networks are designed.

This technological field is known as signalling, and under this name , its reputation has not yet grown as wide as other fields such as modulation, signal processing, and other transmission techniques. Many people consider it simply as a branch of software engineering, though it is in fact at the centre of the design of complex systems, where tasks can be executed only through the co-operation of distinct machines. This is the case of telecommunications, where machines are by essence distinct and distant. The study of GSM signalling is therefore of prime importance to understand how the system operates, and the reader should not be surprised that half of the book is devoted to the signalling interchanges.

Signalling is often the juxtaposition of many simple and more or less inter-dependent procedures, and its complexity stems mainly from the number and the diversity of small issues. We have already seen in Chapter 2, concerning the architecture, that the basic methodology to tackle such issues is "divide and conquer". Pervasive in the specifications of GSM signalling is a split in three functional domains : Radio Resource Management, Mobility Management and Communication Management. The management of the calls is the upper plane, relies on the Mobility Management functions for dialling with the mobility of the users and with security-related functions. The management of the radio resources groups functions specific to the radio interface.

A major difference between a radio mobile telecommunication network and a network with fixed links is the management of the access resources. In a fixed system, a dedicated communication medium exists continuously between the user terminal and the infrastructure, ready to be used when a call needs to be established. On the contrary in a cellular system like GSM, a dedicated channel over the radio interface is provided to the mobile stations only on demand and for the duration of the call, under the control of the infrastructure. This calls for functions which bear no equivalent in ISDN for instance. Even if 64 kbit/s channels are allocated dynamically in the case of an ISDN multi-terminal installation, this resource management is rather than limited compared to the one in a full GSM cell. Moreover, in ISDN, a signalling channel is always ready for use by any terminal. The matter is quite different in GSM, where the signalling capabilities offered to a mobile station in idle mode, that is to say when not allocated a radio channel to its private usage, are limited to the absolute minimum. The consequence is that a host of new procedures are needed.

Besides dynamic channel allocation, another feature of GSM (or cellular systems in general) compared with fixed networks is the handover. The problem consists in providing a dedicated channel from mobile station to MSC at every moment during a call, despite the movements of the user. This calls for a complex measurement and decision process to trigger the transfer of the communication at the right moment and toward the right cell. In a cellular system, the handover process is very important, since it impacts significantly the quality of the communications as perceived by the users, as well as the spectral efficiency.

This chapter will be devoted to these topics, that is to say to the functions required to co-ordinate the mobile stations and the infrastructure so as to provide the suitable transmission means over the radio interface, whatever the telecommunication service requires, and whatever the user's movements. These functions form a well defined area, which we presented in Chapter 2 as the RR (Radio Resource management) functional plane. They are spread among four entities in the canonical GSM architecture : the mobile station (of course!), the two base station sub-system components (BTS and BSC), and a small part in the MSC. All the higher layer functions, described in Chapter

6, are basically managed directly between mobile station and MSC, the base station sub-system (BSS) acting for these functions as a single complex functions implies the existence of signalling procedures between the involved infrastructure machines : this is the purpose of the signalling protocols on the A (BSC-MSC) and Abis (BSC-BTS) interfaces, which will therefore be described in this chapter.

The chapter is basically composed of two parts. After preliminary architecture considerations needed to introduce some specific concepts, the major requirements which drove the design of the RR protocols will be looked at. Then, after a section to present the protocol architecture, the various procedures needed to fulfil these requirements will be developed.

Preliminary Architecture Considerations

A section in the middle of this chapter will be devoted to the architecture and the protocols in the Radio Resource management domain. However, some basic notions are necessary for the understanding of the first part of the chapter. It concerns mainly the notion of anchor MSC.

The major roles are played in this chapter by the components of the BSS, the BTS and the BSC. The MSC intervenes a little, to deal with handovers between cells managed by different BSCs. Some handovers may even transfer the mobile station from a cell within one MSC area to a cell in another MSC area, thus involving two MSCs. The roles of the two MSCs are different. In no case does the MSC in charge of the communication relinquish its control to the new MSC. This MSC is called the anchor MSC for the connection. This is an important design choice of GSM, with numerous consequences on the procedures. Several arguments justify this choice: a compelling one is the charging problem, since toll ticketing is much simpler when one MSC follows a call from its beginning to its end.

A consequence of this architectural choice is that after an inter-MSC handover, two MSCs (and at most two) may be involved in the connection. The transmission chain between the mobile station and the interworking point with external networks is then composed of a BTS, a BSC and either two MSCs, a relay MSC and the anchor MSC, or of a single MSC (see Figure 5.1.). to ease the problem of terminology, a convenient approach is to consider the notions of relay MSC and of anchor MSc as functional, and to admit that when there is only one MSC, it is at the same time the relay MSCV and the anchor MSC. Thus the term relay MSC will be used to refer to the MSC in direct contact with the BSC, even if it is also the anchor MSC; and the term anchor MSC will be used to refer to the MSC in charge of upper layer treatments, even if it is also the relay MSC.

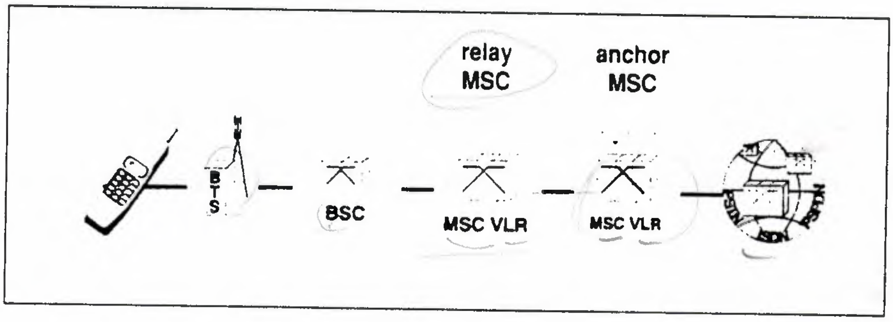


Figure 5.1. The concepts of anchor and relay MSCs.

The transmission chain may involve two (and at most two) MSCs;
the anchor MSC in charge of the communication management and the relay MSC in
charge of the BSS with which the mobile station is in contact

5.1. RR Functions

In this first half of this long chapter, we will study the different Radio Resource management aspects from the requirement side. While implementation issues will often be addressed, in particular for the distribution of the tasks between the involved machines, the details of the signalling procedure will be entirely in the second part.

The functions covered by the management of radio resources are centred on the management of transmission paths over the radio interface, and more exactly between the mobile station and the anchor MSC. To develop these functions, we will use the concept of RR-session, which will be presented first. After a small passage concerning the access and the paging, through which things start, we will deal with the handling of the main properties of the transmission chain, such as whether signalling, speecj or data is transports, and whether ciphering is applied or not.

The next issue will be on how handovers are decided. The handover execution itself will be treated mainly in the procedural section. Addressing the handover preparation issue will take us deep into considerations about the measurements performed by the mobile station and the base station. These measurements are the basic information upon which handovers can be decided.

Next, two ancillary functions of the transmission over the radio interface will be looked at, the management of the transmission power and of the timing advance.

Finally we will deal with the management of the radio channels on the radio interface as a whole set. The two main facets are the handling of the configuration of the radio channels in each cell, and the allocation strategy of the dedicated radio channels (TACH(8s and TACH/Fs).

5.1.1. The Concept of RR-Session

Most of the functions in the Radio Resource management plane relate to the management of the transmission between the mobile station and the anchor MSC. For each mobile station engaged in a communication, there exists a transmission path, as well as a signalling path, between itself and the anchor MSC. As seen from a mobile station, such a path is set up when it enters the dedicated mode (i.e. when it leaves the idle mode), and is released when the mobile station goes back to idle mode. In the infrastructure, a transmission path exists for all this period, but can be thoroughly modified, especially by handovers. We will refer to what is managed during this period of time as an RR-session. As a minimum, an RR-session must include means to transmit signalling between the mobile station and the anchor MSC through the BTS, BSC and a relay MSC, including a dedicated radio channel, references to manage it on the BTS-BSC interface and the BSC-MSC interface, and means in the BSS to the radio connection and take handover decisions when necessary. This set suffices only in the cases where the transfer of circuit-type user data is not such as for location updating, short message transfer or supplementary service When circuit-type user data needs to be transmitted, then a complete connection between mobile station and anchor MSC is also part of the RR- as shown in figure 5.1.1.

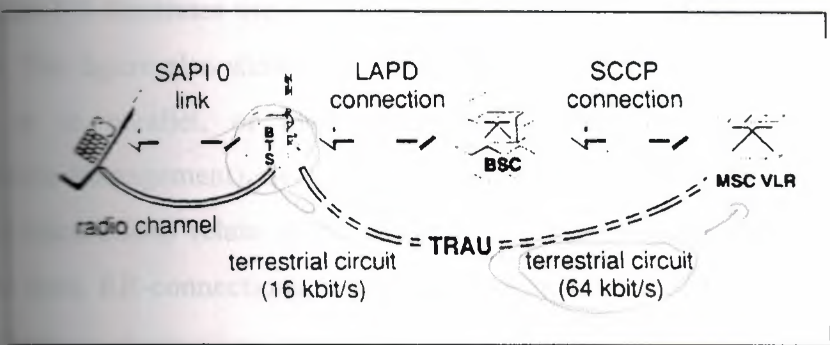


Figure 5.1.1. Contents of an RR-session

An RR-session contains both the signalling resources between mobile station and anchor MSC, included a dedicated channel on the radio path, as well as the user data circuits if need be

For instance, a speech call requires the use of a signalling connection as well as a speech-carrying connection between mobile station and NISC. This last connection makes use of dedicated resources such as the speech transcoder transforming the GSNI-specific speech representation into the 64 kbit/s representation used in fixed networks.

An RR-session has many different characteristics which have to be managed by procedural means. First, two different kinds of dedicated channels exist on the radio interface: they have been referred to as TACHIS and TACHIF in Chapter 4 (there will be three when 'half-rate' channels are included: TACH/8. TACH/H and TACHIF). Second. when a circuit for user data is present. it can be used according to different mission modes. Finally. some other less important transmission peculiarities cterise RR-sessions. An example is whether ciphering is applied or not. All these cteristics may change during the lifetime of an RR-session.

In the Specifications, the term "RR-connection" refers to what is managed g the period of connection to a given BSCT A change of BSC (e.g.. at inter-BSC lover) entails a change of the RR-connection. The Specifications do not have a ific term covering what is managed for the whole period where the mobile station is edicated mode between two periods of idle mode, i.e., what we call here an RR-ion. Figure 5.3 illustrates the concepts of RR-connection and RR-session, and their tionship. The figure also shows that an RR-session can be used for several calls in cession or in parallel, or more generally several CM-transactions (CM for nmunication Management), as will be described in Chapter 8. The beginnings and s of CM-transactions relate to the usage of the transmission, and are completely ependent from RR-connections, whose succession relates to the movement of the bile stations.

The RR-session is the bond between the two domains of radio resource nagement and communication measurement. It represents the views or the mobile tion and of the anchor NTSC. An RR-session starts when the mobile station goes to dicated mode (the access. when the initial assignment of dedicated channels is rformed). and disappears when the mobile station goes back to idle mode.

The life of an RR-connection is punctuated by intra-BSC handovers and anges of radio channels. and this defines anchor subdivision in the lifetime of a RR-ession. At the lower level. the channel connection corresponds to the continuous usage f the same radio channel by the same mobile station. A channel connection starts either rough an initial assignment. a subsequent assignment (a change of channel done. e.g.. ecause the allocated channel is no more of the needed type) or an incoming handover f any kind. It disappears either through the release of the RR-session, the assignment of nother channel or an outgoing handover. Channel connections represent the view of BTSs. By design choice, a BTS considers a change of radio channel within the same cell as two independent channel connections. When a channel connection is cleared, the related data is wiped out in the BTS, regardless of whether the mobile station is allocated a new channel of the same BTS or in another one.

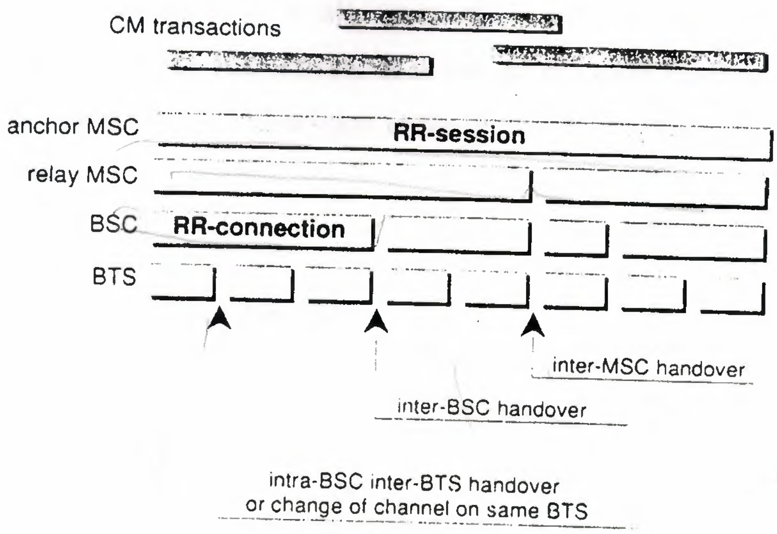


Figure 5.1.2. The concepts of RR-session and RR-connections

From the concept of radio connection (bottom line) to the one of RR-session (top line) different levels of transition awareness may be defined.

The specifications use the concept of RR-connection, which corresponds to the BSC view

CM-transactions may run in parallel or in tandem during the lifetime of an RR-session, as shown on the top of the figure

There are very few stable characteristics of the RR-session beside the corresponding contexts in the mobile station and in the anchor MSC, especially when one recalls that the physical path of the transmission may change thoroughly when a handover occurs. At a given moment in time, a given RR-session is managed by one BTS, one BSC, an anchor MSC and sometimes in addition a relay MSC as shown in figure 5.1.3.. Each of these machines maintains some context related to the RR-session. When a handover occurs, the configuration changes. some contexts must be erased and others must be created in other machines. Based on the corresponding configurations, the functional architecture of GSM distinguishes three kinds of handover. In an intra-BSC handover only the radio channel, the context in the BTS and possibly the BTS are changed. In an (intra-MSC) inter-BSC handover. the BSC is changed in addition to the radio channel and the BTS. Finally in an inter-VTSC handover. the relay MSC is either created, replaced or suppressed. In all cases. the anchor MSC remains in place throughout all the life of the RR-session. It is indeed the only machine sure to be a constant. and the context of the RR-session in the anchor MSC is indeed the anchorage point of the session.

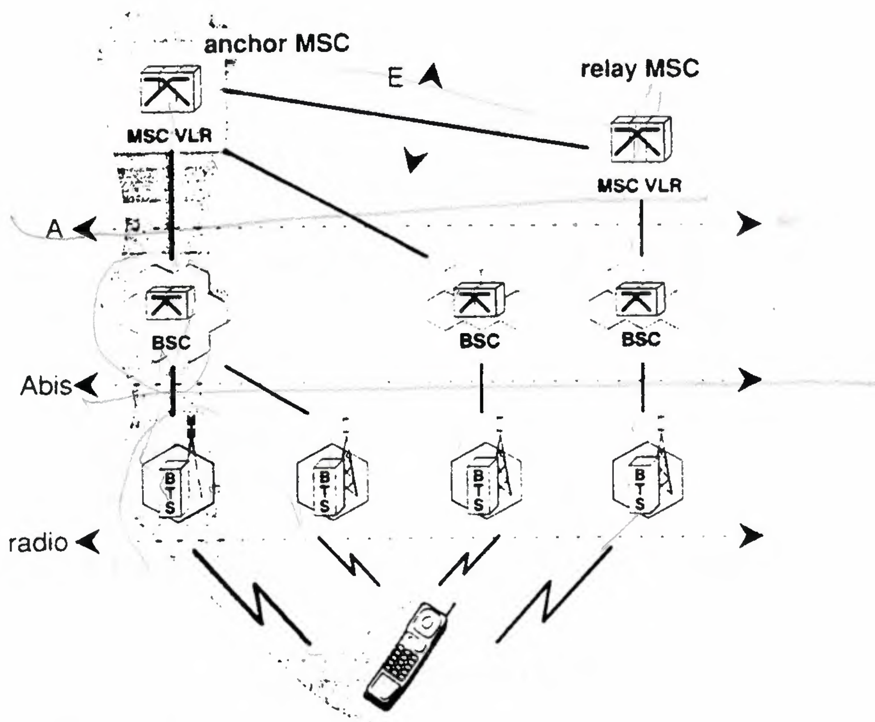


Figure 5.1.3. Configurations changes for an RR-session

During its lifetime, an RR-session may go through different transmission configurations If the initial configuration is for example the left one, it may be changed to any of the others shown. The anchor MSC remains in charge of the upper layers.

5.1.2. The Contents of the RIL3-RR Channel Request Message

The RIL3RR CHANNEL REQUEST message, sent on the RACH, is very short indeed. Its useful signalling information consists of just 8 bits! This capacity is obviously insufficient to carry all the information the mobile station would want to transmit, such as the subscriber's identity the reason for requesting a channel, the characteristics of the mobile equipment,... All of this information is in fact included in the "initial messages which will be the first information transmitted on the dedicated channel, once allocated.

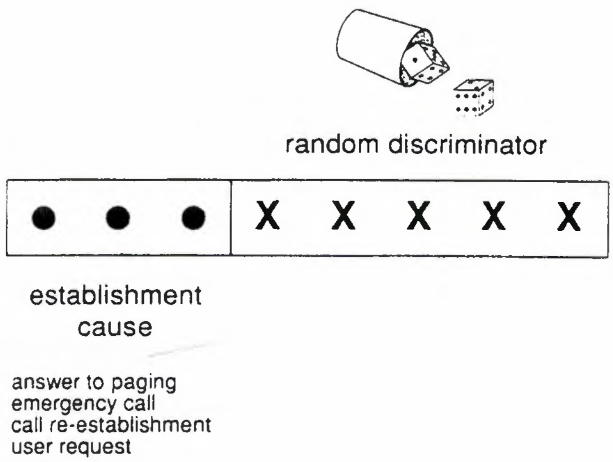


Figure 5.1.4. Useful contents of an access burst on the RACH

Only 8 bits are available in this short type of burst.
Three of them indicate the reason for the access, and the remaining five serve as a random discriminator

But the most critical usage of a discrimination between random access attempts is not to provide information to the network. A given mobile station must be able to correlate An initial assignment from the network with its own request. with as little ambiguity as can be achieved. For this purpose 5 bits among the total of 8 are chosen randomly by the mobile station. reducing drastically the probability that two mobile stations send identical messages during the same slot. which may in case of capture lead to an ambiguity as to which of the two requests is being granted.

Three bits remain, which are used as shown in figure 5.1.4. to provide a minimum indication of the reason for accessing the network. This first rough indication may be useful for discriminating rejections in case of congestion. and also to choose the best type of channel to allocate.

5.1.3. The Initial Channel Assignment

After the BTS has correctly decoded a channel request. it indicates it to the BSC through an RSNI CHANNEL REQUIRED message, with one important piece of additional information: an estimate of the transmission delay (this indication is critical to initialise the timing advance control). A field of unspecified content (PHYSICAL INFORMATION) allows the manufacturer to add more information. such as the reception level.

In normal load situations, the BSC then chooses a free channel (TACH/8 or TACH/F), activates it in the BTS, and, when the BTS has acknowledged this activation. builds an initial assignment message to be sent on the PAGCH.

The activation process requires the BTS to prepare for the access of the mobile station on the newly allocated channel. The timing advance is initialised based on the transmission delay estimate which the BSC indicates (back!) to the BTS. Even though this estimate .has been initially calculated by the BTS, this going back and forth between BTS and BSC is necessary, since the BTS has no means to correlate the messages received on the RACH with corresponding channel assignments.

The initial assignment indication sent to the mobile station on the PAGCH contains the description of the allocated channel. the initial timing advance to be applied. the initial maximum transmission power. as well as a reference allowing all the mobile stations expecting such a message to know whether they are being addressed or not. This last point is worth some more explanation.

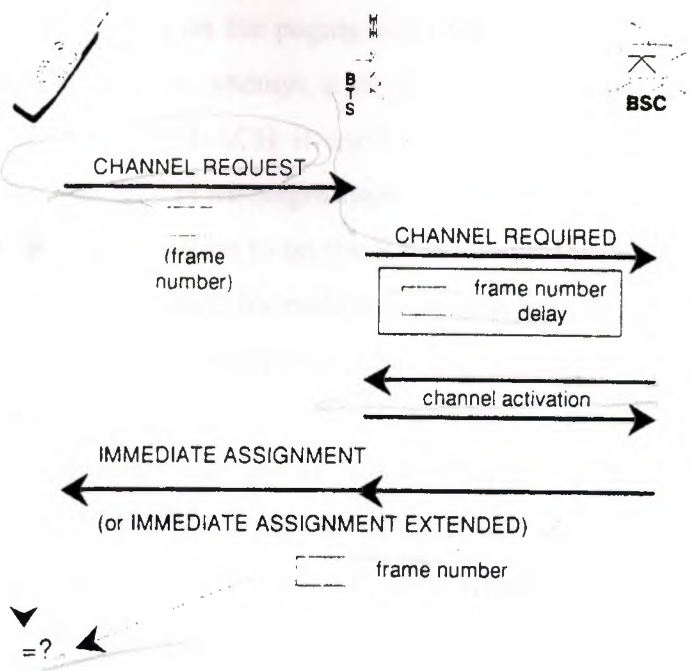


Figure 5.1.5. Initial assignment procedure

After the activation handshake on the Abis interface, the BSC prepares an initial assignment indication containing the 8 bit discriminator as received in the correctly decoded RIL3-RR CHANNEL REQUEST message, as well as the frame number in which it was received.

This enables the mobile station to check whether it is concerned with the message or not.

Addressing is done by including in the initial assignment indication the exact contents of the RIL3-RR CHANNEL REQUEST message which is being answered. plus the time reference of the slot in which it was received (such a time reference exists thanks to TDMA). This allows mobile stations to check whether they are actually concerned by each initial assignment. by comparing these values with the ones they have stored when sending the RIL3-RR CHANNEL REQUEST message, as shown in figure 5.1.5.

Besides. answers to RILS-RR CHANNEL REQUEST messages can be sent in any block of the PACCH. even on the paging sub-channels As a consequence. once a mobile station has made an access attempt. it should monitor the whole PAGCH (of the same Timeslot Number as the RACH it used for access) for an answer from the network. Furthermore. the BCCH messages must be decoded continuously during this period. in order for the mobile station to set the RACH control parameter values in real-time. This phase is very constraining for mobile stations in terms of reception (40 bursts every 51 x 8 burst periods). almost comparable to TACH/F reception.

Let us examine some side issues. It may happen that the reaction of the infrastructure to an RIL3-RR CHANNEL REQUEST message is too slow to avoid a repetition from the mobile station. In such inefficient situations a given mobile station may be allocated a channel twice (or even more times), since the infrastructure has no means of knowing whether an RIL3-RR CHANNEL REQUEST. Is the repetition of a previous one or not. The mobile station will use the channel allocated in the first initial assignment message it decodes, and the other ones will have been blocked for a few seconds in vain. Nevertheless, the Specification require that the mobile station be able to accept the network answer to any of it last three RIL3-RR CHANNEL REQUEST messages, in order to get service from such not-so-efficient BBS equipment.

In congested situations, when no channel is free for allocation, the BSC may choose not to answer to an RIL3-RR CHANNEL REQUEST message, or to send back a rejection indication. The first choice is not very efficient, since the mobile station will repeat its attempt. Explicit rejection is done through an RIL3-RR IMMEDIATE ASSIGNMENT REJECT message, containing a time indication during which the mobile station is forbidden to make any more attempts on the RACH (the WAIT INDICATION parameter). If the overload situation does not concern the RACH, the value of the WAIT INDICATION can be null, otherwise, it is a useful mechanism to help reduce RACH load.

The PAGCH is an important potential bottleneck of the system. In order to improve its efficiency, both initial assignment indications and channel request rejection indications can be grouped together to form messages. There are two assignments in an RIL3-RR IMMEDIATE ASSIGNMENT EXTENDED message, and up to four rejections in an RIL3-RR IMMEDIATE ASSIGNMENT REJECT message. Though this point is unclear in the Specifications, the original intention was for the BTS to perform the grouping and the basis of the individual indications provided by the BSC. The BSC has the possibility to provide the BTS with immediate assignment indications which are not ready-to-send messages (using the RSM IMMEDIATE ASSIGN COMMENT message). The BTS must then build the corresponding RIL3-RR messages. The fact is that the BSC can also perform the grouping and build the messages, and can provide ready-to-send messages to the BTS (this is, by the way, the only possibility for assignment rejection message). However, nothing really precludes the BTS to un-build these messages and to build others grouping the requests differently. The same debate exists for the paging indications.

In no case does the BSC schedule the transmission of the assignment indications: this is much easier for the BTS. The counterpart is that, despite grouping, congestion may happen. This is resolved simply by the BTS, which drops messages which are the excess of achievable throughput. While this resolves the BTS congestion problem, for the BSC it results in TACHs allocated but not "assigned". To avoid a worsening of the congestion situation through this effect, the BTS can indicate the non-sending of a message with an RSM DELETE INDICATION message. A second effect of the message is to indicate the overload to the BSC.

5.1.3. The Initial Message

Once it has received an initial assignment indication. the addressed mobile station modifies its reception and transmission configuration to adopt it to the frequency and time characteristics of the new channel. In phase 1, this new channel can be a TA4CH/S or a TACHIF, always in "signalling only" mode. The transmission level is set to a value broadcast on the BCCH (or to the maximum transmission level of the mobile station. whichever the- smaller). and the transmission starts with the timing advance value specified by the BSC.

The first thing the mobile station does on the new-channel is to transmit a link layer SABM frame for SAPI 0, i.e.. the frame used to establish in acknowledged mode the link layer connection for signalling messages.

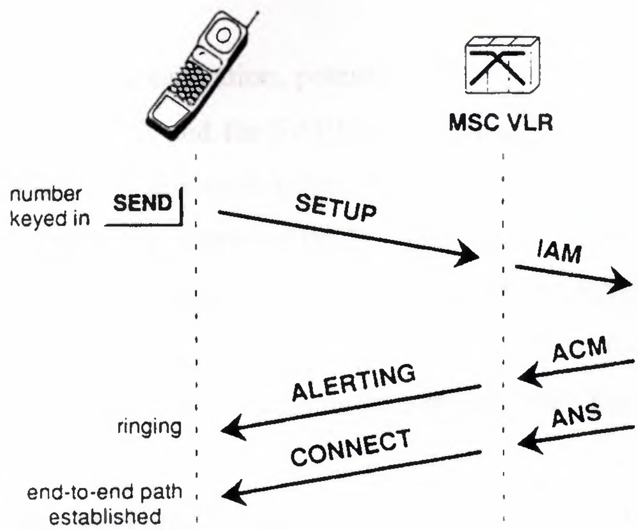


Figure 5.1.6. Contention resolution at link establishment

In rare cases, more than one mobile station may find itself on the same dedicated channel.

The transfer of a non-ambiguous initial message in the SABM-UA exchange allows each mobile station to know whether the channel is for its own use or not

In standard HDLC protocols, an SABM frame does not carry any information other than the one necessary for the link layer level. In GSM, the SABM frame sent within the initial access procedure contains a signalling message. the "initial message". The reasons for departing from standard usage are twofold. The first reason is efficiency, though this was not the leading cause. The other reason comes from the fact that the reference used in the initial assignment to address the mobile station is not fully unambiguous. It may indeed happen (though this is a rare case. around 1% in high load situations) that two mobile stations simultaneously send RIL3-RR CHANNEL REQUEST messages with exactly the same contents and that one of them is correctly received and answered by the BSS. The ensuing channel assignment will be understood by both mobile stations as their own, and both mobile stations will access the "dedicated" channel. Therefore, until the point where mobile stations identify themselves in a non-ambitious way, there is not a 100% guarantee that a single mobile station will access the channel.

As a consequence of this situation, potential collisions must he detected as soon as possible on the new channel, and the SABM-UA exchange provides this facility by including ("piggybacking") unequivocal information on these link layer frames. Mobile stations check the piggybacked contents of the UA frame. if one mobile station receives a UA containing something different from the contents of the SABM it sent. it must leave the channel and start the access procedure all over again; thereby enabling the "right" mobile station to stay undisturbed on its own channel (see figure 5.1.6.).

One obvious way to obtain an unambiguous SABM content is to use an identity unique to the mobile station. This would be enough to serve the purpose of collision detection. However, the efficiency criterion was also taken into account and the SABM includes more than just this identity. and includes a full "initial message".

The initial message comes in four different brands. depending on the reason why the access was triggered (see table 5.1.1.). All these messages contain an identity of the mobile station: the classmark, a held indicating some key characteristics of the mobile equipment. including the maximum transmission power: and complementary information making the reason for access more precise when need be. All but the first of those messages belong to the RIL3-MM protocol. The first belongs to the RIL3-RR protocol. but could have as well been put in RIL3-MM. If there had been enough room. these messages would have been properly formatted. with a part for the RIL3-RR protocol (including the classmark). and the rest in an RIL3-MM pan.

Table 5.1.1. Possible initial messages

Reason for access	Initial message
Response to an paging	RIL3-RR PAGING RESPONSE
Normal location updating, periodic location updating, "IMSI attach"	RIL3-MM LOCATION UPDATING REQUEST
IMSI detach	RIL3-MM IMSI DETACH
All other cases (call set-up, short message transmission, supplementary service management,...)	RIL3-MM CM SERVICE REQUEST

The different types of signalling messages may be used as the "initial message" depending on the reason having led to a channel request

There is usually but one choice for this "'initial message". One case of collision may arise, when. after having sent an RIL3-RR CHANNEL REQUEST message. the mobile station receives a paging indication and then the answer to its RIL3-RR CHANNEL REQUEST. What should the initial message be in such a case? Should it be consistent with the reason for sending the original RIL3-RR CHANNEL REQUEST message. or should it be the RIL3-RR PAGING RESPONSE message? It would seem fair to choose the RIL3-RR PAGING RESPONSE message if for instance the request from the mobile station concerned periodic location updating. but no priority scheme is defined in the Specifications, and the whole matter is left open for mobile station manufacturers.

Once an "initial message" has been received by the BTS (and sent back without any modification inside the UA frame), it is passed to the BSC in an RSM ESTABLISH INDICATION message. At this point. the mobile station classmark is stored for further use (e.g.. to choose the power control loop parameters). and the BSC then sets up an SCCP connection towards the MSC. This is done through an SCCP CONNECTION REQUEST message. on which the initial message may optionally be piggybacked. Only then does the MSC become aware of the contact with the mobile station. The initial message. whether piggybacked or sent after the SCCP connection establishment. is carried in a BSSMAP COMPLETE LAYER 3 INFORMATION message independently from its protocol (RR or MM). It contains enough information for the MSC to trigger required actions in the upper layers; (MM. CC. ...}. But this comes out of the scope of the access procedure.

When the access procedure ends, the RR-session is fully established with a complete signalling path between the mobile station and the MSC. With the establishment of the SCCP connection. the MSC takes control of the decisions concerning the transmission characteristics of the RR-session. and the BSS is at the ready whilst monitoring the transmission and performing handover decisions.

6. COMMUNICATION MANAGEMENT

Managing calls that is to say principally establishing and releasing transmission paths through meshed networks. is not by far a new topic. The development of fixed networks. mainly the PSTN. and the ISDN nowadays, has seen an important evolution of signalling techniques. Public cellular networks in general. and GSM in particular. are basically access networks for these general telecommunication systems. As such. the design of their communication management signalling is very dependent on existing techniques, and offers; few novelties. For instance. the signalling exchanges at the interface between GSM and the external networks are imposed by these networks. The influence goes even further, and the communication management signalling procedures defined between mobile stations and the GSM infrastructure are simplified and somewhat adapted copies of that specified for ISDN access. A part of what is described in this chapter is therefore not really specific to cellular networks, and could equally apply to fixed terminals. However. it is part of the Specifications and this description is included to complete the full picture of the system.

Still, call control is not inherited totally from ISDN. Cellular systems bring some problems of their own. due to the mobility of users and to the tact that there is no fixed link between each user installation and the infrastructure. The central issue is the establishment of calls toward users that move around. One facet of the problem is the way in which the system must follow the movement of users between calls, in order to readily find them when need be. The other facet is the routing of the call through networks to a GSM user. This is part of call control and will be detailed in this chapter.

The technical issues related to the establishment of calls toward mobile users is the reason for a number of novel approaches for telecommunication networks. One of these important new concepts is the notion of the GMSC (Gateway MSC, an ill-coined term from the conceptual point of view, since the "GMSC" function has nothing to do with the MSC function even if they are often grouped in the same equipment). which is the central actor in call routing. This notion is deemed to be important in the future, not only for mobile networks. but for all public telecommunication networks. Their future lies undoubtedly with the development of Personal Communication, i.e., the concept that calls are not directed to locations, as it is the case in PSTN and ISDN at the moment, but to people. who can be reached through different telecommunication means. according to their present whereabouts. Mobility is an essential point for Personal Communication. and the GSM technology, provides if not a ready-for-use solution. at least a case of pioneer work in the subject.

Another area where the signalling protocols developed for ISDN do not fit all the needs of GSM is the management of the radio interface. with the specific problems raised by the complex transmission scheme over the radio path. and by the movements of the users during a communication. As far as this chapter is concerned. the path between the MSC and the mobile station is considered to be a simple fixed link. This is again an application of the "divide and conquer approach. This kind of presentation allows a more consistent description not going this way and that between different subjects (on the other hand. the complete picture of how a call is established in GSNI can only be grasped by considering both points of view).

GSM
=
ISDN-like call control
+ Mobility Management
+ Radio Resource Management

Figure 6.1. GSM functional split

The split of signalling aspects into three major domains enables a description of a communication without taking into account the management of user's mobility and of radio resources

Thanks to this approach, illustrated in figure 6.1. the substance of this chapter can really focus on the notion of end-to-end communication.

The key "objects" that call control signalling deals with are communications, and we will organise the presentation in this chapter on how communications come into and out of existence, and how they change. We will first try to define what a communication is. and what are its attributes, in an attempt to identify the key functions of call management.

6.0.1. The Communication

Basically a communication is a temporary relationship between telecommunication users. for the purpose of exchanging information. A communication makes use of a transmission chain established through networks between users. The only communications of interest here are the ones involving at least one GSM user. A communication is by essence a "distributed" object. which exists over distances. Many of its characteristics must be managed co-operatively by the different machines which appear along the transmission path. This management concerns both static attributes, set at the beginning of the communication, and dynamic ones, which are modified during the lifetime of the call, usually at the request of the user.

6.0.1.1. The Partner Network

More generally. another important characteristic of the communication is the network to which the other party (other than the GSM user under consideration) belongs. This network will be referred to as the terminating network, for an MO communication, or the originating network, for an MT communication. GSM is designed to allow communications between its users and users from a variety of external networks, such as the PSTN, the ISDN, CSPDNs and PSPDNs. Establishment procedures in particular, and communication management in general will be heavily influenced by the nature of the other network. We will concentrate on the PSTN and ISDN cases, since they represent the most important part of the traffic by far.

Also important when involved in a communication link are transit networks. Many different network configurations can be imagined. There again. for practical reasons, we will address mainly simple cases. typically when two networks only are involved. one being a GSM, PLMN, the other the PSTN or the ISDN.

6.1. Call Control

Before describing the architecture and the procedures involved in call control, we will analyse how calls to a mobile subscriber can be routed, since this aspect underlies much of the network architecture, and is of foremost importance in understanding many aspects of the system. Such as, e.g., how to call a GSM user or charging

6.1.1. The Routing Of Mobile Terminating Calls

For a Mobile terminating call, the number given by the calling party does not refer to a telephone line or a location; but points to a record in some HLR. The first digits of a GSM directory number are sufficient to indicate that the number is a GSM number and furthermore to designate the operator with which the subcription is held. The structure of the GSM directory number, also called "MSSIDN" because it is part of the same numbering plan as ISDN numbers, is defined in CCITT recommedation E, 164 and is shown in figure 6.1.2. The HLR holding the record of the subscriber can be determined by the analysis of the first digits of this number.

The HLR record contains information necessary for finding the final destination of the call, i.e.. the MSC where teh GSM user is currently visiting. As a consequence, the final routing can be done only after the interrogration of the HLR. This splits the call establishment into two parts: before the interrogration, and after the interrogration. This corresponds also to a clear division of the call route into two parts: from the call originating point to the interrogation point, and the rest, as represented in figure 6.1.1.

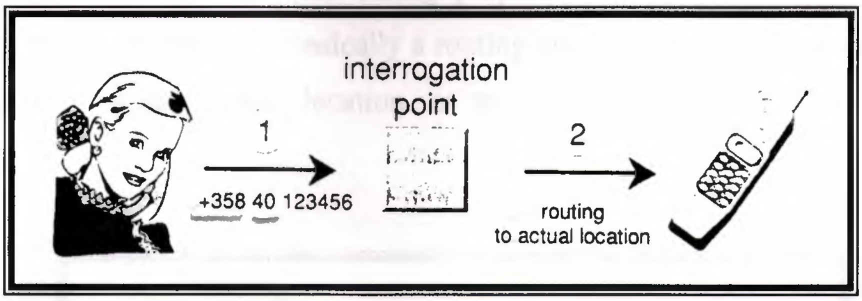


Figure 6.1.1. The two parts of a mobile terminating call routine

The call route consists of two parts: the first part is based on the called directory number, and continues to the point where the actual location of the called GSM subscriber is taken into account for the rest of the route

What follows applies to Mobile Terminating calls requiring the establishment of a circuit. The routing of Mobile Terminating Short Messages is similar, but bears a few differences and will be described separately in the section dealing with Short Message Services, Another point is that is specified in GSM applies mainly, if not only to the cases where the second part of the route goes through the PSTN or the ISDN not through a packet-switched or circuit-switched data network (PSDN or CPDN). An MT call coming from these latter networks must then either enter directly GSM before reaching the interrogation point or use the PSTN or the ISDN as a transit network. We will then restrict our presentation to MT circuit calls involving only general networks, such as the PSTN .ISDN or GSM .

The first part of the routing is done only with the information that can be derived from the called number (the MSISDN) independently from the called party location. This routing is done as for any ISDN number with tables in each of the intervening switches. The routing tables are normally set so as to reach rapidly the switch which is able to interrogate the corresponding HLR. Not only does such a switch include the software necessary for running the interrogation procedures but it also holds a table which relates an MSISDN with the corresponding HLR. This function is referred to as the GSM (Gateway MSC)function ,and its role is shown in figure 6.1.3. The interrogation of the subscriber (for billing purposes),and the information for next routing step. This information is basically a routing number pointing either to the called GSM subscriber in his current location .or to a third user in the case of call of forwarding.

CC (country code)	NDC (national destination code)	subscriber number
+44	802	UK GSM number (Cellnet)
+44	385	UK GSM number (Vodafone)
+44	956	UK DCS 1800 number (Mercury P.C.)
+44	973	UK DCS 1800 number (Hutchinson Microtel)
+358	40	Finnish GSM number (Telecom Finland)
+358	50	Finnish GSM number (OY Radiolinja)

Figure 6.1.2. The structure of a GSM directory number

A GSM "MSISDN" looks like a standard PTSN or ISDN number, but the knowledge of the "National Destination Code" (NDC) identifies an operator within the country, and not an area code.

The first digits following the NSC are used to identify the relevant subscriber's HLR within the home PLMN.

This simple description raises several interesting. First what is exactly a switch capable of the GMSC function? Second where does the routing number come from? Third, who pays for what ? The last question being the key to the first one, let us start with the charging issue.

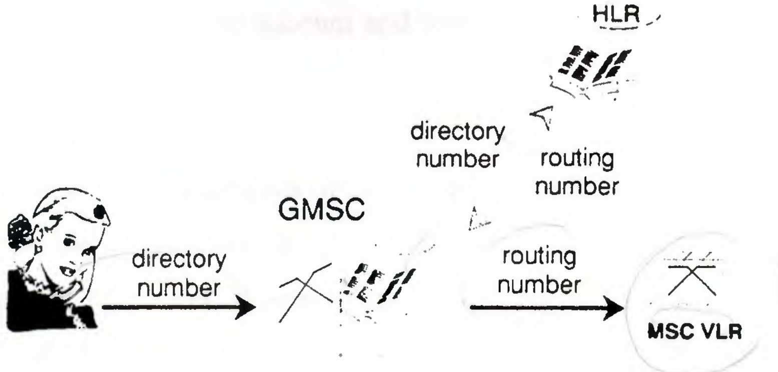


Figure 6.1.3. The key role of the GMSC

A mobile terminating call is first routed towards a GMSC.
i.e., a switch able to interrogate the mobile subscriber's HLR to know where to actually route the call

6.1.1.1. Who Pays What?

The charging and terrifying policies are to a large extent outside the realm of the Specifications. They are an operators' issue, dealt with in the GSM MoU meetings. However, the answer has some consideration. The charging principles which follow have been extracted from various public conferences of the GSM MoU operators.

In the case of a non-forwarded call, two parties are involved and may be charged: the calling party (say Peter, who is a subscriber in the Netherlands, but who happens to be currently travelling in Spain and obtains service from the Spanish TELEFONICA GSM network). Several networks are entitled to demand some part of the call charge. These are the German PSTN, the transit PSTNs or ISDNs along the way, and the PLMN of TELEFONICA which will establish the final segment of the path. Woldemar will receive his bill from the German PSTN, and Ptere, if any charge is levied on him will be billed by his home PLMN, i.e. in the Netherlands and in Dutch currently. The other networks are not able to bill these subscriber directly. The matter is then not simple. And accounting transfer between some of the involved networks will be necessary, as well as specific agreements between the operators. Some practical considerations must be taken into account and some rules must be fixed to simplify the combinations.

The total cost of a call depends obviously on the location of the GSM subscriber. It is also clear that the calling party would like to know in advance how much he will be charged for the call. Moreover, it seems that GSM subscribers are not willing to have anybody calling them know where they are located, even with little accuracy. These arguments lead to the principle that the charge levied on the calling party is independent from the actual location of the called party. This philosophy is in line with the similar problem of a forwarded call in the PSTN; in this case, the calling party is usually levied the same charge as if the call was not forwarded, and the party which asked for the forwarding is charged for the forwarding is charged for the complement.

This principle still leaves room form different solutions. In a first extreme solution the rate of the charge levied do the calling party can be designed so that on average it covers the cost of the call, wherever the location of the called party (and including the charge for the radio segment); and the called party never pays anything. Another extreme, and opposite, solution consists in charging the calling party with only the minimum charge (the minimum being taken on all possible locations of the called party), and have the called paty pay systematically for the rest or the cost.

But levying a charge on the called party is not particularly easy since it means raising charges on calls that the user would sometimes have preferred not to receive. In fixed networks and in most cellular networks in operation today, the entire charge is supported by the calling subscriber (except in the case of forwarding as explained above). It seems then good practice to do the same when the subscriber is located in his home PLMN (in our example if peter was still in the Netherlands). i.e.. to charge only the originator of the call. This will still require compensation mechanisms between operator, but in a more deterministic way.

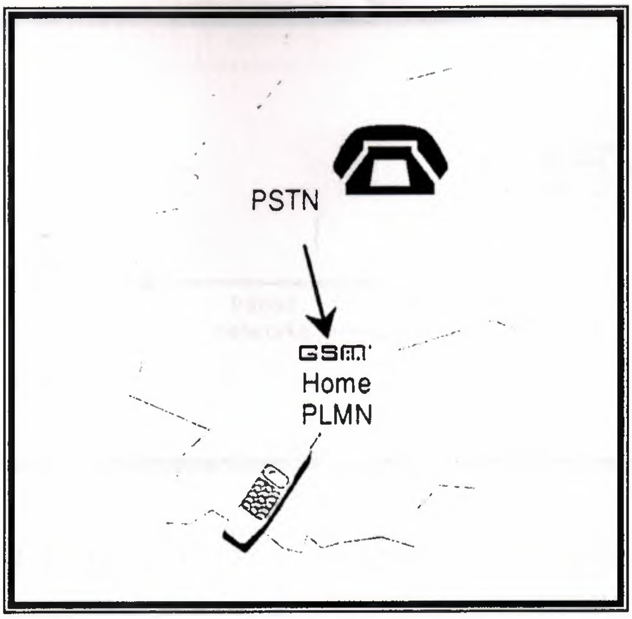


Figure 6.1.4. Routing of an MT call within one country

Ehrn the called GSM user roams in his home PLMN, a call from a fixed user in the same country is simply routed to the home PLMN through e.g. the PSTN

Let us consider several examples to understand better how this principle applies. Starting with the simplest case, let us see what happens when Woldemar (using the PSTN) calls a German GSM subscriber, say Frieder, who is roaming in his home PLMN. Figure 6.1.4. illustrates the case, Woldemar is the only one charged. The amount of compensation paid by the German PSTN to Frider's home PLMN, if any, is a matter of agreement between the operators. It seems logical that this compensation should cover the costs concerning the "radio" segment. (MSC to MS) at least, and possibly more depending on the point of entry in the GSM network. However, since for Mobile originating calls the situation is reversed (the charge is recovered by the GSM operator, who should give some money to the German PSTN for routing the call toward the called user), there is not necessarily a positive flow of money at the end of the month from the PSTN operator to the GSM operator! In any case, the German telecommunication networks taken as a whole cannot obtain any more money for a Mobile Terminating call than what is levied on the calling user by his PSTN.

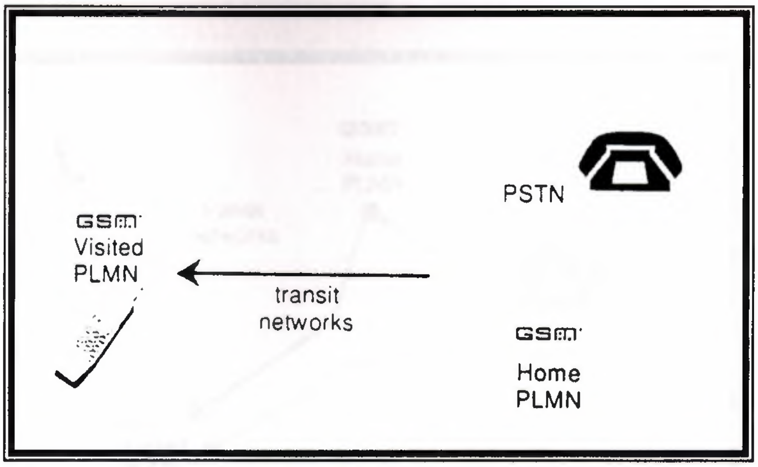


Figure 6.1.5. MT call towad a roaming subscriber

The calling party is unaware of the actual location of the called subscriber (for him, everything is the same as infigure 6.1.4.)
The cost of the international leg can be recovered by levying a charge on the called GSM user.

Let us now offer a trip to Frieder, say in Protugal. Woldemar still dials the same number to call Frieder (i.e.. a German number), and the interrogation point (the GMSC) automatically re-routes the call from Germany to Portugal, without Woldemar being aware of it (principle of location confidentiality). He should then pay the same charges as in the previous case. But more networks have been involved in the call, which now includes an international leg, as shown in figure 6.1.6. the standard compensation mechanism in use between fixed networks for international routing will apply to the German PSTN, so that the transit networks get their share. In such cases, an additional charge to cover the international leg will have to be raised, and it can be levied on nobady but Frieder if the princpile of location confidentialty is retained.

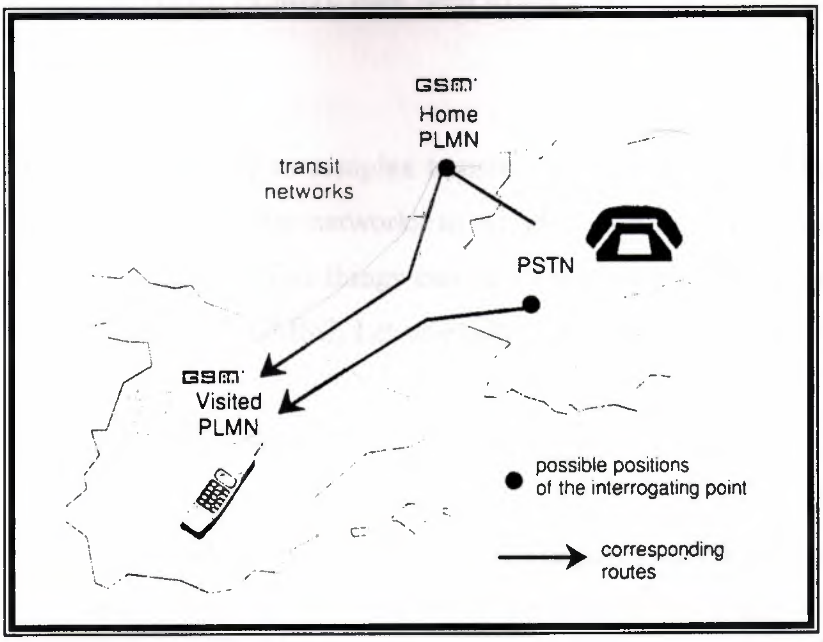


Figure 6.1.6. MT call involving three countries

Depending on the location of the interrogating point, the call may be routed through one or several international legs, without the calling user being aware of it

Let us now take the most complex case, where the originating network, the home PLMN of the called user and the visited PLMN are all potentially in different countries. This corresponds to our original example, when Woldemar calls Peter (GSM subscriber of the dialling an international number starting with 31 (the country code for the Netherlands), and as such is prepared to pay the cost of an international call towards the Netherlands, and possibly at little more if he knows that the number he is calling is the one of a mobile subscriber (and if the German PSTN also knows this information, which is not by far an obvious matter). Depending on the location of the interrogation point (the GMSC able to interrogate Peter's HLR in the Netherlands to know where he is located), the call may experience one or several international legs as shown in figure 6.1.6. Besides the German PSTN, which gets the money from Wolderamr, Peter's home PLMN may charge him for some amount linked with his current location, but it is the GMSC which holds the relevant information on the routing cost of the second segment o2f the call (from GMSC to VMSC). As for the other networks through which the cal is routed (including the visited PLMN), they need to get some compensation for the costs incurred.

All this leads obviously to complex transfer mechanism. The GMSC appears as an important participant, and the networks to which it belong heavily influences the complexity of the problem. In fact things can be simplified if some choices are made concerning the position of the GMSC, Let us examine this issue in more detail.

6.1.1.2. The GMSC Function

GMSC

GMSC the GMSC function requires only a switching capacity, and special software (including the ability to establish a toll ticket for the second branch of the call). It must hold a table linking MSISDNs with HLRs. There is no reason why the table should be complete. Indeed . a given GMSC may fulfil its function only for the subscribers of some Home PLMNs or even for a single Home PLMN, As will be seen, a simplification is to have the GMSCs specific to one PLMN.

By nature, the GMSC function is independent from the radio access function provided by a PLMN, and can be implemented as a part of any network through which the call is routed, typically the PSTN or the ISDN, as specified in the Specifications. The GMSC function can be implemented in any switch of the PSTN, the ISDN or directly connected to those networks. From a service point of view, the greater the density of GMSCs the better. Indeed, the closer the interrogation point to the calling avoided is what is called in technical jargon a "trombone" that is to say the case where the second branch begins by backtracking the path of the first branch. This is obviously not efficient and is prevented if GMSCs are everywhere.

This ideal view is not so ideal when taking into account the charging considerations. The billing record established by the GMSC covers the second segment of the call (from GMSC to called party). This record must be transferred towards the home PLMN of the called subscriber, in order to bill him the amount not covered by the fixed charge of the calling subscriber. A first problem is that this transfer of biling records must be organised with all Home PLMNs for which the GMSC is able to interrogate the HLR. A second problem is that the GMSC if the GMSC is not in the home country o the called party, a part of the charge for the second branch may have to go to the calling party if we want to follow the principles developed in the previous section. An important simplification for charging management is obtained if the first leg is always charged to the calling party . now, applying this principle is in contradiction with the search for an efficient routing. As explained in the previous paragraphs, in a network configuration optimised for routing there should always be a GMSc close to the calling user, and he would not pay for, e.g.. the international path if the called user is a foreigner. To solve this dilemma the choice was made in favour of simplifying the aim the interrogation function for the subscribers of a given PLMN is performed only by switches of this PLMN, the GMSC is always in the home PLMN (and hence in the home PLMN country), and the two segments of the call can be considered as "from calling party to home PLMN", and "from home PLMN to visited PLMN" and each charged to one end of the rouet. The resulting principles affear in figure 6.1.7.

Let us consider the consequences of this approach. If the called party is in his home PLMN (Peter in the Netherlands) the second branch of the call is reduced to a minimum, and this is consistent with not levying the called user in this case (Woldemar pays the whole amount and peter is not charged). Conversely when the called party is roaming his home PLMN to his actual location. This also does not raise an charge is easy to predict (it consists in the price of an international call between the home PLMN country and the visited country). Moreover, a GSM subscriber has the possibility to ask that incoming calls are not completed when he is roaming abroad: to do that, he must activate the supplementary service "Barring of Incoming Calls when roaming outside the home PLMN country" (BIC – roam).

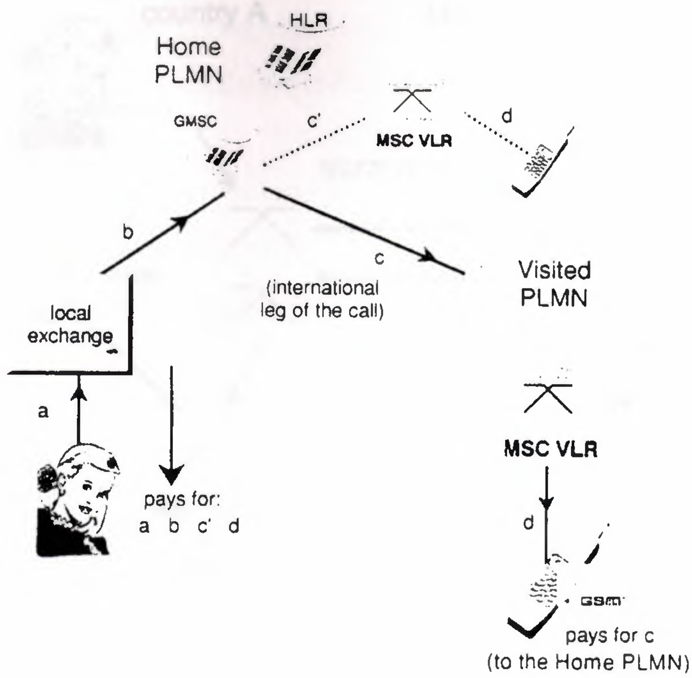


Figure 6.1.7. Charging principles of mobile terminating calls

The charging principles retained by GSM MoU operators are based on the following principle; the originator of the call is charged as if the called GSM subscriber was in his home PLMN.

The international leg, if any, is paid by the called GSM user. Conversly the GSM subscriber is not charged for incoming calls when located in his home PLMN.

The originator is unaware of the actual location of the GSM subscriber, and will always know how much she will pay.

Call charge transfer is much simplified. The same network (the home PLMN of the called subscriber) generates the charging information concerning the second portion of the route, collects the corresponding charges if any from its subscriber, and receives the invoice of the other intervening networks further down the call route, for compensation.

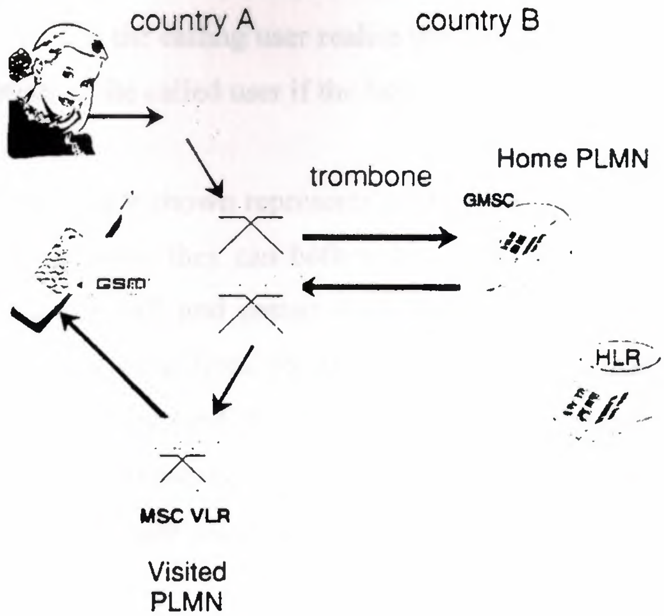


Figure 6.1.8. The “tromboning” effect

When the originator of the call and the mobile subscriber happen to be geographically close to each other, a “trombone” may appear in the routing of the call when the home PLMN

When the originator of the call and the mobile subscriber happen to be geographically close to each other, a “trombone” may appear in the routing of the call when the home PLMN of the mobile subscriber is in another country.

It consists in routing the call back and forth between two countries, leading to two of international legs instead of a national (or even local) call.

The only problem with this approach is that “tromboning” is not suppressed. Routing optimisation can still be achieved on a national basis, if the density of GMSCs is high, but this is not possible when the calling user and the home PLMN are in different countries. The lack of efficiency is obvious when the called GSM subscriber is roaming in a foreign country A and is called by somebody from country A possibly a few meters away. Then the call is routed to the home PLMN country and back as shown in figure 6.1.8. Two international calls are then established where a local route would have been sufficient (this is the problem of "tromboning"). The two international calls are being paid for (one by the calling party, the other by the called party), and only in a small number of cases will the calling user realise the actual situation, since he remains unaware of the location of the called user if the latter does not tell him explicitly.

Of course the example shown represents an extreme and fairly marginal case. If both users realise the situation they can both save on their bills by having the mobile user immediately end the call and restart from his end, since all the problems of subscriber location do not exist from the mobile user to fixed user. More generally speaking the lack of efficiency (and therefore the additional cost) in certain situations such as the one described represents the price to pay for location confidentiality in an international environment where biling and accounting is a complex and constraining issue.

To conclude this long discussion, let us just recall the result; the GMSC function, though designed to be used in a versatile way wiyy be at least in the first years of GSM always co-located with an MSC, and limited to the interrogation of the HLRs inside the same PLMN.

6.1.1.3. Where Does The Routing Information Come From?

The answer to the question of where the routing information comes from is strongly linked to how subscribers are identified in GSM, as well as in ISDn and PSTN, for routing purposes. We have already encountered the MSISDN, which is the "directory number" used to call GSM subscriber The MSISDN is part of the E.164 numbering plan. Also part of this numbering plan is the MSRN, or Mobile Station Roaming Number, which is the routing number used on the second leg of an incoming call between the GMSC to visited MSC: The MSRN is not visible to GSM users or to calling parties, but is used solely between infrastructure machines. It is not allacated permanently to a subcsriber, and it is geographically integrated into the numbering plan of the fixed networks, since its purpose is for routing towards the visited MSC.

A third type of identity, is the IMSI which is used as the main subscriber key in the GSM location databases, Both MSISDN and IMSI contain an identification of the country and of the network within this country. Table 6.1.1. (overleaf) gives a few examples of the correspondence between the "country code" (CC) of the international telephone (or ISDN) numbers and the "mobile country code" (MCC) of the IMSI.

MSISDN			IMSI		
CC	NDC		MCC	MNC	

Country	CC	MCC
Denmark	45	238
Finland	358	244
France	33	208
Germany	49	262
Italy	39	222
The Netherlands	31	204
Norway	47	242
Portugal	351	268
Spain	34	214
Sweden	46	240
Switzerland	41	228
UK	44	234

Table 6.1.1. Correspondance between CC and MCC for a few European countries

The country code (CC9 known to subscribers in the telephone numbering plan is defined in Recommendation CCITT E.164 whereas the mobile country code (MCC) used in the first few digits of the IMSI is defined in Recommendation CCITT E.212

The problem now is how the GMSC gets the MSRN to point to the MSC where the subscriber is located. To sum up. The HLR stores within each subscriber record some location information, including at least an address of the visited MSC/VLR which can be used by the SS7 signalling. The HLR record may also include an MSRN usable for the routing of the second branch of the call, if the visited MSC/VLR has provided such a roaming number when updating the location information in the HLR. In this case, the answer to the interrogation can be readily given, but when the record Is limitet to the visited MSC/VLR address the HLR has first to interrogate (using this address) the visited MSC/VLR to get the routing information.

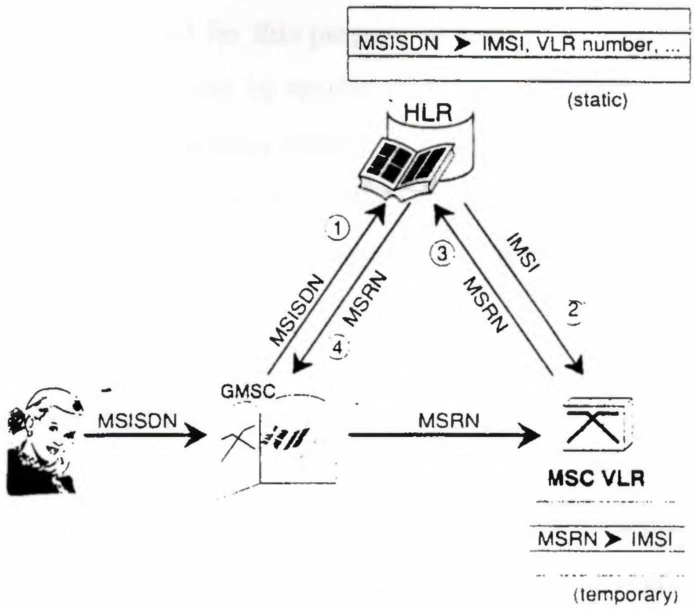


Figure 6.1.9. The provision of the MSRN

The HLR requires the visited location register to provide an MSRN which will be used to route an incoming call towards the correct MSC

The flow of information is shown in figure 6.1.9. When receiving the message, the visited MSC/VLR chooses the roaming number from a pool of free numbers, and links it temporarily with the IMSI. When the call eventually reaches the visited MSC using the roaming number as the address, the MSC cn retrieve the IMSI from its records, and can go ahead with the establishment of the call towards the mobile station. The roaming number can then be freed as soon as the call is fully established.

The two scenarios for the provision of the MSRN (available continuously at the HLR or provided by interrogation at each incoming call) triggered a long debate in the specification committees. The issue is now settled in favour of the provision of the

MSRN on a call Per call basis, though both solutions still appear in the phase 1 specifications.

One of the fundamental arguments is the consumption of directory numbers. If the MSRN is allocated to a subscriber for the whole duration of its stay in the VMSC, the quantly of numbers reserved for this purpose in each MSC must be about equal to the number of subscribers who may be registered at the same time under this MSC. By comparison. A call Per call allocation requires a quantity of numbers which are about equal to the number of simultaneous call establishments by the MSC a problem is taken into account (see next section). In the following only the call Per call allocation of the MSRN is considered.

6.1.1.4. The Problem of Multi service

GSM is designed to provide different telecommunication services, most of them such as speech fax or data bearer services, can be provided though the PSTN or the ISDN. A GSM mobile station, or an ISDN terminal, is designed so that the user can indicate which service it requires amongst the supported ones. This is not so easy for a calling party using a PSTN access line. The problem is for instance how to make the network understand that a fax call is requested instead of a speech call? In addition. even if the information is provided by a GSM or ISDN calling party, an intervening PSTN is unable to carry it. The problem exists if any part of the route is via the PSTN.

A solution exists if the first branch does not include any analogue PSTN segment. This applies if the call is issued from GSM or from ISDN. and if the GMSC can be reached via ISDN using SS7. Then the information can be transferred up to the GMSC. and from then on to the HLR. and thence to the VMSC (in substance bypassing a possible passage of the second branch through the PSTN).

The problem remains when the only information received by the GMSC is the called number (the N4SISDN). Two solutions of general application have been put on the table. and were the source of intensive debates. The first solution consists in letting service be chosen bv the called party. The message setting the call from the network the mobile station does not specify the service. and the mobile station indicates it in urn. This solution imposes the requirement that the service is set by the user in the obile station before the actual start of the communication. A typical scenario to send a x to a GSM subscriber is then first to phone him (speech communication). asking him set the mobile station so that the next call will be treated as a fax call. then hang up d re-dial to establish the fax call.

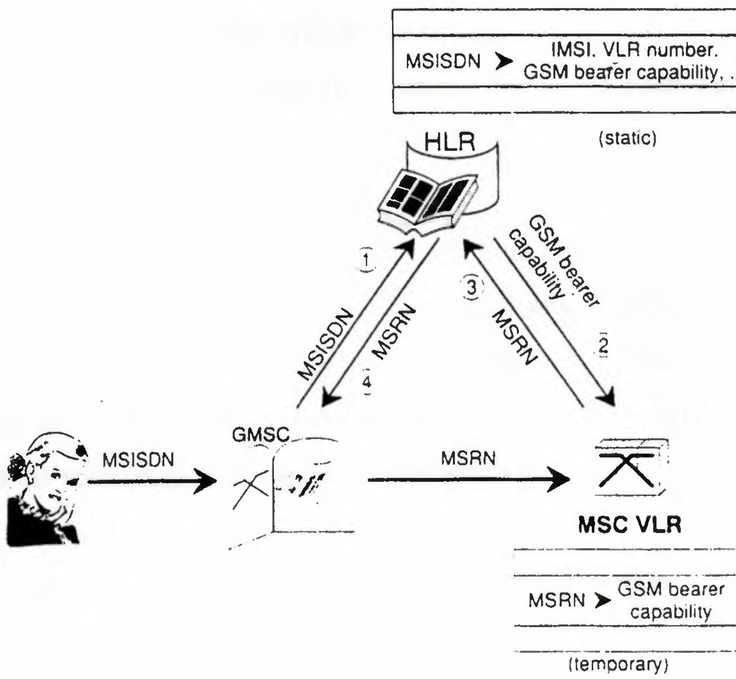


Figure 6.1.10. Routing of a service information

Vhen an incoming call arrives from the PSTN, the only way to convey the information n the service (based on the dialled directory number) up to the visited MSC/VLR is to carry a service profile (bearer capability) between the HLR and visited MSC/VLR.

This service profile is stored in the HLR in association with each MSISDN

This solution has minimal impact on the network. but is not very convenient for e users. This is why an alternative solution was proposed. consisting in providing a SM subscriber with as many ~1S1SDNs as services for which he wishes to receive coming calls (for instance a speech number and a fax number). The service can then be chosen by the calling party, by using the right number. The relationship between numbers and services is held in the HLR. The next issue is to convey the information to the VMSC. This is simply done in the procedure used to get the roaming number, which allows once again to bypass the networks intervening between the GMSC and the MSC/LR. The HLR sends the reference of the requested service to the MSC, which stores it against the provided roaming number. When eventually a call with this roaming number reaches the VMSC, it proceeds with the call establishment for the requested service. This scheme is shown in figure 6.1.10. Its only drawback is the consumption of directory numbers for the MSISDNs, which can be unacceptable in some countries. On the other hand. it enables the provision of a service no worse than that which PSTN subscribers are familiar with.

The two schemes are provided for and both must be supported by all MSC/VLRs. The choice of the scheme is done by the Home PLMN of the called party. When no service information is provided by the HLR, the MLSC/VLR can either choose the service or let the mobile station decide. The first approach is acceptable only in the case when the subscription is limited to one service.

6.1.2. The Mobile Originating Call Establishment Procedure

As seen by the calling user. the establishment of a communication follows a number of steps. which he perceives through displayed or audio information. The rough sequencing for basic telephony is well known by anybody. An originating call starts by the user lifting the receiver, an action which the network answers with some tone. Then the called number is keyed in. Once the last digit is entered. there happens in some cases nothing for some time; in other cases. a waiting tone can be heard. Eventually'. an answer comes. It can be a busy tone. indicating that the called party is already engaged in a communication. an announcement stating for which reason the network was unable to fulfil the request (e.g.. congestion. or non-existent number). or. sometimes. a tone indicating that the called party is being alerted. Then. in the last case, it may happen after a while that the other party lifts the receiver, and this is perceived by the speech path being connected. thus allowing the person to person discussion to take place.

Cellular telephony does not introduce many modifications to this basic scheme. The main distinction is that the number is keyed in before establishing the contact with the network. It is in addition displayed, and a very useful consequence is that it can be modified if need be before transmission. Another difference, applicable to multi-service networks such as GSM, is that some additional information may be exchanged between the user and the network at the beginning of the call. such as the type of service or in the future the type of channel. In most of the cases, these issues will be dealt with automatically by the mobile station. with default or implicit values.

For all these interactions, the mobile station intervenes between the user, with which it exchanges messages according to a man-machine protocol. and the infrastructure, with which it exchanges, information by electronic means. As far as Call Control is concerned. these two flows of information are for the important parts in a one to one correspondence with each other: The mobile station acts as a protocol translator. It receives orders from the user in the form of key pressing for instance. and translates them into signalling messages for the network according to the RIL3-CC protocol. In the other direction. the MSC provides its answers or indication to the mobile station in the form of RIL3-CC messages. and the mobile station translates them into signals understandable by the user. such as audio signals (tones) or visual signals (lights. alphanumerical display. ...)

This modern approach. which is also the one of the ISDN (from which the RIL3-CC protocol has been derived). is different from the one still used in the PSTN. In the PSTN, all the information provided to the user are in the form of audio signals (tones, announcements), and are generated by the network. There is no digital message to be found coming from the network to the telephone set. Inside the PSTN. the transmission of the information ultimately destined to a user may be done in the guise of messages (digitally encoded), in which case the translation in tones or announcements is done at the end. by the switch in charge of the user line. Yet there are still old machines not supporting this scheme. and then the tones and announcements can come from switches far away from the user. the role of the local switch being limited to their transmission. as during the actual communication phase. The consequence for GSM (and for ISDN) is that an escape mechanism ("progress" information) must be provided so that the dialogue between the user and the network can be done in the old fashioned way.

There is yet another approach, which is specified (besides the "functional" protocol described above) in the ISDN terminal to network protocol for simple terminals. and is called the "stimulus" mode. In this mode. the network has the full detailed control of the signals (display or audio) provided to the terminal, and instructs it through simple messages. directly giving the signals to be provided to the user. This approach is not used in GSM,. but some traces of it can be found. in the guise of the SIGNAL information element. It should be understood that this is not a full stimulus protocol, but a modified usage of the information element to qualify a message.

We will describe first the modern scheme, using messages, and then address the escape mechanisms.

The RILB-CC protocol between the mobile station and the MSC is entirely specified in the Specifications, but the man-machine interface between the user and the mobile station is left very much open. Similarly, the interworking between the RIL3-CC protocol and the man-machine protocol is not specified in detail, though the functions to be fulfilled are often described in the RIL3-CC specifications. The description which follows corresponds to an invented man-machine protocol. derived from the knowledge concerning existing terminals. It does not cover, by far, all the possibilities. This area is one of the most important for the marketing of GSM terminals (and GSM), and it is desirable that terminal manufacturers have many ideas about the man-machine interface and protocol. to help and guide the user through the tortuous route of GSM. Eventually sophisticated methods will appear. such as voice recognition. enabling the user to give his commands orally. menu driven protocols. and other let to be imagined features making the use of technique more convivial for human beings.

6.1.2.1. The Signalling Flow for a Mobile Originating Call

Let us take the example of a GSM user (Ansgar) calling a fixed user (Remi). In the first part of the call establishment. Ansgar indicates. through the mobile station. what he wants (directory, number of Remi, type of service requested-speech. fax. ...). and the mobile station will pass the information to the MSC. In GSM, as in ISDN. the user has to provide this basic data before the mobile station contacts the network. This can be done in different ways. depending on the terminal. but also on the type of service. For instance. for emergency calls. this can be reduced to the pressing of one or two keys (typically an "emergency" key and then a "send" key). Some effort was made to prevent the keyed information depending too much on the place where the user is. One of the issues is the format of the called number, and we will come back later to this topic.

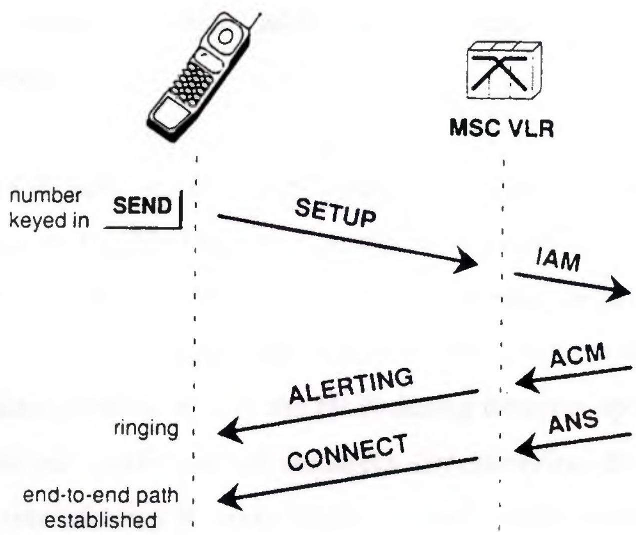


Figure 6.1.11. Mobile originating call establishment

A successful MO call establishment leads to a signalling sequence as shown, each step being related to the following events :

- the man-machine interface between user and the mobile station;
- the protocol between MSC and external networks (ISUP for example)

Explicitly giving the characteristics of the service will often be skipped. For example, for a speech only terminal, speech is obviously the default service. and then has not to be explicitly requested. Typically, Ansgar enters first the called number. which is shown on a display. may correct it. and then presses a "send" key. Only when the "send" key is pressed does the mobile station trigger the effective establishment phase with the network. shown in figure 6.1.11.

Then the mobile station runs the procedure needed for call establishment. This starts by the access procedure (RR level). and the MM connection establishment. Their effect as far as Call Control is concerned here is to establish a suitable signalling connection between the mobile station and the MSC. Then the mobile station sends to the MSC an RIL3-CC SETUP message or an RIL3-cc EMERGENCY SETUP message which is the translation of the Ansgar 5 request. This message contains Remi's number and a description (which can be long and detailed in the case of data services) of the required service. It may contain additional information. in relation with the supplementary services.

When the MSC receives the setup message, it analyses the request and checks whether it can accept it. Whether it is accepted depends on the capacity of the MSC/VLR to provide this service (in a compatible way with the mobile station capacity) on Ansgar's subscription characteristics (this is determined locally thanks to the subscription characteristics sent by the HLR during location updating, and stored in the MSC/VLR), on the availability of resources (interworking devices, a free circuit with the external network etc) If some of these checks fail, the call establishment is aborted by sending an RIL 3 CC RELEASE COMPLETE message to the mobile station before the release of the lower layer connections returning the mobile station to the idle mode. Of everything is network (for instance by sending an ISUP INITIAL ADDRESS message (IAM) in the case of a communication with the ISDN) and on the other hand sends an RIL CC CALL PROCEEDING message to the mobile station which indicates simply that the request has passed the MSC test, and that the MSC is proceeding with it, More pragmatically. It tells Ansgar (if so indicated by some signal on the man-machine interface) to be patient.

Sooner or later, the MSC will receive from the external world a report of the requested call establishment as seen by the switch in charge of the called party. Such a report may indicate that Remi is being alerted indication) because it failed for some reason (congestion. Or because Remi is already busy, or not reachable. Etc.) In the case of ISUP the report of a successful alerting takes the form of an ISUP ADDRESS COMPLETE message. The MSC reacts respectively by passing the instance to an alerting tone by the mobile station or by aborting the call establishment. If an RIL 3 CC CALL PROCEEDING message has been sent before the abortion is done by sending an RIL 3 CC DISCONNECT message, which will be answered by the mobile station with an RIL 3 CC RELEASE message, it self acknowledged by the MSC with an RIL 3CC RELEASE COMPLETE message (explanations on these exchanges will be found in the section concerned with call release). Only then can the lower layer connections be released, if not used for some other context.

Some time can then elapse before the end to end answer i.e before the acceptance of Remi (lifting the receiver in the case of plain telephony) At ti0hs stage the mobile station (and Ansgar) is still waiting for the result of the request. The ball is on Remi side, who may answer or not. In the case of a no answer, the call is aborted by the network after some time (SAY 3 minutes), provided Ansgar has not decided to abort it himself, usually by pressing an "end" key. The acceptance of the call by Remi leads to an ISUP ANSWER message being received by the MSC Transmission path is completed between the two end users. And this is station reacts first by stopping the alerting indication if any. Secondly by answering the network with an RIL 3CC CONNECT ACKNOWLEDGE message, and thirdly by connecting the circuit transmission on the radio path with the suitable terminal. In the case of speech, this consists of completing the speech path to the microphone and the loudspeaker. The call then enters the connected phase ("is connected" in short). Charging starts and effective bi-directional transmission is provided between the two users in the case of speech, or between the terminals or the terminal on the mobile side and the IWF in the cases of data calls. In the last case the establishment procedure goes on in band

Let us now see some of the variants from this basic scenario.

6.1.2.2. Automatic Answering

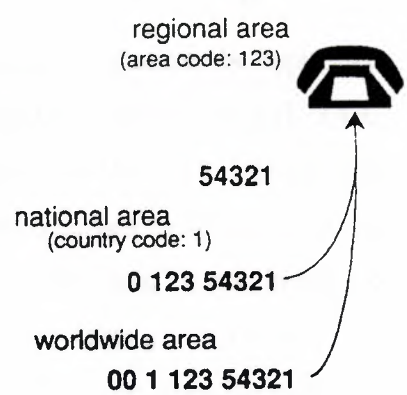
When the called party is a machine, which is often the case for data calls, some of the steps can be merged. For instance, the connect through can happen without any alerting indication. In some even more expedient cases the connect-through can happen as an immediate answer to the request. In the latter case, the RIL 3 CONNECT message is sent directly as an answer to the RIL 3 CC SETUP message. In the former case it directly follows the RIL 3 CC CALL PROCEEDING message.

6.1.2.3. Entered Number

One of the problems specific to an international mobile communication network like GSM. Where each user can establish calls from different countries is the format of the called number. As PSTN users, we usually know three basic types of formats: a local format: which refers to one destination only some regional area: an interurban or national format used when the destination is the same country but in a different region, and an international format used to call abroad. The distinction is done by the use of prefixes. Figure 6.1.12. shows the three formats in an example using the standardised (but not yet universally applied) prefixes 0 and 00.

Figure 6.1.12. Dialling number formats

Depending on the relative positions of the calling and called users.



PSTN numbers can be entered in three different ways which are distinguished by different prefixes

The way in which a number is dialled in the PSTN then depends on the location of the calling party. This is quite unfortunate in a mobile environment. To avoid the difficult task of requiring the user to know in which telephone area he currently is and how numbers are to be entered the issue must be treated differently in a cellular network. The ISDN protocol an RIL 3 CC support many different number formats by qualifying the digit string by two indicators the "type of number" (TON) and the Numbering Plan Identification (NPI) However, the Specifications are not very informative on the subject of what the user has to enter. There is however one mechanism that must mandatory be supported concerning the international format. In order not to ask the user to know all the prefixes in all the countries where he wishes to roam an international number can be entered by pressing the "+" key followed by the country code etc. Moreover a switch must be able to correctly treat a number presented as international even if it addresses a destination in the same country. This behaviour is by no means obvious: for instance in the French PSTN, keying an international call with country code 33 (France) from a PSTN telephone set results in a call to the international directory enquiries for the country whose code followed 33. A different treatment is required from the MSC This minimum specification enables the user to enter the number of a specific destination totally independently from his location: it is therefore advisable for GSM subscribers who travel abroad to always store their abbreviated dialling numbers in this format (e.g,, +44 701 123...), instead of a national format.

As already mentioned this format is only one among many that can be supported by the DIM 3 CC protocol. The translation between the number sent in the RIL 3 CC message is not specified by the Specifications nor is the way the MSC must interpret the numbers it receives. This is unfortunate, because what the user has to key does some standardisation will appear. Though other schemes cannot at this date be precluded the most widespread format for user dialling (besides the "+" key format) seems to be the national format of the visited country usually without prefixes to be used for numbers within the country.

6.1.2.4. Off Air Call Setup

In the basic call set-up scenario the transmission path between the MSC and the MS if fully operational for the requested service before the MSC is aware that both parties are ready to converse. The call can be connected through immediately A variant is possible called Off Air Call Setup OACSU) where the allocation of the suitable radio resource is delayed as much as possible in an attempt to save radio resources In this case the MSC has first to request the BSS to establish the circuit connection with the right transmission mode before connecting the communication For the called party the instant of connection is not delayed but the connection may be temporarily routed to an announcement machine until the actual connect through happens.

6.1.2.5. Progress

The basic set-up scenario assumes that signalling messages are received from the originating network.. Now this is not always guaranteed since an important proportion of the PSTN equipment are not using the advanced protocols It may then happen that the indications.

Concerning the progress of the call come as audio tones or announcements. This case is foreseen and coped with in the ISUP protocol and the MSC will know when he call establishment must go on in the old fashioned way. The MSC reacts by performing a kind of oneway through connection in advance to the normal point so that the speech path from the called party to the GSM user is established and the tones and announcements can be heard. (Note that this is not easy with data services except as alternate speech/data services) This early connection is commanded by way of the RIL 3 CC PROGRESS message or the PROGRESS information element included in messages such as RIL 3 CC CALL PROCEEDING or ALERTING depending on the moment it is known that the call is not ISDN compatible all the way. In addition if the RIL 3 CC PROGRESS message is used there are no more messages for call two way through connection.

The possibility of such a "null" IWF with ISDN was an argument for the peculiar management of the E1, E2 and E3 bits in the V.110 frame.

Once the existence of a V.110-like connection is confirmed the RLP protocol must be initialised in the NT cases. This is done by an exchange of SABM and UA frames as explained in Chapter 5. The next step involves the modem control signals, and is done in the same way as between a terminal and a modem.

The next an last sepes depend on the service and are normally carried between the terminals themselves. They correspond to the establishment of the upper layer protocols for transmission or even to further call establishment as with the double numbering cases we have seen for PSPDN or CSPDN access They do not generally affect GSM but there are some exceptions. For instance the IWF is involved in the procedure in the case of facsimile which consist mainly in a negotiation concerning various characteristics of the transmission and in particular the modem speed. Another example is for dedicated PAD or packet access, where establishment procedure through the PSPDN is taken care of by the IWF.

It should be noted that part of the procedures presented in the previous paragraphs duplicate the basic call establishment or could be supported by it. The trend is to limit them at least for the user as exemplified by the introduction of the dedicated PAD and packet access services.

6.2. Short Messages

All communications of a circuit nature, such as speech or data transfer, are established released and generally managed by the procedures described in the previous sections. Let us now see how communications of another nature are treated in GSM. The only GSM services not requiring the end-to-end establishment of a traffic path are the Short message services. As a consequence, short message transmission may take place even if the mobile station is already in full circuit communication.

A short message communication is limited to one message, or in other words the transmission of one message is a communication all by itself. The service is already in full circuit communication is limited to one message, or in other words the transmission of one message is a communication all by itself. The service is then asymmetric, and Mobile Originating Short Message transmission. This does not prevent a real dialogue, but the different messages are considered to be independent by the system. The transmission of a message is always relayed by a Short Message Service Centre (SM-SC), considered to be outside GSM. The consequence is that the transfer of a short message always takes place between a mobile station and some SM-SC from the point of view of the GSM infrastructure. However for the user, the message has also an ultimate destination or origin identified by some field in the message but relevant only for the user and the SM-SC not for the GSM infrastructure.

6.2.1. Mobile Terminating Short Messages

A short message addressed to a GSM subscriber must first be routed from the sender to a Short Message Service Centre, and from then be routed to the actual destination. The way the message first reaches the SM-SC is once again out of scope of the Specifications. There also a variety of solutions can be imagined to enable PSTN users to send messages towards GSM users using human operators printer-working with other services such as videotext.

When the SM-SC has a message to send to some GSM subscriber it builds a SM-TP SMS DELIVER message containing various pieces of information for the benefit of the recipient. This information includes in particular the user content, the identification of the original sender and a time-stamp indicating when the message was received by the SM-SC Similarly with the Mobile Originating case, the SM-TP SMS-DELIVER message will be transferred on various interfaces using the capabilities of lower layer protocols. In particular to convey the acknowledgement back to the SM-SC.

Before the SM-TP SMS-DELIVER message can reach its destination (the mobile station) its actual routing must be derived using the interrogation functions of

MAP/C. This is achieved in the following manner. The SM-SC conveys the short message to an SMS gateway to which the service centre is connected which it chooses depending on he subscriber it wants to reach since most oten a gateway will be able to country or some of the subscribers (for instance those o some country or some operator) The subscriber is then identified by his directory number ((the same MISSION as for telephony, , typically) entered gateway to identify the relevant HLR and interrogate it. The interrogation is done by sending a special message the MAP/C SEND ROUTING INFO FOR SHORT MESSAGE message. This is answered by either the corresponding MAP/C SEND ROUTING INFO FOR SHORT MESSAGE RESULT message, which contains an SS7 address pertaining to the MSC/VLR where the subscriber is visiting, or by a rejection message if the subscriber is known not to be reachable at this instant. There is no need for a specific roaming number as for circuit calls since the short message uses only SS7 signalling means to be transported to the visited MSC.

The SMS-gateway makes use of the SS7 address to forward the message to the relevant MSC which delivers it to the mobile station after setting if need be a signalling connection as for the mobile originating case. The delivery to the mobile station does not involve the user.

The message can be stored until the user decides to discard it after reading More precisely, it is stored in the SIM, and then can be kept in storage even after the mobile station has been switched off, or even read on another mobile station. The limited memory capacity of the SIM however, raises a small problem when for any reason the emery is full In phase 1, a message delivered to a SIM with no free memory could be lost. In phase 2, a mechanism has been specified to enable a crude sort of flow control by the mobile station, which will then be able to indicate to the network when the memory is full, or conversely when it is back to a state where messages can be accepted again.

An important variation from this basic scenario corresponds to cases when the mobile station cannot be reached. To provide a satisfying taken so that this message, and possible. Since the message if not acknowledged is still stored for some time in the

SM-SC, it can be sent anew as soon as the subscriber resumes contact with the network. This requires the GSM network to store the lack of delivery condition and the address of the SM-SC and to start a procedure to "alert" the SM-SC when the subscriber pops up again. The HLR is obviously the "focal the different situations in which delivery fails.

Three different kinds of non reachability can be identified similar to what we have seen with circuit calls. The HLR can know beforehand that the subscriber is not reachable for the moment the VLR can know it but not the HLR ad finally it can be discovered after failure of the effective attempt to deliver the message by the MSC/VLR. Whe interrogated by an SMS gateway the HLR my know immediately that delvey cannot take place because it already holds a non empty list of service centres which have not succeeded in transmitting messages and waiting to be alerted.

It then adds if possible the new SM-SC identity to this list. In such situations it will usually indicate the problem to the SMS-gateway with a negative answer to the MAP/C SEND ROUTING INFO FOR SHORT MESSAGE message. The same course is taken when the HLR knows that the subscriber is not reachable for instance because the subscriber is not entitled to get service in the geographical area where are he is currently located .There is however one exception to this rejection ait the HLR level. A priority indication is linked With each message, and it is used to bypass the straightforward rejection of the HLR when some messages are still undelivered. The HLR will answer positively if the message is of high priority, and the potential delivery problem, if it still exists, will be detected by the MSC/ VLR.

In the cases where the MSC/VLR is given message but is not able deliver it a failure indication is first sent to the SMS-gateway as an answer then sends on one hand a negative report to the SM-SC and on the at her a MAP/C STE MESSAGE WAITING DATA message to the HLR, which acknowledges the updating of its atable by a MAP/C SET MESSAGE WAITING DATA RESULT message.

This state of affairs is stored by both the MSC/VLR and the HLR in the subscriber record. In addition, as already mentioned, the HLR maintains for each subscriber a list of addresses for the SM-SC holding messages in wait. The sequence of events is represented in figure 6.2.1.

Eventually the subscriber surfaces again. This may be known for instance by a contact instance by a contact with the MSC/VLR where the subscriber was located (e.g., a mobile originating call attempt). When such an event happens, thanks to the stored indication of a previous delivery failure, the MSC/VLR notifies the HLR with a MAP/D NOTE MS PRESENT message.

The mobile station may also reappear within coverage of another MSC in which case the HLR will be directly aware of its state of things thanks to the mobility management procedures. In any case the HLR then sends an indication of the subscriber's reappearance to all the SM-SCs whose identities are stored as holding a message for this subscriber. This is achieved by sending a MAP/C ALERT SERVICE CENTRE message to the suitable SMS gateway for each service centre. The whole sequence as described above is show in figure 6.2.2.The SMS –gateway will convey the relevant information to the service centre to trigger a new transfer attempt.

The alerting mechanism must be supported by all MSC/VLR but it is an operator's option to store the list of SCs in the HLR and to alert them.

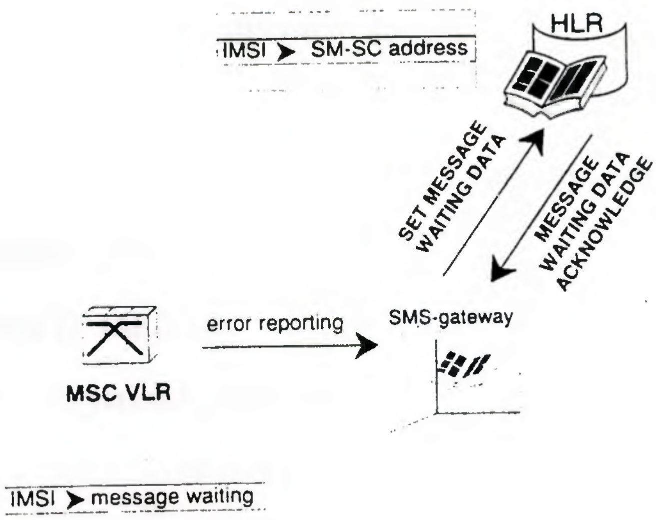


Figure 6.2.1. Short message failed delivery management

When the MSC is unable to deliver a short message to the mobile station This state of things is stored in both MSC/VLR and HLR, for triggering retries when the mobile station reappears.

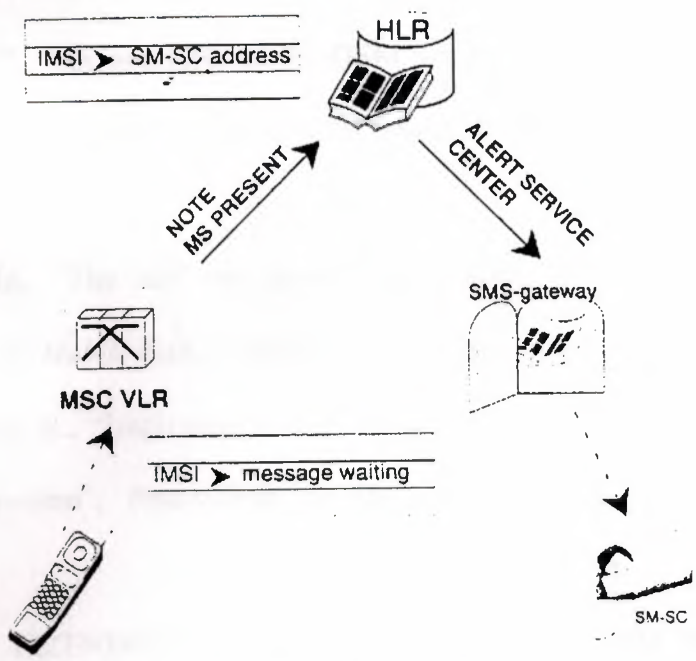


Figure 6.2.2. Service Centre alerting

When a subscriber becomes reachable, the HLR alerts all the service centers which are known to hold messages not delivered to the given subscriber

BIBLIOGRAPHY

1. E.A. LEE, D.G. MESSERSCHMITT, *Digital Communication,* Kluwer Academic Publishers (Boston), 1998
2. J.D. PARSONS, D. JARDINE, J.G. GARDINER, *Mobile Communication Systems*, Blackie (Glasgow), Halsted (New York), 1989
3. G. CALHOUN, *Digital Cellular Radio*, Artech House, 1988
4. T. HAUG, "The GSM Programme, a pan-European effort", *Proceedings of the Mobile Radio Conference*, Nice, November 1991
5. A. FOXMAN, "Activities of the group of Signatories of the GSM MoU", *Proceedings of the Mobile Radio Conference*, Nice, November 1991
6. A. HADDEN, "Development of the DCS 1800 standard", *Proceedings of the Mobile Radio Conference*, Nice, November
7. B. MALLINDER, "An overview of the GSM system", *Proceedings of the Third Nordic Seminar on Digital Land Mobile Tadio Communication*, Copenhagen, September 1988
8. R. MONTAGNA, "The half rate speech voice codec for the GSM system", *Proceedings of the Mobile Radio Conference*, Nice, November 1991
9. I. DITTRICH et. al., "Implementation of the GSM data services into the radio mobile radio system", *Proceedings of the Mobile Radio Conference*, Nice, November 1991
10. G. CRISP, A. EIZENHÖFER, "Architectural aspects of Data and Telematic Services in a GSM PLMN", *Proceedings of the third Nordic Seminar on Digital Land Mobile Radio Communication*, Copenhagen, September 1988

11. A. CLAPTON et. al., "Supporting Facsimile in the GSM PLMN", *Proceedings of the fourth Nordic Seminar on Digital Land Mobile Tadio Communication*, Oslo, June 1990

12. P. VARY, R. HOFMANN, "Sprachcodec für das Europaische Funkfernsprechnetz", *Frequenz*, vol. 42, no. 2-3, February 1988

13. U. JANSSEN, P. BRUNE, "The Mobile Application Part protocol", *Proceedings of GSM Seminar*, Budapest, October 1990

14. M. B. PAUTET, M. MOULY, "GSM protocol Architecture : Radio Subsystem Signalling", *Proceedings of VTC'91*, Saint-Louis, Missouri, May 1991

15. I. BRINI et. al., "European roaming related technical problems", *Proceedings of the Mobile Radio Conference*, Nice, November 1991

16. R. THOMAS, M. MOULY, et. al., "Performance evaluation of common control channels in the European digital cellular system", *Proceedings of the Digital Mobile Radio Communications Conference*, Venice, July 1987