Overview of GSM, GPRS, and UMTS

The Cisco Mobile Exchange (CMX) architecture provides mobile wireless solutions for operators using General Packet Radio Service (GPRS) and Universal Mobile Telecommunication System (UMTS) access technologies. This chapter provides an overview of these technologies and their roles in the evolution from second-generation (2G) to third-generation (3G) mobile wireless networks.

- Global System for Mobile Communications (GSM)—A digital, mobile, radio standard developed for mobile, wireless, voice communications
- GPRS—An extension of GSM networks that provides mobile, wireless, data communications
- UMTS—An extension of GPRS networks that moves toward an all-IP network by delivering broadband information, including commerce and entertainment services, to mobile users via fixed, wireless, and satellite networks

This chapter includes the following sections:

Global Systems for Mobile Communications

In the early 1980s, many countries in Europe witnessed a rapid expansion of analog cellular telephone systems. However, each country developed its own system, and interoperability across borders became a limiting factor.

In 1982, the Conference of European Post and Telecommunications (CEPT), an association of telephone and telegraph operators in Europe, established a working group to develop a new public land mobile system to span the continent. Because their working language was French, the group was called the Groupe Speciale Mobile (GSM).

The GSM group proposed the following criteria for the new mobile wireless system:

- good speech quality
- low cost for terminals and service
- international roaming
- handheld terminals
- support for introduction new services
- spectral efficiency
- compatibility with Integrated Digital Services Network (ISDN)

In 1989, the responsibility for GSM development was transferred to the European Telecommunications Standards Institute (ETSI), and phase 1 of the GSM specification was published in 1990. The first commercial service was launched in 1991.

When the official language of the GSM group changed from French to English, GSM was changed from Groupe Speciale Mobile to Global System for Mobile Communications.

In 1994, phase 2 data/fax services were launched, and in 1995, the GSM phase 2 standard was completed. The first GSM services in the United States were launched.

GSM uses a combination of both the time division multiple access (TDMA) and frequency division multiple access (FDMA) technologies. With this combination, more channels of communications are available, and all channels are digital.

The GSM service is available in four frequency bands:

- 450-MHz—Upgrade of older analog cellular systems in Scandinavia
- 900-MHz—Original band used everywhere except North America and most of South America
- 1800-MHz—New band to increase capacity and competition used everywhere except North America and most of South America
- 1900-MHz—Personal communications service band used in North America and much of South America

The higher frequency bands provide additional capacity and higher subscriber densities.

One of the unique benefits of GSM service is its capability for international roaming because of the roaming agreements established between the various GSM operators worldwide.

GSM Technology Differentiator

One of the advantages of GSM is that it offers a subscriber identity module (SIM), also known as a smart card. The smart card contains a computer chip and some non-volatile memory and is inserted into a slot in the base of the mobile handset.

The memory on the smart card holds information about the subscriber that enables a wireless network to provide subscriber services. The information includes:

- The subscriber’s identity number
- The telephone number
- The original network to which the subscriber is subscribed

A smart card can be moved from one handset to another. A handset reads the information off the smart card and transmits it to the network.

GSM Network Elements

A GSM network consists of the following network components:

- Mobile station (MS)
- Base transceiver station (BTS)
- Base station controller (BSC)
- Base station subsystem (BSS)
- Mobile switching center (MSC)
- Authentication center (AuC)
- Home location register (HLR)
- Visitor location register (VLR)

Mobile Station

The mobile station (MS) is the starting point of a mobile wireless network. The MS can contain the following components:

- Mobile terminal (MT)—GSM cellular handset
- Terminal equipment (TE)—PC or personal digital assistant (PDA)

The MS can be two interconnected physical devices (MT and TE) with a point-to-point interface or a single device with both functions integrated.

Base Transceiver Station

When a subscriber uses the MS to make a call in the network, the MS transmits the call request to the base transceiver station (BTS). The BTS includes all the radio equipment (i.e., antennas, signal processing devices, and amplifiers) necessary for radio transmission within a geographical area called a cell. The BTS is responsible for establishing the link to the MS and for modulating and demodulating radio signals between the MS and the BTS.

Base Station Controller

The base station controller (BSC) is the controlling component of the radio network, and it manages the BTSs. The BSC reserves radio frequencies for communications and handles the handoff between BTSs when an MS roams from one cell to another. The BSC is responsible for paging the MS for incoming calls.

Base Station Subsystem

A GSM network is comprised of many base station subsystems (BSSs), each controlled by a BSC. The BSS performs the necessary functions for monitoring radio connections to the MS, coding and decoding voice, and rate adaptation to and from the wireless network. A BSS can contain several BTSs.

Mobile Switching Center

The mobile switching center (MSC) is a digital ISDN switch that sets up connections to other MSCs and to the BSCs. The MSCs form the wired (fixed) backbone of a GSM network and can switch calls to the public switched telecommunications network (PSTN). An MSC can connect to a large number of BSCs.

Equipment Identity Register

The equipment identity register (EIR) is a database that stores the international mobile equipment identities (IMEIs) of all the mobile stations in the network. The IMEI is an equipment identifier assigned by the manufacturer of the mobile station. The EIR provides security features such as blocking calls from handsets that have been stolen.

Home Location Register

The home location register (HLR) is the central database for all users to register to the GSM network. It stores static information about the subscribers such as the international mobile subscriber identity (IMSI), subscribed services, and a key for authenticating the subscriber. The HLR also stores dynamic subscriber information (i.e., the current location of the mobile subscriber).

Authentication Center

Associated with the HLR is the authentication center (AuC); this database contains the algorithms for authenticating subscribers and the necessary keys for encryption to safeguard the user input for authentication.

Visitor Location Register

The visitor location register (VLR) is a distributed database that temporarily stores information about the mobile stations that are active in the geographic area for which the VLR is responsible. A VLR is associated with each MSC in the network. When a new subscriber roams into a location area, the VLR is responsible for copying subscriber information from the HLR to its local database. This relationship between the VLR and HLR avoids frequent HLR database updates and long distance signaling of the user information, allowing faster access to subscriber information.

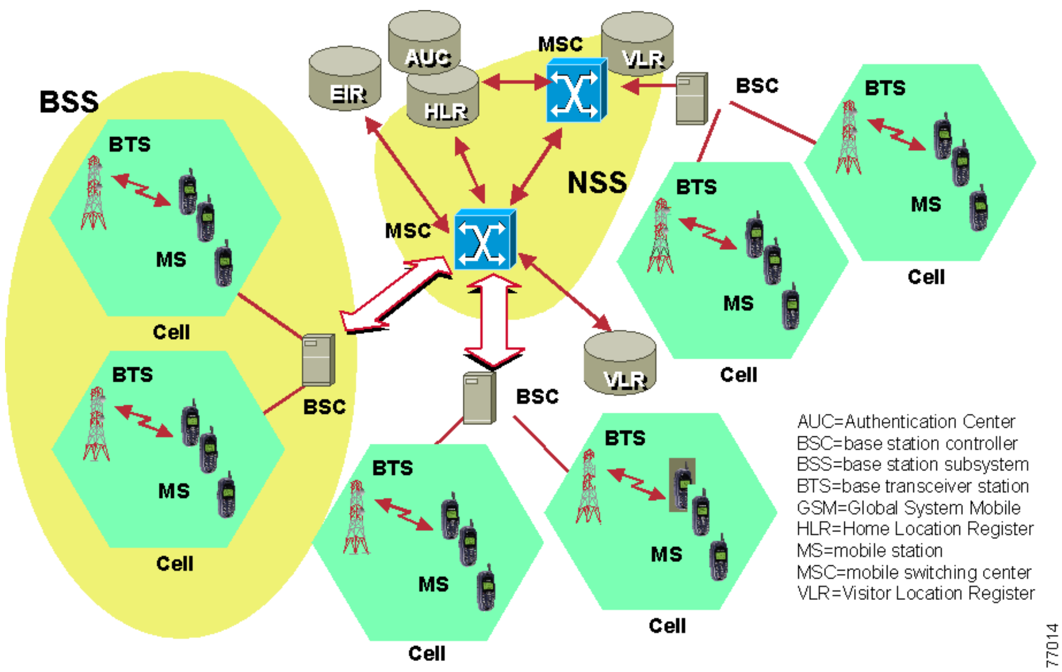
The HLR, VLR, and AuC comprise the management databases that support roaming (including international roaming) in the GSM network. These databases authenticate calls while GSM subscribers roam between the private network and the public land mobile network (PLMN). The types of information they store include subscriber identities, current location area, and subscription levels.

Network and Switching Subsystem

The network and switching subsystem (NSS) is the heart of the GSM system. It connects the wireless network to the standard wired network. It is responsible for the handoff of calls from one BSS to another and performs services such as charging, accounting, and roaming.

Figure 2-1 shows a GSM network and the network elements it contains.

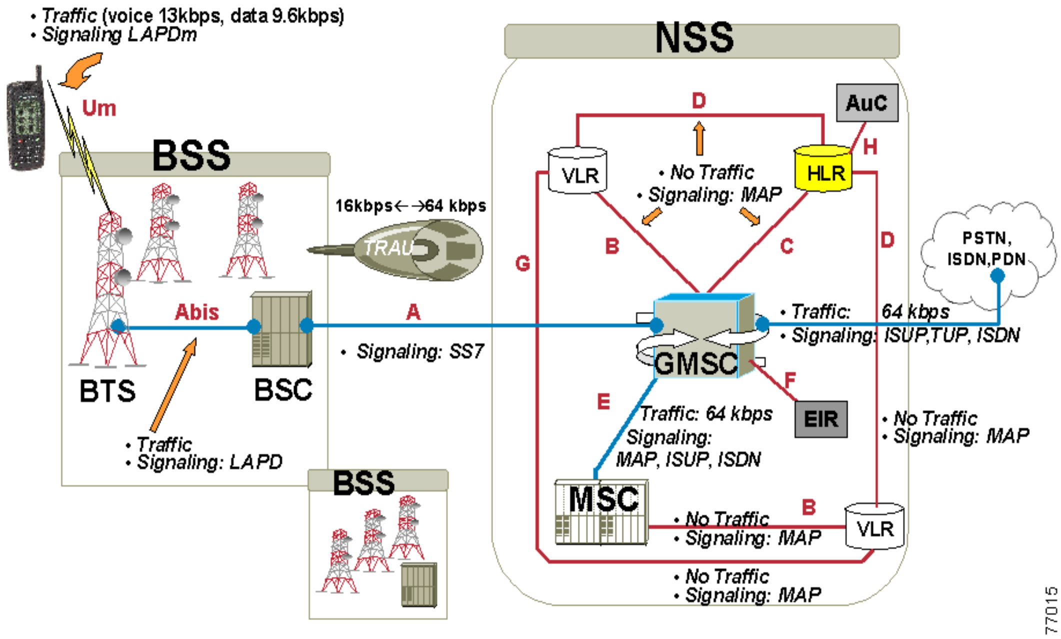
Figure 2-1 GSM Network Elements



GSM Interfaces

The GSM uses various interfaces for communication among its network elements. Figure 2-2 shows these interfaces.

Figure 2-2 GSM Interfaces



Mobile wireless communication occurs over the interfaces between the network elements in a sequential manner. In Figure 2-2, the MS transmits to the BTS, the BTS to the BSC, and the BSC to the MSC. Communications also occur over the interfaces to the management databases (HLR, VLR, AuC, and EIR). Communications might traverse multiple MSCs but ultimately must reach the gateway MSC (GMSC). The GMSC provides the gateway to the public switched telephone network (PSTN). A separate interface exists between each pair of elements, and each interface requires its own set of protocols.

In the BSS block, mobile communication occurs over the air interface to the BTS using the ISDN Link Access Procedure-D mobile (LAP-Dm). This traffic channel carries speech and data. In this example, voice operates at full-rate 13 kbps (supported by LAP-Dm), and data operates at full-rate 9.6 kbps.The BTS communicates to the BSC over the *Abis* interface using the ISDN LAP-D signaling protocol. The BSC communicates to the GMSC via the transcoder rate adapter unit (TRAU), which translates between 16 kbps on the BTS side to 64 kbps on the GMSC side. This interface uses the signaling system 7 (SS7) protocol, which defines call set-up and call services across the interface.

At the NSS, the GMSC is the central node. Link-level traffic and signaling control occurs over the interface between the GMSC and MSC and the interface to the external network (PSTN, ISDN or PDN). Different signaling protocols are used on the interfaces. Some NSS interfaces involve only control signaling protocols with no traffic. For example, no traffic is generated on the interfaces between the GMSC, HLR, and VLR. Instead, these interfaces carry only signaling using the Mobile Application Part (MAP) of the SS7 protocol. The MAP is specified in IS-41 and defines the application layer, signaling protocols, and procedures for registering mobile users and handling handoffs between cellular systems. The GMSC establishes call traffic (at 64 kbps) onto the PSTN via the ISDN user part (ISUP), which is an SS7-based protocol. The GMSC and MSC exchange traffic (over LAP-D at 64 kbps) and use SS7 (MAP and ISUP) control.

GSM Data Services

GSM networks handle both voice and data traffic requirements of the mobile communication by providing two modes of operation:

- Circuit switched (high-speed circuit switched data)
- Packet switched (GPRS)

Circuit switching provides the customer with a dedicated channel all the way to the destination. The customer has exclusive use of the circuit for the duration of the call, and is charged for the duration of the call.

With packet switching, the operator assigns one or more dedicated channels specifically for shared use. These channels are up and running 24 hours a day, and when you need to transfer data, you access a channel and transmit your data. Packet switching is more efficient than circuit switching.

The standard data rate of a GSM channel is 22.8 kbps.

General Packet Radio Service

The general packet radio system (GPRS) provides packet radio access for mobile Global System for Mobile Communications (GSM) and time-division multiple access (TDMA) users. In addition to providing new services for today's mobile user, GPRS is important as a migration step toward third-generation (3G) networks. GPRS allows network operators to implement an IP-based core architecture for data applications, which will continue to be used and expanded for 3G services for integrated voice and data applications. The GPRS specifications are written by the European Telecommunications Standard Institute (ETSI), the European counterpart of the American National Standard Institute (ANSI).

GPRS is the first step toward an end-to-end wireless infrastructure and has the following goals:

- Open architecture
- Consistent IP services
- Same infrastructure for different air interfaces
- Integrated telephony and Internet infrastructure
- Leverage industry investment in IP
- Service innovation independent of infrastructure

Benefits of GPRS

The GPRS provides the following benefits:

- Overlays on the existing GSM network to provide high-speed data service
- Always on, reducing the time spent setting up and taking down connections
- Designed to support bursty applications such as e-mail, traffic telematics, telemetry, broadcast services, and web browsing that do not require detected connection.

By implementing Cisco GPRS products and related solutions, mobile service providers can optimize their networks to deploy high quality mobile voice and data services. They can also benefit from new operating efficiencies, peer-to-peer IP-based architecture for scalability, and IP standard interfaces to billing and customer support.

GPRS Applications

GPRS enables a variety of new and unique services to the mobile wireless subscriber. These mobile services have unique characteristics that provide enhanced value to customers. These characteristics include the following:

- Mobility—The ability to maintain constant voice and data communications while on the move
- Immediacy—Allows subscribers to obtain connectivity when needed, regardless of location and without a lengthy login session
- Localization—Allows subscribers to obtain information relevant to their current location

The combination of these characteristics provides a wide spectrum of possible applications that can be offered to mobile subscribers. The core network components offered by Cisco enable seamless access to these applications, whether they reside in the service provider's network or the public Internet.

In general, applications can be separated into two high-level categories: corporate and consumer. These include:

- Communications—E-mail; fax; unified messaging; intranet/Internet access
- Value-added services—Information services; games
- E-commerce—Retail; ticket purchasing; banking; financial trading
- Location-based applications—Navigation; traffic conditions; airline/rail schedules; location finder
- Vertical applications—Freight delivery; fleet management; sales-force automation
- Advertising

Communications

Communications applications include those in which it appears to users that they are using the mobile communications network as a pipeline to access messages or information. This differs from those applications in which users believe that they are accessing a service provided or forwarded by the network operator.

Intranet Access

The first stage of enabling users to maintain contact with their offices is through access to e-mail, fax, and voice mail using unified messaging systems. Increasingly, files and data on corporate networks are becoming accessible through corporate intranets. These intranets can be protected through firewalls by enabling secure tunnels or virtual private networks (VPNs).

Internet Access

As a critical mass of users is approached, more and more applications aimed at general consumers are being placed on the Internet. The Internet is becoming an effective tool for accessing corporate data and manipulating product and service information. More recently, companies are using the Internet as an environment for conducting business through e-commerce.

Email and Fax

E-mail on mobile networks may take one of two forms. E-mail can be sent to a mobile user directly or the user can have an e-mail account maintained by the network operator or their Internet service provider (ISP). In the latter case, a notification is forwarded to the mobile terminal and includes the first few lines of the e-mail, details of the sender, the date and time, and the subject. Fax attachments can also accompany e-mails.

Unified Messaging

Unified messaging provides a single mailbox for all messages, including voice mail, faxes, e-mail, short message service (SMS), and pager messages. Unified messaging systems allow for a variety of access methods to recover messages of different types. Some use text-to-voice systems to read e-mail or send faxes over a normal phone line. Most allow the user to query the contents of the various mailboxes through data access such as the Internet. Others can be configured to alert the user on the device of their choice when messages are received.

Value Added Services

Value-added services refer to the content provided by network operators to increase the value of services to their subscribers. Two terms that are frequently used to describe delivery of data applications are *push* and *pull*, as defined below.

- *Push* describes the transmission of data at a predetermined time or under predetermined conditions. It also refers to the unsolicited supply of advertising (for example, delivery of news as it occurs or stock values when they fall below a preset value).
- *Pull* describes the request for data in real time by the user (for example, checking stock quotes or daily news headlines).

To be valuable to subscribers, this content must possess several characteristics:

- Personalized information that is tailored to the user (for example, a stock ticker that focusses on key quotes and news or an e-commerce application that knows a user's profile)
- Localized content that is based on a user's current location and includes maps, hotel finders, or restaurant reviews
- Menu screens that are intuitive and easy to navigate
- Security for e-commerce sites for the exchange of financial or other personal information

Several value-added services are outlined in the following sections.

E-commerce

E-commerce is defined as business conducted on the Internet or data service. This includes applications in which a contract is established for the purchase of goods and services and online banking applications. These applications require user authentication and secure transmission of sensitive data over the data connection.

Banking

The banking industry is interested in promoting electronic banking because electronic transactions are less costly to conduct than personal transactions in a bank. Specific banking functions that can be accomplished over a wireless connection include balance checking, money transfers between accounts, bill payment, and overdraft alert.

Financial Trading

The immediacy of transactions over the Internet and the requirement for up-to-the-minute information has made the purchasing of stocks online a popular application. By coupling push services with the ability to make secure transactions from the mobile terminal, a service that is unique to the mobile environment can be provided.

Location-Based Services and Telematics

Location-based services provide the ability to link push or pull information services with a user's location. Examples include hotel and restaurant finders, roadside assistance, and city-specific news and information. This technology also has vertical applications. These allow, for example, tracking vehicles in a fleet or managing the operations of a large workforce.

Vertical Applications

In the mobile environment, vertical applications apply to systems using mobile architectures to support the specific tasks within a company. Examples of vertical applications include:

- Sales support—Configuring stock and product information for sales staff, integrating appointment details, and placing orders remotely
- Dispatching—Communicating job details such as location and scheduling and permitting information queries to support the job
- Fleet management—Controlling a fleet of delivery or service staff and vehicle, monitoring their locations, and scheduling their work
- Parcel delivery—Tracking the locations of packages for customers and monitoring the performance of the delivery system

Advertising

Advertising services are offered as a push information service. Advertising may be offered to customers to subsidize the cost of voice or other information services. Advertising may be location sensitive. For example, a user entering a mall can receive advertisements specific to the stores in that mall.

GPRS Architecture

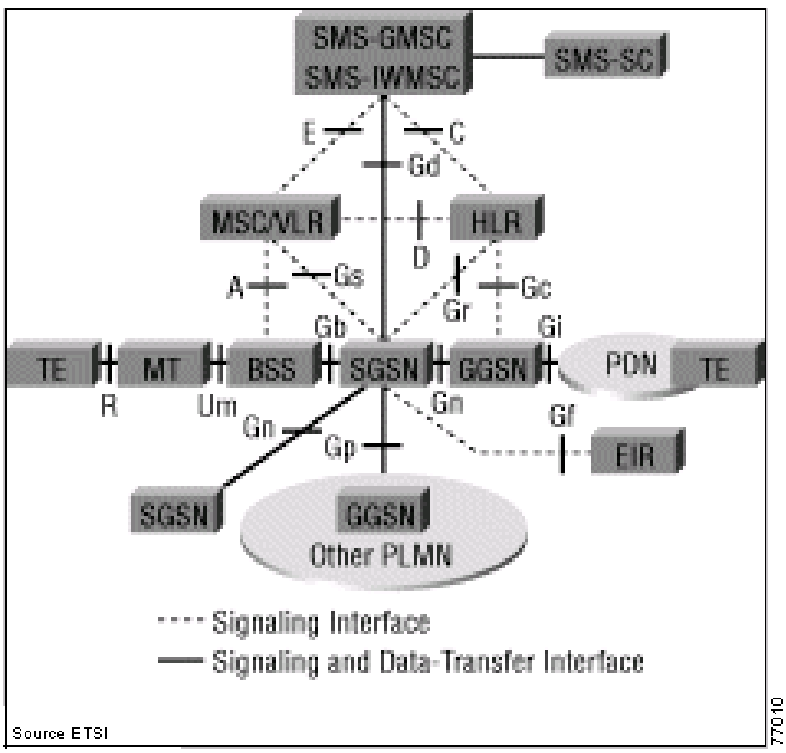
GPRS is a data network that overlays a second-generation GSM network. This data overlay network provides packet data transport at rates from 9.6 to 171 kbps. Additionally, multiple users can share the same air-interface resources simultaneously.

GPRS attempts to reuse the existing GSM network elements as much as possible, but to effectively build a packet-based mobile cellular network, some new network elements, interfaces, and protocols for handling packet traffic are required. Therefore, GPRS requires modifications to numerous network elements as summarized in Table 2-1 and shown in Figure 2-3.

Table 2-1 GPRS Network Elements

GSM Network Element	Modification or Upgrade Required for GPRS.
Terminal Equipment (TE)	New terminal equipment is required to access GPRS services. These new terminals will be backward compatible with GSM for voice calls.
BTS	A software upgrade is required in the existing base transceiver site.
BSC	The base station controller (BSC) requires a software upgrade and the installation of new hardware called the packet control unit (PCU). The PCU directs the data traffic to the GPRS network and can be a separate hardware element associated with the BSC.
GPRS Support Nodes (GSNs)	The deployment of GPRS requires the installation of new core network elements called the serving GPRS support node (SGSN) and gateway GPRS support node (GGSN).
Databases (HLR, VLR, etc.)	All the databases involved in the network will require software upgrades to handle the new call models and functions introduced by GPRS.

Figure 2-3 GPRS Reference Architecture



GPRS Subscriber Terminals

New terminals are required because existing GSM phones do not handle the enhanced air interface or packet data. A variety of terminals can exist, including a high-speed version of current phones to support high-speed data access, a new PDA device with an embedded GSM phone, and PC cards for laptop computers. These terminals are backward compatible for making voice calls using GSM.

GPRS Base Station Subsystem

Each BSC requires the installation of one or more PCUs and a software upgrade. The PCU provides a physical and logical data interface to the base station subsystem (BSS) for packet data traffic. The BTS can also require a software upgrade but typically does not require hardware enhancements.

When either voice or data traffic is originated at the subscriber terminal, it is transported over the air interface to the BTS, and from the BTS to the BSC in the same way as a standard GSM call. However, at the output of the BSC, the traffic is separated; voice is sent to the mobile switching center (MSC) per standard GSM, and data is sent to a new device called the SGSN via the PCU over a Frame Relay interface.

GPRS Support Nodes

In the core network, the existing MSCs are based on circuit-switched central-office technology and cannot handle packet traffic. Two new components, called GPRS support nodes (GSNs), are added:

- Serving GPRS support node (SGSN)
- Gateway GPRS support node (GGSN)

Serving GPRS Support Node

The SGSN delivers packets to mobile stations (MSs) within its service area. SGSNs send queries to home location registers (HLRs) to obtain profile data of GPRS subscribers. SGSNs detect new GPRS MSs in a given service area, process registration of new mobile subscribers, and keep records of their locations inside a predefined area. The SGSN performs mobility management functions such as handing off a roaming subscriber from the equipment in one cell to the equipment in another. The SGSN is connected to the base station subsystem through a Frame Relay connection to the PCU in the BSC.

Gateway GPRS Support Node

GGSNs are used as interfaces to external IP networks such as the public Internet, other mobile service providers' GPRS services, or enterprise intranets. GGSNs maintain routing information that is necessary to tunnel the protocol data units (PDUs) to the SGSNs that service particular MSs. Other functions include network and subscriber screening and address mapping. One or more GGSNs can be provided to support multiple SGSNs. More detailed descriptions of the SGSN and GGSN are provided in a later section.

GPRS Terminals

The term *terminal equipment* is generally used to refer to the variety of mobile phones and mobile stations that can be used in a GPRS environment. The equipment is defined by terminal classes and types. Cisco's gateway GPRS serving node (GGSN) and data network components interoperate with GPRS terminals that meet the GPRS standards.

Three classes of GPRS terminals are provided: Class A, Class B, or Class C.

Class A Terminals

Class A terminals support GPRS and other GSM services (such as SMS and voice) simultaneously. This support includes simultaneous attach, activation, monitor, and traffic. Class A terminals can make or receive calls on two services simultaneously. In the presence of circuit-switched services, GPRS virtual circuits are held (i.e., placed on hold) instead of being cleared.

Class B Terminals

Class B terminals can monitor GSM and GPRS channels simultaneously but can support only one of these services at a time. Therefore, a Class B terminal can support simultaneous attach, activation, and monitor, but not simultaneous traffic. As with Class A, the GPRS virtual circuits are not disconnected when circuit-switched traffic is present. Instead, they are switched to busy mode. Users can make or receive calls on either a packet or a switched call type sequentially, but not simultaneously.

Class C Terminals

Class C terminals support only sequential attach. The user must select which service to connect to. Therefore, a Class C terminal can make or receive calls from only the manually selected (or default) service. The service that is not selected is unreachable. The GPRS specifications state that support of SMS is optional for Class C terminals.

GPRS Device Types

In addition to the three terminal classes, each handset has a unique form (housing design). Some of the forms are similar to current mobile wireless devices, while others will evolve to use the enhanced data capabilities of GPRS.

The earliest available type is closely related to the current mobile phone. These are available in the standard form with a numeric keypad and a relatively small display.

PC cards are credit card-sized hardware devices that connect through a serial cable to the bottom of a mobile phone. Data cards for GPRS phones enable laptops and other devices with PC card slots to be connected to mobile GPRS-capable phones. Card phones provide functions similar to those offered by PC cards without requiring a separate phone. These devices may require an ear piece and microphone to support voice services.

Smart phones are mobile phones with built-in voice, nonvoice, and Web-browsing services. Smart phones integrate mobile computing and mobile communications into a single terminal. They come in various form factors, which may include a keyboard or an icon drive screen.

The increase in machine-to-machine communications has led to the adoption of application-specific devices. These *black-box* devices lack a display, keypad, and voice accessories of a standard phone. Communication is accomplished through a serial cable. Applications such as meter reading utilize such black-box devices.

Personal digital assistants (PDAs), such as the Palm Pilot series or Handspring Visor, and handheld communications devices are data-centric devices that are adding mobile wireless access. These devices can either connect with a GPRS-capable mobile phone via a serial cable or integrate GPRS capability. Access can be gained via a PC card or a serial cable to a GPRS-capable phone.

Data Routing

One of the main requirements in the GPRS network is the routing of data packets to and from a mobile user. The requirement can be divided into two areas: data packet routing and mobility management.

Data Packet Routing

The main functions of the GGSN involve interaction with the external data network. The GGSN updates the location directory using routing information supplied by the SGSNs about the location of an MS. It routes the external data network protocol packet encapsulated over the GPRS backbone to the SGSN currently serving the MS. It also decapsulates and forwards external data network packets to the appropriate data network and collects charging data that is forwarded to a charging gateway (CG).

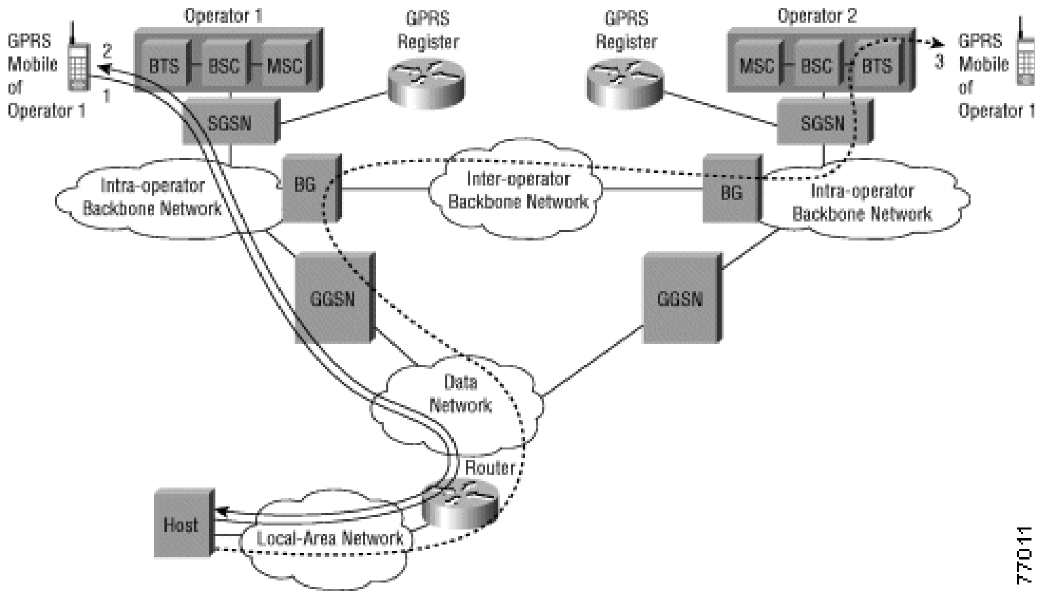
In Figure 2-4, three routing schemes are illustrated:

- Mobile-originated message (path 1)—This path begins at the GPRS mobile and ends at the Host
- Network-initiated message when the MS is in its home network (path 2)—This path begins at the Host and ends at the GPRS mobile
- Network-initiated message when the MS roams to another GPRS network (path 3)—This path is indicated by the dotted line

In these examples, the operator's GPRS network consists of multiple GSNs (with a gateway and serving functionality) and an intra-operator backbone network.

GPRS operators allow roaming through an inter-operator backbone network. The GPRS operators connect to the inter-operator network through a border gateway (BG), which can provide the necessary interworking and routing protocols (for example, border gateway protocol [BGP]). In the future, GPRS operators might implement quality of service (QoS) mechanisms over the inter-operator network to ensure service-level agreements (SLAs). The main benefits of the architecture are its flexibility, scalability, interoperability, and roaming attributes.

Figure 2-4 Routing of Data Packets between a Fixed Host and a GPRS MS



The GPRS network encapsulates all data network protocols into its own encapsulation protocol called the GPRS tunneling protocol (GTP). The GTP ensures security in the backbone network and simplifies the routing mechanism and the delivery of data over the GPRS network.

Mobility Management

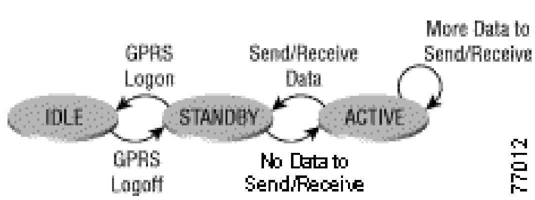
The operation of the GPRS is partly independent of the GSM network. However, some procedures share the network elements with current GSM functions to increase efficiency and to make optimum use of free GSM resources (such as unallocated time slots).

An MS has three states in the GPRS system (Figure 2-5):

- Active
- Standby
- Idle

The three-state model is unique to packet radio; GSM uses a two-state model (idle or active).

Figure 2-5 GPRS States in a Mobile Station



Active State

Data is transmitted between an MS and the GPRS network only when the MS is in the active state. In the active state, the SGSN knows the cell location of the MS.

Packet transmission to an active MS is initiated by packet paging to notify the MS of an incoming data packet. The data transmission proceeds immediately after packet paging through the channel indicated by the paging message. The purpose of the paging message is to simplify the process of receiving packets. The MS listens to only the paging messages instead of to all the data packets in the downlink channels. This reduces battery usage significantly.

When an MS has a packet to transmit, it must access the uplink channel (i.e., the channel to the packet data network where services reside). The uplink channel is shared by a number of MSs, and its use is allocated by a BSS. The MS requests use of the channel in a random access message. The BSS allocates an unused channel to the MS and sends an access grant message in reply to the random access message. The description of the channel (one or multiple time slots) is included in the access grant message. The data is transmitted on the reserved channels.

Standby State

In the standby state, only the routing area of the MS is known. (The routing area can consist of one or more cells within a GSM location area).

When the SGSN sends a packet to an MS that is in the standby state, the MS must be paged. Because the SGSN knows the routing area of the MS, a packet paging message is sent to the routing area. On receiving the packet paging message, the MS relays its cell location to the SGSN to establish the active state.

The main reason for the standby state is to reduce the load in the GPRS network caused by cell-based routing update messages and to conserve the MS battery. When an MS is in the standby state, the SGSN is informed of only routing area changes. By defining the size of the routing area, the operator can control the number of routing update messages.

Idle State

In the idle state, the MS does not have a logical GPRS context activated or any packet-switched public data network (PSPDN) addresses allocated. In this state, the MS can receive only those multicast messages that can be received by any GPRS MS. Because the GPRS network infrastructure does not know the location of the MS, it is not possible to send messages to the MS from external data networks.

Routing Updates

When an MS that is in an active or a standby state moves from one routing area to another within the service area of one SGSN, it must perform a routing update. The routing area information in the SGSN is updated, and the success of the procedure is indicated in the response message.

A cell-based routing update procedure is invoked when an active MS enters a new cell. The MS sends a short message containing the identity of the MS and its new location through GPRS channels to its current SGSN. This procedure is used only when the MS is in the active state.

The inter-SGSN routing update is the most complicated routing update. The MS changes from one SGSN area to another, and it must establish a new connection to a new SGSN. This means creating a new logical link context between the MS and the new SGSN and informing the GGSN about the new location of the MS.

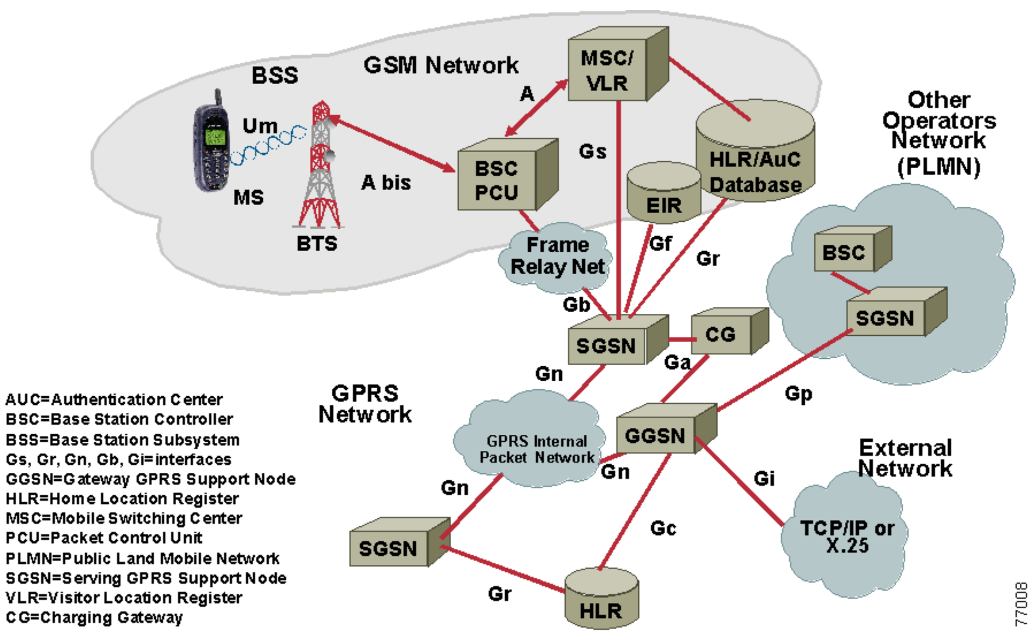
GPRS Interfaces

The GPRS architecture consists of signaling interfaces with various protocols that control and support the transmission of packets across the networks and to the mobile stations. The interfaces in a GPRS network are:

- Ga—Interface between GSN nodes (GGSN, SGSN) and charging gateway (CG)
- Gb—Interface between SGSN and BSS (PCU); normally uses Frame Relay
- Gc—Interface between GGSN and HLR
- Gi—Interface between GPRS (GGSN) and an external packet data network (PDN)
- Gn—Interface between two GSN nodes, i.e., GGSN and SGSN; this connects into the intra-network backbone, for example, an Ethernet network
- Gp—Interface between two GSN nodes in different PLMNs; this is via border gateways and is an inter-PLMN network backbone
- Gr—Interface between SGSN and HLR
- Gs—Interface between SGSN and the MSC/VLR
- Gf—Interface between SGSN and EIR

Figure 2-6 shows these interfaces.

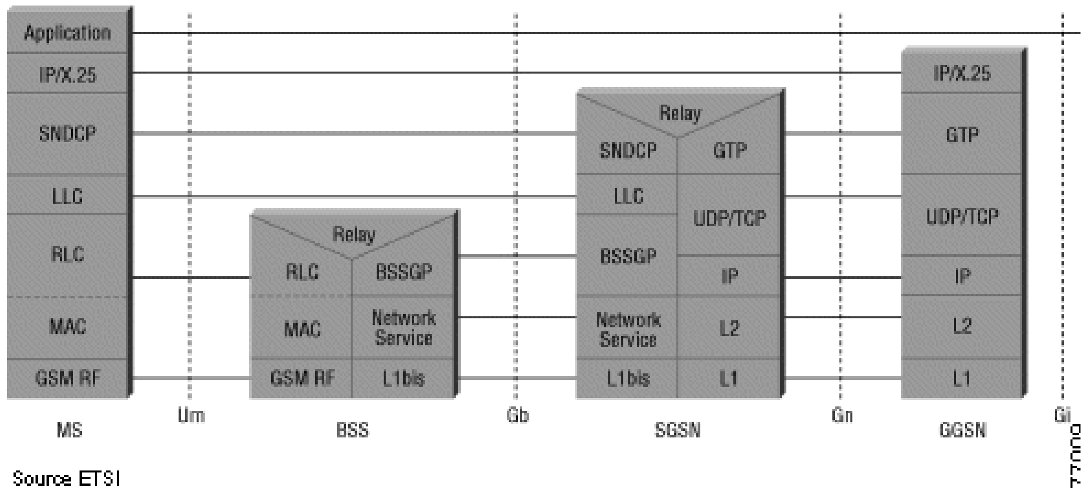
Figure 2-6 GPRS Interfaces



GPRS Protocol Stacks

Figure 2-7 shows the GPRS protocol stack and end-to-end message flows from the MS to the GGSN. The protocol between the SGSN and GGSN using the Gn interface is GTP. This is a Layer 3 tunneling protocol similar to L2TP.

Figure 2-7 GPRS Network Protocol Stack



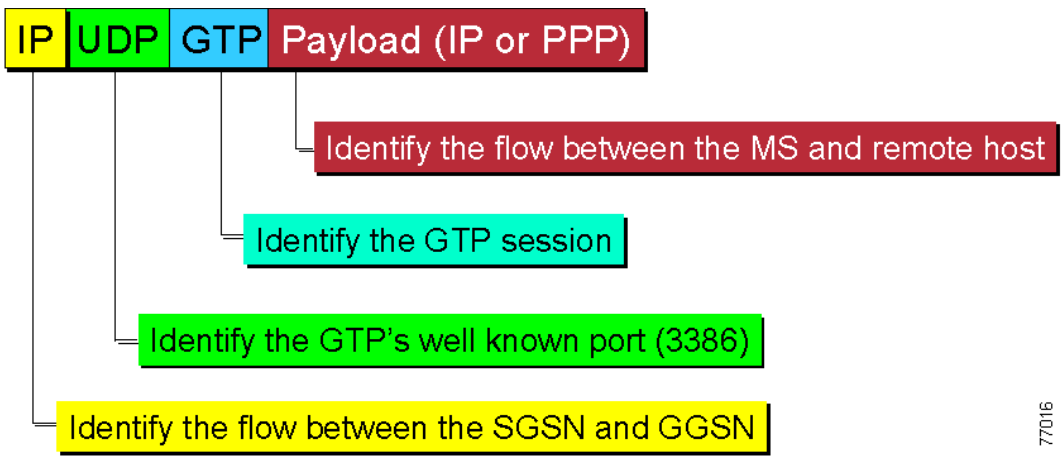
Although Figure 2-7 defines the Gn and Gi interface as IP, the underlying protocols are not specified, providing flexibility with the physical medium. The GGSN software runs on a Cisco 7206VXR hardware platform, which provides a wide range of supported physical interfaces and a high port density. The GGSN software uses a virtual template interface, which is a logical interface within the router and does not depend on the physical medium directly. A list of supported physical interfaces for the 7206VXR can be found at this URL:
http://www.cisco.com/univercd/cc/td/doc/product/core/7200vx/portadpt/index.htm.

The most common physical interface used with GPRS is Fast Ethernet. This interface provides high bandwidth, low cost, and universal connectivity to other vendor equipment. For the Gi interface, common interfaces are Serial, E1/T1 or Ethernet. Running over the physical WAN interfaces can be a wide range of protocols including Frame Relay, ISDN, and HDLC.

GPRS Tunneling Protocol

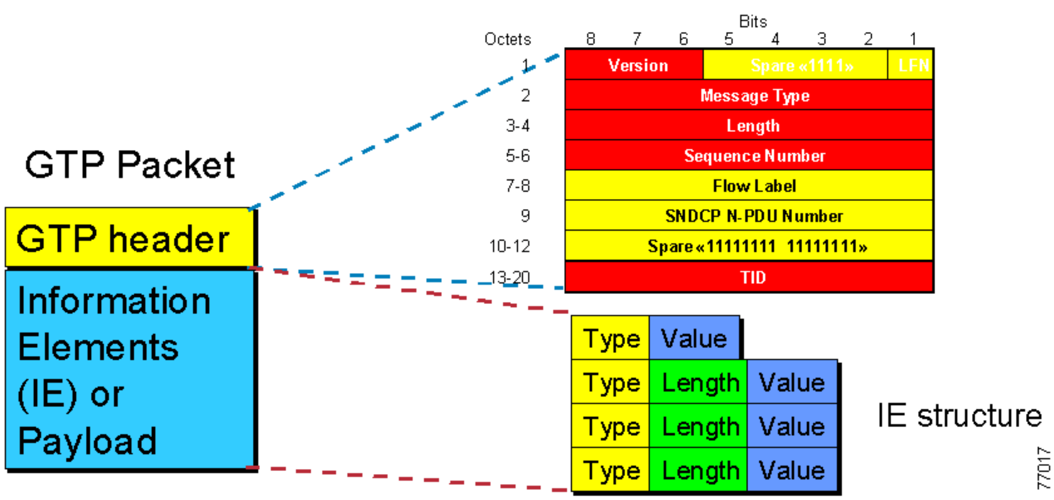
The GTP tunneling protocol is a Layer 3 tunneling protocol. The IP header identifies a session flow between the GGSN and SGSN. The UDP header identifies the GTP application protocol (Port 3386). The GTP header identifies the GTP tunnel session. The payload identifies the session flow between the mobile station and the remote host. See Figure 2-8.

Figure 2-8 GPRS Tunneling Protocol



The GTP packet structure, like any other packet, typically has a fixed-size header and other information called payload or information elements. Currently, bits 1-5 of Octet 1 and Octets 7-12 are not in use. TID is the tunnel ID that identifies a tunnel session. The length field of GTP is different from the length field of IP. In IP, the length includes the header; in GTP, length indicates only the GTP payload. See Figure 2-9.

Figure 2-9 GTP Packet Structure



GPRS Access Modes

The GPRS access modes specify whether or not the GGSN requests user authentication at the access point to a PDN (Public Data Network). The available options are:

- Transparent—No security authorization/authentication is requested by the GGSN
- Non-transparent—GGSN acts as a proxy for authenticating

The GPRS transparent and non-transparent modes relate only to PDP type IPv4.

Transparent Mode

Transparent access pertains to a GPRS PLMN that is not involved in subscriber access authorization and authentication. Access to PDN-related security procedures are *transparent* to GSNs.

In transparent access mode, the MS is given an address belonging to the operator or any other domain's addressing space. The address is given either at subscription as a static address or at PDP context activation as a dynamic address. The dynamic address is allocated from a Dynamic Host Configuration Protocol (DHCP) server in the GPRS network. Any user authentication is done within the GPRS network. No RADIUS authentication is performed; only IMSI-based authentication (from the subscriber identity module in the handset) is done.

Non-transparent Mode

Non-transparent access to an intranet/ISP means that the PLMN plays a role in the intranet/ISP authentication of the MS. Non-transparent access uses the Password Authentication Protocol (PAP) or Challenge Handshake Authentication Protocol (CHAP) message issued by the mobile terminal and piggy-backed in the GTP PDP context activation message. This message is used to build a RADIUS request toward the RADIUS server associated with the access point name (APN).

GPRS Access Point Name

The GPRS standards define a network identity called an access point name (APN). An APN identifies a PDN that is accessible from a GGSN node in a GPRS network (e.g., www.Cisco.com). To configure an APN, the operator configures three elements on the GSN node:

- Access point—Defines an APN and its associated access characteristics, including security (RADIUS), dynamic address allocation (DHCP), and DNS services
- Access point list—Defines a logical interface that is associated with the virtual template
- Access group—Defines whether access is permitted between the PDN and the MS

The Cisco GGSN is based on the routing technology, Cisco IOS. It integrates GPRS with already deployed IP services, like virtual private data networks (VPDNs) and voice over IP (VoIP).

The mobile VPN application is the first service targetted for business subscribers that mobile operators are offering when launching GPRS networks. In GPRS, the selection of the VPN can be based on the same parameters that are used in VPDN applications:

- Dialed number identification service (DNIS), i.e., the called number
- Domain, e.g., user@domain
- Mobile station IDSN (MSISDN) number, i.e, the calling number

In GPRS, only the APN is used to select the target network.The Cisco GGSN supports VPN selection based on the APN.

GPRS Processes

This section describes the following basic processes used in GPRS networks:

- Attach process—Process by which the MS *attaches* (i.e, connects) to the SGSN in a GPRS network
- Authentication process—Process by which the SGSN authenticates the mobile subscriber
- PDP activation process—Process by which a user session is established between the MS and the destination network
- Detach process—Process by which the MS *detaches* (i.e., disconnects) from the SGSN in the GPRS network
- Network-initiated PDP request for static IP address—Process by which a call from the packet data network reaches the MS using a static IP address
- Network-initiated PDP request for dynamic IP address—Process by which a call from the packet data network reaches the MS using a dynamic IP address

GPRS Attach Process

When a mobile subscriber turns on their handset, the following actions occur:

1. A handset attach request is sent to the new SGSN.
2. The new SGSN queries the old SGSN for the identity of this handset. The old SGSN responds with the identity of the handset.
3. The new SGSN requests more information from the MS. This information is used to authenticate the MS to the new SGSN.
4. The authentication process continues to the HLR. The HLR acts like a RADIUS server using a handset-level authentication based on IMSI and similar to the CHAP authentication process in PPP.
5. A check of the equipment ID with the EIR is initiated.
6. If the equipment ID is valid, the new SGSN sends a location update to the HLR indicating the change of location to a new SGSN. The HLR notifies the old SGSN to cancel the location process for this MS. The HLR sends an insert subscriber data request and other information associated with this mobile system and notifies the new SGSN that the update location has been performed.
7. The new SGSN initiates a location update request to the VLR. The VLR acts like a proxy RADIUS that queries the home HLR.
8. The new SGSN sends the Attach Accept message to the MS.
9. The MS sends the Attach Complete message to the new SGSN.
10. The new SGSN notifies the new VLR that the relocation process is complete.

Figure 2-10 and Figure 2-11 show the GPRS attach process (the numbers in the figures correspond to the numbered steps above).

Figure 2-10 GPRS Attach Request Procedure

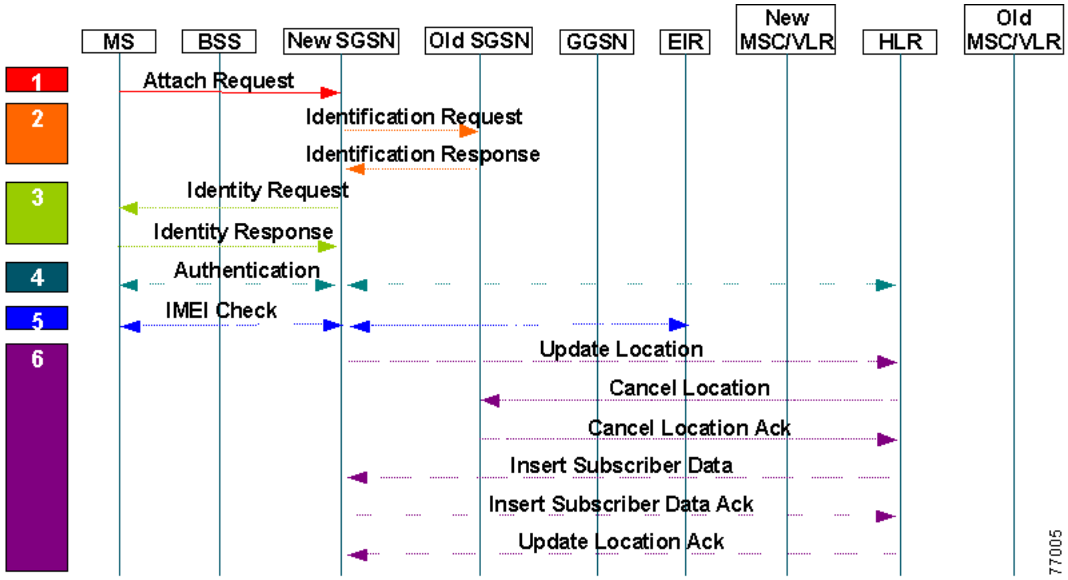
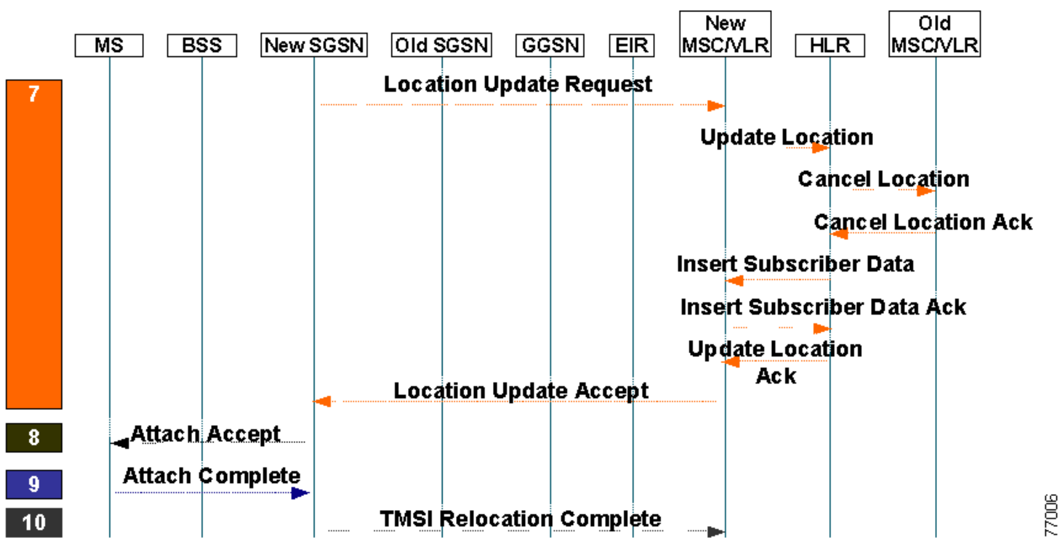


Figure 2-11 GPRS Attach Request Procedure (continued)



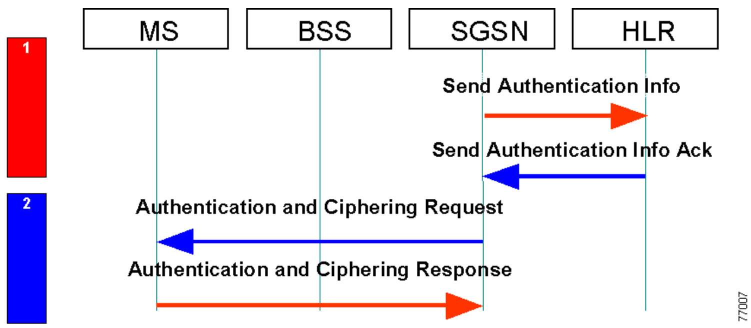
GPRS Authentication Process

The GPRS authentication process is very similar to the CHAP with a RADIUS server. The authentication process follows these steps:

1. The SGSN sends the authentication information to the HLR. The HLR sends information back to the SGSN based on the user profile that was part of the user's initial setup.
2. The SGSN sends a request for authentication and ciphering (using a random key to encrypt information) to the MS. The MS uses an algorithm to send the user ID and password to the SGSN. Simultaneously, the SGSN uses the same algorithm and compares the result. If a match occurs, the SGSN authenticates the user.

Figure 2-12 describes the GRPS authentication process that the MS uses to gain access to the network (the numbers in the figure correspond to the numbered steps above).

Figure 2-12 GPRS Authentication Procedure



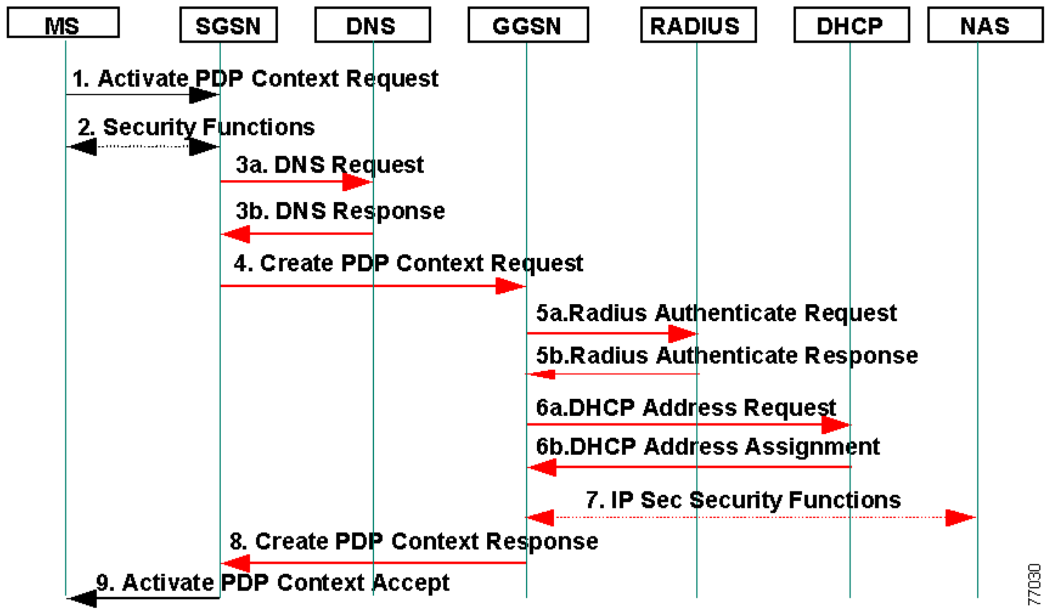
PDP Context Activation Process

The events in the PDP context activation process are described next.

1. The SGSN receives the activation request from the MS; for example, the MS requests access to the APN *Cisco.com*.
2. Security functions between the MS and SGSN occur.
3. The SGSN initiates a DNS query to learn which GGSN node has access to the *Cisco.com* APN. The DNS query is sent to the DNS server within the mobile operator's network. The DNS is configured to map to one or more GGSN nodes. Based on the APN, the mapped GGSN can access the requested network.
4. The SGSN sends a Create PDP Context Request to the GGSN. This message contains the PAP information, CHAP information, PDP request, APN, and quality of service information.
5. If operating in the non-transparent mode, the PAP and CHAP information in the PDP request packet is sent to the RADIUS server for authentication.
6. If the RADIUS server is to provide a dynamic IP address to the client, it sends a DHCP address request to the DHCP server. In transparent mode, the RADIUS server is bypassed.
7. If IPSec functionality is required, security functions occur between the GGSN and network access server (NAS).
8. The GGSN sends a Create a PDP Context Response message to the SGSN.
9. The SGSN sends an Activate PDP Context Accept message to the MS.

Figure 2-13 shows the PDP context activation procedure. The red arrows indicate the communication between the SGSN and GGSN. The numbers in the figure correspond to the numbered steps above.

Figure 2-13 PDP Context Activation Procedure



Detach Process Initiated by MS

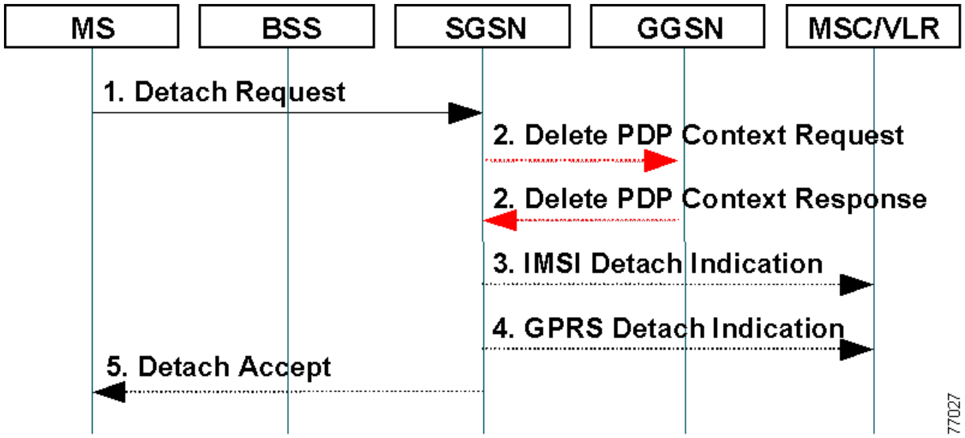
When a mobile subscriber turns off their handset, the detach process initiates. The detach process is described below.

1. The MS sends a Detach Request to the SGSN.
2. The SGSN sends a Delete PDP Context Request message to the serving GGSN.
3. The SGSN sends an IMSI Detach Indication message to the MSC/VLR indicating the MS request to disconnect.
4. The SGSN sends a GPRS Detach Indication message to the MSC/VLR.
5. The SGSN sends the Detach Accept message to the MS.

Note The GSN nodes must always respond to the detach request with a positive delete response to the MS and accept the detach request requested by the client. The positive delete response is required even if the SGSN does not have a connection pending for that client.

Figure 2-14 describes the detach process initiated by the MS. The numbers in the figure correspond to the numbered steps above.

Figure 2-14 MS Initiate Detach Procedure



Network Initiated PDP Request For A Static IP Address

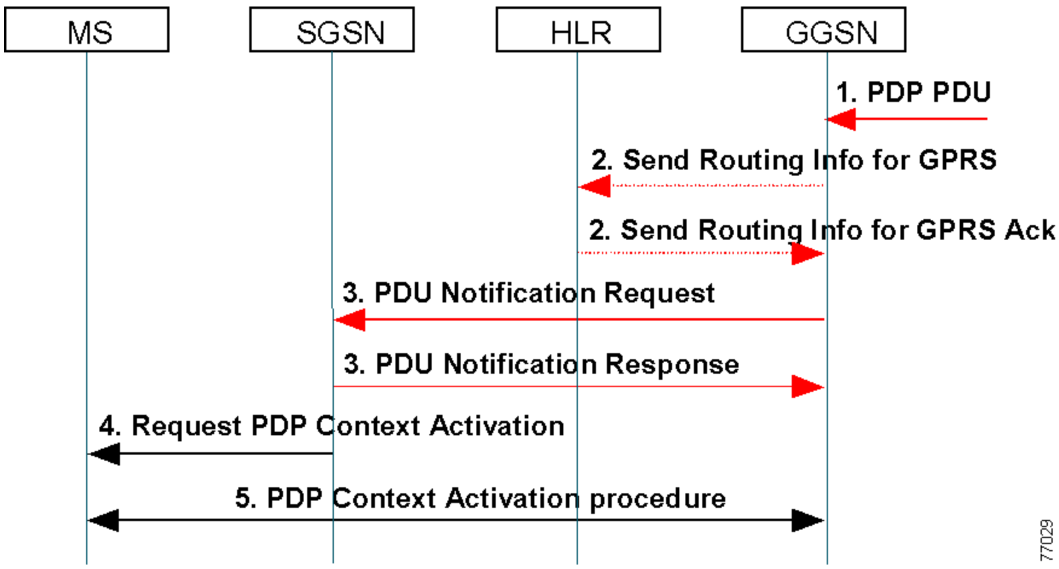
The PDP protocol data unit (PDU) initiated from the network side is not fully specified by ETSI standards. A connection request generated from the Internet/intranet site specifies only the IP address of the client in the IP packets destined for the MS. The requesting host provides no indication of the mobile device IMSI (i.e., the MAC address of the MS). In mobile communications, all communications are based on the MS MAC address called the IMSI. The IP address must be mapped to an IMSI to identify a valid GTP tunnel. Cisco's GGSN implementation provides a mapping table via command line interface (CLI) that allows the operator to key in the MS IMSI and the associated static IP address.

The following steps describe a PDP request initiated from the network side when the client has been assigned a static IP address.

1. When the GGSN receives a packet, it checks its mapping table for an established GTP tunnel for this packet.
2. When the GGSN locates the IMSI associated with this IP address, it sends a Send Routing Information message to HLR through an intermediate SGSN. The intermediate SGSN notifies the GGSN of the actual SGSN currently serving this client.
3. On locating the appropriate SGSN, the GGSN sends a PDU Notification Request message to the serving SGSN.
4. The SGSN sends a Request PDP Context Activation message to the MS and notifies it of the pending connection request.
5. If the MS agrees to accept the call, it enters the PDP Context Activation procedure with the requesting GGSN.

Figure 2-15 shows a PDP request initiated from the network side when the client has been assigned a static IP address. The numbers in the figure correspond to the numbered steps above.

Figure 2-15 Network Initiate PDP (Static IP Address)



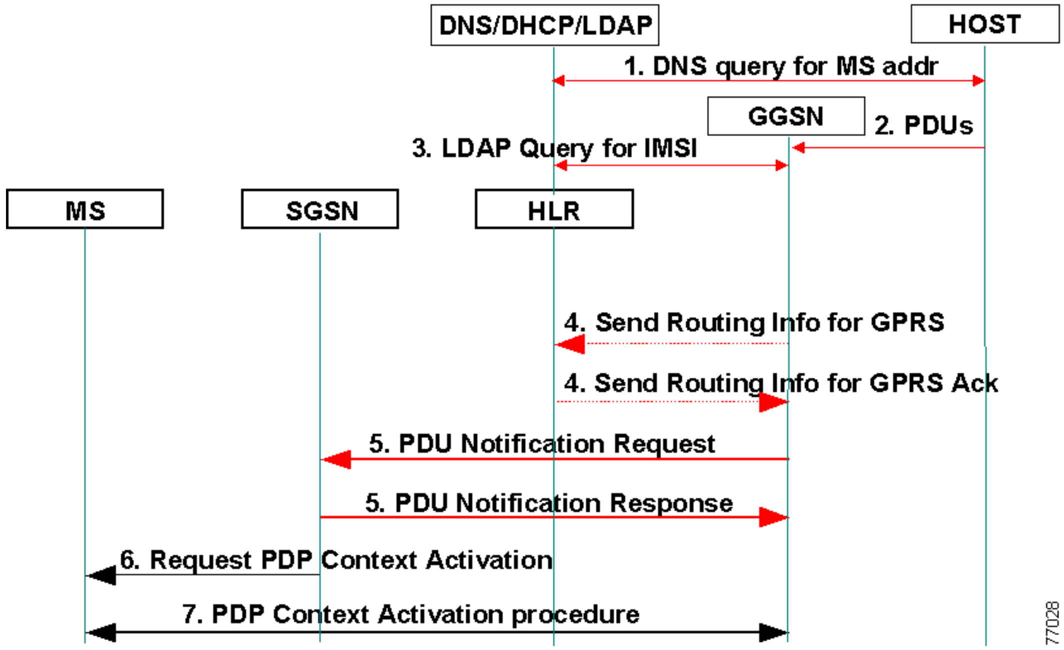
Network Initiated PDP Request For A Dynamic IP Address

The ETSI standards do not fully specify requirements for a network-generated PDP request when the client is dynamically assigned a temporary IP by a DHCP server. The following message sequence is Cisco's implementation for this scenario. This method uses Cisco's Network Registrar (CNR), which includes a DHCP, DNS, and an LDAP server.

1. The host initiates a DNS query to obtain the IP address of the MS from a DNS server. The DNS server resolves the client's name to an IP address previously assigned to the client by the DHCP server.
2. The host sends a request to the GGSN for a connection using this IP address.
3. The GGSN queries the LDAP server to obtain the MS IMSI. The LDAP server stores a record for the MS with the client IMSI, name, and IP address.
4. The GGSN sends a PDU Notification Request message to the serving SGSN.
5. The SGSN sends a Request PDP Context Activation message to the MS and notifies it of the pending connection request.
6. If the MS agrees to accept the call, it enters the PDP Context Activation procedure with the requesting GGSN.

Figure 2-16 describes a PDP request initiated from the network side when the client has been assigned a dynamic IP address. The numbers in the figure correspond to the numbered steps above.

Figure 2-16 Network Initiate PDP (Dynamic IP Address)



Universal Mobile Telecommunication System

The Universal Mobile Telecommunication System (UMTS) is a third generation (3G) mobile communications system that provides a range of broadband services to the world of wireless and mobile communications. The UMTS delivers low-cost, mobile communications at data rates of up to 2 Mbps. It preserves the global roaming capability of second generation GSM/GPRS networks and provides new enhanced capabilities. The UMTS is designed to deliver pictures, graphics, video communications, and other multimedia information, as well as voice and data, to mobile wireless subscribers.

The UMTS takes a phased approach toward an all-IP network by extending second generation (2G) GSM/GPRS networks and using Wide-band Code Division Multiple Access (CDMA) technology. Handover capability between the UMTS and GSM is supported. The GPRS is the convergence point between the 2G technologies and the packet-switched domain of the 3G UMTS.

UMTS Services

The UMTS provides support for both voice and data services. The following data rates are targets for UMTS:

- 144 kbps—Satellite and rural outdoor
- 384 kbps—Urban outdoor
- 2048 kbp—Indoor and low range outdoor

Data services provide different quality-of-service (QoS) parameters for data transfer. UMTS network services accommodate QoS classes for four types of traffic:

- Conversational class—Voice, video telephony, video gaming
- Streaming class—Multimedia, video on demand, webcast
- Interactive class—Web browsing, network gaming, database access
- Background class—E-mail, short message service (SMS), file downloading

The UMTS supports the following service categories and applications:

- Internet access—Messaging, video/music download, voice/video over IP, mobile commerce (e.g., banking, trading), travel and information services
- Intranet/extranet access—Enterprise application such as e-mail/messaging, travel assistance, mobile sales, technical services, corporate database access, fleet/warehouse management, conferencing and video telephony
- Customized information/entertainment—Information (photo/video/music download), travel assistance, distance education, mobile messaging, gaming, voice portal services
- Multimedia messaging—SMS extensions for images, video, and music; unified messaging; document transfer
- Location-based services—Yellow pages, mobile commerce, navigational service, trading

UMTS Architecture

The public land mobile network (PLMN) described in UMTS Rel. '99 incorporates three major categories of network elements:

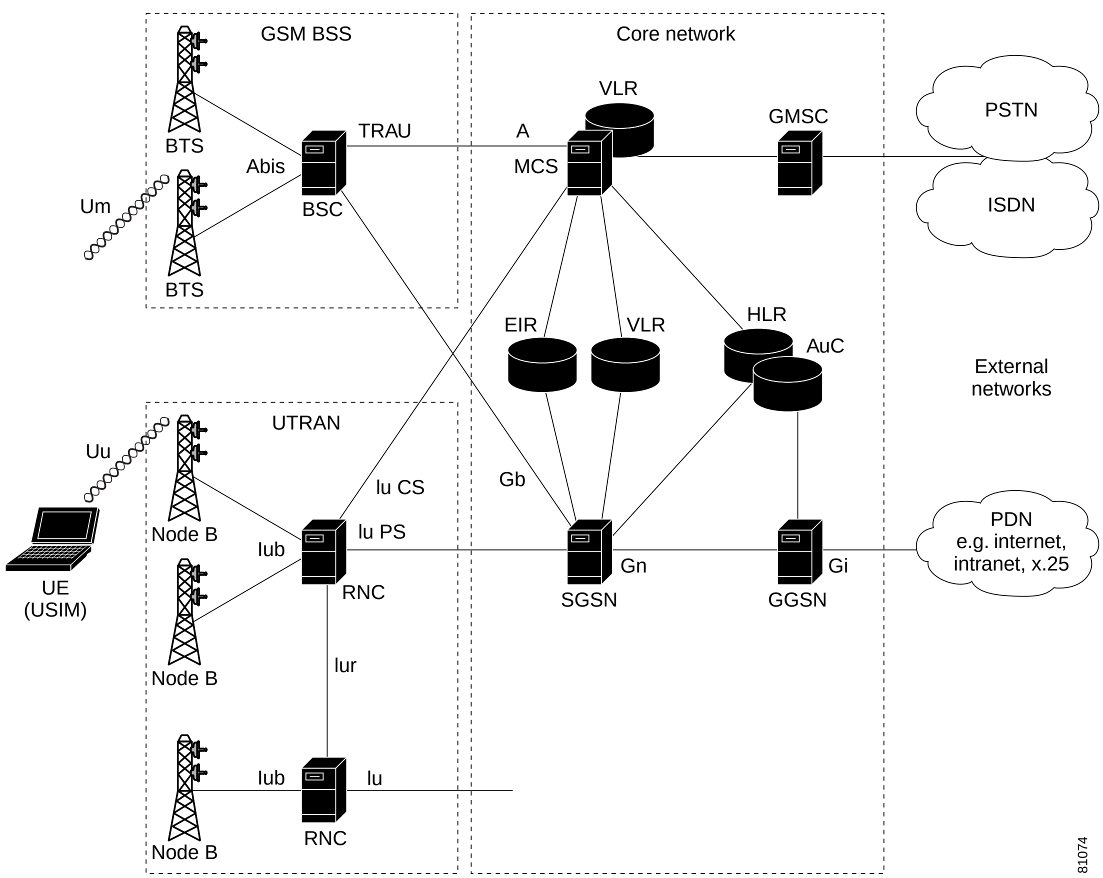
- GSM phase 1/2 core network elements—Mobile services switching center (MSC), visitor location register (VLR), home location register (HLR), authentication center (AuC), and equipment identity register (EIR)
- GPRS network elements—Serving GPRS support node (SGSN) and gateway GPRS support node (GGSN)
- UMTS-specific network elements—User equipment (UE) and UMTS terrestrial radio access network (UTRAN) elements

The UMTS core network is based on the GSM/GPRS network topology. It provides the switching, routing, transport, and database functions for user traffic. The core network contains circuit-switched elements such as the MSC, VLR, and gateway MSC (GMSC). It also contains the packet-switched elements SGSN and GGSN. The EIR, HLR, and AuC support both circuit- and packet-switched data.

The Asynchronous Transfer Mode (ATM) is the data transmission method used within the UMTS core network. ATM Adaptation Layer type 2 (AAL2) handles circuit-switched connections. Packet connection protocol AAL5 is used for data delivery.

The UMTS architecture is shown in Figure 2-17.

Figure 2-17 UMTS Architecture



General Packet Radio System

The General Packet Radio System (GPRS) facilitates the transition from phase1/2 GSM networks to 3G UMTS networks. The GPRS supplements GSM networks by enabling packet switching and allowing direct access to external packet data networks (PDNs). Data transmission rates above the 64 kbps limit of integrated services digital network (ISDN) are a requirement for the enhanced services supported by UMTS networks. The GPRS optimizes the core network for the transition to higher data rates. Therefore, the GPRS is a prerequisite for the introduction of the UMTS.

UMTS Interfaces

The UMTS defines four new open interfaces (see Figure 2-17):

- *Uu* interface—User equipment to Node B (the UMTS WCDMA air interface)
- *Iu* interface—RNC to GSM/GPRS (MSC/VLR or SGSN)
  - *Iu-CS*—Interface for circuit-switched data
  - *Iu-PS*—Interface for packet-switched data
- *Iub* interface—RNC to Node B interface
- *Iur* interface—RNC to RNC interface (no equivalent in GSM)

The *Iu*, *Iub*, and *Iur* interfaces are based on the transmission principles of aynchronous transfer mode (ATM).

UMTS Terrestrial Radio Access Network

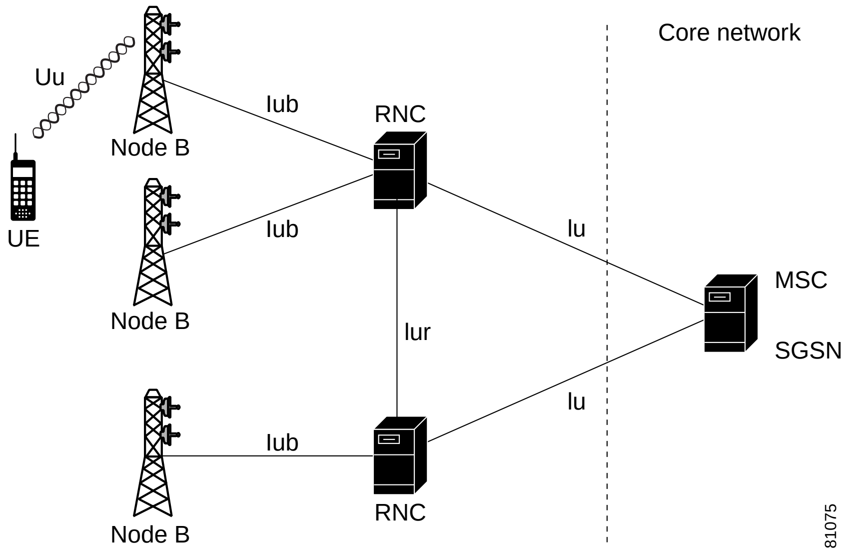
The major difference between GSM/GPRS networks and UMTS networks is in the air interface transmission. Time division multiple access (TDMA) and freqency division multiple access (FDMA) are used in GSM/GPRS networks. The air interface access method for UMTS networks is wide-band code division multiple access (WCDMA), which has two basic modes of operation: frequency division duplex (FDD) and time division duplex (TDD). This new air interface access method requires a new radio access network (RAN) called the UTMS terrestrial RAN (UTRAN). The core network requires minor modifications to accommodate the UTRAN.

Two new network elements are introduced in the UTRAN: the radio network controller (RNC) and Node B. The UTRAN contains multiple radio network systems (RNSs), and each RNS is controlled by an RNC. The RNC connects to one or more Node B elements. Each Node B can provide service to multiple cells.

The RNC in UMTS networks provides functions equivalent to the base station controller (BSC) functions in GSM/GPRS networks. Node B in UMTS networks is equivalent to the base transceiver station (BTS) in GSM/GPRS networks. In this way, the UMTS extends existing GSM and GPRS networks, protecting the investment of mobile wireless operators. It enables new services over existing interfaces such as *A*, *Gb*, and *Abis*, and new interfaces that include the UTRAN interface between Node B and the RNC (*Iub*) and the UTRAN interface between two RNCs (*Iur*).

The network elements of the UTRAN are shown in Figure 2-18.

Figure 2-18 UTRAN Architecture



Radio Network Controller

The radio network controller (RNC) performs functions that are equivalent to the base station controller (BSC) functions in GSM/GPRS networks. The RNC provides centralized control of the Node B elements in its covering area. It handles protocol exchanges between UTRAN interfaces (*Iu*, *Iur*, and *Iub*). Because the interfaces are ATM-based, the RNC performs switching of ATM cells between the interfaces. Circuit-switched and packet-switched data from the *Iu-CS* and *Iu-PS* interfaces are multiplexed together for transmission over the Iur, Iub, and Uu interfaces to and from the user equipment (UE). The RNC provides centralized operation and maintenance of the radio network system (RNS) including access to an operations support system (OSS).

The RNC uses the *Iur* interface. There is no equivalent to manage radio resources in GSM/GPRS networks. In GSM/GPRS networks, radio resource management is performed in the core network. In UMTS networks, this function is distributed to the RNC, freeing the core network for other functions. A single serving RNC manages serving control functions such as connection to the UE, congestion control, and handover procedures. The functions of the RNC include:

- Radio resource control
- Admission control
- Channel allocation
- Power control settings
- Handover control
- Macro diversity
- Ciphering
- Segmentation and reassembly
- Broadcast signalling
- Open loop power control

Node B

Node B is the radio transmission/reception unit for communication between radio cells. Each Node B unit can provide service for one or more cells. A Node B unit can be physically located with an existing GSM base transceiver station (BTS) to reduce costs of UMTS implementation. Node B connects to the user equipment (UE) over the *Uu* radio interface using wide-band code division multiple access (WCDMA). A single Node B unit can support both frequency division duplex (FDD) and time division duplex (TDD) modes. The *Iub* interface provides the connection between Node B and the RNC using asynchronous transfer mode (ATM). Node B is the ATM termination point.

The main function of Node B is conversion of data on the *Uu* radio interface. This function includes error correction and rate adaptation on the air interface. Node B monitors the quality and strength of the connection and calculates the frame error rate, transmitting this information to the RNC for processing. The functions of Node B include:

- Air interface transmission and reception
- Modulation and demodulation
- CDMA physical channel coding
- Micro diversity
- Error handling
- Closed loop power control

Node B also enables the UE to adjust its power using a technique called downlink transmission power control. Predefined values for power control are derived from RNC power control parameters.

UMTS User Equipment

The UMTS user equipment (UE) is the combination of the subscriber's mobile equipment and the UMTS subscriber identity module (USIM). Similar to the SIM in GSM/GPRS networks, the USIM is a card that inserts into the mobile equipment and identifies the subscriber to the core network.

The USIM card has the same physical characteristics as the GSM/GPRS SIM card and provides the following functions:

- Supports multiple user profiles on the USIM
- Updates USIM information over the air
- Provides security functions
- Provides user authentication
- Supports inclusion of payment methods
- Supports secure downloading of new applications

The UMTS standard places no restrictions on the functions that the UE can provide. Many of the identity types for UE devices are taken directly from GSM specifications. These identity types include:

- International Mobile Subscriber Identity (IMSI)
- Temporary Mobile Subscriber Identity (TMSI)
- Packet Temporary Mobile Subscriber Identity (P-TMSI)
- Temporary Logical Link Identity (TLLI)
- Mobile station ISDN (MSISDN)
- International Mobile Station Equipment Identity (IMEI)

- International Mobile Station Equipment Identity and Software Number (IMEISV)

The UMTS UE can operate in one of three modes of operation:

- PS/CS mode—The UE is attached to both the packet-switched (PS) and circuit-switched (CS) domain, and the UE can simultaneously use PS and CS services.
- PS mode—The MS is attached to the PS domain and uses only PS services (but allows CS-like services such as voice over IP [VoIP]).
- CS mode—The MS is attached to the CS domain and uses only CS services.