2

The General Packet Radio Service

2.1 GPRS Objectives and Advantages

As we have seen, there were definite trends leading from speech towards data transmission and from fixed to mobile networks. Thus the introduction of a data transmission service for the GSM standard appeared to be imperative. This service would have to achieve certain objectives, both for users (U) and for network providers (NP). It must:

(1) enable access to company LAN and the Internet (U),
(2) provide reasonably high data rates (U),
(3) enable the subscriber to be reachable at all times – not only for telephone calls but also for information such as new emails or latest news (U),
(4) offer flexible access, either for many subscribers at low data rates or few subscribers at high data rates in order to optimize network usage (NP),
(5) offer low cost access to new services (U + NP).

At the time of its introduction, GPRS was the only service to achieve all of these goals. Other GSM or non-GSM technologies were failing in one or more of the points given above. Some examples are the high speed circuit switched data service offered by GSM (HSCSD), and the 3G universal mobile telephone service (UMTS). HSCSD, because of its circuit switched nature, can neither provide continuous reachability nor flexibility as in points 3 and 4 respectively. UMTS does offer continuous reachability but was not ready at the time of the introduction of GPRS and, because of high licence and investment costs, was struggling with the cost aspect – point 5.

The major advantages of GPRS arise from the fact that it uses *packet* switching:

- since devices are able to handle packet traffic, data can be exchanged directly with the Internet or company intranets;
- packets from one user can be transmitted via several air interface time slots (bundling);
- time slots are not used continuously and can be shared between users;

GPRS Networks. G. Sanders, L. Thorens, M. Reisky, O. Rulik, S. Deylitz
 ISBN: 0-470-85317-4

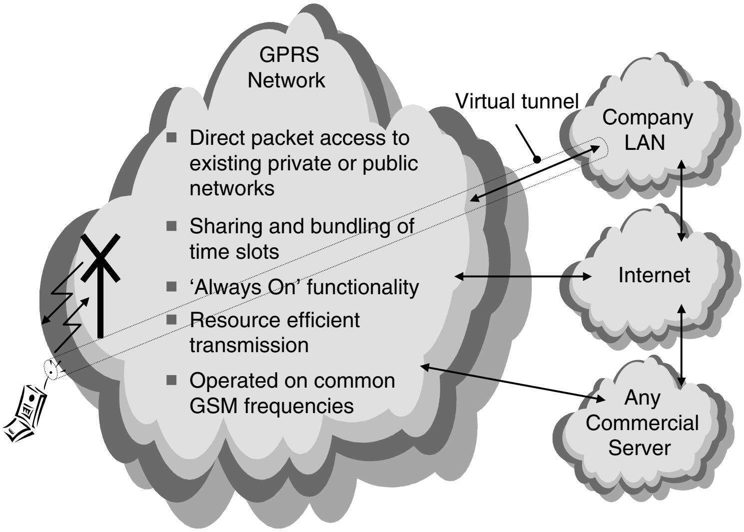


Figure 2.1 GPRS applications and strengths

- even if users are not sending or receiving, they can still remain connected/reachable (e.g. to/from their company LAN) and in doing so, they do not use up any resources;
- users are only allocated resources when they actually need them;
- GPRS is implemented within the GSM standard so no new frequencies have to be used.

Initially, the only major problem faced by GPRS was that the existing circuit switched GSM networks had to be extended to include packet switching equipment. Also, existing mobile devices were not capable of supporting packet oriented data transfer and time slot bundling so new GPRS mobiles had to be developed.

The strengths and applications of GPRS are summarized in Figure 2.1.

2.2 GPRS Architecture

The standard GSM network has had to be altered to be able to support GPRS, since it is not able to transmit data in packet switched mode. Let us first have a look at the original Phase 2 GSM network in order to be able to recognize the differences in a GPRS enabled Phase 2+ network.

2.2.1 GSM Network

A GSM network consists of many functional entities and their respective standardized interfaces. In part, a mobile network is similar to a fixed telephone network (PSTN) – the main difference to fixed networks being the almost unlimited mobility that is offered to subscribers by the use of the air interface.

A GSM network is known as a PLMN (public land mobile network) and can be divided into various parts known as 'subsystems', each of which carries out different functions. The first major division is between the interconnected network of digital telephone exchanges and the network of radio transceivers that enable wireless communication with the mobile. The network of exchanges is known as the ‘network switching subsystem’ (NSS) and is similar to a PSTN but differs in that it also includes the databases necessary for managing the mobility of the subscriber. The transceiver network (and associated network elements) is known as the ‘base station subsystem’ (BSS). The BSS together with the mobile devices and the air interface is known as the ‘radio subsystem’ (RSS).

The various elements described in the following sections, and their relationships to each other, are illustrated in Figure 2.2.

2.2.1.1 The Mobile Station

On the user side there is a mobile device know as the ‘mobile station’ (MS) which consists of the ‘mobile equipment’ (ME) and the ‘subscriber identity module’ (SIM) – as illustrated in Figure 2.3.

The ME is the device known as a cellphone or mobile telephone, and can even refer to fixed terminals such as those found in cars or remote phone booths. It consists of speaker, microphone, display, keypad and, most importantly, sender/receiver unit and antenna. The ME is able to establish both speech and data connections. For GPRS, the ME has to be equipped with packet transmission abilities.

There are three different classes of GPRS equipment:

- Class A: equipment that can handle voice calls and packet data transfers at the same time;
- Class B: equipment that can handle voice or packet data traffic (but not simultaneously) and can put a packet transfer on hold to receive a phone call;
- Class C: equipment that can handle both voice and data, but has to disconnect from one mode explicitly in order to enable the other.

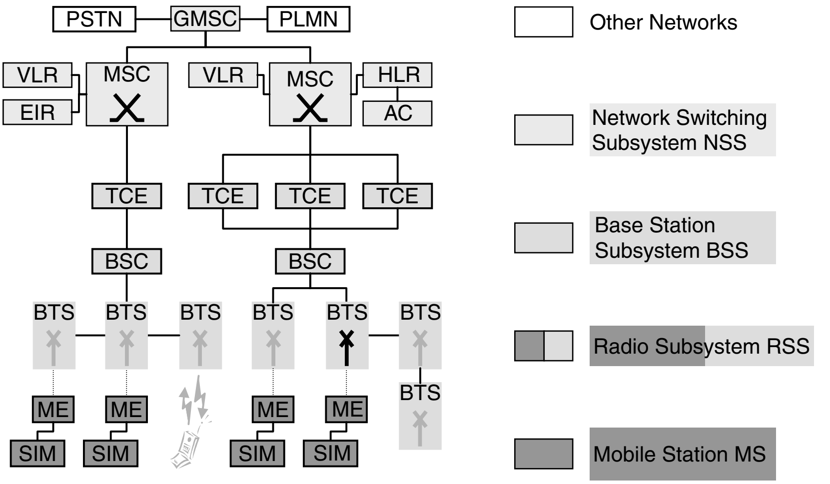
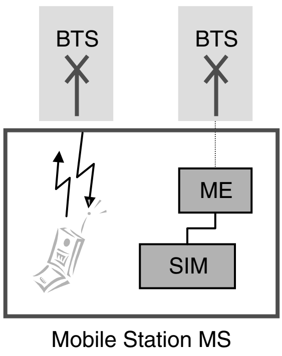


Figure 2.2 GSM network overview



Mobile equipment (ME), device as:

1. User interface
   - speech input and output
   - data handling (received e.g. via CSD or GPRS)
   - dialling and storing numbers
   - advanced menu functions
2. Transmitter/receiver to/from BTS

Subscriber identity module (SIM), for:

1. Authentication key and algorithms
2. Ciphering algorithms
3. User identification
4. Additional telephone book

Figure 2.3 GSM/GPRS mobile stations

The SIM is provided by the network operator and usually consists of a small chip card (the SIM card), which contains important data concerning the subscriber and additional information, such as:

- International mobile subscriber identity (IMSI), which serves as a user name to identify the user towards the network. It consists of the mobile country and the network code (obviously provider dependant) and the MSIN (mobile station identity number).
- An individual subscriber authentication key, Ki, which serves as a password for the subscriber and is stored only on his SIM card and in the related database in the network (the authentication centre AC, see below).
- The ciphering key, Kc, which is calculated anew for every individual connection in order to prevent eavesdropping.
- Some mathematical procedures (algorithms) to calculate the respective keys.
- User phone book.
- A processor (SIM toolkit) for running provider specific software to enable new services such as mobile banking applications, stock market information, and more.

2.2.1.2 Base Transceiver Station

All communication (voice, fax, data or signalling) between the mobile station (MS) and the PLMN must go via a base transceiver station (BTS) – often referred to as a 'base station'. As each BTS serves a cell of limited size, a large number of them has to be deployed across the country in order to provide (ideally) seamless coverage. The speech channel capacity of a GSM network, in any particular area, is determined by the number of frequency carriers per cell and the cell density within that area. The former can vary between one and 16, with eight speech channels per frequency carrier (minus signalling channels). The latter depends highly on population density, with many more cells per unit area in more densely populated regions and fewer BTS (with relatively high output power) in rural areas (see Chapter 8, Planning and Dimensioning for more information).

Thus a BTS has to fulfil certain requirements:

- reliability in order to keep service and maintenance costs low for remote sites;
- low price as thousands of them have to be deployed;
- rugged design to withstand wind and weather on site.

The BTS has to serve as a counterpart for the mobile station (MS) on the air interface. The BTS must also interface with the PLMN via the base station controller (see next section). The main tasks of a BTS are:

- channel coding (using FR, HR or EFR),
- ciphering and deciphering (for circuit switched connections only),
- synchronization of several MS in time and frequency,
- burst formation, multiplexing and HF modulation for all active MS,
- evaluation and optimization of the UL and DL transmission quality (using own measurements and MS measurement reports).

The main roles of the BTS and its place in the BSS architecture are shown in Figure 2.4.

2.2.1.3 Base Station Controller

It would be impractical to equip each BTS with the capability to evaluate all information necessary for the handover of a particular subscriber from one cell to another. A BTS can monitor the received signal strength of the subscriber as it decreases and (theoretically) it should be possible for the BTS to estimate which cell would be most appropriate to hand over to (based on information from the measurement report transmitted by the MS). It would be impractical to equip each BTS with the capability to evaluating all information necessary for the handover of a particular subscriber from one cell to another. This would require every BTS to:

- assess the measurement report transmitted by every mobile during every call to determine whether a handover was necessary.
- have access to information on the availability of resources in each neighbouring BTS;
- be able to communicate with the neighbouring BTS to request resources for each mobile needing a handover, and to coordinate the handover itself;
- coordinate the handover with the telephone exchange so that it knows which BTS is handling each call.

There is thus a need for another network element that controls several cells (i.e. several BTS) and all the active MS in those cells. This element is the base station controller (BSC). Apart from the hand over decision mentioned above, the BSC has to carry out several other functions:

- evaluation of signalling between MS and exchanges (core network);
- performing paging in a group of cells for every mobile terminating call (MTC);
- radio resource management for each BTS;
- switching of time slots from the core network to the right BTS (and vice versa);

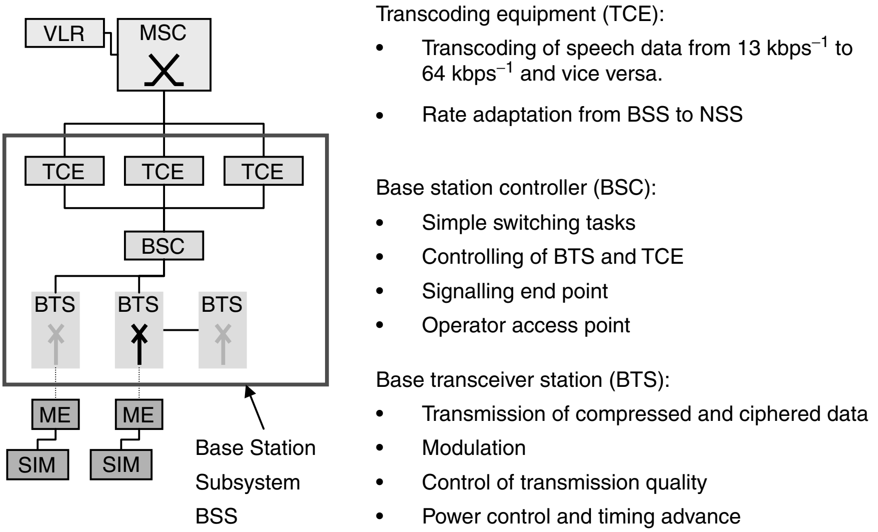


Figure 2.4 Base station subsystem in detail

- providing a central operation and maintenance access point for the whole BSS;
- storage of configuration data for all elements in the BSS.

The main roles of the BSC and its place in the BSS architecture are shown in Figure 2.4.

2.2.1.4 Transcoding Equipment

There is another network element in the RSS that has not been introduced yet. As mentioned in the introduction to the GSM network, there is a conventional network switching subsystem (NSS) that consists mainly of standard fixed network components (we will consider the differences later). On the other side there is the radio subsystem (RSS), which includes the air interface between antennas and mobile devices.

Because of certain physical restrictions, the data rates for speech on the air interface are much lower than the data rate used to transport fixed network speech (the well known 64 kbit s^{-1} from ISDN). The low rate can only be achieved by compression of the original digitized speech data. The MS digitizes speech at a rate of 104 kbit s^{-1} (160 samples of 13 bits every 20 ms) which is compressed to 13 kbit s^{-1} using full rate (FR) coding, 12.2 kbit s^{-1} using enhanced FR (EFR) and 5.6 kbit s^{-1} using half rate (HR).

The high compression ratios required for transmission via air interface time slots are a problem because the NSS uses standard digital exchanges and PCM30 lines, which make use of 64 kbit s^{-1} time slots (it should be noted that in ISDN networks the telephones compress speech into 64 kbit s^{-1} time slots; and for analog telephones the signal is digitized and compressed at the exchange). Because of the above, there needs to be another piece of equipment which converts the coding used for the air interface into the coding used for ISDN (and vice versa) and in doing so, converts the 13, 12.2 or 5.6 kbit s^{-1} bit rate of air interface time slots to the 64 kbit s^{-1} used in the NSS and other networks. This is the transcoding equipment (TCE), often referred to as the transcoding and rate adaptation unit (TRAU) as it is a data rate adaptor as well.

The question of where to position the TCE is an interesting one. The transmission of compressed data is theoretically only necessary between MS and BTS (this is where the physical restriction occurs). However, it is also advantageous to transmit compressed data through the network rather than uncompressed data, because of the line capacity which is saved. Thus the decompression does not take place in the BTS.

On the other hand, the conventional equipment in the NSS is not able to switch data compressed using non-ISDN coding (i.e. only 64 kbit s^{-1} is allowed), so the transcoding must be performed before the time slot reaches the first telephone exchange. Hence, transcoding will be done right between the BSC and the exchange – i.e. as late as possible, as early as necessary. The TCE belongs logically to the BSS (see Figure 2.4) but is usually placed close to the telephone exchange. This is done because for every single PCM30 line coming from the BSC requires four PCM30 lines from the TCE towards the network so operators can save three PCM30 lines by positioning the TCE at the exchange.

This fact in itself is not important for GPRS. What is important is the way the PCM30 lines are utilized. Each PCM30 timeslot can transport data at a rate of 64 kbit s^{-1}. Air interface time slots transport digitized speech data at a maximum rate of 13 kbit s^{-1}. This means that if the PCM30 time slots are sub-divided into 4×16 kbit s^{-1}, then four connections can be carried by one time slot. This approach combined with positioning the TCE at the exchange, is very cost effective for the network operator but causes problems when GPRS is implemented.

The problem arises because although the 16 kbit s^{-1} data rate is sufficient for all speech codecs (coder/decoder algorithms) in GSM, it is not sufficient for all codecs used in GPRS – i.e. since the GPRS data are sharing these lines at least as far as the BSC, GPRS has to be restricted to a maximum of 16 kbit s^{-1} as well. As mentioned earlier, there are new coding schemes (CS) for GPRS that are able to transmit up to 21.4 kbit s^{-1} per channel (so called CS 4) over the air interface. Obviously this can not be implemented in such a line saving environment. Even CS 3 with only 15.6 kbit s^{-1} per channel can not be used because signalling has to be transmitted on these lines as well.

As a result, early GPRS networks will only be able to use CS 1 and CS 2 (9.05 and 13.4 kbit s^{-1} per channel) until the ‘bottle-neck’ (illustrated in Figure 2.5) between the BTS and the BSC is removed.

2.2.1.5 Mobile Services Switching Centre

Until now we have only considered the elements in the base station subsystem (BSS). The role of the BSS is to provide access to the network switching subsystem (NSS) that enables connections to be set up between subscribers via interconnected telephone exchanges. A telephone exchange is a functional unit in the network which can:

- evaluate the dialling information (digit translation),
- find a route to the destination,
- switch the call towards the destination.

This functional unit is known as a switch or exchange, and in GSM it is known as the mobile services switching centre (MSC). Apart from the call processing tasks mentioned above, it also:

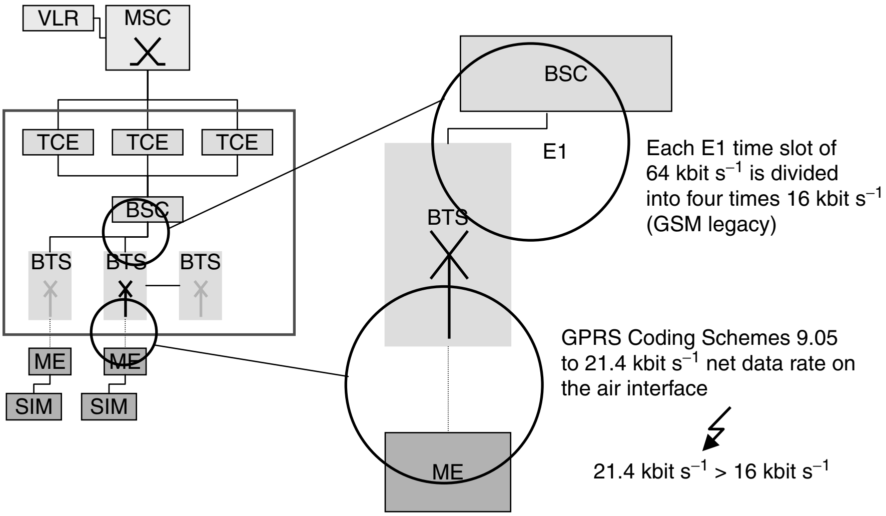


Figure 2.5 Restrictions on early GPRS networks because of GSM legacy

- creates and stores charging data records for subscriber billing,
- performs call supervision and overload protection,
- supports operation and maintenance functions,
- enables legal interception.

In case a call is made to or from another network (such as a fixed telephone network or another GSM network) there needs to be an exchange which serves as a gateway between the two networks. In this case, a special MSC will serve as a gateway MSC or GMSC.

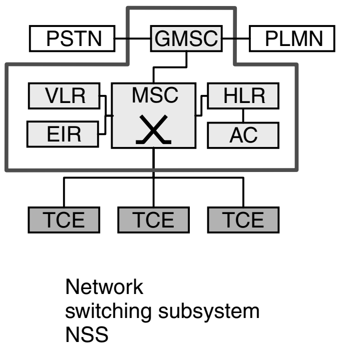
The main roles of the MSC and its position in the NSS architecture are shown in Figure 2.6.

2.2.1.6 Location Register

So now it should in principle be possible to make a call from one subscriber to another, because we are able to transport circuit switched speech traffic via BTS, BSC, TRAU to an MSC and back to another subscriber – but there remains one very significant problem to overcome: the mobility of the subscriber.

In a fixed network, a call is possible as soon as we provide access to the network (as the RSS does for the mobile subscribers) and install an exchange (like the MSC). In mobile networks however, extra infrastructure has to be put in place to enable the correct MS to be located and contacted when a call is made to the subscriber.

So how do we locate the destination MS of a mobile terminated call (MTC) when the subscriber can move whenever and wherever he wants, and all we have is his telephone number (MSISDN)? The most common answer to this question is: ‘The mobile books itself into a cell.’ This statement in itself is partly correct, but it does not answer the question. The network must be able to find any MS, no matter where it is, and to do this some location information must be stored somewhere in the network for every MS.



Mobile services switching centre MSC:
- Switching of the users
- Number translation for routing and charging
- Traffic observation
- Lawful interception
- Operator access point
- Transition to other networks (as gateway MSC)

Home location register HLR:
- Static user data
- Temporary user data like current VLR area

Authentication centre AC (associated with HLR):
- User authentication and delivery of the coding key on request of VLR

Visitor location register VLR (associated with MSC:
- Temporary user data like location area
- Requests triples from AC and stores them temporarily

Equipment identity register (optional) :
- Authentication of equipment
- Supervision of suspicious devices

Figure 2.6 Network switching subsystem in detail

A common follow up to the first statement is, ‘we need a central database, which is informed (updated) whenever a user moves to another cell and can provide the destination for incoming calls’. This is again partly correct but it is important to understand the real system in order to appreciate later how GSM and GPRS can work together to manage the mobility of the subscribers.

There actually is a central database that can give information about the location of a subscriber. This database will always be interrogated when a particular subscriber is called. It is called the home location register (HLR). Apart from temporary data related to the current location of a subscriber, it also contains semi-permanent user data such as name, address, etc., for a certain MSISDN.

There are two further problems which must be addressed with regard to the storage and retrieval of location information. Firstly, a large network provider covers a whole country and may have upwards of 25 million subscribers. So if there really were only the one central database, there would be signalling messages originating throughout the whole country and concentrating at the database site whenever any of the 25 million subscribers moved to a new cell. The above scenario is unacceptable as a model for network signalling traffic as well as for database efficiency. The following presents the solution used by GSM networks.

Initially, several HLR could be introduced, each of which is responsible only for a certain fraction of all subscribers. This approach is used in GSM networks, so that each HLR has to maintain the data of about 1–4 million subscribers, depending on the preferences of the network provider. The particular HLR in which a particular subscriber is stored is coded in the MSISDN number – i.e. when a GMSC receives a telephone number it can immediately contact the HLR of the subscriber. The ID of a subscriber's HLR is given by the first two digits following the network provider code – e.g. in the number 0175 9316585 the digits '93' indicate which HLR is responsible for the subscriber.

Let us now consider how a MS is located so that the subscriber can receive an incoming call. If the network knows which cell the MS is currently in, a signalling message can be broadcast in that cell and, if the MS is switched on, it will receive the message and respond with an acknowledgment. Such signalling is known as 'paging'. The problem with this scenario is that the MS must update the location database every time it changes cell. If the operator has 25 million subscribers, the signalling load will be far too high. At the other extreme, if mobiles never perform updates, then each MS must be paged in every cell in the network whenever a call comes in. This is also an unacceptably high signalling load. The solution is to define groups of cells known as 'location areas' and each cell broadcasts the identity (LAI) of the location area it belongs to. The MS only performs an update when it moves to a new location area (determining the optimum size of a location area is explored in Chapter 3, Section 3.5.1.2). This leads us to the second question which is whether or not the HLR really needs to know whenever any of its 1 to 4 million users moves to another location area?

In reality, there is another network element which is responsible for storing the current location area of each MS. This is the visitor location register (VLR). Every MSC has a VLR which stores the LAI of every MS which is 'visiting' the region covered by the MSC. When an MS registers with the network and its LAI is stored in the VLR, the VLR identity is sent to the HLR, i.e. the HLR always knows which VLR the MS is stored in, no matter where in the world that VLR is located.

So we can now appreciate that:

- a country is divided into (several thousand) cells;
- cells are grouped into location areas;
- a number of location areas form the service area of one MSC/VLR;
- whenever a user changes a cell within a location area nothing will occur;
- when a location area border is passed the MS will notify the VLR;
- if the MS changes MSC/VLR service area, the new VLR identity is sent to the HLR.

For an incoming call, all that is known is the MSISDN of the called party, so the HLR will be asked for the location. All the HLR knows is the identity of the current VLR to which it will redirect the call set-up request. The VLR will know the location area of the called subscriber and will direct the BSC to page him in this group of cells.

The main roles of the HLR and VLR and their positions in the NSS architecture are shown in Figure 2.6.

2.2.1.7 Authentication Centre

If an MSISDN was all that was needed to make a phone call, it would be like leaving an open door to fraudulent network usage, i.e. it would simply be a case of faking the MSISDN of any known subscriber and calls would be charged to that subscriber's bill. In order to prevent such abuse, GSM implements a very thorough authenticity check of a user or, more precisely, of the SIM card in use. As with most computer systems, a

SIM card needs to log on to the GSM system with a user name and a password. To keep this system secure, the user name and password are changed after every authentication so that even if fraudsters manage to somehow intercept these two identities, they will not be able to make use of them. The procedures and identities used are discussed in detail in Chapter 3, Section 3.1.1. For the purposes of this section, it is enough to know that the authentication procedure is based on sets of keys called ‘triples’, which are used by both the VLR and the MS for authentication purposes. These triples are generated by a network element known as the authentication centre (AC). The triples from the AC are used to authenticate all users who attempt to make calls, receive calls, send an SMS, or even just switch their equipment on. Only emergency calls require no authentication.

The main roles of the AC and its position in the NSS architecture are shown in Figure 2.6.

2.2.1.8 Equipment Identity Register

In GSM, it is also possible, to check the status of the mobile equipment (ME) whenever it accesses the network. The word ‘status’ can be taken to mean legitimate, stolen or suspect. The checking is carried out by the equipment identity register (EIR) by comparing the international mobile equipment identity (IMEI), provided by the ME, against several lists. The EIR contains either a black list, where stolen equipment can be listed and eventually blocked, or a white list where legitimate equipment can be listed for unlimited use. If an ME (IMEI) does not appear on any of these lists, it can be stored on a so called ‘grey’ list. ME on the grey list can be restricted to certain services and all activity can be logged in detail. In case it later appears on the black list the logged information could be used to determine the location or user of the possibly stolen equipment.

Very few network providers use the EIR, as checking the IMEI of every ME every time the network is accessed generates a large volume of signalling which represents unpaid overhead for the operator. Added to which is the fact that most countries have several GSM operators, and if one operator chooses not to check for stolen equipment, then stolen ME from other operators can be used on that network simply by replacing the SIM card. Hence, many providers chose not to incur the extra signalling load required for EIR use.

The main roles of the EIR and its position in the NSS architecture are shown in Figure 2.6.

2.2.2 New Elements for GPRS

As mentioned, the general packet radio service (GPRS) is completely different from standard (circuit switched) data or speech data transmission, as the data is sent in the form of packets. Whenever a packet needs to be transmitted, the resource (bandwidth on the air interface) is allocated and the respective data packet delivered to the recipient (which for example, can be another mobile user or a server on the Internet). The network elements we have introduced so far are traffic switching, signalling nodes or both. In GSM, traffic is circuit switched with certain data rates depending on the application. Signalling is packet switched but with a much lower data rate. However, none of the nodes we have seen so far are able to transmit packet switched traffic at high data rates as required by GPRS.

2.2.2.1 GPRS Support Nodes

Clearly, a completely new set of switches is needed for GPRS, capable of routing packet switched data in a separate parallel network. There are two different main functional entities to be considered: the gateway to external data networks, and the node that serves the user.

The first is the access point for an external data network and is known as the gateway GPRS support node (GGSN). The GGSN is capable of routing packets to the current location of the mobile. Therefore it has to have access to the HLR in order to obtain the required location information for mobile terminating packet transfers.

The second is the node that serves the mobile station's needs and is known as the serving GPRS support node (SGSN). The SGSN establishes a mobility management context for an attached MS. It is also the SGSN's task to do the ciphering for packet-oriented traffic, since the base transceiver station is only responsible for ciphering circuit switched traffic. From the specifications, the two functionalities can be combined into one physical node (useful for early deployment of GPRS), separated in the same or even different networks.

Both of the GPRS support nodes (GSN) also collect charging data. Details about the volume of data (kbytes or Mbytes) transferred by the user are collected by the SGSN. Charging information detailing subscriber access to public data network (PDN) is collected by the GGSN.

2.2.2.2 Data/Voice Differentiation in the BSS

A GPRS cell phone will transmit data in a packet switched mode, but voice will be transmitted as circuit switched calls. This means that there has to be a network element that distinguishes between the two different kinds of transfer and diverts each to the correct core network, i.e. towards the MSC or towards the SGSN. This is one of the main functions of the packet control unit (PCU). Other PCU functions include:

- packet segmentation and reassembly, on the downlink and uplink respectively,
- access control,
- scheduling for all active transmissions including radio channel management,
- transmission control (checking, buffering, retransmission).

The PCU can be placed at various different positions in the network: at the BTS site, in the BSC or right before the switch, the preferred location being at the BSC.

2.2.2.3 Channel Control Unit

There is one more issue to address, namely that of how the BTS differentiates between GPRS users and non-GPRS users. There has to be some recognition of the user's needs as he/she will make normal phone calls, send and receive bursty packet traffic, share time slots with other GPRS users and will make use of GPRS coding schemes CS 1 to CS 4, depending on the quality of the air interface. These coding schemes also use their own special error correction in order to match the gross data rate on the air interface.

When implementing GPRS in an existing GSM network, every BTS requires a software upgrade so as to be capable of performing GPRS specific functions such as the use of CS 1 to CS 4, support of the GPRS radio channel measurements. This software upgrade

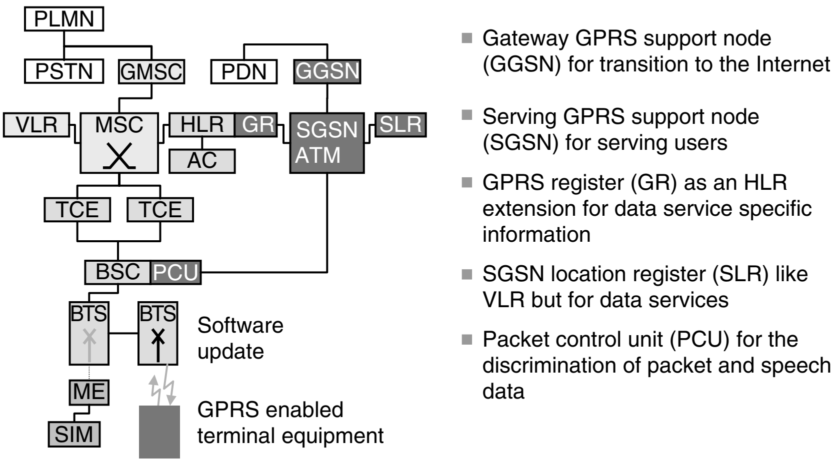


Figure 2.7 Overview of changes in the base station subsystem and core network

is specified as the channel codec unit (CCU) and is always located in the BTS. It can be necessary that the CCU takes over some of the PCU radio management functions if the latter is not located in the BTS itself.

The overview of a Phase 2+ network described above is illustrated in Figure 2.7. and the functionality of all the new elements will be covered in detail in Chapter 6.

2.2.3 Interfaces of the Network

Now that we have introduced all the new network elements, we can have a look at the whole picture. Obviously, the new network elements require new interfaces to be implemented. The new interfaces interconnect both new and existing network elements. Some of these interfaces are essential for the introduction of GPRS while others are specified but will be implemented in later releases of the GPRS standard as new functionality is introduced, e.g. early releases do not implement the Gc interface as they do not support mobile terminating packet transfer. All the interfaces in GPRS are labelled 'G' followed by a letter to indicate a particular interface.

The main interfaces for GPRS data transmission as illustrated in Figure 2.8 are:

- Gb, the interface between the BSS (PCU) and the SGSN,
- Gn, between all SGSN and also to the GGSN in a mobile network,
- Gi, the interface from the GGSN to a public data network,
- Gp, an interface to the GGSN of another mobile network.

There are also several signalling interfaces between GPRS nodes and also towards standard GSM network elements.

- Gr, the interface from SGSN and Gc, from the GGSN to the HLR for the exchange of user subscription, service and location data (in case of GPRS roaming, the HLR can be in a different network or even country);

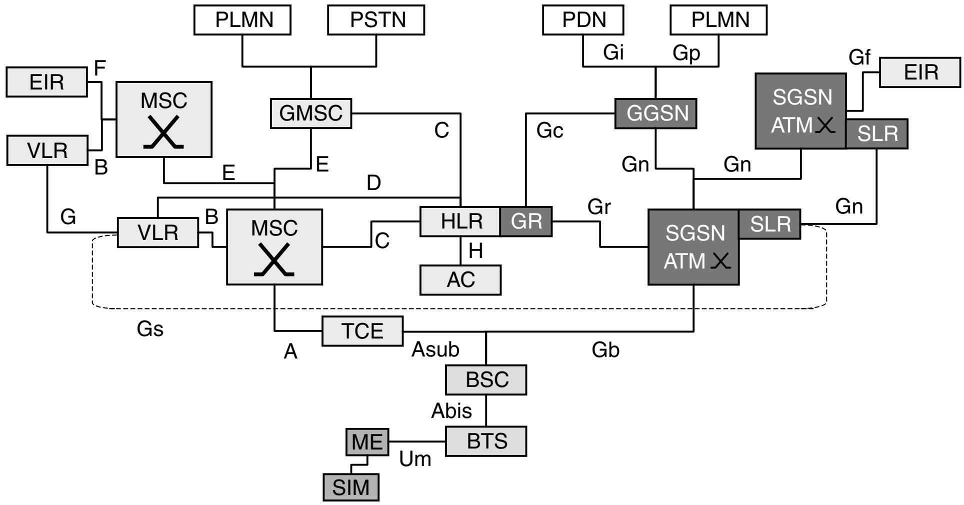


Figure 2.8 Overview of GSM and GPRS network interfaces

- Gd, an interface towards an MSC or GMSC with a connection to the SMS centre;
- Gs, an interface to an MSC to enable common mobility management (explained in detail in Section 2.4 Logical Functions);
- Ga, connection between SGSN, GGSN and the charging gateway function for the collection of billing data;
- Gf, the optional interface to an EIR for an extra equipment check in GPRS.

The interfaces, their functionality and the protocols used will be covered in much more detail in Chapter 3 Interfaces and Protocols.

2.3 Characteristics of a GPRS Connection

Let us have a look at the characteristics of GPRS data transmissions, in particular at how data is transmitted in GPRS, how higher data rates than with classical circuit switched data (CSD) are achieved, and especially how we can still keep it flexible enough to provide acceptable quality of service (QoS) for many users. This section aims to give an overview of the requirements and specified mechanisms in order that the reader can better understand the physical details that follow in Chapter 3 Interfaces and Protocols.

One of the major advantages of using packet switched transmission is the flexibility it offers to both subscribers and operators. When a single user needs to download data, he or she can make use of the entire available bandwidth. If there are other users, they can share the bandwidth according to their individual needs or priority level. In a circuit switched connection, a resource (the line) is allocated to just one user for the duration of a call or data transfer, regardless of whether he actually makes use of it or not. This usually means that there is a call set-up procedure before the call (a procedure to allocate the resources so that they are actually reserved for the user for the duration of the call). When the connection has been established (set-up), there is nothing further for the network elements to do but maintain the call, perform handovers if necessary, and collect charging data. This is completely different from a packet switched connection in which the resources are shared and there is no guarantee that there will not be delays in the transmission (making such guarantees forms part of implementing QoS). For example, one user could be using the resources when another one tries to access it and unless the second user has a higher priority his or her transmission will be delayed. Another difference is the addressing, i.e. defining the origin and destination. In the absence of a call set-up procedure, the network must have a method of determining the recipient of the current data packet. Hence, each packet has to carry additional addressing information. The subscribers must be charged for the use of network resources. Since each time slot can be used intermittently and shared between several users, the billing cannot be based on time. Thus a new billing model has had to be developed based on the volume of data transmitted by each subscriber.

2.3.1 Coding Schemes and Management of Radio Resources

Now let us have a look at how the packet transmission is realized in GPRS. As we have already seen, GSM uses eight time slots per carrier frequency. If an MS transmits data with the classical GSM circuit switched data (CSD), then one of these time slots is allocated to the MS for the whole duration of the data connection. Each time slot offers a gross data rate of 22.8 kbit s^{-1} but not all of this bandwidth is available to transfer user data. To ensure that the data is transmitted reliably, a copy of the data is made and sent separately across the air interface, i.e. each time slot contains both original data and a back-up copy of data from other time slots. As a result of this built-in redundancy, the net rate of a CSD connection is 9.6 kbit s^{-1} (or less if the quality of the air interface is poor).

2.3.1.1 HSCSD

A higher data rate could be achieved by increasing the net data rate, by bundling several time slots, or both. A higher data rate means using a new coding scheme that will necessarily send more original data and as a result will have less redundancy. Theoretically, ‘bundling’ means that up to eight time slots can be made available to one user as this is what is available on one frequency carrier.

The above measures have been implemented in GSM and the result is known as high speed circuit switched data (HSCSD). Using HSCSD enables faster data transfer (14.4 kbit s^{-1}) but this is still just a faster type of circuit switched connection with all the associated advantages and disadvantages. In principle, HSCSD could deliver 8×14.4 kbit s^{-1} (115.2 kbit s^{-1}), but since it is circuit switched, it will be transmitted via the classical nodes (BTS $\rightarrow$ BSC $\rightarrow$ TCE $\rightarrow$ MSC $\rightarrow$ provider) and therefore can not exceed the MSC data rate of 64 kbit s^{-1} per connection. As a result HSCSD is practically limited to 57.6 kbit s^{-1}, i.e. four time slots using 14.4 kbit s^{-1} (Figure 2.9).

2.3.1.2 Coding Schemes in GPRS

To some extent, GPRS makes use of a similar approach to HSCSD in order to achieve high data rates, i.e. GPRS also offers channel bundling of up to eight time slots and increased net rates per time slot by sacrificing redundancy. Despite a superficial similarity, however, GPRS is fundamentally different from HSCSD as can be seen in the following text, which considers the differences in the GPRS channel coding.

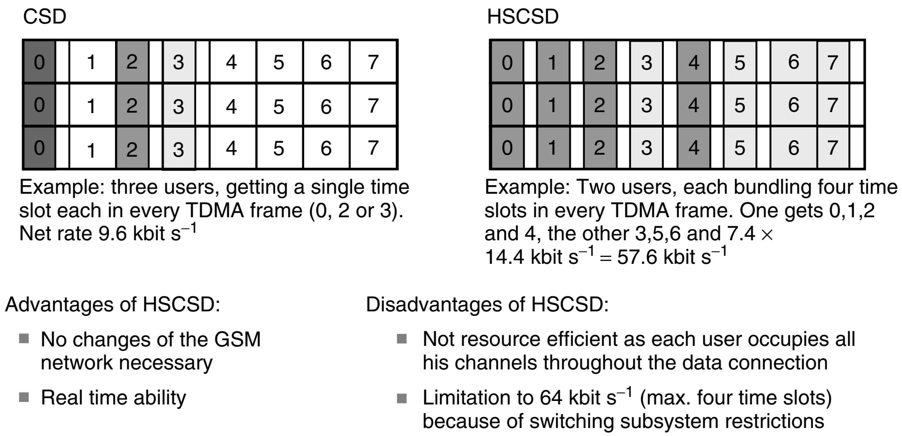


Figure 2.9 HSCSD and classical GSM CSD time slot functionality

The design of GPRS takes into consideration the variability of the air interface and the effect this has on transmission quality. When the air interface exhibits good transmission characteristics, e.g. a very low retransmission rate, little or no redundancy is required. When, however, the rate of retransmissions increases, coding schemes with more redundancy must be utilized. As a result, GPRS was designed with four new coding schemes: CS 1 to CS 4. Coding scheme 1 has a net rate of 9.05 kbit s^{-1}. Obviously this is even less than the CSD rate of 9.6 kbit s^{-1}, but CS 1 is extraordinarily stable, even under the worst air interface conditions, because of its high redundancy convolutional coding. Coding schemes 2 and 3 offer 13.4 and 15.6 kbit s^{-1} respectively. For the actual coding they use the same algorithm as CS 1, but with the high redundancy of CS 1 they would exceed the gross data rate of 22.8 kbit s^{-1}, which is impossible. So following the convolutional coding, a process called 'puncturing' is carried out. This means that certain bits of the copy (redundancy) are deleted in a prescribed manner in order to achieve the desired data rate. As a result, CS 2 offers less reliability under bad air interface conditions (that is when the copy might be needed more often) than does CS 1. Obviously CS 3 is even more dependent on good transmission conditions.

CS 4 offers a net rate of 21.4 kbit s^{-1}, which means that this coding scheme carries no redundancy at all. Consequently, CS 4 should only be used when the air interface offers close to perfect conditions, since every transmission error results in a retransmission, which reduces the effective data rate. In practice, CS 4 will only be used when the MS is close to the base station and is either stationary or moving very slowly.

It is worth noting here that all time slots contain a checksum so that transmission errors can be detected by the receiver. This checksum is the reason that CS 4 can not use the full 22.8 kbit s^{-1}. The GPRS coding schemes and the time-slot sharing described in the next section, are illustrated in Figure 2.10.

2.3.1.3 Radio Resource Management

Because of the packet switching mode, there is no connection set-up in GPRS as there is in CSD and HSCSD data calls, and all speech calls. Furthermore, it is not known simply

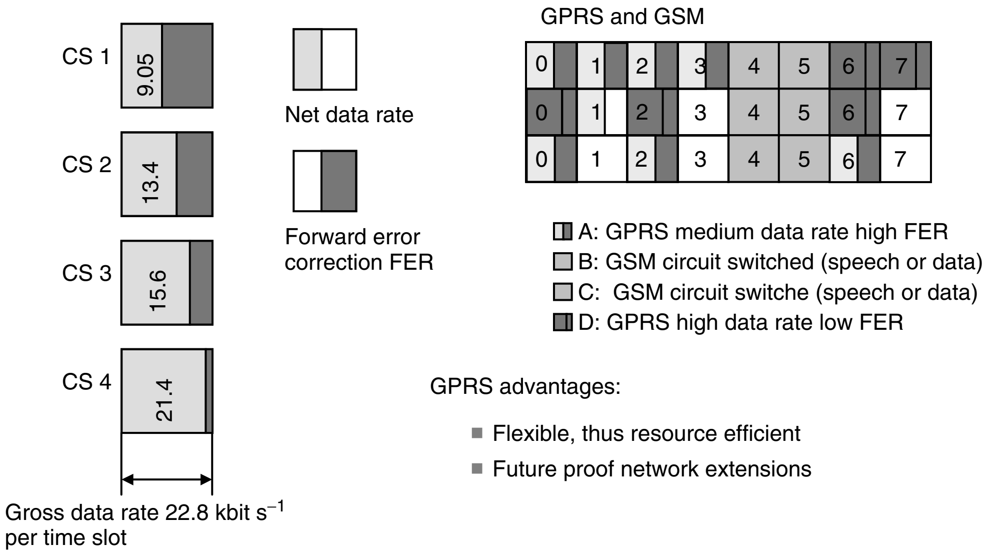


Figure 2.10 GPRS time-slot sharing and coding schemes

from the frequency carrier and the time slot number to which subscriber the transmitted data belongs, i.e. in GSM a user channel is defined by a time slot number and frequency pair (UL and DL) and this is not the case in GPRS. The assignment of carrier and time-slot number(s) to a certain user has to be managed dynamically in GPRS and this applies to both UL and DL, i.e. the allocation of UL and DL resources is managed separately, which means that while one subscriber makes use of a particular DL time slot, another can be using the corresponding UL time slot. This is done by the radio resource management functions, which will be described in more detail in Chapter 3 and is mentioned here only to introduce the concept and highlight the difference to the fixed bundling method of HSCSD.

In HSCSD, the number of time slots to be made available for the subscriber is decided at the beginning of a data call and is then allocated for the entire duration of the call. In GPRS however, following an initial 'logging on' procedure known as a 'GPRS attach', there will be no channel assignment whatsoever until the user is ready to send/receive some data. Only when there is actually a transmission request to or from the attached subscriber will it be decided how many time slots he or she will receive. If there are no other GPRS users, he or she might actually be allocated all eight time slots for the duration of the data transmission. However, if there are many users, he or she might be allocated only one or two time slots. In order to enable maximum flexibility, this decision can be revised after every four TDMA frames, i.e. one user might have several time slots continuously *or* one time slot has several users who must each wait for permission (from the PCU) to use it. GPRS makes use of a mechanism called media access control (MAC), which is described in detail in Chapter 3. It is important to note here that the MAC protocol uses 3 bits to allocate UL resources and 5 bits to allocated DL resources, which means that there can be a maximum of eight subscribers sharing an uplink time slot and 32 subscribers sharing a downlink time slot (the number actually implemented will depend on the network operator). This is what makes GPRS much more flexible than HSCSD: up to 64 subscribers can share or bundle the eight time slots of one DL carrier frequency depending on their individual needs. Theoretically they never actually have to detach themselves from the service as they do not block any resources as long as there is no transmission request. This is the famous ‘always on’ functionality more accurately known as a ‘logical link’. The logical link is analogous to the relationship between your local post office and your home. Once the post office knows your address, a link exists between them and you which only exists as a logical concept but is utilized physically when somebody sends you something. In GPRS, the user is attached to the network but resources are only allocated when required (see Figure 2.11). This forms the basis of the GPRS ‘pay for use’ or ‘volume billing’ approach.

2.3.2 GPRS Subscriber Profile

The GPRS subscriber profile contains information about the content of a subscriber’s contract. When quality of service is implemented, the operator will have to ‘package’ the QoS parameters in such a way that the subscriber has a simple choice to make rather than be confronted with complex parameters, i.e. the customer cannot be expected to choose a mean throughput class, reliability class or a delay class, etc. The customer must be offered levels of service, e.g. bronze, silver and gold, each of which has a certain QoS associated with it. The parameters underlying each service class will be stored in the subscriber profile along with other aspects such as IP address, screening profile, etc. Usually a service request from the user will be validated against the user’s subscription profile. The profile is stored in the HLR extension known as the GPRS Register (GR). It is also possible to renegotiate a request, if the network can not fulfil the users demands.

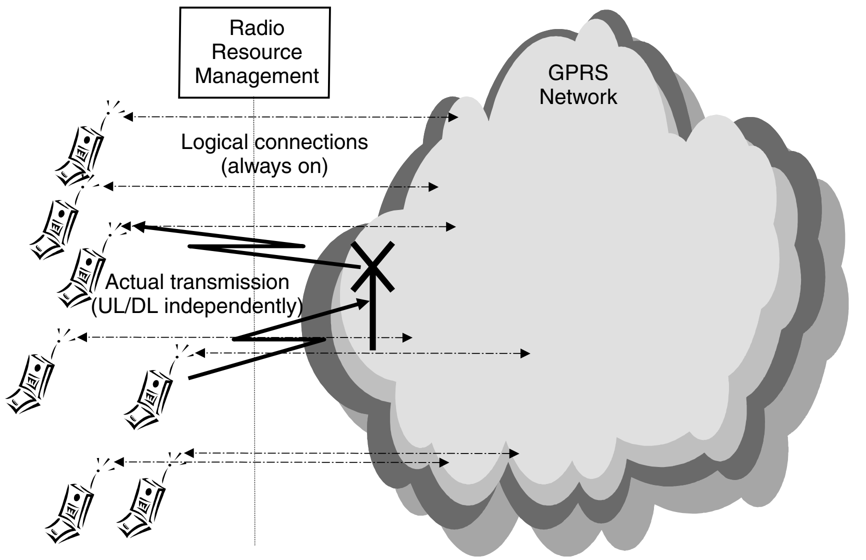


Figure 2.11 Radio resource management and ‘always on’ functionality

2.3.2.1 Transmission Services

Another characteristic of GPRS is that it actually offers more than just a single data service. A lot of different data services and different fixed data networks are already in existence, and GPRS was designed to give access to these networks. So just as GSM had to be adapted to support interconnection to established voice call services in digital ISDN or analog POTS networks, GPRS needs to be able to inter-work with existing networks.

In GPRS, two network services are supported (Figure 2.12):

- Point-to-point connectionless network service PTP-CLNS
- Point-to-point connection orientated network service PTP-CONS

The term 'Point to Point' (PTP) is used because data packets are always transferred from a source to a destination, only one of which has to be a GPRS subscriber, i.e. data can be transferred between users or between a user and a server (e.g. on the WWW, an intranet, etc.). In the PTP-CLNS, each data packet is routed independently of all others through the network. Therefore, no connection has to be established beforehand and each packet can find its own way through the network. This is very useful for bursty traffic, since it avoids the congestion often associated with single dedicated paths. This is a so called 'datagram' service, such as that used in the Internet or more generally in TCP/IP networks. The support of the Internet protocol IP of the TCP/IP suite was the main intention of this service. The PTP-CONS is obviously connection oriented, which means that a dedicated path has to be established between two users before they start the actual transmission. The connection may last only a few seconds or even hours. This should not be confused with a circuit switched connection (as for a voice call). PTP-CONS is just like PTP-CLNS used for bursty traffic so the relationship between the two subscribers is only a logical connection and only when they actually have to transmit data,

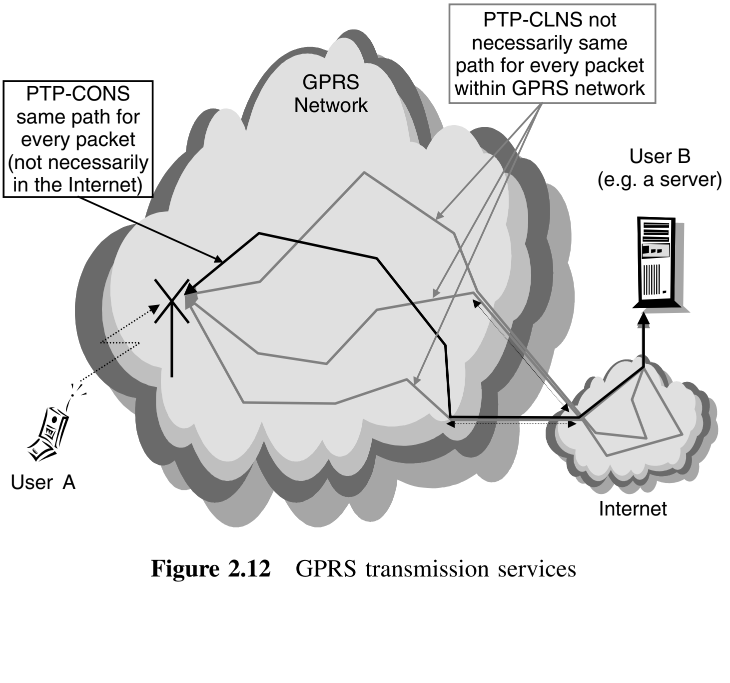


Figure 2.12 GPRS transmission services

it becomes a physical connection on the dedicated line. The circuit switched connection, in turn, is a physical link from the beginning to the end. The connection orientation can be useful for interactive applications where the delay for a reaction to some action of the user has to fulfil certain requirements. These can be planned more easily on a dedicated path (the connection orientation) than with a datagram service. It is also used when an action of the user triggers a transaction with certain requirements for reliability, safety and system response.

GPRS also supports broadcast functionality to several subscribers, which is known as 'point to multi-point'. This system takes a packet from the source and sends copies of the data in the packet to different destinations, i.e. each copy has a different address in the destination address field of the IP header but the payload is the same.

2.3.2.2 Quality of Service (QoS)

As we are already aware, GPRS is designed to allocate the scarce (radio) resources dynamically in order to be able to respond efficiently and effectively to local traffic conditions and subscriber demands. The more users there are transmitting data at any one time, the less data that can be transmitted by each one individually. The more users sharing a time slot, the longer the (potential) transmission time for each user.

The active user (of the subscribers sharing a time slot) can be changed frequently by the PCU (actually every four TDMA frames – for an explanation, see Chapter 3) so that each MS transmits in turn, i.e. many users sharing a time slot are not blocked by one user sending a large file. Even though the MS may request a certain bandwidth, the network will allocate resources on the basis of providing the best QoS to all current users. As a result, users cannot expect to get a consistently small delay or fixed data rate as with circuit switched connections. They also have to cope with the higher probability of data loss. However, this is the price that must be paid to provide maximum flexibility for a much larger number of subscribers. There are several quality of service classes that provide predictability of the data transmission for the certain applications. The class will be 'chosen' by the subscriber for the whole contractual period and has to be negotiated by every new application and the network.

Precedence Class (priority)

The precedence class (Figure 2.13) is used to further enhance the flexibility of GPRS. Packets can be marked with their respective precedence class, to indicate whether they need to be transmitted with high priority or may be discarded when necessary (in order to relieve network congestion for example), i.e. low priority.

There are three possible values:

- High precedence: Service commitments will be maintained ahead of all others, which means a premium can be charged for this service level.
- Normal precedence: Will fulfil the average requirements of the average subscriber.
- Low precedence: Service commitments will be maintained after all higher precedence levels have been fulfilled, which means that this service level will be cheaper but will carry the risk that packets may be deleted on route if necessary.

Data packets in GPRS networks are stored (buffered) for a very short time in each network node along the route between source and destination. These nodes, such as the

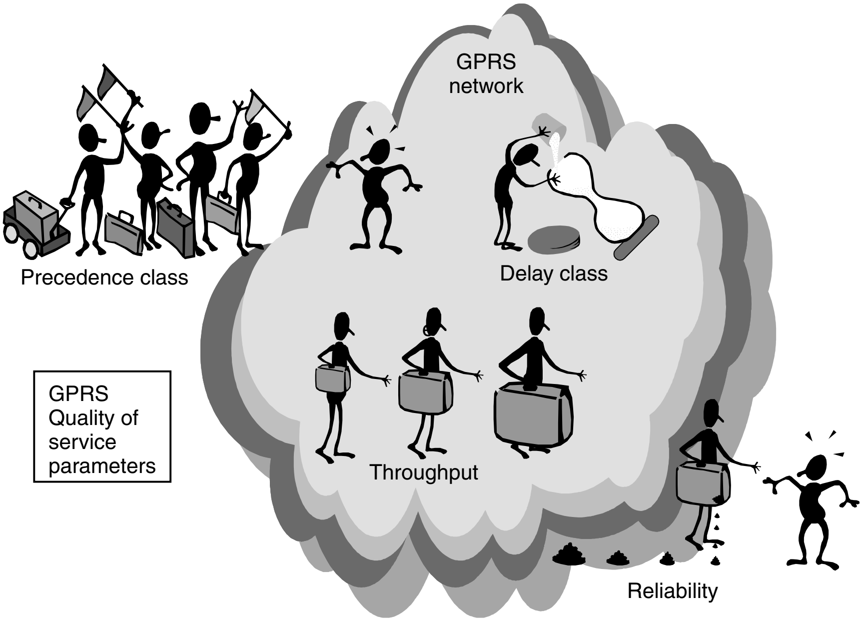


Figure 2.13 Quality of service classes

PCU, SGSN and GGSN, must continuously make decisions based on the precedence class of the subscriber, as to whether packets should be forwarded immediately, buffered until resources are available or deleted. In the case of the PCU, a subscriber's precedence class can be used to give access to shared resources on the air interface, i.e. if all subscribers sharing a time slot have the same precedence class, then statistical multiplexing can be used, whereas if one user has a higher priority, then the other users must wait.

Reliability Class

Low precedence indicates that data may be discarded in case of network congestion, but high precedence does not guarantee flawless transmission. Different applications rely differently on the reliability of the transmission. For example, the TCP in TCP/IP controls the data flow very thoroughly and initiates a retransmission in case of a packet loss. Thus, applications on top of TCP/IP, such as http or ftp, can count on perfect transmission regardless of whether packet loss occurs or not. On the other hand, an application based on UDP relies heavily on the correct transmission and should (ideally) be operated via reliable connections only. It is not easy to define different grades of reliability as common sense dictates that a transfer or connection is either reliable or not, but it is possible to define some reliability parameters and with a statistical approach over many transmissions a relative reliability can be defined, i.e. the statistical probability that a so called service data unit (SDU) will not be delivered 'problem free'. Typical examples of reliability problems are:

- The possibility, that an SDU gets lost, e.g. discarded due to a transmission error. (lost SDU probability);

- The possibility that an SDU will be delivered twice (duplicate SDU probability);
- The possibility that a set of consequent SDU arrive in the wrong order; (out-of-sequence probability);
- The possibility that a user receives a corrupted SDU without the error being detected. (corrupted SDU probability).

Again, GPRS offers a flexible solution to different needs by introducing three different reliability classes, which are certain combinations of error probabilities based on the four points listed above. The various error probabilities of the different classes are compared in Figure 2.14.

Reliability Class 1 means that, statistically, each of the four SDU error probabilities is below one faulty SDU out of every billion transmitted SDUs. Thus the probability equates to 10^{-9}.

Reliability Class 2 sets a limits from 1 in 10 000 for lost SDUs up to 1 in 1 000 000 for corrupt SDUs.

Reliability Class 3 allows the Corrupted SDU probability and the Lost SDU probability to go up to one faulty SDU out of 100 transmitted SDU, whereas the probability of duplicate or out-of-sequence SDUs is the same as Class 2: 1 in 10 000.

It is up to the application to demand the required reliability class and the above values have been selected to enable applications with different error sensitivity – e.g. streaming video and telemetry data – to request the appropriate service class. For example, Class 1 would be necessary for applications that are error sensitive but which do not offer any error correction themselves. Applications with good error tolerance capabilities or limited error correction functionality (in case they are error sensitive) could cope with Class 2.

Reliability class	Lost SDU probability (a)	Duplicate SDU probability	Out-of-sequence SDU probability	Corrupt SDU probability (b)	Example of application characteristics
1	10^{-9}	10^{-9}	10^{-9}	10^{-9}	Error sensitive, no error correction capability, limited error tolerance capability.
2	10^{-4}	10^{-5}	10^{-5}	10^{-6}	Error sensitive, limited error correction capability, good error tolerance capability.
3	10^{-2}	10^{-5}	10^{-5}	10^{-2}	Not error sensitive, error correction capability and/or very good error tolerance capability.

Figure 2.14 Reliability classes in GPRS

Class 3 would only be suitable for applications that are either completely insensitive to errors or which offer a sufficient level of error correction themselves.

The network may or may not be able to grant the requested reliability class. In the case of high system load or if the requested class has not been implemented (e.g. early GPRS networks operate on a 'best effort' basis) then the request has to be renegotiated.

Delay Class

As mentioned previously, there will always be some delay in transporting the data through the network (even short storage in switching devices), and as the packets from one application can be routed via different paths, the delay will not be constant (this is known as 'jitter'). Again there are some applications that are delay sensitive, e.g. real-time applications, or interactive applications (where delays in system responses to user input cause frustration and customer dissatisfaction), and some that are not, e.g. short messages (SMS), as the user will not usually be aware of delays of up to several minutes or even hours.

In GPRS an application can select one of four different delay classes that are specified within the GPRS network only. The values actually specified are the mean delay and the time when 95 % of the data are most likely to arrive. So 5 % of all data are not explicitly specified, only implicitly, because overall the data have to match the mean delay parameter. The GPRS specifications refer to two different packet sizes – these being 128 octets and 1024 octets respectively. The delay is usually measured from user to user, or user to GGSN (i.e. not from hop to hop but for the whole transmission path) but can be guaranteed only within the GPRS system, since delays in external networks such as the Internet are unpredictable. So when external networks are accessed, the GPRS delay is valid from the GPRS gateway support node (GGSN) via the whole network and the radio interface to the GPRS mobiles.

The delay Classes 1 to 3 offer certain guaranteed delay values (see Figure 2.15). Class 4 does not specify any values, it just offers best effort transmission and thus is heavily dependent on the network load. Class 4 is the only class that has been implemented right from the start by all network operators.

Delay class	Delay (maximum values)			
	SDU size:128 octets		SDU size: 1024 octets	
	Mean transfer delay (sec)	95 percentile delay (sec)	Mean transfer delay (sec)	95 percentile delay (sec)
1 (predictive)	<0.5	<1.5	<2	<7
2 (predictive)	<5	<25	<15	<75
3 (predictive)	<50	<250	<75	<375
4 (best effort)	Unspecified			

Figure 2.15 Delay classes in GPRS

Throughput

The throughput is divided into two different parameters, maximum bit rate and mean bit rate.

The theoretical maximum bit rate specified in GPRS is 171.2 kbit s^{-1}, i.e. one subscriber using coding scheme 4 (21.4 kbit s^{-1}) on all eight time slots on one carrier. Firstly, the transmission of eight time slots is not necessarily supported by all GPRS mobiles (see Chapter 7 for details). At the time of writing, most mobiles on the market offer four time slots on the downlink and one or two on the uplink. Secondly, CS 3 and 4 are not supported by all network providers. At the time of writing, CS 2 with a maximum transfer rate of 13.4 kbit s^{-1} is the limit. Solutions are being developed to overcome the bottleneck on the Abis interface and enable the use of the higher data rates of CS 3 and CS 4, with minimal changes to the access lines.

With or without the maximum rates offered by eight time slots and CS 4, the fact remains that network load will change dynamically according to subscriber demand and therefore it would not be in the best interests of the subscribers or the operators to provide each user with a fixed throughput guaranteed for the duration of the GPRS session. One of the main goals of GPRS has always been flexibility, and this also applies to the throughput made available to each subscriber at any particular point in time. Accordingly, the maximum throughput has to be negotiated (between the MS and the network) when the mobile attaches to the network, i.e. the mobile requests certain resources and the network decides what resources will actually be made available. As the negotiated throughput could be significantly lower than the level desired by the user (or the capabilities of the mobile), the network should also be able to alter this value at any time if a change in network load occurs. The maximum bit rate is not a very useful value for a packet switched service, as it is not actually used for long duration transmissions and does not describe the bursty nature of most packet switched applications. For this reason, a mean bit rate is also negotiated. This is a long term level of throughput that the user can rely on. It includes both short term transmissions at maximum throughput and transmission pauses when no data are pending. The mean bit rate can also be renegotiated during a session, in case the conditions in the network and especially the air interface have changed.

2.4 Logical Functions

Both GSM and GPRS have the same goal: to enable the mobile subscriber to establish communication links from a mobile device to a second party. As this common goal is enabled by the same mobile device via the same radio subsystem (RSS), we can expect that they will also have many similarities in terms of the logical functions necessary for operation, such as network access control, mobility management, radio resource allocation and collection of billing data. This is true to some extent, but we must also expect that the requirements and procedures of some of these functions can differ significantly due to the different nature of the communication involved, i.e. circuit switched versus packet oriented. Added to this, GPRS has many functions not found in circuit switched GSM. This section explores the logical functions of GPRS so that the reader can appreciate the common ground, the similarities and the differences as compared with circuit switched GSM.

2.4.1 Network Access Control

Network access can simply be described as the way in which a user becomes connected (or 'attached') to a network in order to use the services and network facilities. The reality is much more complicated than this simple statement implies, since users not only have to be registered but also be checked for their authenticity, their subscriber profiles, whether there are enough network resource to handle another user and so forth.

The term 'GPRS attach' summarizes all the necessary steps to get access to the network and can be thought of as 'logging on' to the network. The opposite expression 'GPRS detach' describes the procedures for 'logging off'. The following paragraphs will explain these functionalities in detail (Figure 2.16).

2.4.1.1 Registration

A mobile has to perform a registration operation in order to make itself known to the network. After successful registration, the network should, for example, know the current routing area of the user, the user's permanent IP address if the user has one (IPv6) and his access point names (APN), be they for the Internet, a company LAN or both. In early versions of GPRS operating under IPv4, the IP address is dynamically allocated during the PDP context activation (see Section 2.4.2.1), but later releases operating with IPv6 will store the permanent IP address of the MS in the GPRS register (GR) as part of the subscriber profile. The outcome of a successful registration should also be a valid service profile (QoS, screening profile) in the SGSN, which could be the static subscriber profile taken from the HLR or, in case of renegotiation, a dynamic one as discussed in the previous section. So as a result of a registration, the mobile station should be ready for activating a PDP context which will determine the transmission of GPRS data between the MS and a data network (access point).

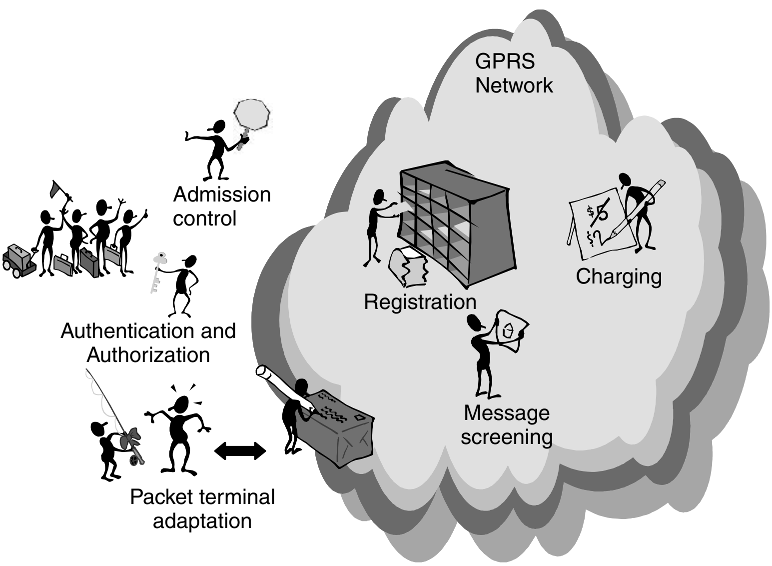


Figure 2.16 Overview of network access control functions

2.4.1.2 Authentication and Authorization

Authentication refers to the necessity for checking the identity of the MS before it is allowed to make use of network resources. Initially this procedure was intended to protect the subscriber from fraudsters who would make illegal use of the network by stealing and using their identities. With the rise of mobile services and m-commerce, the authentication procedure is seen as even more important as the SIM card is being used more and more as a credit card, e.g. it is already possible to purchase goods and pay taxi fares with your mobile telephone. The procedure involves an exchange of electronic keys between the VLR and the MS, and will be covered in detail in Chapter 4, GPRS Procedures. During the procedure an electronic signature (actually a bit sequence) is sent to the VLR from the AC. The VLR then sends another key to the MS with which the MS should be able to generate the same electronic signature as that supplied by the AC. If the two signatures (bit sequences) are the same, the authentication is successful and the MS is provided with access to the network, otherwise access is denied.

Authorization refers to the fact that following a successful authentication, the VLR is provided with a copy of the subscriber data which specifies what types of calls and services the subscriber is allowed to make, i.e. has paid for. So when a user attempts to make a phone call or use a service, the VLR checks the subscriber profile to see whether or not the user is allowed to do so.

Thus it can be said that authentication gives a user access to the network and authorization enables this user to make use of the network. A good example of this is SMS in Europe. Although many operators in different countries have not yet made roaming agreements with each other, it is possible in most countries to send and receive SMS, i.e. the roaming MS are authenticated and therefore have access to the networks but the subscriber profiles forbid them from making telephone calls as there is no roaming agreement, the only agreement is for SMS.

2.4.1.3 Admission Control Function

The admission control function is the one that calculates the network resources needed to fulfil a certain subscriber's request. It then has to check whether these resources are actually available (or not) and then allocate the available resources to the particular subscriber. The checks that the admission control function has to perform not only cover quality of service (QoS) aspects such as mean throughput, reliability class and priority, but also have to be associated with the radio resource management in order to allow for the frequency and time slot requirements for every GPRS subscriber in the cell.

2.4.1.4 Message Screening

The message screening function is concerned with filtering unauthorized and unsolicited messages. It should be implemented on a per packet basis as in a firewall, and is really just a basic filtering function for packets from undesirable sources or with undesirable content. This does not serve as the one and only security measure in a GPRS network. It can not find malicious code on an application level or prevent different kinds of attack from the Internet or from within the mobile network itself, say via a compromised GSN.

The message screening function is usually implemented on a network basis and might later be upgraded for an individually customizable screening function for every single user.

2.4.1.5 Packet Terminal Adaptation

Very often the mobile will be connected to another device on which an application is running. The mobile must therefore be able to adapt to the variety of interfaces which can be used to exchange data between itself and the external device. The packet terminal adaptation function fulfils this role, i.e. packets received via the air interface must be stored and adapted before being sent to the application running on the external device.

2.4.1.6 Charging Data Collection

There are two different functions necessary for collecting charging data in GPRS: use of the GPRS network belonging to a subscriber's HPLMN, and access to external networks.

The billing for use of the GPRS network to which a user subscribes can be based on various models, the most popular of which seems to be a subscriber's traffic volume (in kbytes or Mbytes). The charging data for traffic within a GPRS network, i.e. user to user or user to gateway, is collected by the SGSN to which a subscriber is attached.

The billing for use of external networks such as the Internet can also have different basis. For example, billing for the use of the Internet is currently based only on access but there is increasing interest in also charging for content. The charging data for access to external networks is collected by the GGSN.

As part of the 'logging on' process (see 'PDP context activation' in Section 2.4.2.1) the GGSN sends a so called 'charging ID' (CID) to the SGSN. During the billing process batches of charging data records (CDR) are regularly sent from each network node to a central billing centre. The CID is used to merge the records from different network nodes which apply to the same subscriber.

2.4.2 Packet Routing and Transfer Functions

The transfer of voice data in GSM is relatively simple to follow as it is circuit switched. There is one-to-one mapping between time slots from the RSS and time slots in the NSS, and these time slots always follow the same path from end to end. The transfer of data packets in GPRS however is not so simple to follow as the packets are very large in comparison with the time slots on the air interface, and packets moving between the same source and destination can be sent via different paths. Added to this is the problem of the different interfaces with different protocol structures over which the packets must be sent. The terminology used to describe the way packets are handled by the protocols and the way the packets are transferred must be understood in order to comprehend fully the process involved. For example, in GSM we speak of 'connections' while in GPRS we speak of 'routes'. The former is a fixed path followed by every time slot. The latter is the path followed by an individual packet which can be the same as or different from, the path followed by the next or previous packet. The logical functions on the transmission path are illustrated in Figure 2.17.

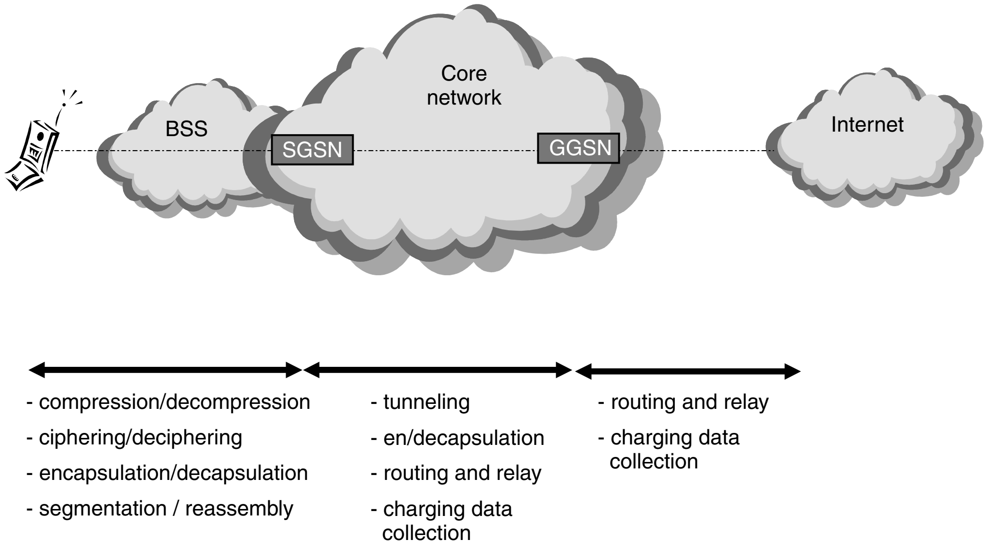


Figure 2.17 Logical functions on the transmission path

2.4.2.1 The Routing and Relay Functions

These terms refer to what happens when a packet travels across a packet data network. Described simply, the packet arrives at a node, the routing function decides which network node the packet should be sent to next in order to reach the destination eventually, the packet is sent to the next node and the above process is repeated. The relay function is the means by which a node forwards the packets received from one node to the next node which is determined by the routing function. These functions are not unique to GPRS – every packet data network must perform these functions.

Routing is often compared with sending packages through the post. In some ways this is a good analogy in that the postal system also makes use of addresses that enable the package to be transported across various infrastructures to the destination. This analogy does not, however, reflect the complexity of the routing decisions which must be made by the routing function in data networks. The movement between two nodes is called a ‘hop’ and routing decisions must be made on a ‘hop-by-hop’ basis according to a set of rules. The rules can dictate that decisions are based on the current load on each hop, the available bit rate on a hop, or other such parameters (see Chapter 8 for more examples). Data transmission between GSNs may also take place via data networks that implement their own internal routing functions, e.g. IP over Ethernet, frame relay or ATM networks (the Gn interface used between GSNs, can for example be realized as IP over ATM).

2.4.2.2 Address Translation and Mapping Function

The previous Section (2.4.2.1) mentioned that packets may be transported across several different networks in order to reach the destination. If the networks concerned do not use the same kind of addressing, then a translation function is required, i.e. the addressing information in the packet header will be used to generate a new header with an address which can be used by the second network. It should be noted that address translation can be used for routing packets within and between GPRS networks, and also between GPRS networks and external data networks.

If, however, the networks concerned do use the same kind of addressing then address mapping rather than translation is used. This means a one-to-one relationship will be established between a network address in one network and another network address in a second network, i.e. an entry in a so called 'look up' table will be made which determines where packets with a certain address will be sent.

2.4.2.3 Encapsulation

This function is one of the most important aspects of the chapter on interfaces and protocols (Chapter 3). When a packet is sent across an interface, some extra information must be attached to the packet. This extra information has various forms and various uses and can be likened to a letter being sent through the post. The letter itself must be put in an envelope with the destination address printed on it. At the local post office the letter is put in a sack which is transported to the central post office. At the central office, the contents of the sack are sorted into other sacks according to destination (city or country) and the sacks for each destination are put into containers for transportation. The transportation may involve several steps before a postal employee, who has a bag containing letters for delivery in the same street or area, finally delivers the letter. The letter, envelope and address are like an IP packet, i.e. information with an address. Putting an envelope in a sack, or a sack in a container, or a container in a boat/plane/train/truck is encapsulation – i.e. the address defines the ultimate destination but the letter is not itself in a suitable form to be individually transported through the postal system. The sacks enable easy transport by the local postal service and the containers enable easy transport to a certain destination by the national postal service. It is important to note that encapsulation can have many layers, i.e. {[(envelope) sack] container}, and that each layer is added for a specific purpose and is removed when that purpose has been achieved. In the case of data packets, additional information is added in the form of headers and check bits which 'encapsulate' the original packet. The removal of this information when it has served its purpose is called 'decapsulation'.

The protocols described in Chapter 3 will clearly demonstrate the importance of en/decapsulation – the Gn Interface between the SGSN and the GGSN is a good example of multiple layer encapsulation of an IP packet.

2.4.2.4 Segmentation and Reassembly

Depending on the physical transport media, i.e. the technology used to realize layer 1, it can be that the datagram is too large to be sent as a single transmission. In that case the datagram must be divided into several smaller segments, each of which receives a header to enable the original datagram to be reassembled on the receiving side. For example, the protocols that prepare an IP packet for transmission over the air interface make use of segmentation, e.g. the IP packets are segmented into smaller frames (called LLC frames) which, in turn, are segmented into radio blocks, each of which is then divided into four time slots (see Chapter 3).

2.4.2.5 Compression

In order to optimize the use of the air interface, i.e. to make packet transmission times as short as possible, the IP packets are compressed before being segmented, ciphered and transmitted. Compression means that the packet is processed and, for example, repeated sequences of zeros and ones are minimized. The compression must reduce the number of bits to be transmitted but still guarantee that the original bit sequence can be reconstructed. This function is carried out by the SNDCP layer, described in Chapter 3, before the IP packet is segmented.

2.4.2.6 Ciphering

The ciphering function in digital systems involves the combination of an information bit sequence with another bit sequence known only to the sender and the receiver (the exact procedure will be covered in Chapter 4). The goal is to ensure that the transmission is secure, i.e. the privacy of the subscriber is guaranteed and signalling information is protected. The ciphering function in GPRS is carried out by the LLC layer mentioned in Section 2.4.2.4, i.e. IP packets are compressed and segmented into LLC frames which are ciphered before being segmented into radio blocks.

2.4.2.7 Tunnelling

When an application on a mobile device sends an IP packet, the header attached to the packet contains the IP address of the MS and the IP address of the destination, e.g. an Internet or intranet address. This packet is sent across the RSS and arrives at the SGSN. The packet must now be sent from the SGSN to the GGSN via the Gn interface. A problem arises here because the Gn interface is realized using an IP network, i.e. the SGSN has an IP address and the GGSN has an IP address, which means that the IP packet from the mobile must be encapsulated in another IP packet for transport across the Gn interface – i.e. IP over IP. The problem arises because the packets from many mobiles are being sent along the same path so an extra protocol is required to differentiate between packets from different mobiles and also between different applications running on the same mobile. The GPRS tunnelling protocol defines the entry and exit points on the Gn interface, i.e. the point of encapsulation to the point of decapsulation. A tunnel is a two-way point-to-point path identified by a tunnel identifier (TID) which defines only the end points.

2.4.2.8 Domain Name Server

When an MS requests access to a data network, it sends the access point name (APN, see Chapter 4 and Chapter 6 for further details) of the requested network – e.g. intranet.tfk.de – to the SGSN. The SGSN does not know what to do with this APN, so sends a request to the domain name server (DNS), which is able to resolve this logical name into the IP address of the GGSN directly connected to the required data network. In this way the SGSN is able to contact the right GGSN.

2.4.3 Mobility Management

Mobility management (MM) has, of course, been a necessary aspect of all mobile networks prior to GPRS, but there are some innovative features in GPRS mobility management

(GMM) functionality which differ from those in circuit switched networks. These will be discussed in the following. An integration of classical GSM CS mobility management with the GPRS functionality is both possible and useful as it reduces signalling traffic on the air interface and across the network. This so called common mobility management (CMM) requires the implementation of the Gs interface from an SGSN to an MSC/VLR in order to allow signalling exchanges between the network elements responsible for MM. If this interface is not implemented, the MS makes location updates for GSM and routing area updates for GPRS separately. If the Gs interface is implemented, the MS only needs to make a single update and both the GSM and GPRS location registers will be updated.

2.4.3.1 Mobility Management States

A GPRS mobile station (MS) can be in one of three different states (Figure 2.18): idle, standby, ready. In the idle state, the mobile station is not known to the GPRS network at all. If the GPRS MS is also a mobile phone, i.e. is not a dedicated GPRS device, then it will be known to the VLR and will be making location updates while switched on. The GPRS network does not hold any location information about the MS in idle state and no paging requests nor data can be sent.

In the standby state the MS and the SGSN have established a mobility management context, i.e. the location (routing area) of the MS is stored in the SLR and the MS performs routing area updates when it passes routing area borders. Thus it can be paged for voice and data calls if necessary. In this state the MS may activate or deactivate a PDP context which is a prerequisite for actually transmitting data.

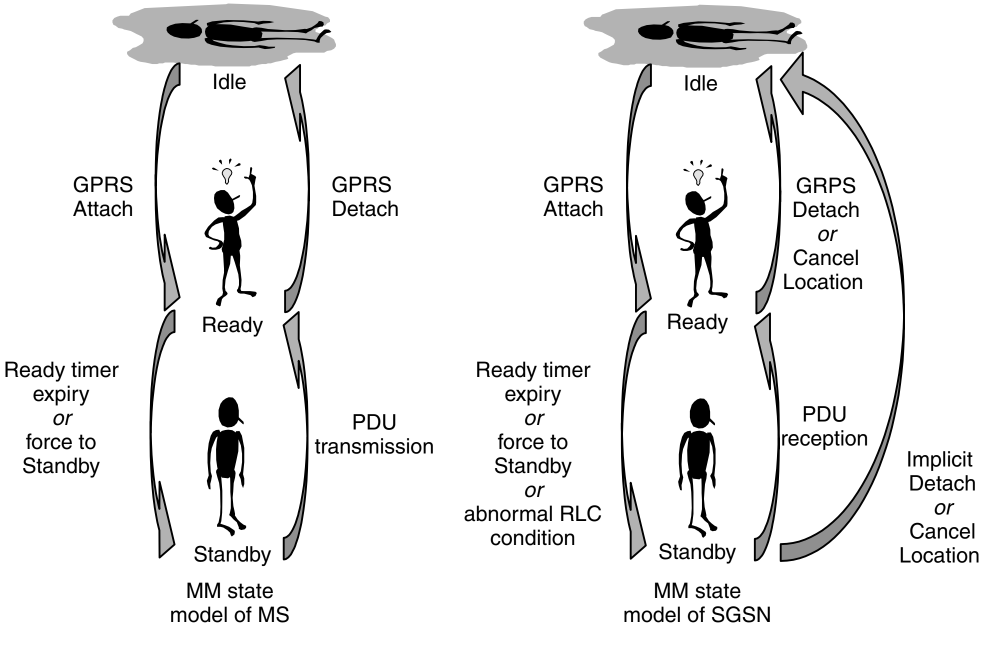


Figure 2.18 Mobility management states

If the MS is paged and responds, it will switch to the ready state. Also in the SGSN the mobility management context will be changed to the ready state. In the ready state, the exact cell location of the MS is known (comparable to an active voice call). Because of the discontinuous nature of GPRS traffic, the cell selection and reselection is usually performed by the MS, i.e. the MS monitors the surrounding base stations and decides for itself which one to use. The requests from the MS to make use of a particular cell may be accepted, or the SGSN may instruct the MS to use a different cell. The MS must be in the ready state in order to activate a PDP context to send and receive packet data. However, being in the ready state is not equivalent to having radio resources allocated to the MS, nor does it mean that a PDP context has to be active.

2.4.3.2 Transitions between Idle, Standby and Ready State

The transition from one state to another depends on the current state. Keep in mind that both the MS and the SGSN have to change to the corresponding state, and that there are some differences in how the transition may be initiated.

The transition from idle to ready state is known as the GPRS attach procedure. The MS requests access to the network, an MM context is established for its particular IMSI and the cell global identifier CGI of the selected cell is stored in the SGSN.

The standby state is usually reached by expiry of the ready timer in the SGSN. This will occur if the MS has neither sent nor received any data within the period of time defined by the timer – a typical value being 40 seconds.

In order to return to the ready state, the MS has to send any data or signalling information; this might be initiated by a page call of the SGSN. The SGSN changes the MS's state to ready only when receiving this data or signalling information, not when sending a paging request.

The transitions from ready to idle state or from standby to idle state are known as the GPRS detach procedure. Such a transition is usually initiated by a GPRS detach request from the mobile. Then the MS is stored as idle in the SGSN, which means for the network that this MS is GPRS disconnected.

There is also a special anonymous access mobility management (AA MM, Figure 2.19) for certain PDP contexts which are handled without storing any user data. In case of an AA MM, only the states ready or idle exist, i.e. if there are no PDUs to be transmitted, no data about the user are stored at all. In this case the transition from idle to ready is not called GPRS attach, but is known as 'AA PDP context activation' and the reverse is 'AA PDP context deactivation'. The expiry of the ready timer will cause a fallback to idle state and all data concerning this user will be deleted. An MS and an SGSN may have several anonymous and/or non anonymous PDP contexts at the same time.

2.4.3.3 Common Mobility Management

As previously mentioned, the common mobility management (CMM) can only be supported if the optional Gs interface between SGSN and MSC has also been implemented (see Figure 2.20). Depending on whether CMM is used or not, a network is said to be operating in a certain network operation mode. There are three network operation modes:

- A network element featuring the Gs interface that supports CMM is operating in the so called network operation mode one (NOM I). In this mode the MS will only have to

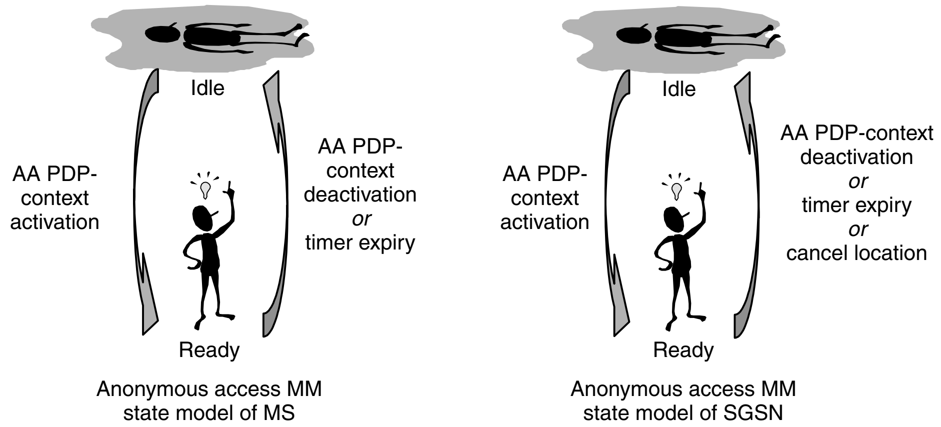


Figure 2.19 Transitions in anonymous access MM state

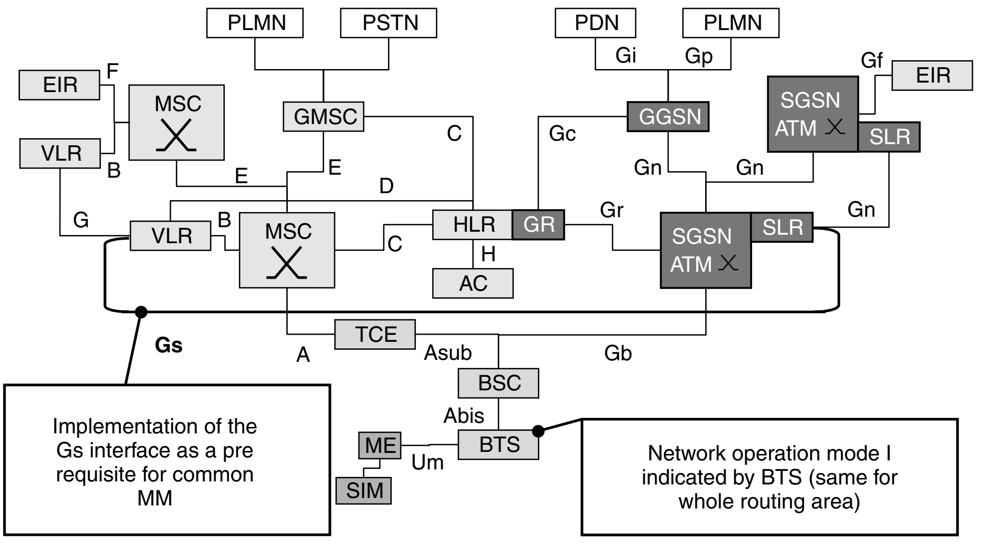


Figure 2.20 Network overview for CMM

monitor one single paging channel (GSM paging channel PCH or GPRS packet paging channel PPCH, see Chapter 5) for CS and PS calls. During a data transfer, the MS will receive paging messages via the assigned packet data channel, i.e. it does not need to monitor a separate paging channel.

- In network operation mode II (NOM II) the MS has to monitor a single paging channel as well (GSM PCH only). Unlike NOM I, the MS continues to listen to the GSM PCH during a data transfer. Paging between SGSN and MSC will not be coordinated.
- In network operation mode III (NOM III) an MS shall monitor the GPRS packet paging channel (PPCH) and the GSM paging channel (PCH).

NOM I, II or III will be indicated by the cell on the broadcast channel and should be consistent throughout a routing area.

If CMM is used, the SGSN/SLR and the MSC/VLR have to be associated with each other which is done by simply storing each others identity (SS7 identifier). There is not necessarily one SGSN for every MSC, so there are usually many VLRs associated with one SLR, especially when GPRS is first implemented in a network. Whenever an MS enters the SLR or VLR area, and the MS is both IMSI attached for voice calls and GPRS attached for data calls, the association supports the following functions in order to save radio resources:

- combined GPRS/IMSI attach and detach;
- coordinated location and routing area updates (location area update forwarded by the SGSN to the VLR);
- paging for CS connections via the SGSN;
- alert and identification procedures.

The association is initiated by the SGSN whenever a mobile requests a combined GPRS/IMSI attach or an additional GPRS attach, or one of several combined LA/RA update cases (this includes changing to a part of the network using network operation mode I from a NOM II or III part of the network), but never updated during a CS connection.

A class 'A' terminal equipment makes RA updates during CS connection but no LA updates (LA updates are never performed during a voice call as the MS cell is always known due to handovers). A class 'B' terminal equipment does not do any updates during a CS connection as it can not provide simultaneous GSM and GPRS functionality (see Chapter 7). The association is removed on either IMSI or GPRS detach as it is no longer needed.

2.4.4 Radio Resource Management

The radio resource (RR) management procedures include the functions related to the management of the common transmission resources, e.g. the physical channels and the signalling channels.

In general, purpose of these RR procedures is to establish, maintain and release RR connections that allow a point-to-point dialogue between the network and a mobile station. This includes the cell selection/reselection and the handover procedures, which will all be discussed below. In addition to standard GPRS, the so-called E-GPRS must also be considered. E-GPRS refers to GPRS combined with EDGE (enhanced data rates for GSM evolution). EDGE is a new modulation scheme, developed to raise the net bit rate offered by one timeslot to 59.2 $\mathrm{kbit\,s^{-1}}$. An E-GPRS mobile station has to comply with GPRS standards, as well as with E-GPRS standards which include new radio access protocol features and new modulation and coding schemes, described in more detail in Chapter 5. Regardless of whether an MS uses GPRS or E-GPRS, everything described in the following section holds true for both.

2.4.4.1 Dynamic Allocation of Resources in GPRS

The definition of a traffic channel is different between CS GSM and GPRS. In GSM a traffic channel (TCH) is bidirectional, i.e. refers to a time slot on the uplink *and* downlink, each with the same time slot number (TN). In GPRS however, a packet data traffic channel (PDTCH) is unidirectional, i.e. either UL or DL.

A cell-supporting GPRS may allocate resources on one or several PDTCHs (UL and/or DL). In order to support the GPRS traffic, those PDTCHs (possibly shared by several GPRS MS), are taken from the common pool of traffic channels available in the cell. The allocation of traffic channels to circuit switched services and GPRS can be done dynamically according to the 'capacity on demand' principles described below.

The common control signalling required by GPRS in the initial phase of the packet transfer is conveyed on GPRS's own packet common control channel (PCCCH) when implemented, or on GSM's common control channel (CCCH). This allows the operator to have capacity allocated specifically for GPRS in the cell only when a packet is to be transferred. Nevertheless the operator can also allocate packet data channels permanently. This is the case for early GPRS networks because of some restrictions in the base station system equipment and the mobile stations (these limitations will be covered later).

Master/Slave Concept

In GPRS, the physical channels are referred to as packet data channels (PDCH). At least one (PDCH) acting as a master, may accommodate the optional packet common control channels (PCCCH) that carry all the necessary control signalling for initiating packet transfer. If the PCCCH is not implemented (the standard case in early GPRS networks), that signalling is carried by the existing CCCH, which has to be present for the GSM circuit switched traffic anyway. Since the PCCCHs do not use much of the line capacity of a PDCH, there will also be user data and dedicated signalling (i.e. packet data traffic channel PDTCH and packet access control channel PACCH) on this PDCH. Other PDCHs, acting as slaves (Figure 2.21), are used for user data transfer and for dedicated signalling only.

Capacity on Demand

The GPRS does not require permanently allocated PDCHs. The allocation of capacity for actual packet transfers can be based on the demand of the subscribers, which is referred to here as the 'capacity on demand' principle (Figure 2.22). The operator can, as well, decide to dedicate permanently or temporarily some physical resources (i.e. PDCHs) for GPRS traffic only. In early GPRS networks this is usually the case for one reason in particular: power control. In GSM the power control is continuous during a circuit switched connection. The BTS measures the signal strength received from the mobile and then signals the mobile to either increase, maintain or decrease its transmission power accordingly. This is the so called 'closed loop' power control, in which sending, receiving, measuring and regulating together build a closed loop.

In GPRS there is no continuous transmission, so no continuous measurement is possible, hence the mobile station can only estimate its required transmission power level, based on the strength of the signal it receives from the BTS. This is so-called 'open loop' power control – no measurements are taken and no feedback is given. For such a system, based on estimates to be workable, the MS requires a reference on which the estimates can be based. Such a reference is provided by the carrier which broadcasts the BCCH, as this is the only carrier that is always transmitted at full power. For this reason, the time slots on this carrier will be permanently allocated to GPRS in early GPRS networks, as long as the MS and BSS are not capable of a more sophisticated power control procedures.

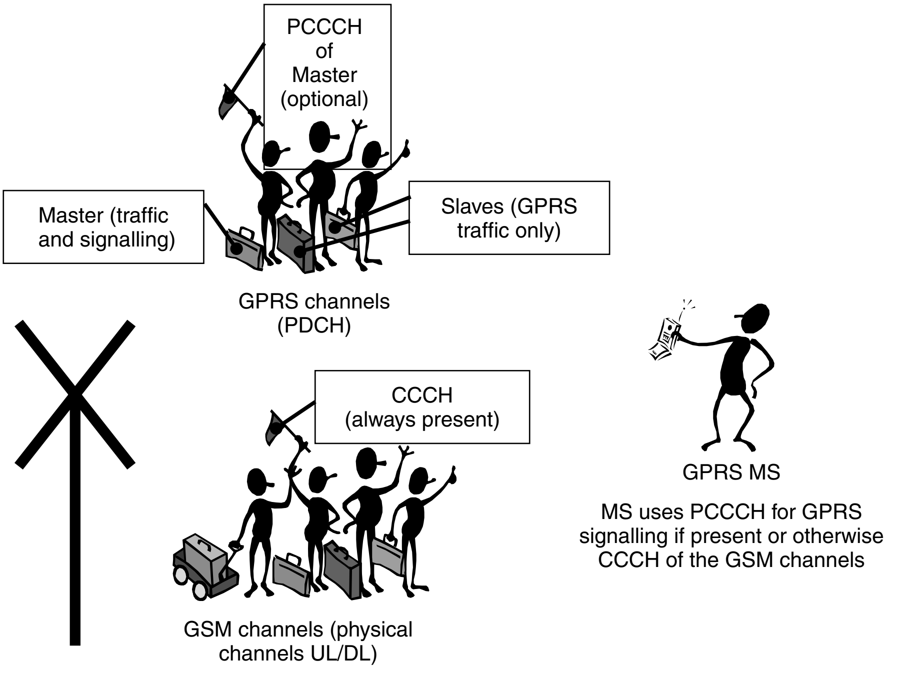


Figure 2.21 Master/slave concept

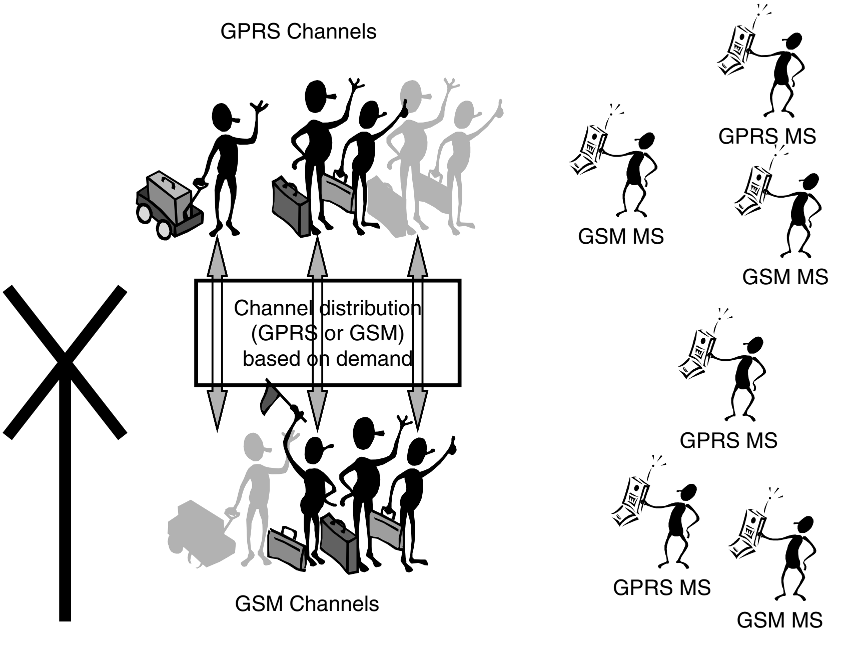


Figure 2.22 Possible combinations for capacity on demand

In later releases of GPRS, the network will be able to allocate more physical channels as PDCHs when the existing channels are congested due to the GPRS traffic load and more resources are available in the cell – however, more sophisticated power control will be needed.

The existence of a PDCH does not, however, imply the existence of PCCCH since the implementation of this channel is optional. When no PCCCH is allocated in a cell, all GPRS attached MSs camp on the CCCH.

In response to a packet channel request sent on CCCH from the MS that wants to transmit GPRS packets, the network can assign resources on a PDCH for the uplink transfer. After the transfer, the MS returns to monitoring the CCCH.

When the PCCCH is allocated in a cell, all GPRS attached MSs camp on it. PCCCH can be allocated either as the result of the increased demand for packet data transfers (which is one reason why it is not implemented in early GPRS networks with relatively low demand) or whenever there are enough available physical channels in a cell, i.e. in order to increase the quality of service. The information about PCCCH is broadcast on GSM's broadcast common control channel (BCCH). When the PCCCH capacity is inadequate, it is possible to allocate additional PCCCH resources on one or several PDCHs. If the network releases the last PCCCH, the MS performs cell reselection to the next valid CCCH.

The number of allocated PDCHs in a cell can be increased or decreased according to demand. The following principles can be used for the allocation:

Load supervision:

A load supervision function may monitor the load of the PDCHs so that the number of allocated PDCHs in a cell can be increased or decreased according to demand. The load supervision function may be implemented as a part of the medium access control (MAC) functionality. The common channel allocation function located in the BSC is used for the GSM services.

Dynamic allocation of PDCHs:

Unused channels can be allocated as PDCHs to increase the overall quality of service for GPRS. Upon resource demand for other services with higher priority, de-allocation of PDCHs can take place.

Conditions for Release of a Packet Data Channel

The fast release of PDCH is an important feature for operators who want dynamically to share the same pool of radio resources between packet and circuit switched services. The following possibilities exist for releasing a PDCH (Figure 2.23):

- wait for all the assignments to terminate on that PDCH;
- individually notify all the users that have assignment on that PDCH;
- broadcast the notification about de-allocation.

The first option does not incur any extra signalling load but could potentially take a long time. Packet uplink assignment and packet downlink assignment messages can be used to realize the second option. The network side has to send such notifications on the packet

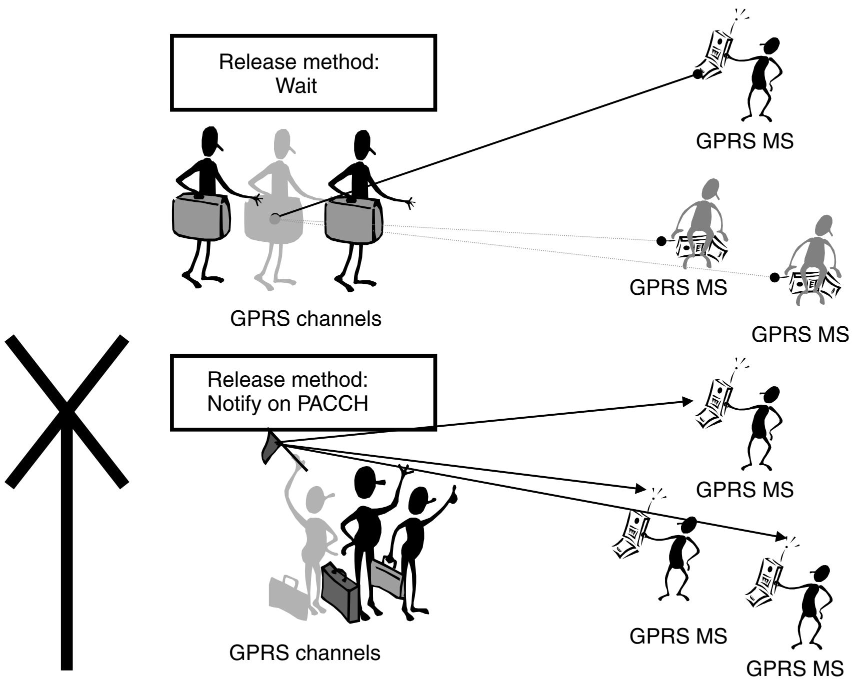


Figure 2.23 Network overview of different release methods

associate control channel (PACCH) individually to each affected MS. This method is very reliable, but slow and resource inefficient.

The 'quick and easy' approach is to broadcast the packet data channel release on all the PDCHs on the same carrier as the PDCH to be released. All MSs monitor the possible occurrences of PACCH on one channel and should capture such notification. In practice, a combination of all the methods can be used.

The situation may occur in which an MS remains unaware of the released PDCH. In such a case, the MS may cause some interference when wrongly assuming that the decoded uplink state flag denotes that it has been allocated the use of the following uplink block period. This 3-bit uplink state flag (USF) is the means by which up to eight mobile stations are able to share an uplink time slot (if the PDCH contains a PCCCH, one USF is reserved for the packet random access channel (PRACH), thus only seven mobile stations can be addressed and share the resource). Each MS will wait, until its personal USF is sent from the BTS and then react accordingly, but after not getting proper response from the network, the MS itself would release the RLC connection.

2.4.4.2 Radio Resource RR Operation Modes

Radio resource (RR) management procedures are characterized by two different RR operating modes. Each mode describes a certain amount of allocated functionality and information. RR procedures and RR operating modes are specified in GSM 04.07.

Packet Idle Mode

In packet idle mode, no temporary block flow exists. The temporary block flow (TBF) is a temporary physical connection used to support the unidirectional transfer of packet data units. Upper layers can require the transfer of a LLC PDU which, implicitly, may trigger the establishment of TBF and thus the transition to packet transfer mode.

In packet idle mode, the MS monitors the packet broadcast control channel (PBCCH) and the paging subchannel for the paging group the MS belongs to (*Note*: in GPRS it is possible to implement more than one paging channel so a mobile can be instructed to monitor a certain group of paging channels). As mentioned earlier, if PCCCH is not present in the cell then the mobile station monitors the BCCH and the relevant paging subchannels.

Packet Transfer Mode

In packet transfer mode, the mobile station is allocated radio resource providing a temporary block flow on one or more physical channels. Continuous transfer of one or more LLC PDUs is possible. Concurrent TBFs may (but do not have to) be established in opposite directions (as defined above, TBFs are unidirectional and do not need a TBF in the opposite direction). When selecting a new cell, the mobile station leaves the packet transfer mode, enters the packet idle mode where it switches to the new cell, reads the system information and may then return to packet transfer mode in the new cell.

While operating in packet transfer mode, a mobile station belonging to GPRS MS Class A may simultaneously enter the different RR service modes defined in GSM 04.08. A mobile station belonging to either GPRS MS Class B or C leaves both packet idle mode and packet transfer mode before entering dedicated mode for a CS call.

Dual Transfer Mode

The packet transfer mode is not applicable to MSs supporting dual transfer mode (DTM), which is defined since release 99. This mode is a subset of the terminal equipment Class A mode of operation (see Chapter 7 for a more detailed description of equipment classes) and is optional for GPRS or E-GPRS mobile stations as well as for the network. In this mode, the MS has an ongoing RR connection and is allocated a temporary block flow on one or more physical channels. Continuous transfer of one ore more LLC PDUs is possible. Concurrent TBFs may be established in opposite directions.

While in dual transfer mode the MS performs all the tasks of dedicated mode. In addition, upper layers can require:

- The release of all packet resources, which triggers the transition to dedicated CS mode,
- The release of the RR resources, which triggers the transition to idle mode and packet idle mode.

When handing over to a new cell, the MS leaves the dual transfer mode, enters the dedicated mode where it switches to the new cell, may read the system information messages sent on the SACCH, and may then enter dual transfer mode in the new cell.

It is clear from the above that there is a difference between a 'normal' Class A device and a Class A device that supports dual transfer mode. The definition of a Class A device assumes a total independence between the circuit switched and the packet switched side. As a result, such a Class A device has to be able to operate on two different frequencies at the same time in case the CS radio resources are allocated on a different carrier to the packet resources. The development of such a device is very complicated. The dual transfer mode offers a solution for this problem by allowing the transfer of packet switched data and signalling via a timeslot set up for a CS connection or other time slots on the same frequency.

Transition between RR Operation Modes

The RR operation modes, and therefore the transition between them, are dependent on the current operation mode as well as the terminal equipment operation mode (Class A, B or C). Class C mobile stations can only be attached to either GSM or GPRS. So they can only change from idle to dedicated mode and back when GSM attached (by establishing or releasing an RR state), or from packet idle to packet transfer mode and back (by temporary block flow establishment or release) when GPRS attached. The user has to change manually between GSM and GPRS attach.

For a Class A mobile station that does not support dual transfer mode, there are four possible states. The MS can be regarded as a combination of two devices, each with two possible RR states:

- On the circuit switched part of the device: idle mode and dedicated mode;
- On the GPRS part of the device: packet idle mode and packet transfer mode.

This makes a Class A MS much more versatile as it is able to swap back and forth between combinations of the operation modes, which are available only separately for the Class C MS.

For a Class B MS or a Class A mobile station that supports dual transfer mode, there are three possible RR modes:

- packet idle mode,
- packet transfer mode,
- dedicated mode.

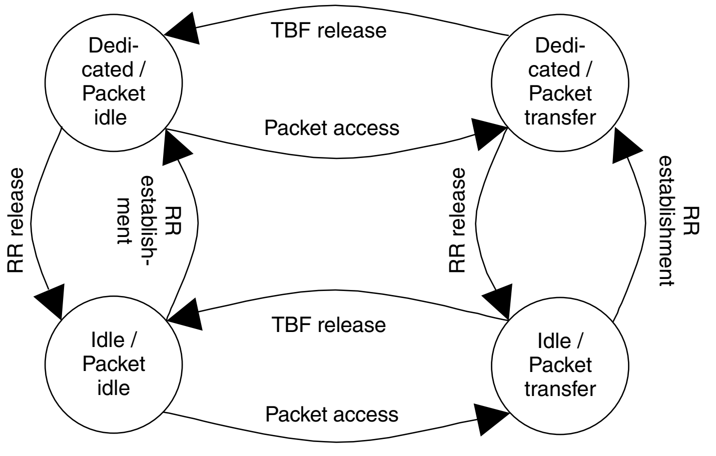


Figure 2.24 RROM and their Transitions for Class A (non DTM)

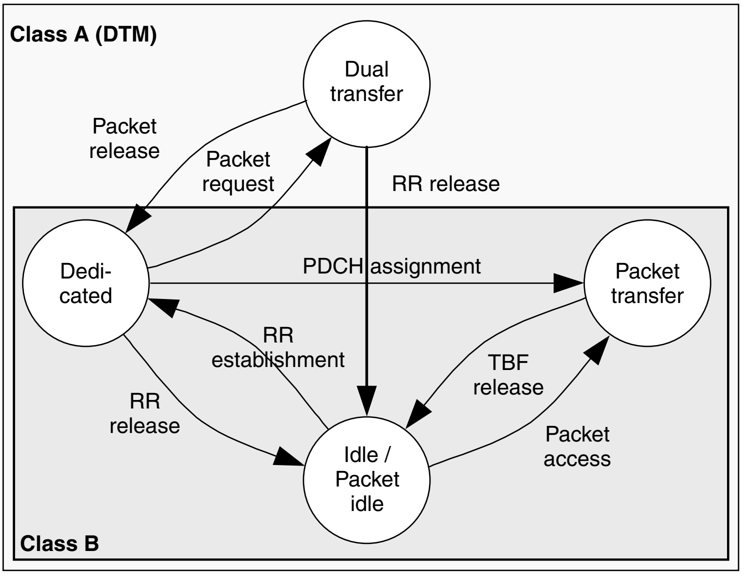


Figure 2.25 RROM and their transitions for Class B and Class A (DTM)

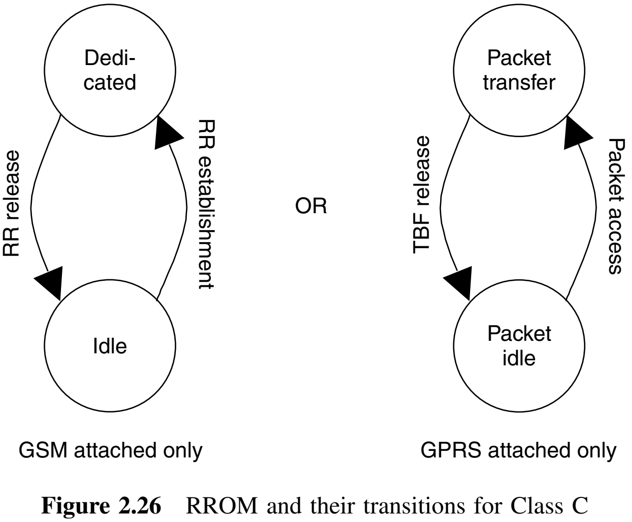


Figure 2.26 RROM and their transitions for Class C

The Class B MS can change between these three like the Class C MS, and in addition can also change from dedicated mode to packet transfer mode when assigned a PDCH. This can only be done by performing a packet request procedure while in dedicated mode. The Class A (DTM) MS, in addition to having the same capabilities as the Class B MS, can also enter the dual transfer mode.

The transitions described above are illustrated in Figures 2.24, 2.25 and 2.26.

Radio resource BSS	Packet transfer mode	Measurement report reception	No state	No state
Radio resource MS	Packet transfer mode	Packet idle mode		Packet idle mode
Mobility management NSS and MS	Ready			Standby

Figure 2.27 Correspondence between RR operating modes and MM states

2.4.4.3 Correspondence between RR Operation Modes and MM States

The mobility management states are defined in GSM 03.60. Figure 2.27 illustrates the correspondence between radio resource states and mobility management states: Each state is protected by a timer. The timers run in the MS and the network. Packet transfer mode is protected by RLC protocol timers.