Charging Requirements for UMTS Packet -Switched Data Services

Report Number: EX 001/2001

By
Taiwo Oyedele

A Thesis work in Mobile Communications for the
Department of Computing and Electronic Engineering
Chalmers University of Technology
Gothenburg, Sweden

Acknowledgement

I would like to express my gratitude and appreciation for the invaluable guidance, helpful suggestions, encouragement and support given by Pete Barron. I would also like to thank Nigel Lobley, Nadir El Yakhlifi, Uduak Ekpe and all the members of the Mobile Network Design group at BT Adastral Park, who have in one way or the other influenced and contributed positively to the project work presented in this thesis.

I also express thanks to all my lecturers both at Chalmers University of Technology and at University College London for the knowledge I gained from the course, which has been invaluable to the completion of this project. Special thanks to Erik Ström for his outstanding guidance and encouragement throughout the project phase.

Gratitude and appreciation is also extended to my family for the support provided to me throughout the duration of my Masters course. Above all I would like to show my thanks and appreciation to God for enabling me to successfully complete my Masters programme.

Abstract

Why do we need new charging mechanisms? Packet-Switched services do not require a dedicated path for the duration of the call. Instead, they rely on a virtual connection with the network setting up a real connection only when there is data to be sent/received, and then dropping the connection until more data is ready to be transmitted or received. As a result, network resources are used only when packets are in transmission, and there is thus no clear notion of “the duration of a call”. Hence, users may no longer wish to be charged in the traditional duration-based manner. Besides, in contrast to today’s dedicated networks, (e.g. PSTN, Frame Relay and even GSM) which carry a single application type at a fixed duration-based price, General Packet Radio Service (GPRS) and UMTS are **multi-service** networks. These require flexible pricing models, where price is a function of data volume, QoS provisioned, packet content, etc. For example, voice and video data is more or less continuous, thus requiring network resources more or less for the entire duration of its transfer. Consequently, we can expect voice and video to have a flat per-minute rate, whereas other services, such as, Resource (e.g. file) Access and fax/email applications may be charged on a volume basis, i.e. charged per MB of data transferred. Moreover, voice and video applications require bi-directional latency and bandwidth guarantees, whereas Resource Access and fax/email applications do not have the latency constraint, but may require significant bandwidth depending on the resource being accessed. Hence, QoS plays a significant role in most new charging mechanisms. In other words, a flexible charging mechanism capable of measuring and rating services under various usage-based metrics is paramount.

Due to the packet-based nature of the underlying network technology, these services require novel charging mechanisms, which in turn require extensive data-extraction from various network service elements. This document is not intended to be a technical specification sufficient for practical implementation of network-usage data extraction among service elements participating in the delivery of UMTS IP-based services. Such a technical specification is left to standards bodies and associations like 3GPP, IPDR.org GBA etc. Instead, it is intended to highlight the possible mechanisms that may be implemented for charging UMTS services. It also identifies the network requirements, the limitations and more importantly the possibilities these new charging schemes open up for Mobile Network Operators also known as PLMN operators. Some of these possibilities include additional revenue streams like acquisition of financial transactions, as opposed to solely being a voice/data ‘carrier provider’, as is the case with traditional PLMN operators. The project involved lots of information gathering from the Internet, technical papers, magazines and suppliers. It also required a need to work with people from across BT, Adastral Park to determine their views on how technology will affect the future user experience, and hence, future charging schemes for packet based services.

Table of Contents

Definitions and Abbreviations

Network Operator: A company providing the network infrastructure required for administering various telecommunications services.

Node: Part of the Network infrastructure. Could be a host, server, switch, router or a base station.

Service Provider: A company administering the services provided by the Network Operator. May or may not be a Network Operator.

R99 and R00: The specifications are grouped into 'Releases'. The original specifications were published by ETSI in 1994 and are now known as 'Phase 1'. The next releases were known as 'Phase 2', and 'Phase 2+' respectively. Within Phase 2+, there have been annual releases since 1996, which are known as R96, R97, R98, R99 and R00.

3G	3^{rd} Generations
3GPP	3G Partnership Project
AAA	Authentication, Authorisation, and Accounting
ADSL	Asymmetric Digital Subscriber Line
AMPS	Advanced Mobile Phone System
ARPU	Average Revenue Per User
ATM	Asynchronous Transfer Mode
AUC	Authentication Centre
BS	Billing System or Base Station depending on context
BSC	Base Station Controller
BSS	Business Support System
BTS	Base Transceiver System
CCCH	Common Control Channels
CDMA	Code Division Multiple Access
CDR	Call Detail Record
CEPT	Conference of European Posts and Telecommunications
CGF	Call Gateway Functionality
CID	Charging Identity
CL	Connectionless
CO	Connection Oriented
CS	Circuit Switched
CSCF	Call Server Control Function
DHCP	Dynamic Host Configuration Protocol
ECSD	Enhanced Circuit Switched Data
EDGE	Enhanced Data rates for GSM Evolution
EIR	Equipment Identification Register
ETSI	European Telecommunications Standards Institute
FDD	Frequency Division Duplex
FTP	File Transfer Protocol
G-CDR	GGSN CDR
GGSN	Gateway GPRS Support Node
GMSC	Gateway Mobile Switching Centre
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GSN	GPRS Support Node
GTP	GPRS Tunnelling Protocol
HLR	Home Location Register
HSCSD	High Speed Circuit Switched Data
HTML	Hyper Text Mark-up Language
IETF	Internet Engineering Task Force
IMEI	International Mobile Equipment Identity
IMSI	International Mobile Subscriber Identity
IN	Intelligent Network
IP	Internet Protocol
IPDR	Internet Protocol Detail Record

ISDN	Integrated Services Digital Network
ISP	Internet Service Provider
ITU-T	International Telecommunications Union –Telecommunications Sector
IWF	Inter-Working Functions
LAN	Local Area Network
M-CDR	Mobility CDR
ME	Mobile Equipment
MGCF	Media Gateway Control Function
MM	Mobility Management
MO	Mobile Originated
MS	Mobile Station
MSC	Mobile Switching Centre
MSISDN	Mobile Subscriber ISDN number
MT	Mobile Terminated or Mobile Terminal depending on context
MTU	Maximum Transmission Units
MoU	Memorandum of Understanding
NAMTS	Nippon Advanced Mobile Telephone System
NAS	Network Access Server
NE	Network Element
NMT	Nordic Mobile Telecommunications
NSS	Network Switching Subsystem
OLO	Other Licensed Operators
OMC	Operations & Maintenance Centre
OSI	Open Systems Interface
PBX	Private Branch Exchange
PCM	Pulse Code Modulation
PDA	Personal Digital Assistant
PDN	Packet Data Network
PDP	Packet Data Protocol
PDU	Packet Data Unit
PLMN	Public Land Mobile Network
PPS	Pre-Paid Service
PS	Packet Switched
PSTN	Public Switched Telephone Network
PTM	Point-to-Multipoint
P-TMSI	Packet-Temporary Mobile Subscriber Identity
PTP	Point-to-Point
QoS	Quality of Service
RA	Routing Area
RAB	Radio Access Bearer
RAI	Routing Area Identity
RAN	Radio Access Network
RADIUS	Remote Access Dial-In User Service
RLA	Remote LAN Access
RNC	Radio Network Controller
RSVP	Resource reSerVation Protocol
SACCH	Slow Associated Control Channel
S-CDR	SGSN CDR
SDU	Service Data Unit
SIM	Subscriber Identity Module
SGSN	Serving GPRS Support Node
SMS	Short Message Service
TACS	Total Access Telecommunications Systems
TAP	Transfer Accounting Procedure
TBF	Temporary Block Flow
TCP	Transport Control Protocol
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TID	Tunnelling Identity
UDP	User Datagram Protocol

UE	User Equipment
UMTS	Universal Mobile Telecommunications System
USIM	UMTS SIM
USSD	Unstructured Supplementary Services Data
UTRAN	UMTS Terrestrial Radio Access Network
VLR	Visitor Location Register
VMSC	Visited Mobile Switching Centre
VoIP	Voice Over IP
VoD	Video on Demand
WAP	Wireless Application Protocol
W-CDMA	Wide-band CDMA
WML	Wireless Mark-up Language

Introduction

High-bandwidth networks with sophisticated security features and fast evolving Quality of Service (QoS) mechanisms, enables Service Providers (both fixed wire-line and wireless) to offer packet-based services, traditionally provided by Circuit-Switched networks. Already, we see services such as, telephone, fax, pay-per-view-movies, video-conferencing, secure remote access, secure corporate network interconnection (VPN services), electronic commercial transactions, TV and radio broadcasts, all carried over IP networks. The Universal Mobile Telecommunications System (UMTS) network providing very high data rates and enabling end-to-end QoS is intended to support all of these services and applications – thus, enabling complete wireless Internet access. Since the existing flat-rate pricing model, usually applied for IP services, simply does not apply to these resource-hungry, performance-oriented applications, a comprehensive IP-metering system coupled with the UMTS-Call Detail Record system is required to bring these services to market. Furthermore, customers need to control their communications expenses and will choose providers not just based on cost, QoS options etc but also on their ability to measure and rate services under various usage-based metrics. Apart from the off-peak discounts existing in traditional billing and rating systems, UMTS billing systems will include a QoS discount, since the network may fail to meet the QoS level specified by the application. This inability of the network to meet the QoS profile requested is likely, particularly in early network deployments.

The UMTS network preceded in most cases by GPRS integrates two of the world's fastest growing industries – mobile communications and the Internet. It is intended to support a wide range of voice, data, and multimedia services and applications. Since GPRS/UMTS is IP-based, it is fully compatible with the Internet and truly provides the capability to extend the Internet to the mobile phone. The UMTS network can be viewed as a sub-network of the Internet with UMTS-capable handsets being viewed as mobile hosts. The implementation of a fully functional and commercially viable UMTS network is however faced by technological and commercial challenges, some of which include

- Issues such as battery life, cost, monitoring of paging channels etc associated with its "Always on" "Always connected" functionality.
- Billing issues - How do we charge for packet-based services and who is billed for Mobile Terminated traffic.
- Security concerns (Spam/Viruses) – Unwanted/unsolicited data, pushing information to UMTS-handsets could result in virus threat, etc.
- Radio resource usage and optimisation

The main subject of this document however, is the billing issue. This document has been divided into four main chapters – GSM, GPRS, UMTS and Packet-Based Charging. For those with prior knowledge/experience in GSM and/or GPRS, refer to chapter 4 for packet-based billing issues.

The GSM chapter gives a brief summary of the history of GSM; includes an overview of general cellular concepts and describes some of the features and services supported by the GSM network infrastructure. Then, a description of the separate but dependent network subsystems and individual components is given before moving on to address the charging scheme employed within GSM. Billing within the Pre-Paid Service is also discussed.

The GPRS chapter aims to give a general overview of the GPRS network, identify the key differences between GPRS and GSM standards and technologies, and briefly discusses the services intended to be supported by GPRS, and the network requirements to support these services.

The UMTS chapter describes the network infrastructure and the new network elements and thus new services and applications supported by UMTS. It also identifies the main differences between GPRS and UMTS, i.e. radio modulation technique and its ability to support real-time- as well as non-real-time applications, as opposed to the constraint imposed on GPRS to only support non-real-time applications. This additional functionality requires enhanced network requirements to enable the production and collection of relevant Call Detail Records (CDRs) for flexible charging mechanisms.

Chapter 4 describes the charging functionality within GPRS and UMTS. GPRS and UMTS intend to support an 'always on, always connected' functionality, a distinguishing property of packet-based services and as such require new and innovative charging mechanisms.

In order to ease the understanding of the need for new innovative charging mechanisms required for GPRS/UMTS packet-based services, I have dedicated a section in each chapter to the Network and Service Element layer (NSE). This section describes the network architecture and the functions of major network nodes and how these inter-operate to enable the provision of IP-based services and also the production of CDRs and IPDRs (Internet Protocol Detail Records) for billing purposes. For example, routers (SGSN and GGSN), access devices (User Equipment – UE), and transmission facilities together provide basic connectivity; e-mail, and file servers provide application services; gateways provide a translation service from circuit to packet voice and so on. Also, the systems that configure and manage these devices like the RSVP policy server, bandwidth management system, H.323 gatekeeper, RADIUS authentication server etc. are part of the NSE. A section of chapter 4 has also been dedicated to mediation, which sits between the Network Elements/infrastructure and the Business Support System /Operating Support System (BSS/OSS). The goal of the mediation system is to capture all usage information required by the BSS systems, thus providing a single interface to the OSS/BSS. The mediation system must in some way determine the devices at the NSE layer and interface with that element to extract the relevant usage information. The BSS consists of the systems deployed by the service provider to support business operations. These include Billing System, Customer Care/Relationship Management, Fraud Detection, Market Analysis etc. Only the Billing System has been considered in this report.

Chapter 1

1 GSM

This section gives a brief summary of the history of GSM, how and where it developed and who was responsible for its development. It also includes an overview of general cellular concepts and goes on to describe some of the features and services supported by the GSM network infrastructure. Then, an introduction to the GSM infrastructure and a description of the separate but dependent network subsystems and individual components is given before moving on to address the charging scheme employed within GSM. Finally, the fastest growing GSM account type, particularly among the younger GSM user base, - the prepaid service and the charging mechanism involved is briefly discussed. The subsequent chapters describe other digital mobile technologies to be employed in the future including GPRS and UMTS. GPRS opens up the mobile market to a wealth of data applications, with over 1 billion people forecast to be accessing the Internet by 2004 – over half of these via wireless devices.

1.1 History

The need for people and businesses to be accessible at all times and outside the offices prompted the idea of manufacturing mobile telephones, which use radio waves as the communication medium. The idea was to provide the user with those facilities normally attributed to 'wire-line' networks, i.e. place and receive calls to/from other users connected either to fixed or mobile networks. Several cellular systems have existed, the very first ones using analogue technology. Some of these systems include TACS adopted in the U.K, AMPS in the U.S.A, NAMTS in Japan, NMT in the Scandinavian countries and Network C in Germany. Below is a brief history of GSM, the world's fastest growing and most dynamic digital cellular technology.

In 1982, Conference of European Posts and Telecommunications (CEPT) formed a new technical committee called the Groupe Special Mobile (GSM), whose task was to specify a new mobile communication system for Europe - The Second-Generation System. The system was given an English title, the Global System for Mobile Communication, thereby retaining the GSM logo.

In the mid 1980s, it was planned that GSM, a digital communications system would be launched in 1991 and that all major cities would be covered by 1993 and major trunk routes by 1995. Due to concerns about meeting this deadline, the introduction of GSM was phased. Phase 1 introduced the basic services in 1991 and other services were delayed until the later release of phase 2.

In 1988, the European Telecommunications Standards Institute, (ETSI) was created and given the overall responsibility for GSM. Unlike CEPT, made up solely of Administrations, ETSI comprised of Network operators, Network manufacturers/vendors, administrations, user organisations etc.

Another important body is the Memorandum of Understanding (MoU) who are concerned with global rather than technical issues such as international roaming, charging strategies, service requirements, security aspects, compatibility of numbering and routing plans, time scales for deployment of systems, etc. An extensive and detailed description of the system known as the GSM Technical Specification TS was produced to facilitate the production of system hardware and software. The TS was divided into 12 aspects including Service Aspects, Network Aspects, Physical Layer in the Radio Path, Audio Aspects etc.

As previously mentioned, GSM introduction was phased with phase 1 specifications defining the following "basic" Services – Telephony, Emergency calls, Short Message Service, and Bearer services 300bps – 9.6kbps. Its Supplementary Services included, Call forwarding and Call Barring. Group 3 Facsimile, Half-Rate CODEC, and other Supplementary Services were proposed for Phase 2

GSM was officially launched on 1 July 1991 as planned but the services were not available until the following Spring due to unavailability of the mobile phones. They existed but had not completed the formal acceptance procedure. A new approval procedure known as Interim Type Approval (ITA) was employed and hence the first operational mobile phones were ITA approved. The first operational Networks offering only Telephony and Emergency Calls appeared in the spring of 1992.

Phase 2 included several Supplementary Services including, Line Identification, Call waiting, Mutli-Party calls, etc. The Phase 2+ terminology was later adopted to indicate an evolving Second-Generation System towards the Third Generation System - UMTS. GSM has been proven to be extremely successful worldwide and today has a user base of over 300 million. This figure is expected to rise within the next three years to over 600 million. [15].

Today, GSM the world's fastest growing and most dynamic digital cellular technology is on the brink of a change more dramatic than any other change in its short history. The circuit switched technology, which has been at the heart of GSM since its inception is about to give way to packet switching, a technology used by GPRS and UMTS. There are several benefits and advantages associated with packet switching (these are discussed in section 2.2.) as opposed to circuit switching but this new technology brings along with it new difficulties and challenges that need to be overcome for a successful transition to take place. One of the many new challenges, which have emerged, is the requirement of novel charging mechanisms, hence this project. The Communication Service Providers (CSPs) and Network operators need to bill accurately and generate revenue based not just on time and distance parameters but also on the basis of content, data volumes, bandwidth, and Quality of Service metrics.

1.2 Cellular Systems

Today's mobile radio systems are engineered around fixed radio base station sites at high altitudes, radiating large amounts of power. This configuration allows large coverage areas but prevents the re-use of a particular radio channel for some considerable distance. Since the radio spectrum is severely restricted and only a small fraction of it is allocated to Public Services, frequency re-use is essential for systems aimed to cater for large populations. This led to the conception of cellular radio systems. In these systems, the area to be covered is divided into a number of cells; a fixed radio site, the Base Station, serves each cell. The cells are grouped into clusters and the frequencies allocated to a particular cell within a cluster are not re-used until the corresponding cell in adjacent clusters. The figure below shows three clusters, each with 7 cells. Other cluster formats may contain 4, 9, 12 or 21 cells.

The Cellular Concept

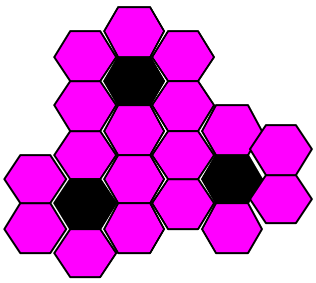


Figure1.1 Three clusters each with seven cells. Each cell is fed by a BTS described below.

The number of channels within a cell, affected by the number of cells in a cluster, dictates the maximum number of calls that can exist within that cell. Reducing the cell size and/or number of cells per cluster increases the system's traffic capability, by reducing the system's re-use distance, but on the other hand enhances interference, specifically co-channel interference. The re-use distance D is obviously dependent on the cell size as well as the cluster size.

$D = R*(3n)^{0.5}$, where R = cell radius and n = number of cells/cluster.

Corresponding cells in adjacent clusters are the source of co-channel interference. Another important quality measure is the C/I (Carrier to Interference) ratio, which is affected by the number of cells/cluster. The greater the number of cells, the higher the C/I ratio. This ratio is also affected by the propagation co-efficient, which is between 3 and 4 for most environments.

To accommodate for increased traffic levels without compromising the Quality of Service to network subscribers, the cells can either be split or sectorised. Cell splitting requires that additional base stations are provided and as a result, the channels need be re-allocated. This method is expensive due to additional hardware requirements and it is also limited by co-channel interference as the re-use distance is reduced considerably. Sectorisation handles the interference problem by making use of directional antennae to divide the cell into 3 or 6 sectors.

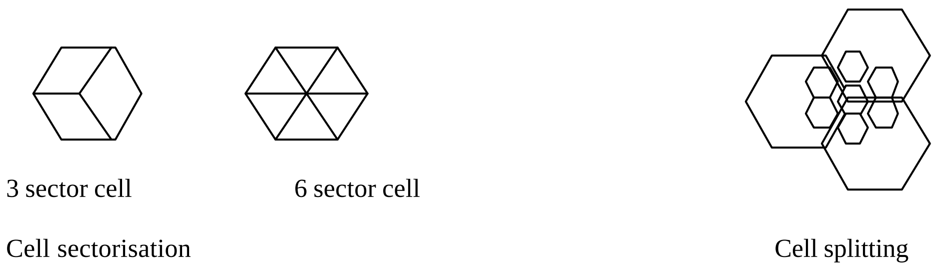


Figure 1.2 Two different methods commonly used to cater for increased traffic levels without compromising the QoS.

A portion of the radio spectrum around 900MHz was allocated for radiotelephone systems by the World Administration Radio Conference, (WARC). These frequencies were used by earlier systems including TACS and NMT 900. GSM today uses these frequencies as shown below.

890 – 915MHz mobile to base station (up-link)
935 – 960MHz base station to mobile (down-link).

Each channel has a bandwidth of 200 kHz and as a result a maximum of 125 channels can exist in each sub-band. Digital Cellular System also known as GSM-1800 makes use of frequencies around 1800MHz, and a couple of networks including one-2-one in the UK use these frequencies.

1.3 GSM Features and Services

Technology is not the first thing that consumers think about when purchasing wireless hardware. Instead, they care about applications, ease of use and cost of available services. GSM has been able to fulfil these customer demands thus explaining its widespread success. In accordance with the ISDN model, GSM networks or GSM PLMNs provide both Bearer Services and Tele Services. These services are identical in concept to the ISDN definitions. These services are not concerned with the Terminal Equipment/Mobile Station (TE/MS) and cover only the lower three layers of the OSI model i.e. the Physical layer, the Link layer and the Network layer. On the other hand, the Tele Services incorporate the MS and are thus concerned with the applications running on the TE. Hence, all the seven layers of the OSI model are of interest here. This difference is illustrated in figure 1.3 below

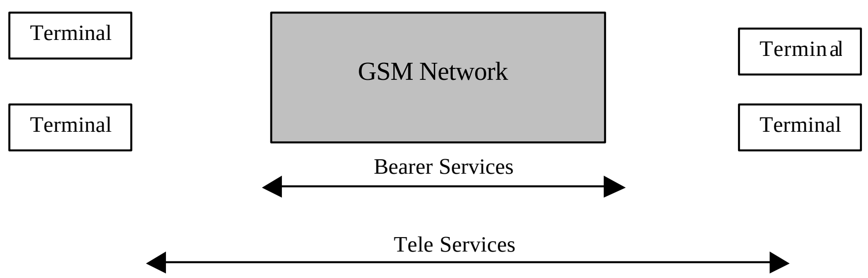


Figure 1.3 Diagram identifying the difference between Bearer Services and Tele services.

1.3.1 Bearer services

The GSM Bearer services cover a range of asynchronous and synchronous data options, up to 9.6kbps for inter-working with the PSTN, ISDN, as well as other circuit- and packet-switched data networks. These services are not concerned with the MS and cover only the lower three layers of the OSI model. Bearer services are always established by a call SETUP message in either direction, i.e. mobile originated or mobile terminated depending on the subsequent flow of traffic. This SETUP message contains bearer capacity information to ensure that the system allocates the correct resources to create a bearer suitable for the service. The message varies between 3 and 11 octets depending on the service required. For speech, only 3 octets are required but all 11octets carry useful information for some other data services.

1.3.2 Tele Services

Tele Services incorporate the MS and are thus concerned with the applications running on the TE. Hence all the seven layers of the OSI model are of interest here. These services are highlighted below.

- Speech
  - Mobile to Mobile Call
  - Mobile to Group Call
  - Mobile Broadcast Call
- Short Message Service, Cell Broadcast
- Data Services
  - Circuit Switched up to 9.6k (14.4k)
  - High Speed Circuit Switched up to 9.6k (14.4k)
  - Packet Data (GPRS)
- Facsimile

In common with other networks, the range of Tele services defined for, and offered by GSM are likely to grow as user needs grow and diversify. This is already apparent and evidenced by the deployment of new network technologies like GPRS and UMTS discussed in chapter 2 and 3 respectively. Emergency calls are defined as a separate service even though they are essentially speech. Short Message Service and Cell Broadcast Service are both described in Appendix 1A.

Apart from the Bearer and Tele Services, there is a host of other services called Supplementary Services, which although not essential for basic communication purposes, enhance the Bearer and Tele Services.

1.3.3 GSM Supplementary Services

These include

Number Identification	Calling Line ID Presentation (CLIP)
Calling Line ID Restriction (CLIR)	Connected Line ID Presentation (COLP)
Call Offering	Call Forward Unconditional (CFU)
Call Forward on No Reply (CFNRy)	Call Forward on Busy (CFB)
Call Completion	Call Waiting
Call Hold	Multi-Party Call
Communication Of Interest	Closed User Groups

Charging e.g. Advice of Charge, Advice of Charge Charging
Call Restrictions E.g. Bar all Outgoing Calls, Bar all outgoing International Call

Most of the basic supplementary services mentioned above are self-explanatory and some of them exist even in the terrestrial fixed-line network. A short description of some of these services can be found in Appendix 1A.

1.3.4 Advanced Supplementary Services

- Call Deflection
- User-User Signalling
- Enhanced Multi-Level Precedence & Pre-emption
- Explicit Call Transfer
- Completion of Calls to Busy Subscriber
- Unstructured Supplementary Service Data

More information on Basic and Advanced supplementary services can be found in [23]

GPRS, a GSM data service, will be discussed in further detail in Chapter 2.

As mentioned above, GSM supports services to other networks like the PSTN and ISDN and therefore need a number of Inter-Working Functions, (IWF). An example of this is the transcoder, (discussed in the next section), which converts GSM 13kbps format to the standard 64kbps PCM used in the Network Switching System of GSM and other speech oriented networks like the PSTN. Speech across the air interface in GSM is coded at a bit rate of just 13kbps for bandwidth conservation reasons. Another IWF could be the Modem to support data transmission between a GSM subscriber and a PSTN subscriber. For a GSM to GSM data transmission, no modem is required, as the data remains digital end to end.

The next section describes the GSM architecture and the functional elements, which lays a foundation that leads us gradually to where and how the charging records are produced and how these records are used to provide a final invoice to the customer.

1.4 GSM Network Architecture and Functional elements

The GSM network comprises basically of three subsystems, the Mobile Station (MS), the Base Station Subsystem (BSS), and the Network Switching Subsystem (NSS), each with a distinct set of responsibilities and network elements. These subsystems work together to operate as a mobile network, a PLMN.

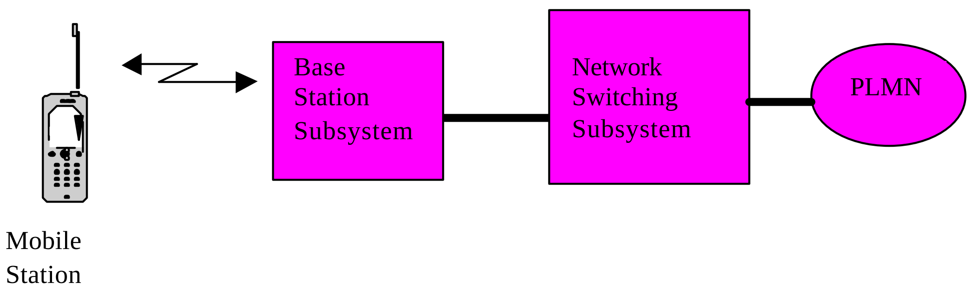


Figure1.4 the GSM Network Architecture

1.4.1 The Mobile Station (MS)

It constitutes the **Mobile Equipment (ME) and the SIM card** and offers a large variety of Tele services as described in section 1.3. This is achieved by the combined functions of the software, the radio hardware and signal processing. The Mobile Equipment is identified by a unique International Mobile Equipment Identity (IMEI) but is not related to a specific subscriber. This IMEI allows the ME to be verified by the Equipment Identification Register (EIR), a functional element of the NSS. The ME without a valid SIM card can make only Emergency calls. The SIM card is specific to a particular subscriber and has an International Mobile Subscriber Identity (IMSI) and the Mobile Station ISDN number, which is the number used to contact it. The main function of the SIM in conjunction with the network is to authenticate the MS.

1.4.2 The Base Station Subsystem (BSS)

This forms the users' interface with the fixed network (the radio interface on the one side and the switching network on the other). The main function of this subsystem in a nutshell is to provide radio coverage across the network. The Base Station Subsystem includes the **transcoder, the Base Station Controller (BSC) and the Base Transceiver System (BTS),** all linked together and connected to the Mobile Switching Centre (MSC).

1.4.2.1 BTS

The BTS is the Radio Base Station providing radio carriers and mapping logical channels to physical channels as well as coding and ciphering the channels. BTSs can be configured in a variety of ways. The main factor influencing the choice of configuration is the relative geographical location of the BTSs and the BSC as well as the cost of transmission between these points.

1.4.2.2 BSC

The BSC controls a number of base stations and handles signalling traffic, and operations and maintenance messages to and from all BTSs under its control. The BSC engages in some switching functionality too. It provides the switching link between terrestrial circuits on the MSC side and radio circuits on the BTS side. Although there is a terrestrial line between the BSC and the BTS, it can be considered "hard wired" to a particular logical traffic channel on the air interface. It also handles intra- and inter-BTS handovers as BTS has no switching function.

1.4.2.3 Transcoder and Rate Adaption Unit

The function of the Transcoder and Rate Adaption Unit (TRAU) has been discussed earlier. That is, conversion from 13kbps to 64kbps voice circuits. It is concerned with traffic channels only. It is logically part of the BSS but is often remote and co-located with the MSC for transmission costs purposes.

1.4.3 The Network Switching Subsystem (NSS)

This comprises of the switches and databases that enable mobility management, subscriber location, interconnecting of speech and data calls etc. The NSS is the most important part of the GSM network with regards to charging records. It consists of the MSC and its associated databases, interfaces and Inter-Working Functions. Operations and Maintenance Centres (OMCs) and transcoders are often co-located with the MSC but are not considered part of the NSS. The main components of the NSS are

- Mobile Switching Centre (MSC)
- Home Location Register (HLR)
- Visitor Location Register (VLR)
- Authentication Centre (AUC)
- Equipment Identity Register (EIR)
- Inter Working Function (IWF)
- Echo Cancelling Device (ECD)
- Short Message Service Centre (SMSC)
- Group Call Register (GCR)

1.4.3.1 MSC

The MSC is responsible for call processing of data and voice communications, data gathering for billing purposes, mobility management functions and supplementary services.

1.4.3.2 HLR

There may be more than one physical HLR but there is only one logical HLR per PLMN. It contains all subscription details for all subscribers registered on the network.

1.4.3.3 VLR

The VLR contains similar information to the HLR, but on a temporary basis for every mobile active in a particular VLR area. There is one VLR per MSC in each network.

1.4.3.4 EIR

The EIR is, just like the IMEI not concerned with the subscriber. It is only concerned with the IMEI. There is only one EIR per PLMN and it provides a 3-category list of IMEIs.

White List	-	Valid IMEIs
Black List	-	Stolen mobiles or IMEIs
Grey List	-	Faulty IMEIs or ME, which do not meet the GSM specifications.

1.4.3.5 AUC

The AUC is responsible for the authentication process.

1.4.3.6 GCR

The GCR is a functional unit containing all the attributes needed for set-up and handling of voice group and broadcast calls.

1.4.3.7 SMSC

The SMSC is simply an information handling and management centre.

1.4.3.8 IWF

The IWF provides the functionality for a GSM terminal to connect to a range of external public and private data networks. This is mostly rate adaption or protocol adaption.

1.4.3.9 ECD

In GSM, the round trip delay of the echo in the side tone is significant (up to 180ms) due to channel coding and the equalisation processes for the air interface. The ECD aims to remove this echo. It is usually placed on the PSTN side of the MSC.

1.4.3.10 OMC

The OMCs (many in a network) are responsible for alarm management, fault management, security management and configuration management. They are split into OMC-S, which looks after MSC functions and OMC-R, which looks after BSC/BTS functions.

A more detailed diagram of the components of the GSM network is depicted below.

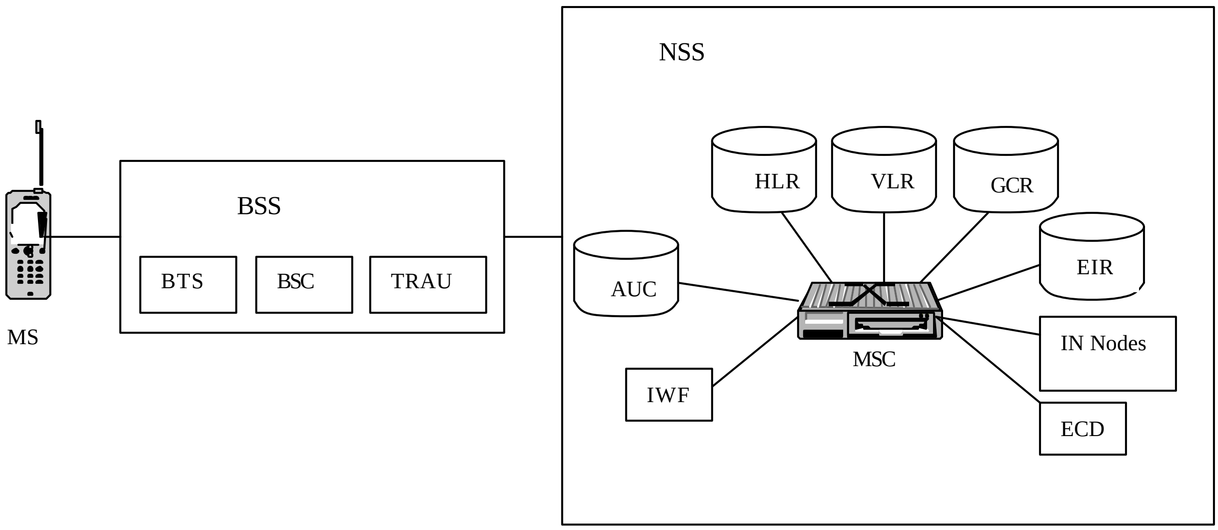


Figure 1.5 More detailed diagram of GSM Network Architecture

1.5 Pre-paid Service (PPS)

This section aims to give a very brief introduction and general overview to the fastest growing GSM account type, particularly among the younger GSM user base, the pre-paid service. Basically, it functions in a very similar manner to the post-paid subscription service, initially introduced for GSM with very few differences. It has however been incorporated into this report due to its vast popularity and slightly different method of charging, - the project's main subject.

According to statistics performed in the U.K, about 40% of the entire GSM user base has a prepaid account despite the fact that it was only introduced into the market about 3 years ago as opposed to the 9 years of GSM existence. It is interesting to note that PPS adoption varies across different areas. For example, in Italy PPS is very popular while the concept is not really considered in Japan. A PPS subscriber can make and receive voice calls, send Short Messages (using SMS) and even in some countries utilise the fax and data services specified for GSM, as long as no call limitations apply. Call limitations include account expiry, i.e. no service credit or service fee period and/or low account value (low calling credit). When a PPS account has expired, it can either be refilled or terminated. A PPS account is also capable of most GSM supplementary services. However, some of these services are not offered to PPS subscribers, most notably international roaming. Most network operators have chosen to ignore the possibility of having in place a roaming agreement with Foreign Network Operators. This may however change in the near future and today, a U.K one-to-one PPS subscriber is allowed international roaming.

A PPS account can be in one of the following service scenarios depending on the account data: account value, Service Fee Period and Supervision Period together with the settings of its Account Service Class.

- Active Service
- Passive Service
- Full Service
- No Service

Regardless of service scenarios, emergency calls can always be made. Passive Service or No Service is most likely due to low account value or Service Fee Period expiry. Hence, only calls to certain toll-free numbers are possible. The operator has the right to manually change the service scenario an account finds itself in.

1.5.1 Benefits of Pre-paid

The benefits of this variant have proven itself by its huge success. The network operator, the service provider and the end user benefit from the pre-paid account type.

1.5.1.1 Network Operator's benefit

PPS increases airtime and hence increases revenue. It also creates PR and goodwill. PPS also ensures better use of network resources, as most PPS users tend not to use their phones at business peak hours. Hence, the network need not be expanded in proportion to the increase of airtime. PPS also improves route to market as it enables operators to slash post processing costs of market, sales, link sales, airtime sales etc compared to post paid.

1.5.1.2 Service Providers benefits

PPS means low financial risk, as the Service Providers never have to fear unpaid bills, as the cost of the call is deducted in real time and an ongoing call is disconnected if the account reaches zero. Increase in airtime also favours the service provider. PPS also ensures better cash management as the call is paid in advance.

1.5.1.3 End user's benefits

The cost control functionality gives financial security to the subscriber. PPS can be shared within the family or it may be used for commercial renting. It's easy to use and makes easy the process of "entering the world of mobile services". Slow and awkward credit processes and subscriber loading are all eliminated.

1.5.2 Network Structure

The pre-paid GSM network infrastructure is slightly more complicated than the ordinary subscription network infrastructure. Apart from the GSM network nodes discussed in section 1.4, the PPS account requires the functionality of an Intelligent Network (IN) subsystem comprising of a variety of nodes with different functions. The reader can find more details on the IN subsystem, its functions and requirements in Appendix 1C.

1.6 Call Scenarios

This section describes very briefly two of the possible call scenarios that exist in GSM. These include

- Mobile Originated call (MO)
  - MO to Other Licensed Operator (OLO)
  - MO to Mobile Terminated

- Mobile Terminated call (MT)
  - OLO to MT
  - MO to MT

Each of these calls may involve an Intra MSC handover, an Inter MSC handover, Roaming Call Forwarding etc, and may terminate on either the Gateway MSC (GMSC) or the Visited MSC (VMSC). To be able to make/receive calls, the MS runs through idle mode procedures, which involve signalling to the VMSC and the transfer of subscriber profile information from the HLR to the VLR serving the VMSC. The details of these signalling messages are not covered here.

Both MO and MT calls are briefly described below. The associated CDRs and CDR Parameters are discussed in section 1.7 below.

1.6.1 Mobile Originated call

When the mobile has completed the idle mode procedures of registering with the network, the HLR will have passed the subscriber profile to the VMSC's VLR. The mobile can then receive calls and also make Mobile Originated calls. When the call request is received from the mobile, the VMSC checks the subscriber profile in the VLR (to ensure the subscription allows such calls to be made) and then the VMSC will route the call to the destination in the same way as a normal switch routes a call. The destination may either be another Mobile Station or a fixed line telephone. The latter case will require the call to first be routed to a GMSC before handing over the routing to the OLO switches. Note that OLO may either be PSTN or other Mobile Networks – i.e. other PLMN. In the course of a call of this nature, several signalling messages are transferred between the MS and various components of the network. These messages are transferred using a host of various transport protocols. The figure below illustrates the outline procedures of the relevant signalling sequences involved.

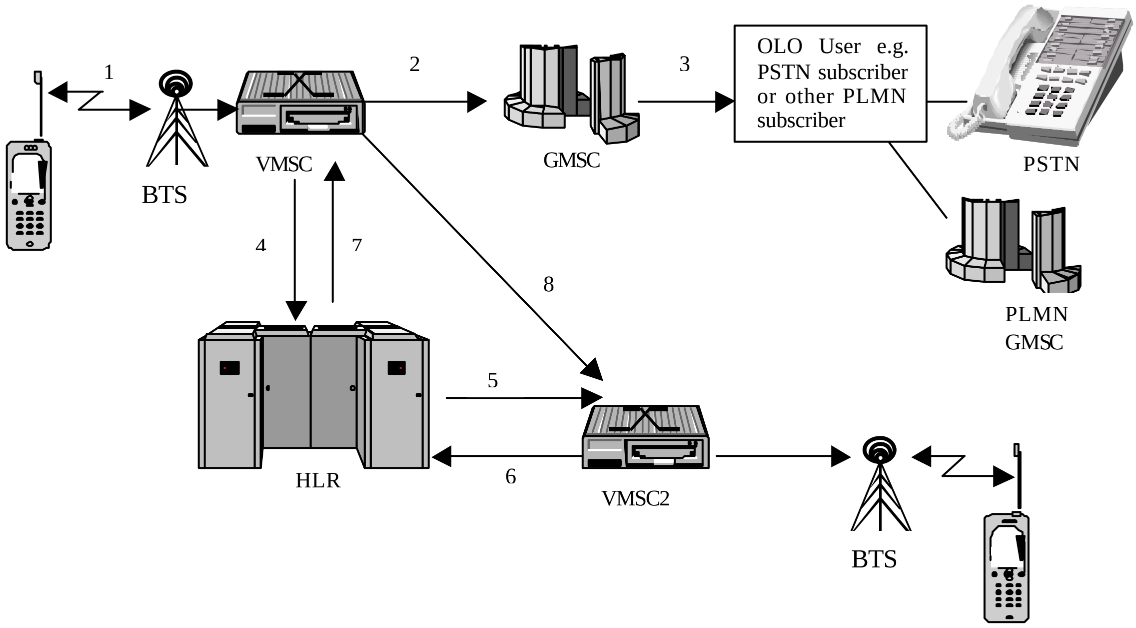


Figure 1.6 Signalling sequences involved in Mobile Originated Calls

The signalling attributes passed are

1. The caller dials the called number, which is either a fixed line number or a mobile number
2. The VMSC routes the call to the GMSC if the called number is an OLO subscriber, which could be a PSTN subscriber or a mobile user belonging to another PLMN.
3. The GMSC hands over the signalling and routing procedures to OLO switches.

If the called number is a mobile subscriber within the same PLMN, then

4. The MSISDN is converted to international format (e.g. +447881977473) to request routing information from the HLR.
5. The HLR looks up the IMSI and requests a roaming number from VMSC2 (The IMSI is used to identify the mobile in the VMSC)
6. VMSC2 returns an MSRN (MSRN stands for Mobile Station Routing Number and is used by the switches to route calls. IT is a routable number used to pass the call to the VMSC)
7. The MSRN is returned to the GMSC
8. The GMSC/VMSC uses the MSRN to route the call to the VMSC

1.6.2 Mobile Terminated Call

The calling party within the OLO dials your mobile phone number, (MSISDN) and since this is a mobile network, you could be anywhere in the world, i.e. the dialled numbers have no geographical significance. The originating network looks up the National Destination Code (NDC) part of the MSISDN and routes the call to a GMSC of the appropriate home network. The GSM network then checks up the Mobile Subscriber Identity Number (MSIN) part of the MSISDN and looks up the HLR to find the address of the MSC on which the MS is registered on. The HLR requests a roaming number (Mobile Station Roaming Number: MSRN) from the VMSC that will allow the Gateway Switch to route the call to the VMSC. When the call arrives at the VMSC, the MSRN also points to a database entry in the serving VLR. The traffic connections have now been made to the destination MSC – the VMSC. All the VMSC needs to do to complete the connection is to identify the location of the subscriber, which is available at the VLR. The MSC generates the paging message to the appropriate BSC. This paging message contains the subscriber’s IMSI. It is a point of note that when Mobile Terminated calls are forwarded from the VMSC (such as call forwarding on busy (CFB)), the VMSC treats the forwarding leg of the call in a similar manner to a mobile originated call.

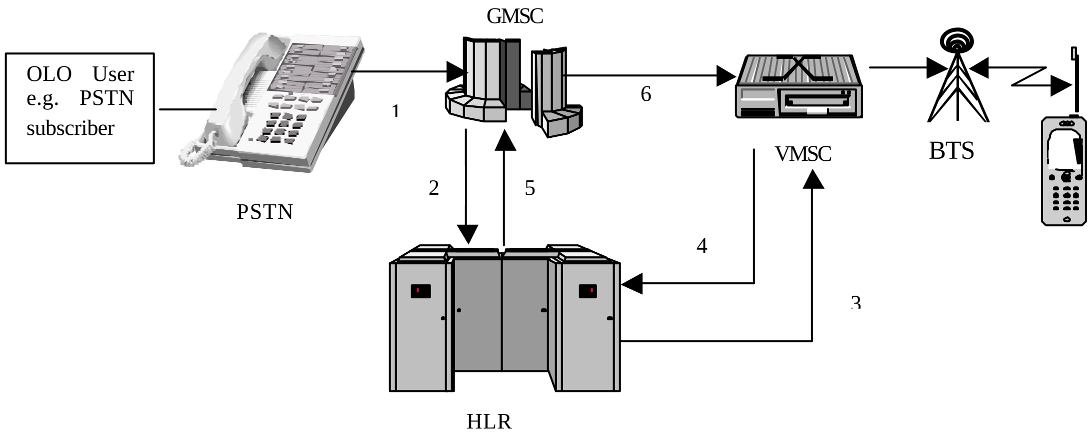


Figure 1.7 Signalling sequences involved in Mobile Terminated Calls

The signalling attributes passed are

1. The caller dials the mobile number (MSISDN) 07881977473.
2. The MSISDN is converted to international format (e.g. +447881977473) to request routing information from the HLR.
3. The HLR looks up the IMSI and requests a roaming number from the VMSC (The IMSI is used to identify the mobile in the VMSC)
4. The VMSC returns an MSRN (A routable number used to pass the call to the VMSC)
5. The MSRN is returned to the GMSC
6. The GMSC uses the MSRN to route the call to the VMSC

The figure above depicts a scenario where the call is terminated on the VMSC. Here, two traffic circuits are held between the ingress from the OLO and the VMSC, one to route the call to the GMSC (for initial interrogation of the HLR), the other to route from the GMSC to the VMSC. The call could also be terminated on the GMSC, if the Mobile Station is currently located in the Routing Area under the control of that GMSC. This simplifies matters as the number of traffic circuits as well as call records is reduced. That is the GMSC could also be the VMSC.

1.7 GSM Charging Architecture

We now have a strong enough foundation to understand which functional elements are used to obtain billing information. This section discusses the relevant platforms and functional units within the NSS related to charging and their role within the network. It also includes what information in the form of Call Detail Records (CDR) is recorded at each of these nodes. Most GSM network operators have an Automated Billing System (ABS), which produces both customer and Service Provider (SP) wholesale charging invoices. There may be more than one logical Billing System per operator. These could be BS for wholesale invoice, which produce invoices sent off to Service Providers – The Link, Carphone Warehouse, Phone4u etc, BS for Interconnect charges, which provide invoices for Other Licensed Operators, BS for roaming mobiles, which provide invoices for Foreign Network Operators etc. Each of these invoices is subsequently sent off to the SPs who apply appropriate tariffing before eventually sending off personal invoices to individual customers. The charging information is derived from call records generated in switches (MSCs) across the network.

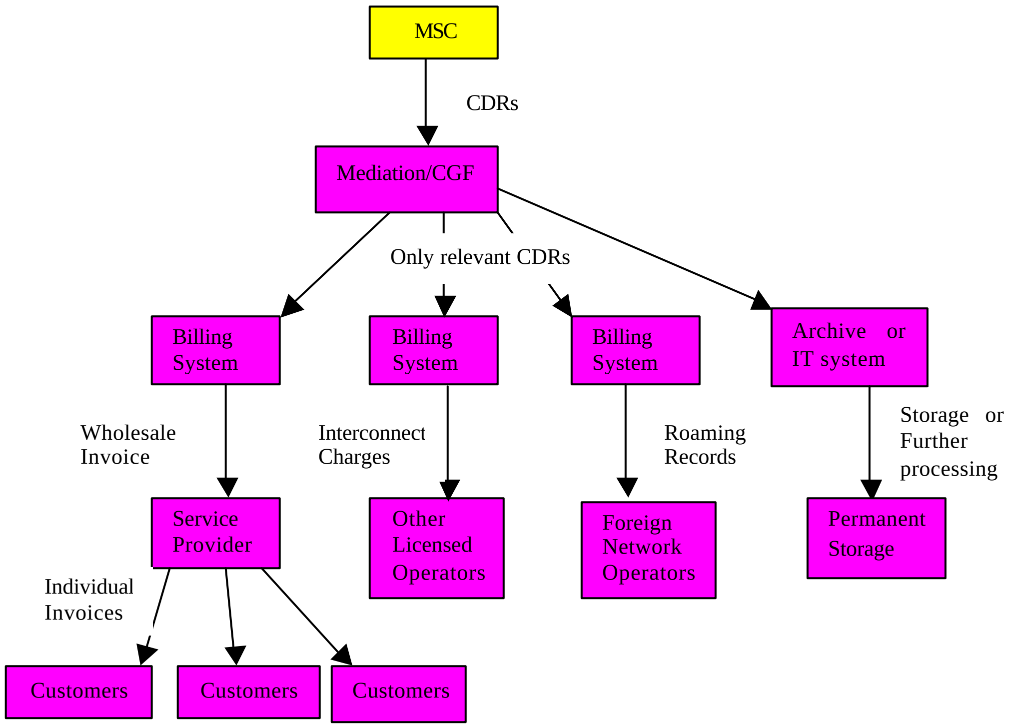


Figure 1.8 Flow of CDRs and invoices within a GSM network

There are several Call Record Types generated at the MSC. These records are produced for every call type and invocation of supplementary services. For certain calls, it is not unusual for more than one record type to be produced, although just one of these records is sufficient for billing purposes. It should also be noted that the call records are produced at one MSC (the anchor MSC) irrespective of whether or not the mobile is in motion during the call. Below is a brief outline of which records are produced and where these records are produced for the call scenarios described in the previous section.

1.7.1 Mobile Originated Calls

For Mobile Originated calls, an MO CDR is produced at the VMSC as illustrated below.

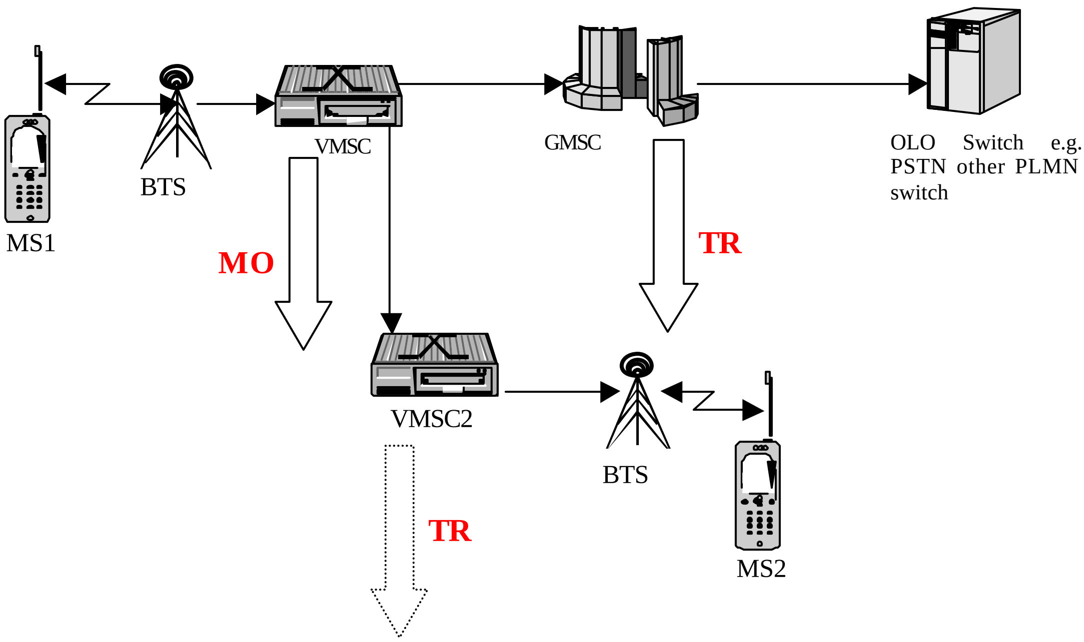


Figure 1.9 Call Records generated for Mobile Originated Calls

If the mobile subscriber is in motion, the call is handed off to the Base Station responsible for the Routing Area in which the Mobile Station is currently located. This handover may be intra- or inter-MSC. It should however be noted that the MSC, which originally handled the call (i.e. the MSC where the call was originated – in this case, VMSC) is still responsible for Call Control and is known as the 'anchor MSC'. In the case illustrated in the above diagram, even if MS1 roams so that VMSC2 is responsible for its channel and time slot allocation, VMSC will still have control over its MO record generation. Hence, no MO records are produced in the secondary MSC (VMSC2 in this case), only a TR record is produced. As soon as the call is terminated, VMSC2 informs VMSC via signalling that the call has ended, and the MO CDR is released.

A Transit Record (TR CDR) is also generated at the GMSC, if the VMSC had to route the call to the GMSC for transfer to the OLO switches. The TR record is however not used for billing purposes, and is not required either as all routing and billing information can be captured from the various parameters contained in the MO CDR. These MO parameters can be found in Appendix 1B.

If the called party is a mobile subscriber within the same network, then an MT record is generated at the MSC (VMSC2 in this case) currently handling the called party's Mobile Station. (MS2). This is described in more detail in Mobile Terminated calls in the next subsection.

1.7.2 Mobile Terminated Calls

For Mobile Terminated calls, a Transit Record (TR CDR) and a Roaming Call Forwarding Record (RCF record) is generated at the GMSC, while the VMSC currently controlling the MS produces both a Transit record (TR) and a Mobile Terminated record (MT). In the case, where the call originates and terminates on the same MSC, (i.e. GMSC), then all three records – TR, RCF, and MT – are generated by the GMSC, as shown in the table below.

MT Call terminates on GMSC	MT Call terminates on VMSC	
GMSC	**GMSC**	**VMSC**
TR	TR	TR
RCF	RCF	MT
MT		

Table 1.1 Call Record types generated for Mobile Terminated Calls

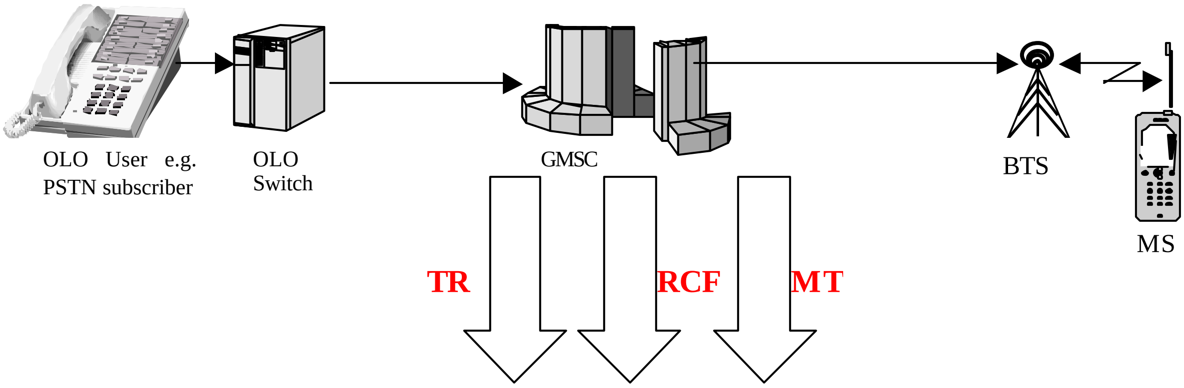


Figure 1.10 Records generated when a Mobile Terminated call terminates on the GMSC

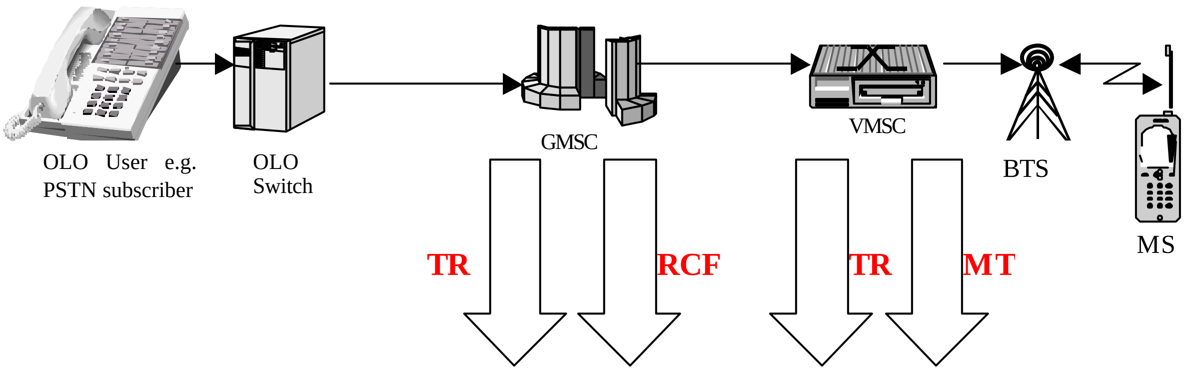


Figure 1.11 Records generated when a Mobile Terminated call terminates on the VMSC

For routing purposes, we cannot from the MT record alone, infer which GMSC the call originated from if the call terminated on the VMSC. Likewise, if an RCF record exists on the GMSC, with no associated MT record, we know the call terminated elsewhere (on some VMSC), and end to end routing information can be obtained from the collation of all the records generated. However, for billing purposes, the called party is not billed (assuming the called party is a European Operator subscriber) and the MT record is therefore never needed for billing and accounting purposes.

If the called party has activated their Call Forwarding function, then the call is forwarded to either their voice bank or the number to which the called party has predefined. In both cases, a Call Forwarding record (CF) is generated at the VMSC. In the case where the call is forwarded to the voice bank, the called party is not charged for this. They are only charged if and when they decide to retrieve their voicemails, and this is treated in the same manner as an MO call. If the call was forwarded elsewhere however, then the called party is charged for the last leg of the call and the CF record generated at the VMSC is used to determine the relevant information required for billing - the call duration among other things.

Each of the Record Types mentioned above contains a lot of information, some of the information is more relevant than others. They are all recorded however to be able to produce a detailed account of the call event should the need arise. A list of Call Record Types (also referred to as Call Detail Records - CDRs), generated at the MSC and required by most network operators is given in Appendix 1B.

1.7.3 Call Data Record Parameters

Each Call Record Type or Call Detail Record comprises of several parameters. All the MT Record Type parameters are listed in Appendix 1B. Refer to [25] for the description and function of each of these parameters. A few of these parameters are briefly described below.

The *Calling Party Number* is used to identify the traffic stream to the VMSC as well as the traffic profile between a call entering a GMSC (from an OLO) to it terminating on another MSC.

The *Internal Cause and Location* parameter along with the *Call Position* parameter are used to determine any call failures. However, none of the MT parameters can be used to determine "Not Routed Calls" (i.e. calls to powered off mobiles). Only Transit Records (TR) are produced to calls directed to mobiles that are powered off.

The *Incoming- and Outgoing Route* simply tell us which MSC the call is coming from and which one it is going to, respectively. These can be used for end-to-end routing by collating all these for each MSC through which the call passed.

The *Tele- and Bearer Service Code* give information on which service is being provided e.g. voice group call, Circuit data, GPRS or facsimile etc.

The *Related Call Number* parameter is used to identify the end-to-end routing within the network.

The *Date and Time for Start of Charge* alongside the *Chargeable Duration* are the basis for determining the **duration** of a particular call, which is the most important parameter for the end customer as this determines the amount he will be charged by his network operator.

From these parameters, the service provider determines the tariffs based on whether the call was made during peak hours or not, the network to which the terminated subscriber belongs to and so on. These parameters are recorded for all Record Types except the MOSM (Mobile Originating Short Message) and the MTSM (Mobile Terminating Short Message) as these are charged for, not on the basis of duration but on the basis of number of messages sent/received. The Date and Time for Start of Charge are however still recorded but there is no notion of duration.

1.7.4 Pre-paid Service Charging

GSM charging was originally based on a post-paid subscription billing process described above. Customers received bills based upon CDRs generated by the network. Due to the huge demand for pre-pay services however, a pre-paid charging mechanism was introduced. The charging mechanism used in the pre-paid domain is slightly different from the charging mechanism used for ordinary subscription services (post-paid). It is still based on the **duration** of the call, taking into consideration parameters like time of day/day of week, destination of the call, network to which terminating node is attached to etc. The only difference is that the charging is done on a **real time** basis. This is of importance as GPRS/UMTS (although, Packet-Switched) are intended to support PPS as well. UMTS will also support real-time applications, further complicating real time charging. GPRS, which will be discussed in the next chapter, is however not capable of supporting real-time applications.

The PPS account balance is a monetary value, which can be increased/refilled by purchasing a voucher, containing an activation code. The *charging interval* predefined by the service provider is the shortest charged time-period of a tariff. Assume it's 10 seconds and the tariff is 30p/min, then a call lasting for 10 seconds will cost 5p, while an 11seconds call will cost 10p. Due to the real-time nature of PPS charging, there are some service features, which do not exist in post-paid contract accounts. These are discussed below.

1.7.4.1 Supervision of Charged calls.

All PPS calls are supervised in real time. Hence a PPS account can never be overdrawn. If the account value is too low to make a call, the call will never be established. If a PPS account runs out of money during a call, the call will be disconnected by the call cut-off procedure, which usually starts off with a warning announcement, usually a tone. After a configurable amount of time, the call is disconnected. Certain toll-free numbers can however still be dialled, depending on the service scenario of the subscriber.

1.7.4.2 USSD Based Account Balance on Demand

This allows the PPS subscriber to get the account balance presented on the display of his mobile phone. USSD is just another mechanism to transfer data between the Mobile Station and the Network. It is similar in concept to SMS discussed in Appendix 1A. First, a USSD message, (account balance inquiry) is sent to the service code, which replies. Alternatively, the cost of an event and the new balance is displayed after every charged event. This USSD message is displayed in the language and currency indicated by the subscriber data and Account Service Class data respectively. An optional feature could be an SMS warning on Service fee expiry, supervision expiry or low account value.

1.7.4.3 Home Based Charging

This enables the PPS subscriber to define a home-zone for differential call charge depending on the area. Calls made within the home-zone will be cheaper just as local calls are cheaper in the fixed line PSTN. A change of location during the call does not affect the charging of that particular call. This feature can be implemented in post-paid accounts as well but has not gained much popularity due to its high cost.

Due to the fact that a caller may be barred from making a particular call, as well as other factors, announcements are important in the PPS domain. There are call set-up announcements, call cut-off announcements already discussed and In-call announcements. These will not be dis cussed any further here.

1.7.4.4 Real time charging – PPS Account Deduction

Calls are charged in real time due to certain call or non-call related events. The account is debited stepwise for every call and even during the call. The first deduction is of a size defined by a service class parameter. The deduction is done in advance. Money that has been deducted from an account but not used is put back after the call is finished. The SDP, SCP and SSF (see Appendix 1C for details) all interact to make the deduction and refunding possible. On the other hand, if the PPS account has not got enough value to continue the call, the SDP feeds the SCP with this information and the SCP orders the SSF to play the call cut-off announcement. Even though, charging is done real time, CDRs are generated for statistical purposes and call history functions. Similarly, for SMS, the message goes from the MS to the MSC, and then to the SMS-IWMSC via the **SDP.** The SDP is involved to ensure that enough credit is available on the PPS account. For post-paid subscribers, the message bypasses the SDP.

An example

A user makes a call to a fixed line telephone. The tariff for such a call is 10p/minute. Assume the first deduction is 20p, which gives the user 2 minutes. Towards the end of his allocated time, another 20p is deducted resulting in a further 2 minutes airtime allowance. If the call is ended after 3 minutes, then 10 p (the amount left over from the last deduction) is added on to the current balance after the end of the call. This is analogous to using a fixed line payphone.

1.8 Summary of the GSM charging scheme

The GSM subscriber is allocated a traffic channel for the call. He and no one else throughout the duration of this call is allowed to use this particular traffic channel. This is familiar from the PSTN, which also allocates dedicated circuits to the user. These circuits are tied down to the subscriber irrespective of whether or not any information is being passed between the end points. This technology is often referred to as “Circuit-Switched”. Since the subscriber “owns” the circuit for the entire duration of his call, the **call duration** is the primary parameter with respect to the charging mechanism employed within GSM or indeed any Circuit-Switched network service. In other words, the subscriber is essentially paying for the link dedicated to him for his call.

It should be noted that the GSM subscriber is only charged if they initiated the call (apart from the American Mobile telephony market and assuming that Mobile Terminated calls have not been forwarded). This is based on the assumption that they are located in their home network (HPLMN) at the time. If they are located in a foreign network (VPLMN, then they are charged both for making and receiving calls, i.e. MO and MT calls. For the MT call scenario described in section 1.6.2, the PSTN user is charged for making a call to the GSM subscriber, as though they were located in their HPLMN. This is fair, as the PSTN user may be unaware of the mobile users current location, and should therefore not be charged by the Foreign Network Operator (FNO). The GSM subscriber pays the charge incurred due to their location. This charge is paid to the FNO for using their network resources to enable the traffic circuit connection. The GSM subscriber however does not pay directly to the FNO. Their Service Provider receives an invoice from the FNO, who in turn includes this in their monthly invoice as described in section 1.7.

A GSM voice call is charged based on the **duration** of the call, taking into consideration parameters like

- Time of day/day of week
- Location of the caller
- Destination of the call
- Network to which terminating node is attached to

SMS, a fundamental feature of GSM services, is however charged on an **event-basis**.

Chapter 2

2 GPRS

This chapter is dedicated to giving an overview of GPRS and identifying the key differences between GPRS and GSM standards and technologies, i.e. the transition from Circuit-Switched to Packet-Switched transmission technology. It also describes the services, which GPRS intends to support and the functions, benefits and limitations of the new network components, (i.e. the newly introduced nodes and platforms – SGSN and GGSN) required to support these services. GPRS systems are only just evolving in Europe and to date, only very few functional commercial GPRS systems have been installed. The User Equipment (UE) or Mobile Equipment exists but in limited number and are still under development by various manufacturers.

Mobile operators Europe-wide are looking for ways to best handle current Circuit-Switched traffic (including High Speed Circuit Switched Data – HSCSD) and new packet data traffic. This system should preferably protect operators' existing investments in network infrastructure and maintain the reliability and quality of current mobile services as well as enable new, reliable and affordable services. GPRS has proven to be the answer. The reasons for this are discussed in sections 2.1 and 2.2 below. GPRS can be seen as the evolution path to the third generation UMTS. Its most important role is to fill the 'gap' between GSM Circuit-Switched data & UMTS. It is crucial in providing a smooth transition from current standards to 3G. Without this transitional step, the move from GSM to UMTS would be costly and extremely complex. By first deploying GPRS, GSM operators will have in place a 3G-capable network, making the transition to UMTS comparatively cheap and pain-free. GPRS is seen as a stepping stone to UMTS due to the following reasons.

- UMTS uses Packet-Switched technology, and will be developed on the packet data networks already installed for GPRS;
- At launch, UMTS coverage is expected to be available in metropolitan areas and the main transport routes in between, rather than providing full national coverage. Outside these areas, customers will continue to use GPRS coverage and services using dual mode UMTS/GPRS terminals;

2.1 GPRS overview

Most mobile networks, including GSM were developed to primarily cater for those users wishing to make and receive voice calls whilst on the move. However, being N-ISDN compatible, GSM was also designed to cater for data transmission as discussed in chapter one of this report. The number of subscribers making use of this data transmission facility to access their e-mails, browse the World Wide Web etc has constantly increased during the last couple of years, and is anticipated to continue in this fashion from roughly 3 – 5% today to an estimated 15% by 2003. [15]. These data services provided by the GSM network are perceived by many of its users as slow, unreliable and expensive when compared to other media used to access the same services using a fixed terrestrial network. Hence, the development of a new mobile radio service called General Packet Radio Service (GPRS), which is intended to improve the speed, reliability, and cost of these data services. The main factors, which have inhibited the market success of non-voice applications, include poor user interface and the lack of "killer" applications. Increased data rates supported by GPRS will enable realistic and useful applications, while technologies such as WAP will provide good user interfaces to the Internet and other services. Currently, GPRS only provides best effort service but there is a growing demand for the support of end-to-end QoS and of real time applications/services such as VoIP and VoD.

GPRS is one of the Data Services defined for the GSM network and as such is essentially an extension of the GSM network. Hence, it is often not perceived as a mobile network system in its own right. However, it relies on an entirely different technology, - "Packet-Switched" technology for the delivery of packets to the intended recipient. This technology does not allocate any particular circuit to the subscriber nor does it require any implicit Call Set-up or Teardown as does GSM or PSTN. There are a couple of other differences between these two underlying technologies, which are discussed in section

2.2. GPRS is set to trigger an explosion in mobile data services, data speeds of up to 171Kbps are attainable in theory. These data rates are however never achieved in practice yet, but may be achievable as GPRS develops. These speeds are even better than what Internet connection via a modem at home can offer. This explains the interest that network operators worldwide have shown in this technology. Such high data rates are required as users demand to access multimedia services rich in audio and video even when they leave their desks. Just to give the reader an idea of the kind of data rates required for multimedia services; if a raw TV picture was to be transmitted to a mobile terminal, the data speed will need to be increased to 155Mbps.

EDGE another system aimed at increasing wireless data rates, developed mainly by Ericsson is also being considered by some mobile operators. EDGE stands for **E**nhanced **D**ata rates for **G**SM **E**volution and is also based on Packet-Switched technology. It makes use of an enhanced radio modulation and channel coding techniques and reuses as much of the physical layer as possible to enable user data rates of up to 384Kbps or even higher in good radio environments. Its concept has been considered for both GPRS (EGPRS) and Circuit-Switched Data (ECSD). EDGE offers an evolutionary path from GPRS to UMTS and provides backwards and forwards compatibility with GSM and UMTS respectively. However, for the purpose of this project, GPRS was chosen as the Packet-Switched technology system to be studied due to its huge popularity among Network Operators. GPRS has 4 different coding schemes as shown

Coding Scheme	Code	Raw Data rate (Kbps)
CS1	½	9.05
CS2	2/3	13.4
CS3	¾	15.6
CS4	1	21.4

Table 2.1Various GPRS coding schemes with their corresponding data rates.

CS1 gives lowest data rate, but most error protection while CS4 gives the highest data rates but no error correction. As a result, the faster coding schemes cannot be used to reach handsets at the periphery of the cell, where corruption and loss due to poor signal-to-noise ratios are commonplace. The Mobile Stations will support all coding schemes but the network will probably only support CS1 and CS2 for launch. The initial handsets are likely to support 1 Up-link timeslot and 2 to 4 Down-link timeslots giving a maximum one-way peak user rate of 4 x CS2 rate = 53.6kbps. Hence, users are likely to experience effective throughput figures of 30 – 40Kbps, comparable to fast PSTN/ISDN modems on a dial-up land line network. The absolute maximum of 171kbps comes from 8 timeslots x 21.4 (CS4).

GPRS has been recognised as the stepping stone in the evolution path to 3G systems. It will provide the backbone for handling data traffic in a W-CDMA environment. Some operators (in particular those who do not expect to apply for or receive 3G licenses) will however consider GPRS as a long-term solution to the predicted explosion in demand for wireless data at speeds comparable to fixed line data rates. GPRS is still currently being standardised by SMG2 within GSM phase 2+ framework. Most mature GSM operators planned to launch GPRS during the latter half of 2000. However, due to the ongoing uncertainty surrounding the availability of GPRS terminals, it seems unlikely that there will be a high take-up of GPRS services among these pioneer operators if at all they are able to launch the service as planned. Nevertheless, besides the benefits of packet data services discussed in section 2.3.3 below, there are a couple of other factors indicating that the take-off of wireless packet data services will happen in the near future. These include

- Current mobile penetration and usage
- Current Internet usage
- SMS penetration and usage
- Usage of business critical applications like E-mail, file transfer, Remote LAN Access
- Laptop and Palm-top penetration and sales forecast etc.

One may ask; "How is the small user interface of the mobile handset going to be dealt with?" The most probable solution will be to link the GPRS-enabled mobile phone to a PDA via a data card or a data suite. The Bluetooth[1] concept is also like to b deployed. Refer to http://www.bluetooth.com for more information. However, people who have mobile phones and laptops/palmtops, but do not want to tie up their mobile phones when transmitting/receiving data may use a Cardphone to provide mobile data functionality without using a mobile phone. For consumer users without a PDA, the GPRS handset may be used on its own. GPRS intends to support an 'always on, always connected' functionality.

As has already been clarified, GPRS is one of the data services specified for GSM. It is a mobile IP-based Packet Data Service designed to operate over the existing GSM infrastructure, supporting interconnectivity with other external Packet Data Networks (PDN). It shares the existing GSM Base Station Subsystem and makes very efficient use of GSM radio interface, but makes use of a separate core network. As a result, it is economical but also future-proof due to its open standards and interfaces. It can therefore be perceived as a stepping stone towards UMTS, described in chapter 3. The efficient use of the GSM radio interface is as a result of users sharing the network resources for the support of 'bursty' type applications. The actual number of users supported depends on the application being used and how much data is being transferred. GPRS offers Packet-Switched services and as such does not require a dedicated path for the duration of the call. It however still works with existing GSM services (Circuit-Switched voice & data as well as SMS). It supports a true end-to-end IP packet data transport, with several protocols running in the layers below. The protocol stack used for communication across the various interfaces from the external PDN to the User Equipment (UE) is outside the realms of this project. Since GPRS is IP-based, it is fully compatible with the Internet and truly provides the capability to extend the Internet to the mobile phone. The GPRS network can be viewed as a sub-network of the Internet with GPRS capable handsets being viewed as mobile hosts.

The implementation of a fully functional and commercially viable GPRS network is however faced by technological and commercial challenges. Some of these challenges or issues concerned with GPRS include

- Inability of initial GPRS services to support Mobile Terminated calls because we need a dedicated IP address to terminate calls. This issue is discussed briefly in section "Address space crisis".
- Terminal issues. – How is a session managed when a call is terminated?
- Security concerns (Spam/Viruses) – Unwanted/unsolicited data, pushing information to GPRS/UMTS handsets could result in virus threat.
- Billing issues - How do we charge for packet-based services and who is billed for Mobile Terminated traffic.
- "Always on" functionality raises issues such as battery life, cost, monitoring of paging channels

Security concerns are currently not viewed as being a major challenge and standards work is currently being done concerning the billing issues. These billing issues are also discussed in chapter 4.

2.2 Migration from Circuit-Switched to Packet-Switched Networks

A reasonable question to ask at this stage is "Why do we need Packet-Switched Data Services when we already have a Circuit-Switched Data Service in existence within GSM?" Current and future telecommunications services such as videoconferencing, remote database access and multimedia file-transfer, require a flexible network with the availability of virtually unlimited bandwidth – they demand dynamic variable bandwidth allocation. The Circuit-Switched approach means that a circuit (with a statically reserved bandwidth) is allocated to the user, for the duration of the call irrespective of whether or not any information is being passed between the end points. This is suitable if the information being transmitted is fairly continuous as in traditional voice telephony. Most data transmission is however inherently discontinuous, with long periods of little or no data transmission. Consequently, packet switching is more appropriate for data transmission. No dedicated path is allocated and hence, no network resources are used if no packets are in transmission, resulting in a

[1] Bluetooth wireless technology is a low-cost, low-power, short-range radio link for mobile devices and for WAN/LAN access points. It offers fast and reliable digital transmissions of both voice and data over the globally available 2.4GHz ISM (Industrial, Scientific and Medical) band.

cost-effective use of network resources thereby driving the cost for operators down. This results in a low-cost service for the end customer compared to the dial-up GSM. Another significant disadvantage of the Circuit-Switched technology over the Packet-Switched technology when transmitting data is in the event of a break in connection at any time, during the transmission. In the Circuit-Switched approach, all data may be lost and we may need to retransmit the entire data, which is expensive and frustrating for the customer. The Packet-Switched approach on the other hand only re-transmits those packets, which have been lost. GPRS is intended to address most of these issues, i.e. speed, expense and reliability of transmitting data over the GSM network.

As no dedicated path is allocated, a dynamic rather than fixed bandwidth is available. With the introduction of GPRS, mobile data services are no longer limited to 9.6kbps as in today's GSM or 14.4kbps in best cases when High Speed Circuit Switched Data Service (HSCSD) is used. HSCSD was however not widely employed, as it required a new mobile handset or a special radio chip if used by a laptop or PDA, without offering the speed offered by GPRS. Also, it is generally more expensive for the customer.

Packet-based transmission technologies easily support "always-on always-connected" functionality by having a constant virtual connection and setting up a real connection only when there is data to be sent/received and then dropping the connection until more data is ready to be transmitted or received. The key difference between this system and a Circuit-Switched connection is that network resources are used only when data is being sent/received. As such, simply logging in at the start of the day may constantly connect the GPRS user, hence avoiding repeatedly dialling up. This is affordable, as the mobile user is not charged for the duration of his constant "virtual connection". They are only charged either for the volume of data carried or based on some other scheme discussed in chapter 4.

Below is a highlight of the key differences between the switching/routing technologies deployed in GSM and GPRS respectively.

Item	Circuit-Switched	Packet-Switched
Dedicated 'copper/radio' path	Yes	No
Bandwidth available	Fixed	Dynamic
Potentially wasted bandwidth	Yes	No
Store-and-forward transmission	No	Yes
Information follows the same route	Yes	No
Call Set-up	Required	Not Required
When can congestion occur	At set-up time	On every packet
Charging	Per minute (time)	Per packet or other means

Table 2.2 Comparison between Circuit-Switched and Packet-Switched technologies

Currently, connectivity to an IP network requires the use of a Network Access Server (NAS), which is responsible for the conversion of data from Circuit-Switched to Packet-Switched. This is shown below in figure 2.4. In GPRS however, there is a special node called the Gateway Node which is essentially an IP router and this communicates with the rest of the network using IP or GTP which will be discussed further later in this chapter. Hence, there is no need for a NAS.

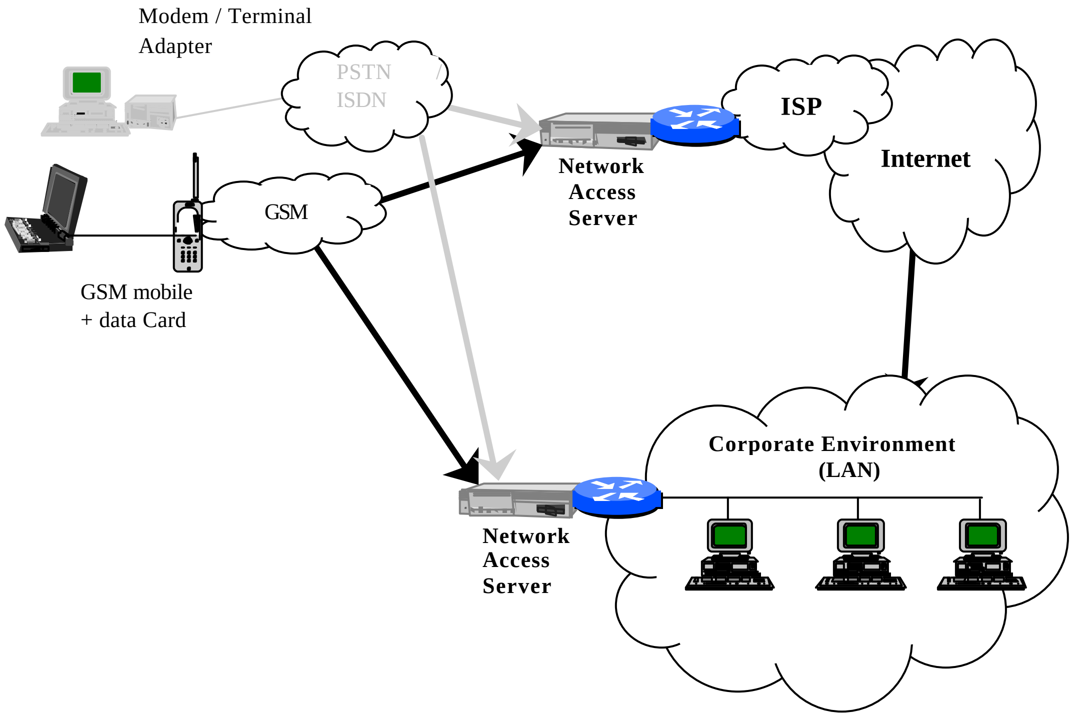


Fig 2.1 Current Connectivity Overview to an IP network

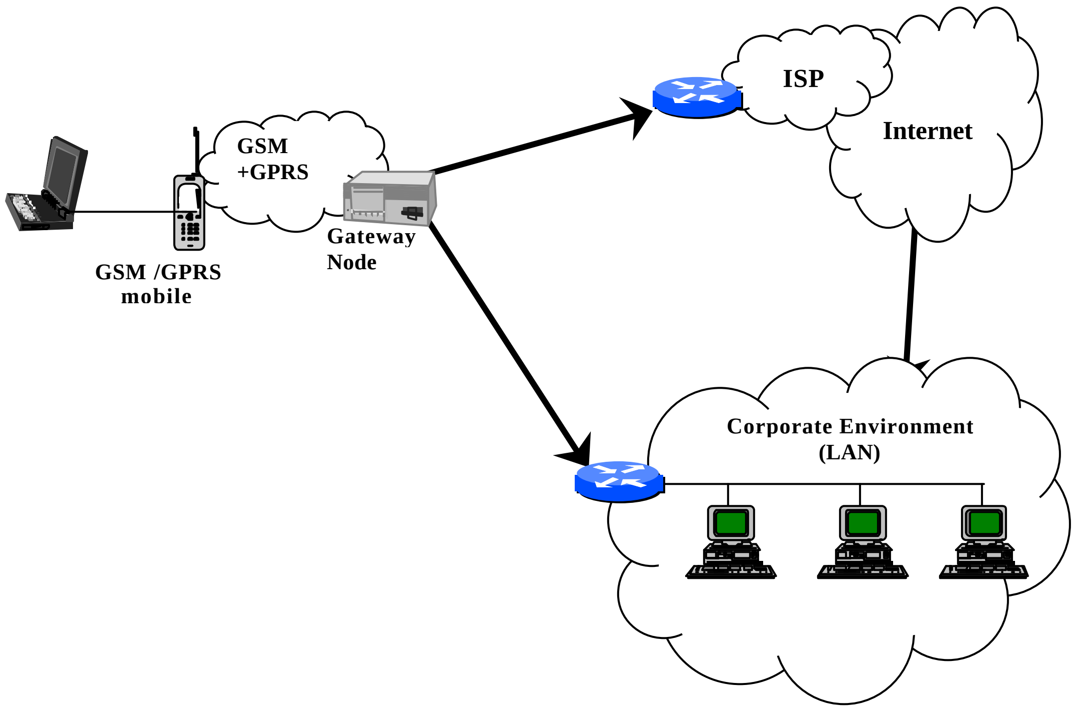


Fig 2.2 GPRS Connectivity Overview to an IP network via a Gateway Node

2.3 GPRS Features and Services, Benefits and Requirements

GPRS and ultimately UMTS will offer a reliable, low cost, high speed, wireless access to corporate networks, the Internet and a range of newly developed applications. This new data communications mechanism in conjunction with the variety of services and applications it provides, bring along a number of user benefits, but also a number of requirements. This section briefly describes the features, services, and benefits, but also briefly highlights the additional requirements, associated with GPRS.

2.3.1 Features and Services

Many potential users simply don't hold the perception of wireless access devices as an essential service. Moreover, mobile operators today face a world of fierce competition where the customer is the king. For this reason, network operators/service providers rapidly need to invent new, innovative and useful applications and services for its potential customer base. Such an application could be file transfer applications, which encompasses any form of downloading sizeable data across the mobile network. This data could be a presentation for a travelling salesperson, an appliance manual for a service engineer or a software package such as Adobe Acrobat Reader to view documents. GPRS provides remote/mobile access to corporate LANs – the "Central Nervous System" of an organisation and supports all of the usual office applications. GPRS will retain SMS and may use a data channel rather than a control channel to allow for traffic growth without compromising the signalling capability of the network. It is feared that SMS may be replaced by other message services as GPRS advances. One distinguishing factor between GPRS and GSM is that it supports an "always on, always connected" feature. It also makes use of data compression, which enhances user experience through faster downloads, web access, e-mail and access to shared drives etc. The potential applications of GPRS can be divided into a horizontal and a vertical market segment.

Horizontal markets

- Wireless Personal Computers
- Electronic Funds Transfer from Point of Sale
- Mobile Office
- Telemetry
- Entertainment

Vertical markets

- Road Transport Information
- Fleet Management - Vehicle Location Services
- Field Service Business – Goods/Supply logistics
- Messaging
- Access to the World Wide Web - E-commerce and booking services, Banking

Two target groups or key users have been identified and the table below depicts the main differences between these groups.

Factor	Key Users	
	Business	**Private/Consumer**
Key Applications	Pure data services: Corporate intranet, email, telemetry	Internet access, e-commerce, WAP–based services
Usage levels	High volume	Low volume
Handset availability	Limited from Q3 2000	Mass production from Q4 2000
Service costs	Same as current GSM data services	Lower than existing GSM data service, almost equivalent to current fixed line services

Table 2.3 User groups

Examples of day-to-day applications, which will become commonplace with the advent of GPRS, include:
Remote LAN Access, calendar and address book functions, e-mail access, file transfer, Lifestyle applications e.g. on-line gaming, chat, m-commerce and so on. By linking the Internet email with an alert mechanism such as SMS, users can be notified when a new email is received. This prevents out-of-office users having to dial in speculatively and periodically to check their mailbox contents. Making restaurant reservation through the Internet becomes a convenient function using a mobile handset due to the much faster set-up times and throughput. Audio files will move quickly via GPRS, hence, journalists will be able to broadcast interviews recorded on mobile phones rather than using tapes or electronic recorders, etc. However, GPRS is unable to support packetised speech, (Voice over IP - VoIP) or indeed any form of real time application.

2.3.2 Benefits

Each of the unique GPRS features mentioned above, bring along with it a couple of user benefits including

- Immediate access to the same processes and systems used by office-based colleagues anywhere, anytime, without any need to reconfigure systems or for users to learn new skills, thus leading to greater productivity and potentially reduced office space.
- Personalised services and Location-based services, Push advertisements etc.
- Real time information allows intelligent decision-making, improved scheduling and time management and ability to respond faster to customer requirements.
- Savings from increased professional efficiency

Other user benefits of the packet/IP nature of GPRS as opposed to GSM are:

- Access to a wider range of services at faster data rates
- The “always on, always connected” feature, enables the user to log-in just once a day with a virtual connection maintained until the user chooses to log-off, hence reducing time for call set-up.
- Increased network business – e-commerce, call centres

Operator benefits are mainly

- New and diverse potential sources of revenue as a result of the Packet-Switched nature of GPRS and the diverse services on offer.
- Reduced network costs - Unified Infrastructure and Management (for both data and voice).
- Possibly improved customer relationship by offering wider range of services and mobile data.
- Improved Network efficiency as more services and applications are offered

2.3.3 GPRS Requirements

The GPRS specifications do not set an upper limit to the data set that can be transmitted per access. Nevertheless, GPRS is required to support frequent transmission of small volumes (up to 500 bytes) and infrequent transmission of medium volumes (up to several Kbytes). It is also required to support Point-To-Point (PTP) Connectionless (CL), PTP Connection Oriented (CO) and Point-To-Multipoint (PTM) services. A PTP service could be non-dialogue traffic, where there is no relationship between the individual data packets or dialogue traffic, in which case a logical relationship exists. A PTM service in a similar fashion to the GSM PTM could either be a Multicast or a Group call.

A GPRS traffic channel is required to be able to occupy one or more TDMA time slots within a GSM carrier frame and it should be possible to increase or decrease the number of physical channels allocated to a GPRS subscriber on a dynamic basis. As in GSM, the logical channels used in GPRS are divided into

Control channels
Packet Common Control Channel
Packet Broadcast Control Channel

Traffic Channels
Packet Data Traffic Channel
Packet Associated Control Channel

Some of these channels are in turn subdivided into a number of channels. The Control channels are intended to carry signalling and synchronisation data between the base station and the mobile station while the traffic channels are used to carry encoded speech or user data.

Due to the Packet-Switched nature of GPRS, there are a number of updates and new network components required compared to Circuit-Switched GSM. These include

- New mobile handset
- Updated radio infrastructure capable of handling packet data.
- New GPRS nodes to handle traffic between the mobile handset and the network – known as Serving GPRS Support Node (SGSN)
- GPRS nodes to handle traffic between the network and the remote data networks – known as Gateway GPRS Support Node (GGSN)
- Connectivity to remote data networks
- Billing systems to charge for GPRS usage.

The SGSN and GGSN are important for the production of Call Detail Records and will be discussed in more detail in sections 2.4.2 below. The updated mobile handset is also discussed below.

2.4 GPRS Network Architecture and Functional Elements

The introduction of a new technology – Packet-Switched technology, poses the need for additional nodes, and network components, but also updates to nodes already existing in the GSM network. This section gives an overview of the GPRS network architecture, as well as describes the new network functional elements.

The GPRS network architecture is basically the same as the GSM network except that the SGSN and GGSN replace the MSC and GMSC respectively. Also, the GGSN is interfaced with an IP-based PDN as opposed to the PSTN in GSM. The GPRS logical network architecture with its several interfaces is shown in Appendix 2A. Refer to Standards document 3GPP TS 23.060 V3.5.0 (2000-10) "General Packet Radio Service (GPRS); Service description Stage 2 (Release 1999)" for further architecture details.

Each of the functional blocks in the GPRS architecture has already been encountered in the GSM network architecture except for the two new GSNs. The elements, which remain unaltered and have a similar function within GSM are not included in this section. A very brief description of the new and modified components of the GPRS network follows.

2.4.1 GPRS Mobile Terminal

The GPRS Mobile Terminal (MT) is usually used in conjunction with a Terminal Equipment (TE), which is either a laptop, a palmtop or some other form of Personal Digital Assistant (PDA). This is however not necessarily the case. The MT and TE are connected either by a cable, bluetooth or by infrared. The basic function of the MT is for communication over the radio interface. Most GPRS handsets are expected to have WAP (Wireless Application Protocol) capabilities to allow access to the Internet and to company Intranets direct from the handsets. Alternatively, a PDA, may be used on its own if equipped with a GPRS Radio Card, which would provide the user direct access to the GPRS service without using a mobile handset. Since GPRS is expected to support existing Circuit-Switched services as well as new Packet-Switched services, GPRS-enabled handsets are required to support a combination of these services and the handsets have been divided into three distinct classes all of which allow a combination of GSM and GPRS services. The terminal may or may not have an IP address allocated to it at the point of sale. If an address is allocated at the point of sale, the terminal is said to have a 'Static Address Registration', if no address is allocated, it is said to have a 'Dynamic Address Registration'. The difference will become more apparent later on in this chapter.

Class A

Can operate in both GSM and GPRS modes at the same time and can send/receive packet data simultaneously with a Circuit-Switched call (e.g. voice).

Class B

Can be attached to both the GSM and GPRS services at the same time, but can only be active in one mode at a time. That is, it supports GPRS and GSM registered sequential operation. The handset can 'listen' for calls on the other service. It automatically switches between GSM and GPRS, GSM Circuit-Switched calls i.e. voice have priority. The GPRS services are placed in suspend mode whilst the Circuit-Switched is used.

Class C

Can operate in GPRS or GSM modes and only be attached to one service at a time. It cannot 'listen' for calls on the other service. The user manually switches between services. A GPRS only mobile with no GSM service support is a subset of Class C. An SMS cell broadcast message can take place simultaneously with GPRS services but not with a Circuit-Switched connection.

The Mobile Station can be in one of three modes.

Idle Mode

In this mode, the GPRS MS is seen as not reachable. Neither Data transmission to/from the subscriber nor paging of the subscriber is possible. The subscriber is not attached to the GPRS mobility management. The MS and SGSN contexts hold no valid location or routing information for the subscriber.

Standby Mode

In STANDBY mode, the subscriber is attached to GPRS mobility management. Pages for data or signalling information transfers may be received. It is also possible to receive pages for the CS services via the SGSN. Data reception and transmission are however not possible in this mode.

Ready Mode

In READY mode, the SGSN MM context corresponds to the STANDBY MM context extended by location information for the subscriber on cell level The MS may send and receive PDP PDUs in this mode. The MS may activate or deactivate PDP contexts while in READY mode.

2.4.2 Serving GPRS Support Node (SGSN)

The SGSN is a fundamental component of the GPRS network infrastructure. It is analogous in many respects to the MSC in a GSM network. It is intended to handle traffic and signalling between the mobile handset and the network. It has a unique IP Address and has a list of GGSN Addresses to which it may establish a connection. It is normally co-located with the GGSN and has several functions including:

- Mobility Management
- State Control – Ready, Standby, Idle
- Session Management or Call Control
- Packet Data Unit routing (location tracking)
- Authentication
- Ciphering
- Compression
- GSM CS interactions – Voice calls, SMS
- Data/Packet counting
- Database functionality

The SGSN achieves many of the above functions by simply producing various call records and passing these records to the Calling Gateway Functionality (CGF) within the network. Data compression is performed for transmission efficiency over the radio network to optimise the use of the radio path capacity. However user data is normally NOT compressed between GSNs.

Mobility Management activity is monitored by a special call record known as Mobility Management – Call Detail Record (**M-CDR**), whereas Data/Packet counting as well as other factors related to the amount of data transmission are recorded in the SGSN - Call Detail Record (**S-CDR**). Records related to SMS services are produced in the **S-SMO-CDR** or **S-SMT-CDR**, which stand for SGSN delivered Short Message Mobile Originated – Call Detail Record and SGSN delivered Short Message Mobile Terminated –Call Detail Record respectively. Each of these records contains charging information and is discussed in section 4.4.

The SGSN also plays an important role in setting up a GPRS Attach as well as a PDP Context Activation, both of which are discussed in the “Connecting to the Network” section. The other SGSN functions are related to security issues and will not be discussed any further in this report.

2.4.3 Gateway GPRS Support Node (GGSN)

This is very similar to the SGSN except that it handles traffic between the mobile core network and the external remote Packet Data Network (PDN) and is analogous to the Gateway MSC in a GSM network. It is essentially an IP router and is itself assigned an IP address. As such, it maintains a routing table, which is updated periodically from the home DNS. The GGSN analyses and converts the PDP addresses to the IMSI of the respective MT. Decapsulation of data packets occurs here. It is a protocol converter and router, and as such, must have a PDP context for each activated subscriber. If the next node is at a different hierarchical level (e.g. GGSN to SGSN), then some protocol conversion must be done. If it is at the same level (e.g. GGSN to GGSN), no protocol conversion need be done. To protect the mobile subscriber from receiving (and paying for) unsolicited packets, it must have screening functionality for mobile terminated packets. Although, the GGSN has several duties in a similar manner to the SGSN, unlike the SGSN, it produces just one type of call record known as the GGSN – Call Detail Record (G-CDR).

Some of its important functions are listed below

- Transport layer routing protocol support
- IP (v4 &v6) + X.25
- PDU Tunnelling via IP GPRS backbone
- Screening
- Data/Packet Counting
- Address mapping, routing tables

It is also involved in completing the connection required for data transfer from the UE to the external PDN as described below. One other important function of the GGSN is fragmentation, which is basically preventing oversize packets from entering the GPRS network. When an IP packet arrives at the GGSN from an external PDN (e.g. Internet), it may be up to 65536 octets long (the maximum for IP). Different networks have different maximum packet sizes (Maximum Transmission Units – MTU), 128 octets for X.25 and over 8000 for some LANs. For GPRS, the MTU is set to 1500 octets. Therefore if a packet arrives at the GGSN from a sub-network which allows 4000 octet MTUs, it must be fragmented before the GGSN can pass it over the GPRS network. IP headers have a "D" flag to indicate 'don't fragment this packet'. If this is set AND the packet length is greater than 1500 octets the packet will be discarded or rejected depending on the configuration of the GGSN. The maximum packet size may increase as network technology advances.

2.4.4 Domain Name Server (DNS)

The DNS function may be incorporated into the SGSN, or it may be a separate entity. It is used to resolve logical GSN names to IP addresses, for routing across the GPRS IP backbone.

2.4.5 HLR

This incorporates GPRS subscription data and routing information, with the already existing GSM subscription data, keeping the same index key (IMSI). The HLR records need to be extended to add the GPRS subscription data. This will take up extra memory, and may therefore require a restructuring of the HLR memory. All GPRS related data is stored in the GPRS Register (GR) and is regarded as part of the GSM-HLR.

2.5 Connecting to the Network

The actual process of connecting to the network involves three main signalling procedures

I. GPRS Attach
II. PDP Context Activation
III. PDU Transfer

2.5.1 GPRS Attach

When the GPRS subscriber switches on their handset, it is initially in the 'Idle' mode. It then starts to listen to the broadcast channels to know which cell to camp onto. Once Cell camping is achieved, it then carries out a "GPRS attach" procedure. During the attach procedure, several signals containing important information are passed to and from the MS, just as in the case of GSM IMSI attach. The 'Attach Request' message contains the following:

- IMSI
- P-TMSI & RAI
- Classmark – Mobile terminal multislot capabilities
- Attach type – Indicates whether GPRS only or combined GPRS/IMSI attach

Some other security related parameters are also transmitted. These are however of no interest here. Usually, if there is a valid P-TMSI (Packet-Temporary Mobile Subscriber Identity) and RAI (Routing Area Identity), then the IMSI isn't included. When this Attach Request message successfully reaches the SGSN, which has little knowledge of the particular user, the HLR is consulted by the SGSN to obtain further information using the received P-TMSI or IMSI as identity. The HLR retrieves the GPRS relevant records, sends them to the SGSN where authentication and other security procedures are performed before a "GPRS Attach Accept" message is returned to the mobile station.

Once the GPRS attach is successful, the MS is now in the 'Ready' mode where it remains until the READY timer expires or it's forced to the 'Standby' mode. If in the 'Standby' mode, it needs to transmit a PDU to return to the 'Ready' mode. Finally, a GPRS detach message is transmitted to finally reach the 'idle' mode once again before the MS is switched off. The sequence of events, which affect the mode of the MS is depicted in the figure below.

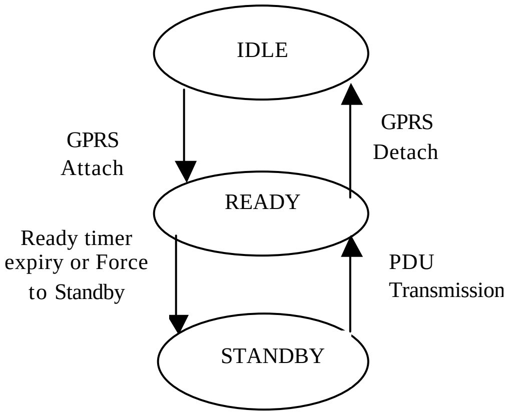


Figure 2.4 Sequence of events affecting the mode of the Mobile Terminal

2.5.2 PDP Context Activation

The MS has only just established a communication link to the SGSN once the "GPRS attach" has been successfully implemented. In order that data may flow to its destination - the PDN, there needs to be a communication path between the SGSN and the GGSN. This could have been accomplished at the 'Attach' stage but in order to make the most efficient use of network resources, it is generally preferred to establish a virtual link as and when necessary. This virtual link is established via the PDP Context Activation. Even though the communication link is between the GSNs (SGSN and GGSN), it is the duty of the MS to initiate its establishment by first transmitting a 'PDP Context Activation request' message to the SGSN. This message contains

- TLLI – Temporary Link Layer Identifier, which enables the SGSN to identify what GPRS services the MS is entitled to.
- PDP Type – Indicates whether the call is X.25 or IP.
- PDP Address – This is the address by which the MS is known when communicating with the PDN or the Internet. This address can either be an X.25 address or an IP Address. IP addresses however seem to be the favoured form. The PDP Address field may however be empty if dynamic allocation scheme is deployed. Otherwise the Mobile Stations' X.25- or IP address is inserted into this field if static allocation is used in which case the address was allocated to the MS at the point of sale.
- Quality of Service – This is basically the Quality of Service required by the user. It is often referred to as "QoS requested". This would be investigated in further detail in section 2.9.2.1 as it has an impact on the charging of GPRS services.
- APN – Access Point Name. This indicates which GGSN the SGSN should connect to, as there is more often than not more than one GGSN within a network. The actual mapping of APN to GGSN is done in a Domain Name Server, which the SGSN 'talks to'. The Address Resolution process would not be covered here.

The SGSN after receiving the 'PDP Context Activation Request' performs an authentication procedure and if successful, attaches a Tunnelling Identity (TID) as well as some other parameters to the original message before forwarding it to the appropriate GGSN. This TID would identify all messages originating or terminating at the MS. If dynamic IP address allocation is deployed, then the GGSN may request for an IP address from the external PDN or use one it has locally. The address allocation server could be DHCP (Dynamic Host Configuration Protocol) or RADIUS (Remote Access Dial-In User Service). It then sends a 'PDP Context Activation Response' message (including the MS IP Address) to the SGSN, which in turn forwards it to the MS. With dynamic addressing, it is possible to reuse the IP address. It is however more difficult to provide tailored "push services" to end-users, as these no longer have unique IP addresses. Push services are perceived as a potential source of revenue for many Mobile Network Operators and as such Static Address allocation seems to be the favoured addressing mechanism. Moreover, with the introduction of IPv6, there will be no shortage of IP addresses as discussed below.

IP Address Space Crisis and IPv6

Every node attached to the Internet has a specific address used for identification. This is its IP address, essential for packet delivery etc. As has been discussed, GPRS-enabled handsets will be connected to the Internet and as such require an IP address, which may be dynamic or static. IP addresses have been partitioned into classes to differentiate usage and this has led to the overall inefficient use of the entire address space. IP addresses are slowly running out especially as more assorted devices (including mobile handsets, TVs and VCRs in the future) are being connected to the Internet. As a result, the Internet Engineering Task Force (IETF) has accepted a proposal to upgrade the current IP (also known as IPv4) address space i.e. 32 bits to a 128-bit address space. This new version of IP addresses is known as IPv6, and a lot of work is currently being done to ensure interoperability between IPv4 and IPv6. Apart from increased number of IP addresses, IPv6 incorporates other advantages including improved security. Ipv6 will help to eliminate address space crisis and enable GPRS to grow with static address allocation without concern about the depletion of addressing resources. Refer to [30] for more information regarding IPv6.

2.5.3 Packet Transfer

Data transfer is classified either as Up-link data or Downlink data depending on the direction of flow.

2.5.3.1 Up-link Data Transfer

Packet Data exchange can now commence! Now that both GPRS Attach and PDP Context Activation have successfully been performed, resulting in a complete 'MS to PDN' communication link, the mobile station is either in the 'Ready' or 'Standby' mode, ready to transmit/receive data. Assume it is in the 'Standby' mode, it first sends a 'Packet Data Channel Request' message to the BSS, which assigns it an Up-link TBF. The data is then wrapped into an IP packet, with the standard format containing the data section and the packet header, which includes various fields like 'source address', 'destination address' etc. This packet is then sent to the SGSN, where a new packet is created having the source address of the SGSN and the destination address of the GGSN with the initial IP packet occupying the data field of the new packet as shown below.

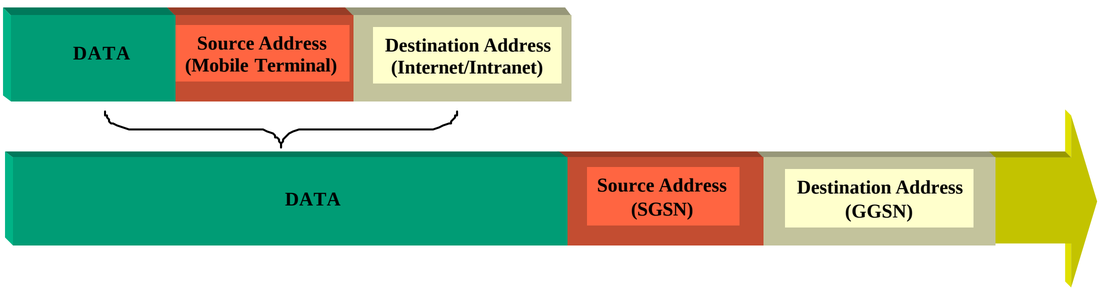


Figure 2.5 Diagram illustrating the Tunnelling procedure, how an IP packet is encapsulated into a GTP packet.

The GSNs communicate via a communication protocol called GPRS Tunnelling Protocol (GTP). The GTP PDU together with the TCP/UDP header form the TCP/UDP PDU, which in turn is combined with an IP header to form the IP Packet Data Unit (PDU).

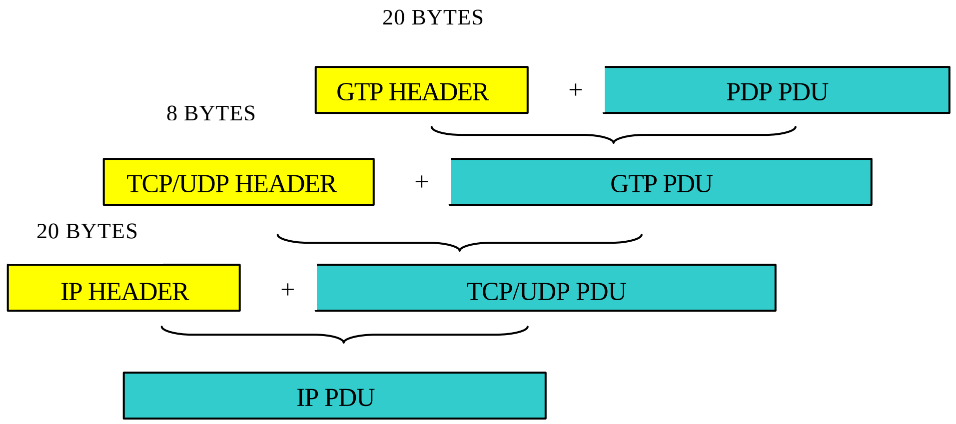


Figure 2.6 Diagram illustrating the various Tunnelling procedures, from an IP packet to a GTP packet.

The process of encapsulating the initial IP packet into a new one is known as tunnelling or encapsulation. At the GGSN, the source and destination address fields as well as other fields constituting the packet header are stripped off and the original packet is routed to the required destination – the PDN or Internet.

2.5.3.2 Down-link Data Transfer

This is basically the same as for up-link data, the difference being in the direction of data flow. Here, the packet arrives at the GGSN from the PDN. The details of how this is done have been left out as it involves a process known as sub-netting. The GGSN looks up the 'destination field' of the packet, compares with its table containing PDP Address/TID/SGSN entries to find out which SGSN to forward the packet to. Remember that these were recorded during PDP Context Activation. A new packet is then created in the same manner as above, with the destination address of the SGSN and the source address of the GGSN, with the data field containing the original message and the TID. This is illustrated in figure 2.7 below.

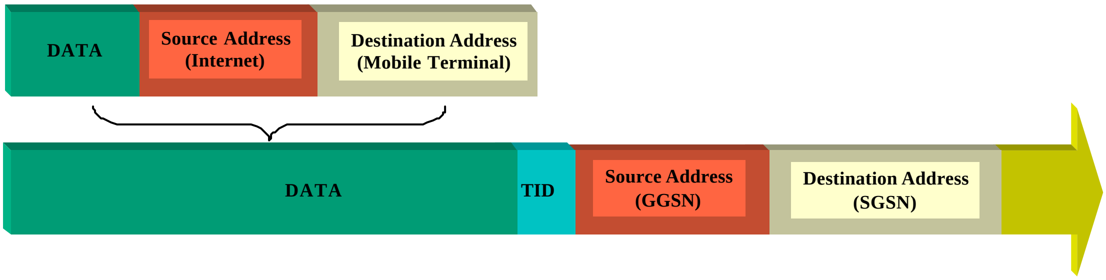


Figure 2.7 Tunnelling by GGSN. Includes TID.

At the SGSN, the packet header is stripped off and the original packet is forwarded to the right MS with the aid of the TID.

During data transfer (up-link or downlink), the SGSN to which the MS is connected to, may change due to mobility, which is managed by the SGSN. This is analogous to the change of MSC being used in GSM.

Data transfer discussed above was mainly between the SGSN and the GGSN. Data traffic between the BSC and the SGSN is carried via a Frame Relay link. One potential problem is that the Frame Relay device often has too few ports for all 2Mbps E1 frame relay links coming from the BSC, at a given site.

Another problem that may arise is that the SGSN may have too few Gb interface ports. However, if EDGE is introduced and integrated into the GPRS network infrastructure, it is likely to move the Gb interface to IP. Hence, it is essential to minimise costs of implementing a frame relay network.

2.6 Call Scenarios

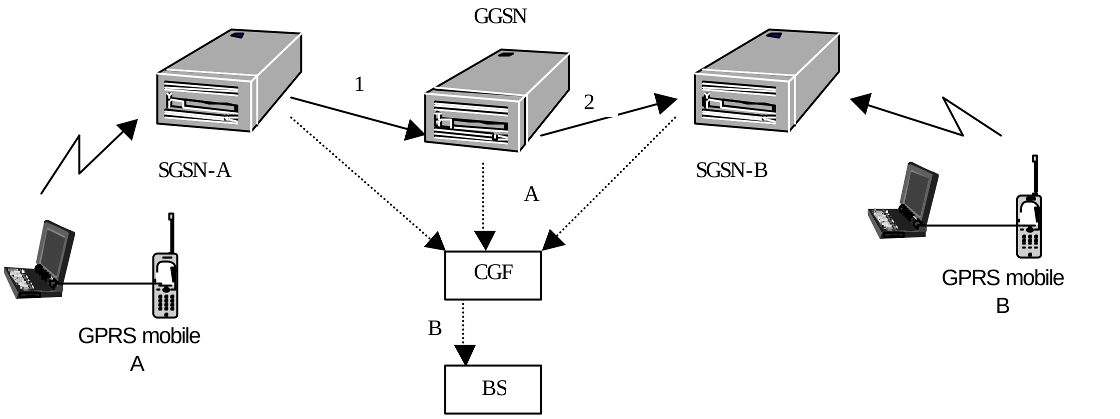


Figure 2.8 GPRS Mobile to Mobile Context

In the figure above, transfer of call records are represented by means of dotted lines and referenced by letter, while network connections and signalling transactions are represented by solid lines and referenced by a number. The A-party related PDP context is activated in the SGSN-A and the GGSN (1). After the location of subscriber B is determined, the B-party related PDP context is activated (2) in SGSN-B and GGSN. PDP PDUs are routed in MO and MT direction. SGSN-A and GGSN create an S-CDR and G-CDR respectively for subscriber A while SGSN-B and the GGSN create S-CDR and G-CDR for subscriber B. Subscriber A and B may or may not use the same GGSN. The records generated are subsequently transferred to the BS (B) via the CGF (A). We shall now have a look at the record contents of the S-CDR produced in the SGSN, but before we do that, a quick overview of the charging scenario of a “GPRS mobile to PDN context” is illustrated.

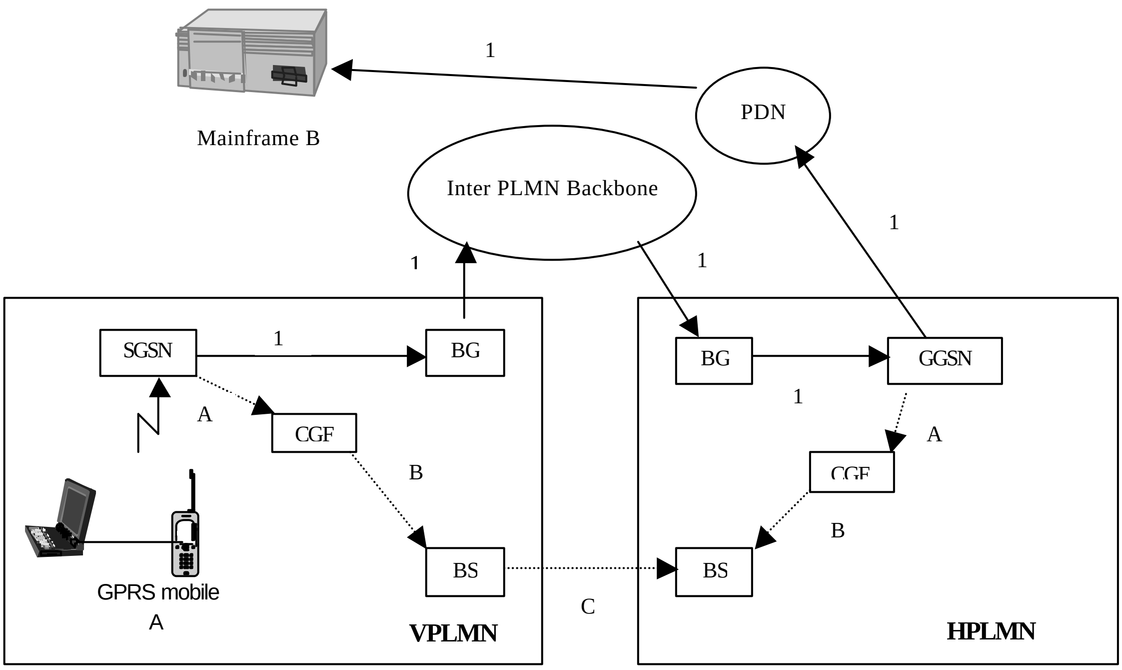


Figure 2.9 GPRS mobile to PDN context while roaming via BG

Figure 2.9 above illustrates an outgoing PDP context from a roaming GPRS mobile subscriber “A” to mainframe “B” via Border Gateway, Inter PLMN backbone and GGSN of the HPLMN. The SGSN creates an S-CDR (VPLMN) and the GGSN in the HPLMN creates the G-CDR for subscriber “A”. The packets are sent from the GGSN via the PDN to the mainframe ”B”. Records generated in the VPLMN are transferred from the BS to the BS of the HPLMN via TAP procedure (C). Record transfers (A) and (B) have already been discussed in the earlier example. TAP stands for Transfer Accounting Procedure and is used extensively by operators across Europe for transferring CDRs from VPLMN to HPLMN.

A typical calling scenario would be as follows

A laptop PC connected to a GPRS enabled handset, requiring connection to a specific host on a remote IP network, e.g. a corporate LAN (Packet Data Network), the BT intranet, say. The handset communicates (using an RLC communication protocol) with a local base station over the radio interface (the Um interface) in a similar manner to that of a GSM voice network. The call is then routed to an SGSN over the Gb interface in a similar (but not identical) manner to a GSM switch.

The SGSN identifies the destination of the connection by looking up an Access Point Name (APN). It is expected that each subscriber should be able to access up to 50 services via APNs provisioned in the HLR. The SGSN routes the connection to the relevant GGSN, which acts as a gateway to the requested service, and allocates a “charging ID” to the session such that it can be uniquely identified. Each GGSN can act as a gateway to a number of services. Now that the connection has been established, data packets are routed between the mobile handset and the remote host via the SGSN and GGSN. This connection is known as a Packet Data Protocol context (PDP context), of which there may be up to 14 established by a single subscriber at any time. The initial handsets are however designed to support just one PDP context but the capability to support multiple contexts will be possible as GPRS develops and the requirements of its users increase.

Chapter 3

3 UMTS

In this, chapter, we aim to give the reader an overview of UMTS, a brief summary of the features and services supported by UMTS and a summary of the UMTS architecture and functional elements. Section 3.1 includes the UMTS operating frequencies, its developments in terms of standards and networks, its intended dates of launch in various parts of the world, etc. It also highlights the main differences between UMTS and GSM/GPRS. Section 3.2 gives a brief summary of the features and services supported by UMTS. It then goes on to discuss the requirements of UMTS and the possibilities it opens up to its users as well as network operators, but also to businesses in a wide range of industries, including Content and Application providers, Banking and finance, the Internet and so on. Finally a section is devoted to the UMTS architecture and functional elements.

3.1 UMTS Overview

Second generation telecommunication systems, such as GSM, enabled voice traffic to go wireless. Their data handling capabilities are limited, however, and third generation systems are needed to provide the high bit rate services that enable high quality voice, data and video to be transmitted and received. Third generation systems are designed for multimedia communication**:** with them, both person-to-person communication and access to information and services on private and public networks are possible. The challenge for third generation mobile systems is to have a **wireless data** penetration as high as GSM voice penetration is today, i.e. as high as 70% in the most advanced countries. One of the third generation systems currently under development is the Universal Mobile Telecommunication System (UMTS). It integrates two of the world's fastest growing industries – mobile communications and the Internet. "It makes your data fly." The two technologies will converge in terms of services, functionality, applications, transmission technologies and pricing. With UMTS, the dream of mobile Internet, delivering the enormous global resources of the World Wide Web to users anywhere, anytime, can be realised. UMTS will not only provide a platform for handheld devices, it will be embedded in laptops, TVs, PDAs and radios. UMTS services and applications are anticipated to cause a significant change in usage and behaviour for many mobile subscribers – as different as PC usage is from telephone usage. It will create new business opportunities not only for manufacturers and operators, but also for providers of content and applications running over these networks. Interest in UMTS is not only confined to traditional telecommunications operators, but from all sorts of industries – banks, broadcasters, newspaper publishers, retailers and manufacturers.

The UMTS core network is similar in structure to the GPRS core network described in chapter 2. They are both IP-based and rely on the functions of routers (rather than switches as is the case for GSM) for the delivery of packets. UMTS will also make use of the GSM core network for the initial support of voice and other Circuit-Switched services, until a reliable and high quality VoIP service is provided. Hence, the UMTS core network will consist of both GSM and GPRS core-network elements, which is also a requirement for backward compatibility.[2] UMTS however, relies on a different radio access network technology from both GSM and GPRS, to ensure faster data rates, as the bottleneck lies in the air interface. Wideband Code Division Multiple Access (WCDMA) has emerged as the most widely adopted third generation air interface, and UMTS will make use of it. Although, faster data rates can be achieved within the existing GSM 900 and GSM 1800 bands using EDGE, WCDMA provides even faster rates and it will be possible to refarm WCDMA to the GSM bands. The World Administrative Radio Conference (WARC) of the ITU[3] at its meeting back in 1992 identified the frequencies around 2GHz as available for third generation mobile systems. Hence, WCDMA will be deployed in Europe and Asia around the 2GHz frequency band. The specific frequencies for UMTS are

[2] Backward compatibility requires that earlier releases of software and hardware are compatible with current software and hardware equipment. In this particular case, a UMTS handset should be able to make use of the GSM Circuit-Switched network.

[3] Within ITU, Third generation mobile systems are called International Mobile Telephony 2000 (IMT-2000). The UMTS Forum currently made up of 200 members and founded in 1996 is part of the ITU's IMT 2000 family.

- WCDMA FDD: 1920 – 1980 MHz (Uplink)
- WCDMA FDD: 2110 – 2170 MHz (Downlink)
- WCDMA TDD: 1900 –1920 MHz as well as 2020 –2025 MHz bands.

UMTS operates using 2 x 60 MHz as opposed to 2 x 25 MHz used for GSM operation. One significant difference however, is that, WCDMA has a carrier spacing of 5 MHz (resulting in just 12 carrier frequencies in each direction) as opposed to GSM, which has a carrier spacing of 200KHz (giving it 125 different carrier frequencies). The larger bandwidth of 5 MHz is needed to support higher bit rates and the fewer carrier frequencies is not a drawback on its capacity as each WCDMA user is allocated the entire 5MHz band. WCDMA employs efficient radio resource management algorithms to provide QoS.

Commercial UMTS networks are scheduled to open in Japan during 2001, and in Europe and elsewhere in Asia at the beginning of 2002. The expected schedule is presented in Figure 3.1.

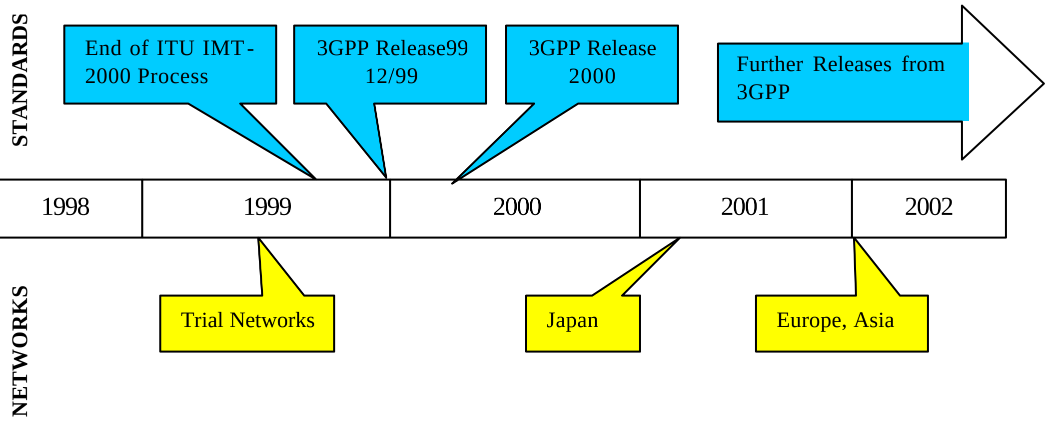


Figure 3.1 Standardisation and commercial operation schedule for UMTS

3.2 UMTS Features and Services, requirements and benefits

This section aims to introduce the various features and services supported by UMTS. The list is by no means exhaustive, but hopefully, it helps readers to understand the variety of different services and gives a flavour of what we will see in just a few years from now. We also discuss the requirements and benefits of UMTS in this section.

3.2.1 UMTS Features and Services

UMTS will offer a range of mobile services that have, until now, been available only on fixed networks. The best known new feature of UMTS is its high user throughputs or bit rates. On Circuit-Switched connections, 384Kbps will be theoretically possible while Packet-Switched connections will enable connections as fast as 2Mbps. These higher bit rates facilitate new services such as video telephony and Remote LAN Access and other Internet services and applications.

Because of different technologies and frequency allocations, **global roaming** will continue to require specific arrangements between operators, such as multi-mode and multi-band handset and roaming gateways between the different core networks. To the user, operator arrangement will not be visible, just as it is today with GSM.

UMTS deployments will start off with voice traffic dominating but **data traffic** is speculated to gradually take over, especially as high data rates (comparable to fixed line connection throughputs) are feasible. The transition from voice to data will also involve a transition from Circuit-Switched connections to Packet-Switched connections, and Packet-Switched connections will eventually handle all traffic including voice. However, not all the QoS functions will be implemented for initial UMTS deployments. As a result, delay-critical applications like voice and video telephony will still be carried on Circuit-Switched bearers.

UMTS allows a user/application to **negotiate bearer characteristics via QoS**. The set of parameters and parameter values should be selected so that negotiation and re-negotiation procedures are simple and unambiguous. More on UMTS QoS can be found in section 4.5.5.

The UMTS network will support all of the services supported by GSM and GPRS. In addition, as UMTS develops and advances, more diverse services and applications will be supported. The killer application is still under speculation but it is most likely going to be the quick access to information and the filtering of this information appropriate to the location of the user.

3.2.2 UMTS Requirements

Some of the new requirements of UMTS include

- Bit rates up to 2Mbps.
- Variable bit rates to offer bandwidth on demand.
- Multiplexing of services with different QoS requirements on a single connection (See 4.5.5.6)
- The UMTS QoS concept should be flexible enough to allow bearer negotiation with yet unknown future applications.
- Delay requirements from delay-sensitive real-time traffic to flexible best-effort packet data.
- Coexistence of GSM and UMTS and inter-system handovers for coverage enhancements and load balancing
- Coexistence of FDD and TDD modes

Since the UMTS core network is similar in architecture and function to the GPRS core network, much of the UMTS core network requirements have already been achieved. On the Radio Access Network on the other hand, WCDMA requires a significant change from GSM and GPRS and the interested may find more information on this subject in [10].

Other UMTS requirements may be found in 3GPP Technical specifications.

3.2.3 UMTS benefits

All of the benefits associated with the packet-based nature of GPRS apply equally to UMTS. These are

- Access to a **wider range** of services at **faster** data rates
- The "**always on, always connected**" feature, enables the user to log-in just once a day with a virtual connection maintained until the user chooses to log-off, hence reducing time for call set-up.
- Increased network business – e-commerce, call centres

UMTS provides even faster data access than GPRS does, and hence offers even greater benefits than GPRS offers. Due to its very efficient channel coding technology, it offers a very high user capacity.

It is also able to support **multiple PDP contexts**, which is not currently possible with GPRS. Also, it supports real-time applications like voice and video telephony, although initially, through the CS part of the core network, it is intended to support these applications via its PS core network. Other benefits include

- Immediate, anywhere, anytime, access to the same processes and systems used by office-based colleagues thus enabling intelligent decision-making, improved scheduling and time management and ability to respond faster to customer requirements.
- Personalised and Location-based services, as well as push advertisements

Operator benefits are

- New and diverse potential sources of revenue
- Reduced network costs - Unified Infrastructure and Management (for both data and voice).
- Improved customer relationship by offering wider range of services, e.g. mobile data.
- Improved Network efficiency as more services and applications are offered on one network.

3.3 UMTS Network Architecture and Functional Elements

This section is concerned with the System Architecture of UMTS. It also gives a brief description of some of the new functional elements defined for UMTS.

The UMTS system consists of a number of logical network elements that each has a defined functionality. All UMTS interfaces are open,[4] and are defined by standards bodies, 3GPP in particular. Functionally, the network elements are grouped into the UMTS Terrestrial Radio Access Network (UTRAN), and the Core Network (CN). The RAN/UTRAN handle radio-related functionality, while the CN is responsible for switching and routing calls and data connections to external networks. To complete the system, the User Equipment (UE) interfaces the user with the RAN/UTRAN. Both the UE and the UTRAN consist of completely new protocols, the design of which is based on the needs of the new WCDMA radio technology.

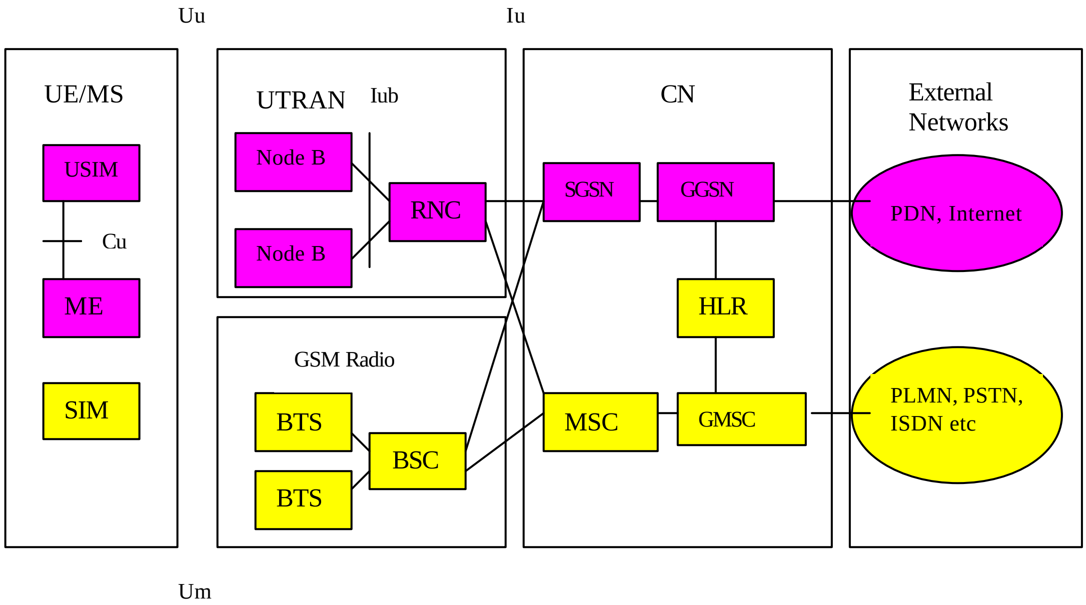


Figure 3.2 Network elements in a UMTS PLMN

3.3.1 UMTS Subscriber Identity Module (USIM)

This is similar in function to the GSM SIM, holding subscriber identity, performing authentication algorithms, storing authentication and encryption keys etc

[4] For an interface to be open, its requirement is that it has been defined to such a detailed level that the equipment at the endpoints can be from two different manufacturers. As a result, the internal functionality of the network elements does not need to be specified in detail.

3.3.2 Node B

This is the UMTS Base Station and is referred to as Node B by the 3GPP specifications. It converts the data flow between the Iub and the Uu interfaces. It also participates in radio resource management and all of the other functions of a Base Station described in chapter 1.

3.3.3 Radio Network Controller (RNC)

The RNC owns and controls the radio resources in its domain (the Node Bs connected to it). RNC is the services access point for all services UTRAN provides the CN, for example management of connections to the UE. One RNC connected to one or more Node Bs make up a Radio Network Subsystem (RNS), which together with other RNSs make up the UTRAN.

The **CN** also consists of Media Gateway and The Media Gateway Control Function as well as the Call Server Control Function.

3.3.4 Media Gateway and The Media Gateway Control Function (MGCF)

These are responsible for handling multimedia transactions between the UMTS Core Network and other external networks like the Internet. The MGCF takes care of the control signals and messages while the Media Gateway itself handles the user data – video, audio, and graphics.

3.3.5 Call Server Control Function (CSCF)

This is in effect **not** a new node. Its function is precisely that of the MSC in the GSM network, i.e. call control, mobility management, switching etc. It takes care of the CS calls made on the UMTS network.

The HLR has also been renamed to **Home Subscriber Server** (HSS).

All other UMTS Network elements have already been encountered either in the GSM network (sections 1.4) or the GPRS network (section 2.4) and have thus been omitted here.

The process of connecting to the UMTS network is similar to that described in section 2.5 for GPRS. The main differences will be in the processes involved in the Radio Access Network.

Chapter 4

4 Packet-Based Charging

The new data communications mechanism (Packet-Switched) in conjunction with the wide variety of data, voice, and video services, which GPRS and UMTS provide, opts for the requirements of novel charging mechanisms described in this chapter. This chapter first of all describes the GPRS/UMTS charging architecture, i.e. the flow of charging information in a GPRS/UMTS network, and then goes on to address the main issues concerned with Packet-Switched billing in GPRS and future broadband wireless systems including UMTS. It also describes how these packet-based services are expected to generate revenue, as we are no longer primarily concerned with the duration of the call, and can therefore no longer rely solely on the revenue generated by airtime usage. This chapter all along emphasises the key differences in the approach to charging between GPRS/UMTS and GSM. Finally, it goes on to describe the various possible charging mechanisms that may be employed for GPRS/UMTS Packet-Switched services, the network requirements, limitations and possibilities that these new mechanisms open up for Mobile Network Operators.

Packet-Based networks have five types of costs associated with their operation. These costs have to be covered by the users of the services provided by the network. New and innovative charging mechanisms, which will be accepted by users, thus need to be introduced to cater for these costs. These costs are

- Fixed cost of providing the network infrastructure, e.g. Links, Switches, Routers etc.
- An incremental cost of connecting to the network, i.e. bringing the network to the user. The user in the form of a *connection cost* pays this.
- Cost of expanding the network capacity. This is borne by users who wish to use the network when it is congested. Users willing to defer their transmission during congested periods should not pay for this.
- An incremental cost of sending a packet. This should be very low, during uncongested periods, since the bandwidth of a broadband network such as GPRS/UMTS is a shared resource.
- A social cost resulting from the fact that the transmission of a packet invariably leads to a delay in other users' packets.

In the Circuit-Switched domain, only the first three costs exist, and possibly also the social cost in networks unable to cater for full capacity. Traditionally, these costs have been condensed into a single price, fixed over long periods of time. They should however be split up into an *access charge, a usage charge, a congestion charge, and a QoS charge*. Access charges are commonplace now, where the user is charged for the privilege of accessing the network. These are imposed to recover the fixed cost of running the network. Usage charges are imposed to recover the variable cost of running the network, while congestion charges induce users to transmit during off-peak periods. All of these exist even in the Circuit-Switched domain. The **new** charge in the **Packet-Switched domain**, the **QoS charge**, arises from the user willing to pay to receive a certain QoS guarantee.

4.1 GPRS/UMTS Charging Architecture

The charging architecture is of utmost importance to network operators so that they are rest-assured that the services they provide are commercially viable. This section reviews the flow of charging information in a GPRS/UMTS network. The GPRS charging architecture is similar to that of GSM, the difference being that various CDR types (rather than just one CDR) are required for the charging mechanisms to be employed for GPRS services. These multiple CDRs are extracted from more than one network node, rather than from just one anchor MSC as is the case with GSM. The GPRS/UMTS charging architecture is illustrated below.

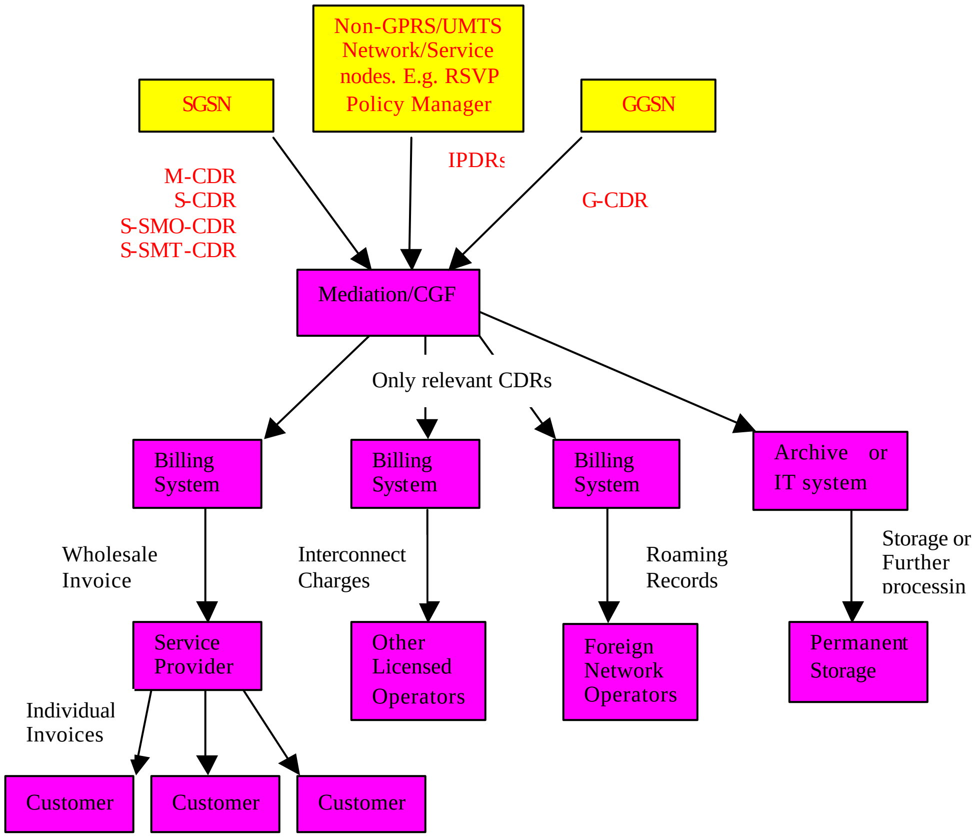


Figure 4.1 GPRS/UMTS charging architecture

4.2 Packet-Based Billing Issues

There are several issues that need to be considered when charging a customer for the use of a Packet-Switched service. Billing issues for multi-service broadband networks have only attracted a minority of researchers in recent years. Even in the late 1980s, research engineers gave pricing concepts little attention, focusing instead on technical issues. This trend however, has to be broken, if today's network operators are to survive. Billing for GPRS/UMTS is constantly cited as one of the most difficult issues that operators are facing in the transition to the next generation of Packet-Switched mobile networks. There are several unresolved issues relating to Packet-Based billing and the big questions on the minds of Mobile Network Operators looking to bill for Packet-Based data services are:

- How would users react to charges placed upon each packet?
- How do I combine voice, data usage, subscriptions and financial transactions into one bill for the customer?
- What changes are needed to my existing billing infrastructure to bill for these new types of data services?
- How do I make these changes quickly enough, to match the changes in the market, with minimum impact on existing systems?
- What types of tariff plans and charges can I make, that will be accepted by customers and content providers, so that I gain maximum market share and return on my investment?
- How is traffic billed where the recipient is the beneficiary, as is the case for VoD or similar information services?
- What services will be viable in order to generate revenue?

The traditional charging models for services in a Circuit-Switched environment are well understood and are typically based on parameters such as duration of call, time of day/day of week, location of called and calling parties, subscriber specific tariffs, and discount plans. Customer Care and Billing Systems on the market are capable of charging for calls on these criteria. One might ask. “Why do ISPs base their charging on the duration of the connection to the Internet when data in the form of **packets** are being transmitted/received?” The simple answer is, the services provided still use Circuit-Switched technology, as the customer connects via the PSTN or ISDN, which make use of Circuit Switching. See figure 2.1 above. For a true end-to-end IP network, where Packet-Switching is used all the way as in GPRS/UMTS, the charging schemes used, have to be different as the customers make use of network resources as and when they transmit/receive data. There is no longer a dedicated physical circuit allocated to the customer for the call and hence, the customer may no longer be charged purely on a duration-basis. GPRS/UMTS Packet-Switched data services mean a move towards the charging models of the data-communications world. These include, duration-based, event-based, volume-based, and content-based varied by location, QoS etc. Several issues face these new charging models. These include:

- Volume and complexity of Charge Data Records
- Complexity and flexibility requirement of Packet-Based Billing Systems
- How/What to charge packets
- How to capture the contents of packets being transferred over the network.

4.2.1 Volume and complexity of Charge Data Records

The volume of charging data per subscriber using GPRS/UMTS is about 20 – 40 times greater than that of voice or Circuit-Switched data in GSM. Even the most basic service, such as e-mail has a number of meaningful cost components associated with each transaction (size, distance, time of day, delivery options, disc storage for queued mail and POP3 queries). Real-time services, such as video, substantially increase the billing complexity, as they require a far more complex set of metrics to characterise the service and the quality delivered. As has already been mentioned, and as illustrated in figure 4.1, several types of CDRs with increased level of complexity are produced at the SGSN and each of these records has a specific function in the production of the final invoice. The GGSN produces the G-CDR too. Other complexities include

- Up-link and Downlink data volumes shall be counted separately.
- Every PDP context shall be assigned a unique identity number (CID) for billing purposes.
- It should be possible to provide reverse charging as a subscription option. This may however not be applicable to certain external PDN protocols

4.2.2 Complexity and Flexibility Requirement of Packet-Based Billing Systems

Due to the complexity and volume of GPRS/UMTS Packet-Based CDRs, flexible and high quality Customer Care & Billing Systems are required for these services and will be the keys to maintaining customer satisfaction and reducing churn. A successful GPRS implementation requires the availability of a sophisticated Billing System that can handle the multiple requirements of data billing. Traditional Billing Systems do not understand packet data. The flexibility requirement of the BS comes from the fact that, UMTS is intended to support **multiple, simultaneous** packet-based data applications. For example, voice and video data is more or less continuous, thus requiring network resources more or less for the entire duration of its transfer. Consequently, we can expect voice and video to have a flat per-minute rate, whereas other services, such as, Resource Access and fax/email applications may be charged on a volume basis, i.e. charged per MB of data transferred. Moreover, voice and video applications require bi-directional latency and bandwidth guarantees, whereas Resource Access and fax/email applications do not have the latency constraint, but may require significant bandwidth depending on the resource being accessed. Hence, QoS plays a significant role in most new charging mechanisms. The goal for service providers is to encourage GPRS service usage not to throttle usage with legacy billing strategies. This can only be achieved by introducing different and creative charging mechanisms capable of measuring and rating services under various usage-based metrics. Hence, an enhanced billing solution is imperative. Since it is difficult to determine exactly what services and applications will be offered in the future, a data collection entity that maximises flexibility and supports changes in pricing plan strategies is essential.

To illustrate the complexity of a UMTS IP-based network, consider a remote user located in London wishing to connect to the corporate network in Ipswich via an ISP. Within the ISP cloud, the London Remote Access Server (London RAS) provides ISDN or PSTN access. Routers (assume Cisco 7000s) are used at the ISP's London and Ipswich network access points. Finally, a number of support systems (e.g. RADIUS authentication server, RSVP manager, H.323 Gatekeeper, network management server, web proxy, and a traffic shaper) are used to manage the network; that is, provide AAA service and guarantee QoS. (Note: a similar paradigm is applicable for a stationary remote office PC.)

Assume the remote user would like to access corporate resources in Ipswich and perform the following network transactions:

- Attend a video conference call with two people from the Ipswich office;
- Search for an item in the corporate inventory database;
- Store a presentation on the Ipswich office file server;
- Send multiple e-mails and a fax to a business partner in another company;
- Browse through Internet web sites; and view an IP video news broadcast from CNN.

For these applications to work to the satisfaction of the mobile/PDA user (or for that matter, a corporate desktop or home user), the carrier has to provide suitable QoS for each transaction type. Exactly how the network determines the QoS requirements depends on the application. For example,

- The video conferencing application would likely issue a connection request to the remote PC through an H.323 gatekeeper - which, in turn, allocates the bandwidth.
- The database inventory access and remote storage application can contact the known peer address directly, and request the bandwidth allocation directly from the RSVP[5] policy server.
- The fax application would indicate its QoS requirements to a border traffic shaper, which uses statistical parameters (i.e., priorities) of packet flows. Alternatively, depending on corporate policy or the type of SLA in the service contract, the bandwidth or priority reservation might be given automatically when the user logs into the network.
- Finally, e-mail and other low-priority messaging applications can be given low priority on the routers and traffic shapers so as not to disturb time-sensitive traffic.

The other two highlighted issues, i.e. “How/What to charge packets” and “How to capture the contents of packets being transferred over the network” are discussed in section 4.5.

4.2.3 GPRS/UMTS Pre-Paid charging

GPRS in a similar manner to GSM is intended to support both contract (post-paid) and pre-paid charging. People who want pre-paid on voice will want pre-paid on data. However, there will be no Intelligent Network subsystem incorporated into the initial GPRS network. In section 1.6, it was mentioned that Pre-Paid Services (PPS) required an IN subsystem to enable real-time charging. Although, this is the preferred method for charging PPS subscribers, it is not the only method. An alternative method is to transfer the relevant CDRs to the BS or mediation (described below) every so often, say every 5 minutes. Another problem coupled with PPS within GPRS is that bandwidth-based pricing is not viable. The services will have to be based on a ‘per-message’ or ‘per-service’ model. Various Packet-Switched charging models are described in more detail in section 2.11.

[5] RSVP is a protocol that, for a designated flow, enables the allocation of bandwidth on each router/network element between two hosts.

4.3 Potential Revenue opportunities for GPRS/UMTS network operators

A smooth running, high-quality-performance GPRS/UMTS network is no good to an operator unless they are clear on potential revenue opportunities. Incumbent GSM operators have long lived on revenue generated by being a 'carrier' of voice and data. Mobile Network Operators of Packet-Based services can however not rely on this as their only source of revenue. Value added services would be the main revenue generator in the future. Today's operators can no longer be differentiated from competition by offering only standard services and therefore have to offer value added services both for competitive reasons as well as for revenue generation reasons. Intelligent services are also a way to improve customer loyalty and to win new subscribers. This section aims to introduce into the minds of GPRS/UMTS mobile operators new and diverse revenue streams for packet data services.

4.3.1 Legacy 'Carrier' or 'Data pipe provider'

A GPRS/UMTS network can be viewed as a 'data pipe' that operates via the mobile phone. However, if the network is used solely as an access 'pipe' to data, then its operators will suffer from the problems of commodity business, in particular, low margins and several competing sources of supply. This is particularly noticeable among ISPs, providing 'data pipes' to consumers. A network operator will be standing on shaky grounds if "providing a data pipe" is its only means of generating revenue. Hence, mobile operators are looking out for additional sources of revenue via their GPRS/UMTS network. Mobile Network Operators should stop thinking like a telecommunications industry and start thinking more as a service delivery industry, offering a rich variety of content products and services. Since all of the services used over the fixed Internet today – Remote LAN Access, File Transfer Protocol, web browsing, e-mail, etc will be available over the mobile network (GPRS), many network operators are considering the opportunity to use GPRS/UMTS to become wireless Internet Service Providers in their own right.

4.3.2 Provider of Personalised and Location-based "push" services

Location-based services and applications will become one of the new dimensions in UMTS. Since the operator knows the current location[6] of its users, it could offer location-based services/advertisements provided by 3rd party content providers. Hence, the notion of "location-based services". These services may either be "push services" from third party advertisers or "pull services" (e.g. localisation of emergency calls). Hence, one area that offers a huge new business opportunity for GPRS/UMTS operators is its ability via the mobile handset to offer a range of personalised and localised services, which is of great value to someone on the move. GPRS/UMTS network operators are therefore looking for means of tapping into this unique possibility especially as the mobile phone is portable and persistently with the user and is therefore a potential means of notifying the user of desired bargains or localised services. These services need to be looked into carefully, to ensure that customers are able to afford and willing to pay for them as well as application/content providers are charged reasonable tariffs for using operator's networks to reach the customers. In order to be able to charge the application/content providers, the services need to arrive at the mobile handset via the GPRS/UMTS network under the control of the operator. This simply means that the operator is aware of the fact that a particular content (information) has been delivered to the user and this will be discussed more under content-based charging in section 4.5.6. This can be achieved for example by having a "Location-based Push Advertisement" (LPA) server in the GPRS/UMTS network, which handles all advertisements from 3rd party content providers. The relevant call records produced at the LPA server are then forwarded to mediation for processing. It is important that mobile operators are **aware** of information flow over their network, as operators would like to avoid the tragic experience of the Palm VII[7], whereby content providers exploited some network operator to deliver various wireless services at no cost. The owners of the network responsible for delivering this U.S-wide wireless data access were

[6] Either the network-assisted GPS method, the cell coverage-based positioning method or OTDOA-IPDL may be used to determine the position of the UMTS user. See 3GPP R99 for more information on OTDOA-IPDL.
[7] The Palm VII provides wireless data services to its holders in the U.S.

unaware of the traffic running over their network. This is a huge threat to mobile operators and should be carefully looked into. Apart from location-based advertisements, audio advertisements, where subscribers receive free call minutes in return for listening to advertisements at the start of or during the call is already proving successful. As GPRS and UMTS develop, video advertisements rich in colour and graphics will attract greater number of users.

An example: Any company wishing to advertise a particular product or service approaches a content provider, which may or may not be the network operator. The content provider charges the company for this service. In the case, where the content provider is a third party, (i.e. a different company from the network operator,) they in turn request the Mobile network operator to use both their knowledge of the location of users and their network resources to deliver the advertisement to the appropriate users. The Operator then claims a percentage of the revenue generated by the content provider (from the advertising company). Alternatively, the operator may simply charge the content provider depending on the content and size of the message and possibly the location of the targeted recipients. The advertising company may be a hotel, a restaurant, etc wishing to advertise special offers to people currently in the vicinity of their business location.

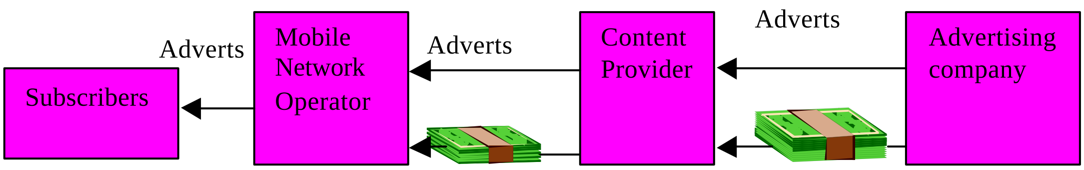


Figure 4.2 Revenue generation from Advertisements

With regards to personalised information, marketing companies are already developing interactive wireless marketing platforms, so that consumers can select the type of goods and services about which they want to receive offers and information. These platforms may be developed to allow highly targeted marketing messages to specific consumers, so advertisers can monitor real time response rates and adjust their messages accordingly. These advertisements, vouchers, and offers will be targeted to consumers by personal profile, business category or physical location.

Unsolicited adverts are likely to annoy uses, who may find them an intrusion into their privacy. It will therefore be important for users to feel in control of what is delivered to them, and incentives such as free offers and discount vouchers may act as 'carrot' to get them to opt in.

4.3.3 Internet Service Provider or Portal-type service provider

Fixed line Internet Service Providers used to make profits from monthly subscription charges as well as a fraction of the revenue generated by the call traffic caused by their customers - the Internet users. These days however, due to fierce competition, monthly subscription charges have been scrapped and ISPs have had to survive solely on the revenue generated from interconnect payment received from the local loop provider in the operating sphere. This hardly keeps the food on the table and most ISPs and portal providers are looking for other means of revenue generation. **Advertising**, **sponsorship** and **e-commerce** have been identified as alternative or rather additional sources of revenue. For example, a share price alert can arrive with a sponsor's message. A Company may pay to be included on a link from the portal or to be included in a search result and finally, a purchase of a book from amazon.com made by an ISP's customer may bring the ISP a cut of the transaction.

Since mobile operators have access to the customers more intimately and for more prolonged periods of time than any one else, this arena of untapped potential revenue seems to be a big opportunity for mobile operators. E-commerce-minded consumer goods vendors will be very keen to reach out to this operator subscriber base and operators should be aware of the worth of customer ownership and shouldn't give away this ownership lightly.

These sources of revenue (advertising, sponsorship and e-commerce) are already common among Internet portals and ISPs on the fixed line Internet for example Yahoo. The idea here is for the mobile operator to be an official Internet portal and persuade mobile Internet users to use their portal for web shopping – mobile e-commerce also known as m-commerce[8] – in order to acquire from these m-commerce transactions. For the portal to be attractive to m-commerce customers, the operator has to become a trusted brand, offering a range of quality-controlled services, backed up by help-desk support and offer a guarantee that any financial transaction performed on the portal is secure and backed by the operator. The strategies required to achieve this marketing challenge is outside the scope of this project. Valued operators already have a trusted brand and can use this to their advantage. For smaller network operators, who are not major portal players or who are unsuccessful in persuading their customers to use their portals, these opportunities - acquisition of financial transactions, advertisements etc may be missed out on completely. They may have to contend themselves with being bit pipe providers and providers of location-based services. Alternatively, the network operator may charge the portals being used by the customer, for any transaction taking place but this does not seem to be a feasible option. The analogous situation in the fixed line network exists. ISPs providing (directly or indirectly) network resources, required to deliver the contents of the web do not charge other Internet portals for any transactions committed via their site. Instead, they rely on interconnect settlement, monthly subscription fees. Hence, mobile operators with a huge loyal customer base have a great opportunity to increase revenue by offering ISP and portal-type services.

4.3.4 Partnerships with Content providers

Mobile operators due to their unique opportunity of access to a large customer base, are just the right people for content providers (looking to reach out to customers), to build partnerships with. These content providers rather than pushing advertisements to customers, provide customers with desired services. Hence, the customer is charged for these services and a share of the revenue is given to the content provider concerned. For example, a game provider may provide opportunities for mobile subscribers to play on-line games. This vastly increases airtime and as a result, operator revenues are increased. The mobile operator then shares this revenue generated with their content partners. Another example may be a businessman on a journey and away from his PC requiring critical business information and news. Reuters may for example provision this service. The scenario may vary but the concept remains. Settlement or revenue sharing can be handled by many existing billing systems and in the future, mobile operator will have to build several partnerships with various providers. There will be a lot of settlement between content partners and a range of options exist, which may be as complex or as simple as the market demands.

4.3.5 Payment Method Provider

The Mobile Equipment together with the SIM card may be used to identify a user and can therefore be used as a payment method just as a credit card is used today. Purchases are divided into two phases: The merchant receives the customer's mobile phone number, and the customer is in turn sent a payment request, which is accepted and authenticated by entering a PIN code, then transmitting through encrypted SMS or encrypted data.

[8] Operators who lead customers to m-commerce revenue generating transactions stand to gain a significant share of the proceeds and m-commerce could eventually be the killer application for GPRS/UMTS operators. The drivers for mobile commerce (m-commerce) are assumed to come from one of three basic needs.

- The need to purchase something quickly, like train tickets while on the move
- The need to buy some new product/merchandise, convenience items, or gifts in free moments during busy days or
- The need to conduct time-critical financial transactions, such as transferring money between bank accounts, paying bills, or stock trading.

Comparison of sources of revenue among GSM operators and GPRS/UMTS operators

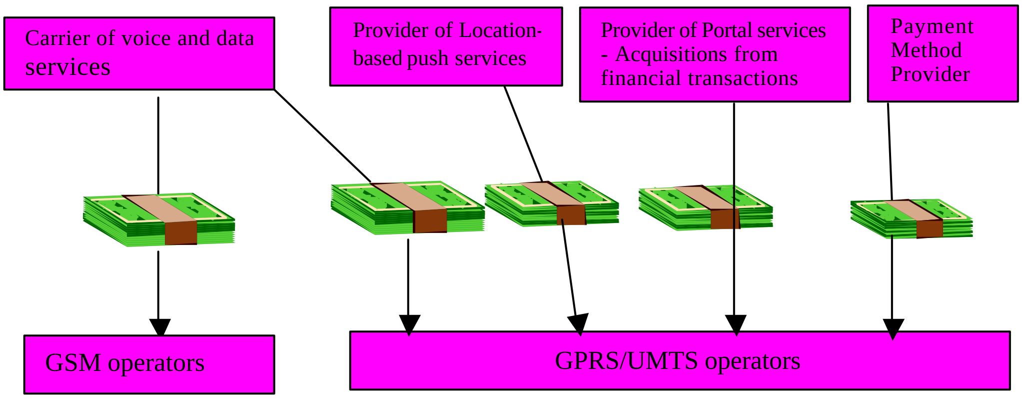


Figure 4.3 Comparison of sources of revenue among GSM operators and GPRS/UMTS operators

The potential revenue opportunities mentioned above bring about questions such as: What business are Mobile operators in? Is it communications, content or commerce? Or is it a combination of all? Who is the customer etc? But more importantly: whom do operators bill for the various services provided? Are there sufficient network resources to provide the data required for billing? These are the issues that will be dealt with in the remainder of this chapter. The revenue-generating potential of 3G services can be difficult to quantify without first looking at an example in detail.

Example: '*Your car needs a 12,000 -mile service - I have booked it into Super-Service garage next Tuesday*

CarMaker Inc makes a deal with Fab Telecom Mobile that they will install in all their new cars a UMTS terminal. Each car has an intelligent control system, which makes numerous continual checks to determine when it needs servicing, for example mileage, fluid levels and brake pad thickness. You have one of these cars, and when any of these 'alarms' are raised, a message is sent through the UMTS network to a 'Lifestyle Management System'. This has access to your diary, and to your preferences - for example your favourite servicing garage. So it sends a message to this garage's booking system and makes an appointment for you to take your car in for servicing the next day you are in the office - say next Tuesday. You receive a (voice) message saying 'Your car needs a 12,000 mile service - I have booked it into Super-Service garage next Tuesday' and ask you to confirm that this is acceptable. You confirm, and the Lifestyle Management system updates your diary and confirms to the garage. On Tuesday morning it will remind you of the appointment. In this example, there have been at least six uses of the network, mostly small amounts of data, and one voice call. How is this charged?

There are many possible ways. For example it could be offered as a free service, or based on a small subscription charge. Or it could be charged on the volume of data sent (probably as part of a total charge per month for all data sent). Or you may not be charged at all for specifically using the network, but instead charged a single fee called 'use of car servicing booking service'. But, what about CarMaker Inc, the Lifestyle Management System Provider, and the Super-Service garage? Do some of them get a proportion of the charge made – i.e. does the transaction involve a bill to the customer and some sharing of the revenue with partners? Or does the operator charge them for giving access to their customer, and therefore bills the customer AND Partners? Suppose this is a service offered entirely by the operator, so in addition to the use of the network, the garage bill for the car service also appears on the 'telephone bill'? In this case the operator has to collect the full debt and gives it back (less commission) to the garage. And maybe the garage sends adverts to those customers who don't subscribe to the service, and the operators charges them every time an advert is sent, or every time someone signs up.

A mobile telecommunications 3G billing system has to recognise the functionality needed to cope with all these scenarios, because no one today can predict the actual charging mechanisms and business models that will actually be put in place.

4.4 Record Types

This section gives an overview of the 5 different CDRs produced in a GPRS/UMTS network, where they are generated, and what types of parameters each of them contains. It should be noted however that, in order to apply, various charging schemes on various services and applications, other records generated outside the GPRS/UMTS network need to be considered. These records include Internet Detail Records (IPDR), currently under development by the IPDR organisation, as well as service specific Call Detail Records (CDRs) generated by service platforms outside the scope of the GPRS/UMTS network. IPDRs may be generated by several network elements (including RSVP policy managers etc), each providing unique and invaluable information regarding the details of the call/event. An example of a service platform may be an m-commerce platform, which generates records regarding the purchase of goods/ tickets, stock options etc. This section however concentrates solely on the records generated within the GPRS/UMTS Packet-Switched network

4.4.1 S-CDR

An S-CDR is opened for each activated PDP context and is produced in the SGSN to hold details of the communication between itself and the GPRS handset. Hence, it records the usage of the radio interface as well as general GPRS resources. It is used to collect charging information related to packet data information transmitted in Mobile Originating and Mobile Terminating directions categorised with QoS and user protocols. Typical data held within the S-CDR includes

- IMSI of MT/UE
- IP Address of SGSN
- Charging ID (CID) allocated by the GGSN
- GGSN used which, together with the CID form a unique key.
- Access Point Name (APN)
- List of Traffic Data Volumes

Most of the data mentioned above is static but 'The List of Traffic Data Volumes' is directly dependent on dynamic GPRS usage. It is a variable length list with a variable number of "containers", each of which records the volume of traffic transmitted/received between the GPRS MT and the SGSN under various charging conditions.

Triggers for change of S-CDR charging conditions

- The Quality of Service changes
- A new tariff period is entered
- S-CDR closure

The S-CDR closure is as a result of one of two of the following

- End of PDP context within the SGSN due to SGSN handover, termination of PDP context, or any abnormal release of the CDR.
- Partial Record reason as a result of data volume limit (configurable), duration limit (configurable), maximum number of charging conditions changes (configurable) or Operations and Maintenance reasons like management intervention.

Thus, each PDP context may be described by a large number of S-CDRs each of, which is variable in length depending upon changes to the charging conditions.

It should be noted that the SGSN counts all downlink data sent to the RNC over the Iu interface. Hence, any data discarded by the RNC before delivery to MS causes inaccurate charging, as the SGSN cannot account for this and subsequently the subscriber is overcharged. To avoid inaccurate charging at the SGSN, the SGSN will always instruct the RNC at RAB set-up to count the unsent downlink data towards the MS. (RAB stands for Radio Access Bearer). The reporting of unsent data by the RNC to the SGSN will only occur at RAB release. This occurs at either the termination of the PDP context or handover. When the SGSN receives a report of unsent data volume from the RNC at RAB release, the SGSN then reports this value to the *'RNC Unsent Downlink Volume'* field in the S-CDR.

4.4.2 M-CDR

An M-CDR, produced at the SGSN is used to collect charging information related to the Mobility Management (MM) of a GPRS MT. An M-CDR is opened for each mobile upon GPRS Attach. A PDP context need not exist for M-CDR records to be produced. Thus, M-CDRs are produced even in the event of no data transfer. The customer is not charged for this but the records are there to help keep track of which MT attached itself to the network and when this occurred. Since an M-CDR does not identify any information related to active PDP contexts, it knows nothing about the connected GGSNs but only contains the address of the current SGSN used. In the GSM Circuit-Switched world, there is always an "anchor switch" through which all CDRs are routed. Hence, all the CDRs for a particular connection originate from the same point, regardless of the mobility of the GSM mobile. For GPRS however, the S-CDRs related to a particular connection may come from various SGSNs if the handset user is in motion, hence the need for M-CDR. Typical data recorded in the M-CDR include

- IMSI of handset
- SGSN Address
- Routing area
- List of changes in routing area

Not all of this data is static. Some of the charging information is directly dependent on GPRS mobility. The "change of location/routing area" attribute of the M-CDR consists of a set of containers, which are added when specific trigger conditions are met. These conditions include

- Mobility Change
- M-CDR closure

M-CDR closure is a consequence of one of the following

- End of MM Context within SGSN due to SGSN change, GPRS detach, or any abnormal release
- Partial Record reason, which constitutes any of the reasons mentioned for the S-CDR case.

If an M-CDR is closed and the GPRS mobile is still known to the SGSN, a new M-CDR will be opened with an incremented Sequence Number.

4.4.3 G-CDR

A G-CDR, produced at the GGSN is used to collect charging information related to packet data information for a GPRS mobile. It is records related to the data traffic between the GGSN and the external PDN that is recorded. Hence, it is the usage of the external PDN that is of interest here. In a similar manner to S-CDR, a G-CDR is opened for each activated PDP context, and the parameters recorded are very similar to those of the S-CDR. One difference is that an SGSN change need not close the G-CDR. The G-CDR field, "SGSN Address used" is simply updated.

4.4.4 S-SMO-CDR and S-SMT-CDR

Short messages transmitted over the GPRS network result in the creation of these CDRs in the SGSN. The record produced depends on whether the message is Mobile Originated (MO) or Mobile Terminated (MT). No active PDP context is required when sending or receiving short messages. If the subscriber has an active PDP context, volume counters of S-CDR are not updated due to short message delivery.

4.5 Record Contents

The record contents of each record type varies slightly, but in effect, they are quite similar except for the Short Message CDRs, which hold much fewer fields and are thus, much less complicated. The QoS-field for Short Message CDRs is non-existent. A table illustrating the important parameters recorded in a G-CDR is included in Appendix 2B. A description of each field of this G-CDR can be found in [25]. However, 'Charging ID', 'List of Traffic Data Volumes', QoS, and Charging characteristics have been briefly described below, as these are closely related to the main subject of this report - charging.

4.5.1 Charging ID (CID)

GSM Circuit-Switched calls can be charged using Call Detail Records from one MSC (the anchor MSC), which is guaranteed by routing all signalling information through the anchor MSC even if the traffic channel of a call is routed through another MSC due to handover. In a GPRS network however, the handling of a complete PDP context can be switched over from one SGSN to another due to RA updates. Consequently, charging records are generated in more than one SGSN. Furthermore, PDP context-related data is recorded in the GGSN as well. Thus, for one activated PDP context, charging records, which are forwarded to the CGF are needed from two or more GSNs. All records belonging to one active PDP context could normally be identified by the Tunnelling Identity (TID). However:

- A Mobile Terminal can activate and deactivate PDP contexts in a very short time interval, and these PDP contexts can have the same TID. Only parallel established PDP contexts have different TIDs.
- Different SGSNs can be involved in the same PDP context and the timing clocks of the SGSNs may or may not be fully synchronised.

It is therefore not advisable/possible for the CGF to gather the records of one PDP context only by using the IMSI, the TID and time. A unique CID is therefore assigned to all records generated for a particular PDP context. This unique CID is generated in the GGSN when the PDP context is activated and is transferred from one SGSN to another following a Routing Area update. Hence, the CID together with the GGSN address, uniquely identifies the PDP context and helps the CGF and BS to correlate all generated CDRs. Different GGSNs allocate the CID independently of each other and may allocate the same numbers. The CGF and/or BS check the uniqueness of each CID together with the GGSN address and if ambiguous, checks the record opening timestamp.

4.5.2 List of Traffic Data Volumes

This list contains one or more containers each of which include the following fields.

- Data Volume (Up-link)
- Data Volume (Down-link)
- Change Condition
- Timestamp

Change Condition defines the reason for closing the container, which may be tariff-time change, QoS change, or closing the CDR. The timestamp is the time, which defines the moment when the change condition occurred. Below is an example of a list, which has three containers, caused by one QoS change and one tariff-time change.

QoS Requested = QoS1 QoS Negotiated = QoS1 Tariff Time = TT1 Data Volume (Up-link) = 1 Data Volume (Down-link) = 2 Change Condition = QoS change Timestamp = TIME 1	QoS Negotiated = QoS2 Tariff Time = TT1 Data Volume (Up-link) = 5 Data Volume (Down-link) = 6 Change Condition =Tariff Time change Timestamp = TIME 2	QoS Negotiated = QoS2 Tariff Time = TT2 Data Volume (Up-link) = 3 Data Volume (Down-link) = 4 Change Condition = Record closed Timestamp = TIME 3

Table 4.1 List of Traffic Data Volumes with Three containers

The first container contains initial QoS and Tariff Time values and corresponding volume counts, and of course the reason for closing the container and the time, which this occurred. Second container contains new QoS values and corresponding volume counts before tariff-time change, while the last container includes volume counts after the tariff-time change. The QoS change is as a result of the network unable to deliver the initial QoS (QoS1) requested, possibly due to an increase in network resource demand. TT1 could be peak hours while TT2 defines off-peak hours.

		Container
QoS1 + Tariff 1	Up-link =1, Down-link =2	1
QoS2 + Tariff 1	Up-link =5, Down-link =6	2
QoS2 + Tariff 2	Up-link =3, Down-link =4	3
QoS1	Up-link =1, Down-link =2	1
QoS2	Up-link =8, Down-link =10	2+3
Tariff 1	Up-link =6, Down-link =8	1+2
Tariff 2	Up-link =1, Down-link =2	3

There are different control mechanisms to ensure that packets transmitted from an old SGSN to a new SGSN does not increase the PDP CDR volume counters in both SGSNs. These control mechanisms are not discussed here. Although, QoS is an attribute of “List of Traffic Data Volumes”, it will be discussed in detail and a complete subsection (4.5.5) has been dedicated to QoS.

4.5.3 Access Point Name (APN)

The network operator provisions the APNs for each subscriber on the HLR, when the subscriber contracts for the service. Types of Access Points include ISPs, corporate networks, mail servers, broadcast services (news, stock exchange etc), telemetry (reading remote meters, controlling remote plants etc), games. These various services differ widely and hence, the data transferred through them will have different content value and will consequently be charged differently. The services from the different APNs would be charged on one or a combination of several of the charging bases discussed in section 4.7 below.

4.5.4 Charging Characteristics

This allows the operator to apply different kinds of charging methods for the CDRs. N indicates normal charging, P indicates prepaid charging, F for flat rate charging and H for Hot billing. The P flag for example, can be used to easily identify whether the customer is pre-paid or not and as such avoid unnecessary transfer of data. One or more of these flags are set according to the charging characteristics received from the HLR.

SPARE BITS ‘0000’	N	P	F	H

Table 4.2 Charging Characteristics flag

4.5.5 Quality of Service (QoS)

In the traditional IP world, QoS cannot be guaranteed. The inherent unreliability of Packet-Switched networks coupled with the weaknesses of wireless networks, make guaranteeing QoS for GPRS/UMTS a challenging task. Two main questions are of interest.

1) How will QoS be implemented technically?
2) How will QoS be measured?

An approach to ensure QoS is to dedicate a special pipe of routers to a transaction. Implementing QoS in current IP networks is a challenging task, as IPv4 has no real QoS features. Though extremely difficult, there are practical solutions such as RSVP, MPLS, Diffserv and various other forms of policy management. Ipv6 will have QoS features built in but we are still some time away from its implementation. Packet Loss, Jitter, and Delay typically measure QoS. These measurements are then compared to a Service Level Agreement (SLA) made with the customer. One issue for consumer implementations is how can these technical measurements be expressed to a consumer in a meaningful fashion? We are however not interested in the mechanisms for implementing end-to-end QoS in GPRS/UMTS. Rather, we are more concerned with how QoS may affect the charging mechanisms employed for GPRS/UMTS services.

As discussed in section 2.5.2, the MS initiates a 'PDP Context Activation' by first transmitting a 'PDP Context Activation request' message to the SGSN. This message contains among other things, a QoS profile requested by the customer. Only mobile terminals can request QoS-based sessions. Network requested QoS-based sessions are not possible. The SGSN to which the MS or UE is connected to, has access to the QoS profile subscribed for by this particular user. These records are obtained from the HLR (or GR within the HLR containing all GPRS-related details) during 'GPRS Attach'. The network then offers the user the requested QoS profile, provided that this requested profile does not exceed the QoS profile subscribed for by the user. If it exceeds the profile subscribed for, then the user is not granted his request. The QoS profile granted by the network is known as "QoS negotiated". In other words, the QoS profile in the UMTS subscription describes the upper limits for the QoS negotiated if the service user requests specific values. If the user requests no QoS, then the subscribed QoS profile may describe default QoS profile requested by the user.

4.5.5.1 GPRS QoS Attributes

The QoS profile defined in the ETSI specifications (GPRS R97/98) comprises of multiple data transfer attributes. These are:

- Precedence
- Delay
- Reliability
- Peak throughput and
- Mean throughput

Precedence

This attribute gives an indication of the relative priority of each service. Under network congestion, the packets to be discarded (low priority packets) can be identified. There are three levels of precedence - High priority, Normal priority, and Low priority

Reliability

This attribute indicates the probability of loss, duplication, mis-sequencing and corruption of each packet, for the end to end transmission of packets. There are five levels of reliability

- Non-real-time traffic, error-sensitive application that cannot cope with data loss.
- Non-real-time traffic, error-sensitive application that can cope with infrequent data loss.
- Non-real-time traffic, error-sensitive application that can cope with data loss.
- Real-time traffic, error-sensitive application that can cope with data loss.
- Real-time traffic, error non-sensitive application that can cope with data loss.

Delay

The Delay attribute specifies the maximum values of end-to-end delay through the GPRS network. It includes, Radio channel access delay (up-link), Radio channel-scheduling delay (down-link), Radio channel transit delay (up-link/down-link), and GPRS access and core network transit delays (multiple hops). It however does not include transfer delays in external networks. There are four levels of delay:

- < 7 seconds
- < 75 seconds
- < 375 seconds
- Best effort

The actual value associated with each class varies according to the packet size and the percentage of packets that are included in the delay calculations.

Throughput

Peak and Mean user data throughput attributes are represented by two negotiable parameters - Maximum bit rate, and Mean bit rate respectively. Specific values are given to each of them. The throughput is measured end to end and is the most important QoS parameter as it is the most noticeable from the user's perspective.

4.5.5.2 R99Extended QoS Profile

R99Extended QoS Profile (as the name implies) includes some other parameters apart from the above mentioned parameters for GPRS R97/98. These are

- Guaranteed bit rate for up-link
- Guaranteed bit rate for down-link
- Traffic handling priority etc

The up-link and down-link bit rates have been separated to support asymmetric bearer services i.e. services requiring variable bit rates for up- and down-link traffic as mentioned in point 4 of the "QoS Technical Requirements". An extract of the "QoS Technical Requirements" can be found in Appendix 4A whereas the complete Technical Requirements can be found in 3GPP Standards Document TS 23.107 V3.2.0 (2000-03), "QoS Concept and Architecture".

4.5.5.3 UMTS Bearer Service QoS

The UMTS Bearer Service consists of two parts, the Radio Access Bearer Service and the Core Network Bearer Service. A Bearer Service includes all aspects, such as control signalling, user plane transport and QoS management functionality, to enable the provision of a contracted QoS. The UMTS Bearer Service is specified by a set of QoS parameters, which include, apart from the "R99Extended QoS Profile" parameters such as:

- Allocation/Retention Priority attribute (not included in Extended QoS Profile located in UE),
- Indication of the UMTS QoS class. (Described below)

Several QoS Profiles are obtainable by the various combinations of the attributes listed above. However a PLMN may support only a limited subset of the possible QoS Profiles as mentioned in point

3 of the UMTS Technical Requirements listed in Appendix 4A. This is achieved by having a list of finite attribute values or allowed value range for each attribute. When a service is defined as a combination of attributes, further limitations may apply; for example the shortest possible delay may not be possible to use together with the lowest possible SDU error ratio. The granularity of the bit rate parameters still remains to be studied by standards bodies. Although, the UMTS network has capability to support a large number of different bit rate values, the number of possible values shall be limited so as not to unnecessarily increase the complexity of user terminals, charging functionality and Inter-Working Functions, for example.

4.5.5.4 QoS Mapping

The QoS management functionality enables translation, **mapping,** and packet classification to maintain the specified QoS characteristics.

A mapping between R99 (UMTS) and R97/98 (GPRS) may be needed. For example, when an R99 MS performs PDP Context Activation to an R99 SGSN while the GGSN is of R97/98, the SGSN performs the mapping of R99 QoS profile to R97/98 QoS profile. Also, as the QoS parameters on Core Network Bearer Service level do not all have the same value as corresponding parameters on Radio Access Bearer level, a QoS mapping/relation is required. Finally, a mapping between Internet QoS and UMTS QoS may be required. The two main Internet QoS concepts, namely Integrated Services (IntServ or RSVP) and Differentiated Services (DiffServ) will be controlled by applications residing in the TE, allowing different application-specific QoS levels for the same PDP context.

Neither PSTN nor ISDN has QoS mechanisms, thus QoS profile Inter-Working/mapping is not needed. All of the above QoS mapping are part of the “QoS Technical Requirements” listed in Appendix 4A.

4.5.5.5 UMTS QoS Classes

Four QoS classes are specifically defined for UMTS. These are:

- Conversational
- Streaming
- Interactive and
- Background

The main factor distinguishing these classes is how delay sensitive the traffic is.

The **Conversational class** is intended to carry extremely delay-sensitive applications. The best-known application of this class is the voice service, but with the Internet and multimedia, a number of new applications (like or Videotelephony and VoIP) will require the QoS profiles defined for this class. Real-time conversation is mostly performed between peers or groups of human end-users and subjective evaluations have shown that for the end-to-end delay to be acceptable, it has to be less than 400ms.

The **streaming class** is one of the newcomers in data communication, raising a number of new requirements in both telecommunication and data communication systems. It is fast becoming important because most users do not have fast enough access to download large multimedia files quickly. With streaming, the client-browser can start displaying the data before the entire file has been transmitted. Streaming is a one way transport and requires the delay variation of the end-to-end flow to be limited, i.e. to preserve the time relation (variation) between information entities of the stream. Streaming applications are very asymmetric and therefore withstand more delay than more symmetric conversational services. Streaming is divided into web broadcast and Video streaming on Demand (VoD), the latter being used by large corporations wishing to store video clips or lectures to a server connected to a high bandwidth local intranet. Several people seldom use these on-demand clips/lectures at the same time. On the other hand, web broadcast providers target very large audiences that connect to a high performance-optimised media server.

The **interactive class** involves either a machine or a human, on-line, requesting data from remote equipment (e.g. a server). Examples of human interaction with the remote equipment are web browsing, database retrieval, server access etc. Examples of machines interaction with remote equipment are polling for measurement records and automatic database enquiries (telemetry). Its most important characteristic is that the content of the packets shall be preserved and transparently transferred with low bit error rate.

The **Background class** is meant to be the least delay-sensitive traffic class, carrying applications such as background download of files (FTP), e-mails or SMS. Also, the downloading and reception of measurement records can be delivered background since the data may not require immediate action. Background class traffic is characterised by the fact that the destination is not expecting the data within a certain time. However, payload content need be preserved but not necessarily transparently transferred. A new application within this class is the electronic postcard.

Due to the looser delay requirements of Interactive and Background classes as opposed to Conversational and Streaming classes, the former provide better error rate using techniques such as channel coding and retransmission. Traffic in the interactive class has higher priority in scheduling than Background class traffic, so background applications use transmission resources only when interactive applications do not need them. This is very important in wireless environment where the bandwidth is lower compared to fixed networks.

QoS requirements especially for Conversational and Streaming classes shall be maintained during Inter-SGSN RA Update within the UMTS network.

Traffic class	**Conversational**	**Streaming**	**Interactive**	**Background**
Fundamental Characteristics	Preserve time relation between information entities Stringent and low delay	Preserve time relation between information entities	Preserve data integrity. Request response pattern	Preserve data integrity. Destination not expecting data within specific time
Example of application	Voice, Videotelephony, Video games	Streaming multimedia	Web browsing, network games	Background download of e-mails

Table 4.3 UMTS QoS classes

The various UMTS QoS classes require a different combination of parameters as well as a different combination of parameter values. For example, conversational and streaming class may have a Transfer delay attribute associated with them while Interactive and Background do not. The *Traffic Handling Priority* parameter for example, exists only for the interactive class. See Appendix 4A for a table and a description of parameters describing the QoS profile for the various UMTS QoS classes.

4.5.5.6 Secondary PDP Contexts

Point 6 of “Technical Requirements for QoS” found in Appendix 4A requires the support of Secondary PDP Contexts, also known as Multiple PDP Concept. A Secondary PDP Context is one that is activated reusing the PDP address, APN and other PDP context information from an already active PDP context, but with a different QoS profile. Procedures for APN selection and PDP address negotiation are not executed. A unique Transaction Identifier (TI) – CID, and a unique NSAPI shall identify each PDP context sharing the same PDP address and APN. The figure below illustrates the Secondary PDP Context concept.

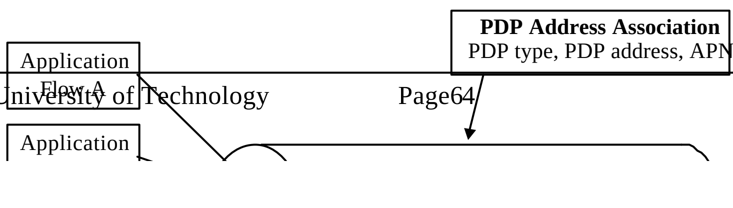


Figure 4.4 Secondary PDP Context concept

The main purpose of the Secondary PDP Context concept is to support applications containing a number of independent flows such as multimedia applications. Indeed, let's assume that a multimedia application combines three different flows of information, i.e. data, voice and video. These three flows originate from the same domain (same APN), terminate at the same UE (same PDP address), but require different QoS levels. The use of Multiple PDP Context is suitable for the support of this type of applications as illustrated below.

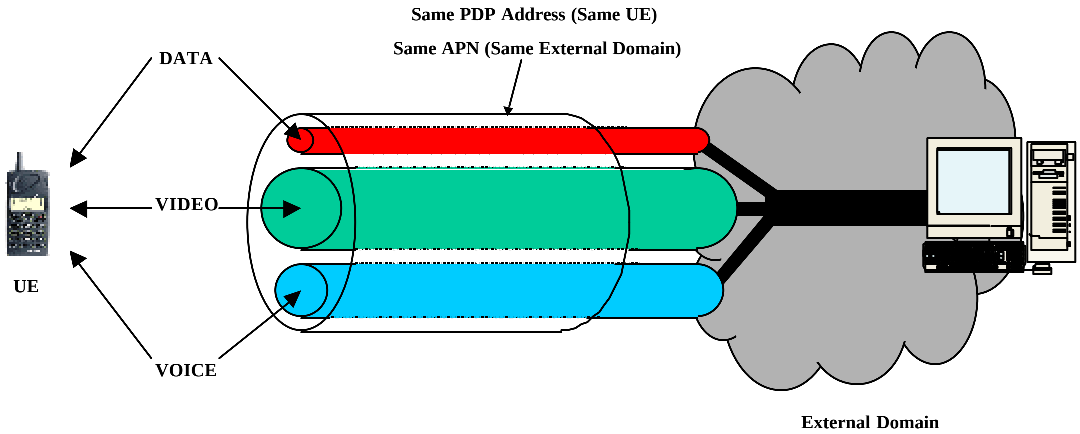


Figure 4.5 Example of Multimedia Support

4.6 Network Infrastructure Required for Billing Purposes

This section describes the function of the Charging Gateway Functionality and the Mediation Device. Although, the SGSN and GGSN are required for billing, they are not discussed here, as they have already been discussed in chapter 2. It should also be noted that as well as the 3G Network specific services, other applications/services will be provided to the 3G subscribers by service nodes outside the scope of the 3G network. These servers (service nodes) responsible for the provision of an application services to a subscriber via the 3G network, can generate a service related CDR to record the details of the service transaction provided. These service-specific CDRs (generated by service nodes) contain details about the specific service. For example, we may have an entertainment server for Music and Video, as well as content like games, an information record server for weather, stock, road/travel, and a transaction server or m-commerce server for tickets (Theatre, Cinemas, Train) and Produce (Vending). The servers generating these CDRs can be positioned either within or outside the Network Operator's network. Where the server is positioned outside the 3G Operators network, compliance, security and financial arrangements between the two parties will need to be agreed.

4.6.1 Charging Gateway Functionality

The Charging Gateway Functionality (CGF) forms an important part of the GPRS/UMTS network infrastructure. It is part of the core network and has a variety of functions, which will be discussed. The CGF provides a mechanism to transfer charging information from the record producing nodes (SGSN and GGSN) to the network operator's BS. It is intended that the protocol stacks and configurations of the GSNs need not be updated if a new Business Support System (BSS) is introduced. This is achieved by introducing the CGF. This CGF concept enables the operator to have just one logical interface between the GSNs and the BS. Hence, only the CGF configurations need be altered. Since most operators looking to offer Packet-Switched services will require new or modified BSS, the CGF is of key to success. The amount of processing required at the CGF before the records are finally passed on to the Operator's Billing System (BS) is quite substantial and hence, the CGF may appreciate the help of a Mediator (an extra processing unit) to interface into the exis ting Billing System. Apart from various Billing Systems illustrated in figure 4.1 above, the CGF-processed CDRs may also be sent to other BSSs like an archive or an IT system for further analysis and evaluation for purposes such as network planning, fraud prevention etc. Because each of the BSSs require variable amount of CDR processing, most operators make use of a mediator (Enhanced CGF), which collects, collates, explodes, filters and transforms the CDRs. The CDRs are transferred to the mediator via the CGF and eventually to the various BSs at intervals determined by one or more of the charging change conditions listed in 4.4.1 above.

The CGF may be supported either

- as a centralised separate network element or
- as a distributed functionality resident in the SGSNs and GGSNs

The chosen configuration is manufacturer dependent and subject to operator agreement. Figure 4.1 illustrates a centralised separate node while figure 4.5 (a) below illustrates the other option.

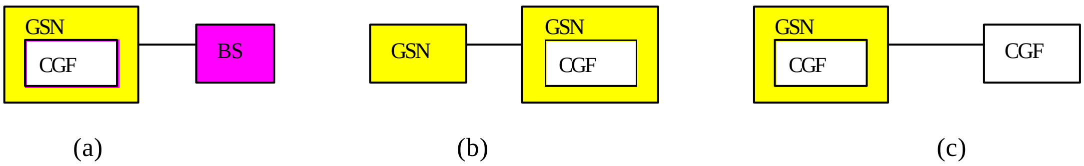


Figure 4.6 GSN with integrated CGF

It is also possible to have a combination of both configurations, i.e. a GSN with integrated CGF also acts as CGF for other GSNs or a GSN with integrated CGF supports transmission of CDRs to external CGFs as shown in figure 4.5 (b) and (c) respectively.

The main functions of the CGF are

- The collection of CDRs from the SGSN and the GGSN
- Intermediate CDR storage buffering and
- Transfer of CDR data to the BS

The CGF may have to support several transmission protocols for data transmission to the BS depending on the BS used. Since, the CGF is such a critical node in the GPRS/UMTS network, it is usually distributed to several physical GSNs or Charging Gateway Entities to facilitate redundancy and resilience to failure. Thus, each GSN has a prioritised list of CGF IP addresses to which it may transfer its records in the case of "primary" CGF node failure or communication link severity.

4.6.2 Mediation

GPRS/UMTS networks produce "billing information ", which is different, both in content, volume and format from what any Circuit-Switched network (fixed or mobile) produces. Hence, a mediation system is needed to take these billing records as they come out of the network, and convert them into forms suitable for legacy billing infrastructure to work with. Thus, the billing mediation device provides the link between the core network and the IT applications (Billing Systems, Fraud Management systems, Data Warehousing, Interconnect administration etc.). Furthermore, in contrast to a GSM network, where the MSC provides AAA (Authentication, Authorisation, and Accounting) and generates CDRs, a UMTS IP-based network is composed of a substantial number of heterogeneous network elements, which provide and provision these services. It is intended that a new Billing System need not be introduced, even though the CDRs collected and processed from the CGF vary in volume and complexity from CDRs produced from traditional GSM voice/Circuit-Switched calls. This is achieved by introducing **Mediation**, capable of interfacing with operator's existing BS and other Business Support Systems (BSS). The mediation also passes provisioning information from the BSS to the network elements but this particular function has not been considered further in this document. The Billing Mediation system has to work with new network equipment, handling complex data formats, which are still being defined. It has to be scalable to match a dramatic market uptake of GPRS and also reliable and secure enough to handle commercially critical data.

Mediation plays an important role in dealing with most of the issues/questions highlighted in section 4.2, including the ability to bill for any type of data services, and the possibility to combine data with voice or any other service into a single bill. The mediator's ability to handle Internet Data Records, otherwise known as IPDRs[9] is of utmost importance if the operator wishes to be able to charge for content, acquire commissions on accomplished transactions, and identify applications, which is important to enable the implementation of differential charging. The mediation device is comprised of several modules for its various tasks. Additional modules may be introduced for additional functionality as required by the operator. The mediator is expected to be capable of processing the various CDRs of voice, video, and data usage coupled with IPDRs from other external IP network elements and providing a consolidated record for the operator's legacy Billing System. Mediation devices today are used for managing advanced network services, not just as protocol adapters. Some of these advanced network services include Fraud management, congestion control etc. They also offer an abstraction layer that allows the BS development to focus on the operator's core business and leave the technical details to the mediation layer. Below is a brief description of the mediators various modules and functionality. The key role of the Mediation device is however to collect data from the network infrastructure and deliver data to the Business Support Systems.

4.6.2.1 CDR Collection

The collection module is responsible for extracting the CDR batch files from the disparate parts of the GPRS/UMTS network as well as

- checking for and eliminating duplicate copies
- transferring the CDR batch files to safe storage (possibly tape)
- forwarding the CDR files for collation

The collection module may also be required to collect data from other external (non-GPRS/UMTS) network nodes.

4.6.2.2 Collation

The collation module then associates the CDRs from the same session/PDP context and extracts useful data, hence cumulatively building a "collated CDR". It may also maintain a database of open PDP contexts and support multiple criteria for closing and forwarding the "collated CDR". CDRs generated

[9] IPDR is the IP equivalent of CDRs generated by telephone switches (MSCs in GSM).

by the SGSN and GGSN are easily correlated as these include a CID, which identifies all records produced from the same PDP context session. The other records produced by other external (non-GPRS/UMTS) nodes are however more difficult to identify.

4.6.2.3 Explosion

As mentioned above, it may be desired to forward the CDRs to several BSs and/or other systems, a data warehouse say, for analysis of service usage. The explosion module is responsible for supplying these various systems with the relevant data required. It examines the input CDRs and determines whether they are appropriate for the systems.

4.6.2.4 Filtering

Some of the data fields in the various CDRs are optional, it should be understood which of these could be discarded to reduce the volume of data storage and transfer to the BS. The filtering module, as the name implies, simply filters the CDRs by removing non-billable or invalid CDRs (for e.g. a failed session) so that these are not forwarded to the BS. The data warehouse may however require these records. Hence, the filtering and explosion modules work hand in hand.

4.6.2.5 Rating

Rating may either be done in the mediator or in the Billing System device itself. The advent of GPRS packet-based data services means a move towards the rating models of the data-communications world, which involve various charging models, including duration-based, event-based, volume-based, and content-based varied by location, QoS etc. As a result, a separate billing engine may be required to avoid integrating too many functions into the mediation device. This is why rating is left to the billing device and not done in the mediation.

The mediator may constitute several other modules, which are operator and manufacturer specific, depending on the requirements of the operator and the capabilities of the manufacturer. These may include roaming, handling of Internet Data Records from Cisco routers, firewalls, mail and web servers, QoS facilitators etc. The ability to handle IPDRs is an important capability that needs to be a function of good mediation systems.

An example (*the Corporate User example*) of a typical UMTS service, which requires the services of a mediation device is described in section 4.3.3, and the roles played by the mediation device for this particular service is described below.

The mediation system must obtain usage details from each of the network elements, and consolidate them into an aggregate billing record (or, aggregate CDR). This aggregate CDR should contain at least the parameters illustrated in Table 2.7 (other parameters may be needed depending on the type of billing schemes to be accomplished).

Initiator customer ID	Initiator and terminator IDs identify customer and supplier

Terminator customer ID	accounts within the billing system database that must be billed for the transaction.
Debitor/creditor ID	Transactions can be billed to the initiator, terminator, or both depending on specific business logic or other transaction parameters. If the transaction goes beyond the boundaries of the IP provider, there may be the issue of inter-carrier settlements. This is especially true when QoS reservations span across multiple IP carriers.
Initiator Location ID/Terminator location ID Distance ID	Initiator and terminator location IDs are associated with the geographical location of transaction end-points. This information is used to produce a distance identifier, which can be used to charge differently for local, long-distance, international and inter-carrier service.
Time of Day/Day of week	These are important if traditional off-peak discounts are used.
Application ID	The Application ID identifies the type of application used in the transaction (e.g., web, voice, or mail). The application information can be obtained from the port numbers within the TCP/IP header (e.g., web traffic is sent on port 80), or can be obtained from the gateway or RSVP policy manager.
Reserved QoS ID	QoS (i.e., traffic specification) reserved for the transaction.
Outgoing bytes	Number of bytes generated by the station.
Incoming bytes	Number of bytes received by the station.
Start	Start time and date of transaction.
End	Transaction end time and date.

Table 4.4 Aggregate billing record (CDR)

To generate an aggregate CDR, the mediation system must first collect and correlate data from the various network elements. In the remote corporate user example illustrated above, the router, RADIUS server, RSVP policy manager, network management node, and routing tables each produce various usage records. Hence, apart from the SGSN and GGSN producing various CDRs of interest, the collection module needs to be capable of collecting usage data from these other non-GPRS/UMTS network elements in order to facilitate content billing. It should be able to extract all relevant data from every IP session from multiple sources at all layers of the network, from the physical layer to the application layer.

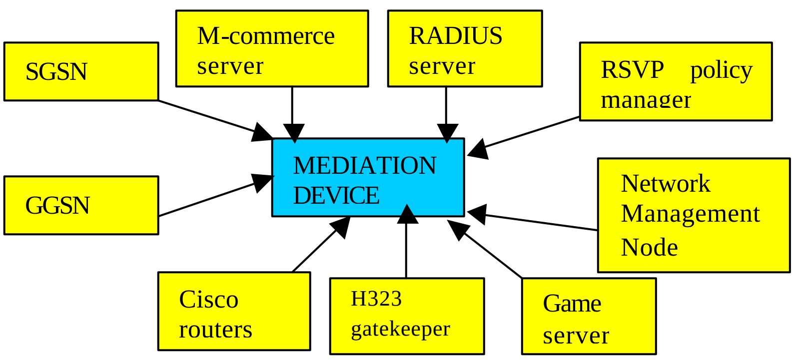


Figure 4.7 Diagram illustrating where charging information is captured

Appendix 4C explains what each network element involved in the *corporate user example* does, and the usage information produced. By carefully correlating the usage records, the mediation system can derive network usage details for each network transaction. It was earlier mentioned, that each CDR produced by the GPRS/UMTS network elements are identified by a Charging ID (CID). Hence, all CDRs generated from the same PDP context can easily be correlated. This CID however, unfortunately does not exist in the records produced by the non-GPRS/UMTS network elements i.e. the Cisco routers, RADIUS servers, RSVP policy manager etc. Hence, other means of identifying which records belong to which transaction must be implemented.

The Internet Protocol Detail Record Organisation (IPDR.org) has taken the initiative to define the essential elements of data exchange between IP network elements, and OSS/BSS. I.e. identify the information required to characterise an IP service (including e-commerce transactions) to the level of granularity desired by a BSS. IPDR and IPDR.org are very important and should no longer be constrained to legacy IP network providers. Mobile Telecommunication Network Operators and content providers should ensure that they have an influence in its developments to enable correlation of records produced by various network elements for each transaction. More information on IPDR and IPDR.org can be found in Appendix 4D or http://www.ipdr.org.

The mediation's final task is to consult the Customer Care and Billing (CCB) database to add account identifiers that are meaningful to the BS (e.g. customer, debitor, and location IDs). After incorporating the CCB information, the mediation system can forward the aggregate CDR to the back-end billing system. The aggregate billing record can now be rated and billed by the BSS depending on the type of contract with the customer. In some cases, based on the type of peering agreement with neighbouring carriers, this transaction may require settlement, transit, or access charges, also known as Interconnect charges.

4.7 Packet Based Charging Models

This section gives an overview of various Packet Based Charging Models, and discusses the requirements, limitations, and possibilities, which these models open up for Mobile Network Operators. Within some of these charging models, various pricing schemes exist. For example, priority pricing, smart-market pricing, and responsive pricing (discussed in section 4.7.3) are all volume-based charging schemes, but vary slightly from each other on the basis of QoS dependency, user perception, billing and accounting overheads, and network requirements. Various pricing schemes can also be considered as new means to control network congestion. These charging solutions mentioned above would partly or wholly apply to the services offered in future broadband wireless and fixed line systems. The ultimate aim of the Packet-Switched technology is to offer a sophisticated range of reliable data services in a cost-effective manner affordable by all.

Various evaluation criteria, for the various charging mechanisms will be used. These include

- Compliance with existing technologies
- Measurement requirements for billing and accounting
- Support for congestion control or traffic management
- Degree of network efficiency, etc.

Compliance with existing technologies is an important indicator of implementation complexity. This criterion is stricter than an evaluation of the scheme's implementation complexity in itself, as a scheme may be easily implemented, but non-compliant with existing technologies like ATM or IP, thus, precluding it from immediate deployment. Obviously, charging schemes that are compatible with existing technologies are easier to adapt than those requiring significant changes in the underlying network technology are.

The amount of *measurements required for billing and accounting* is a related criterion indicating the implementation complexity of the pricing schemes. Some schemes require no measurements at all, thus simplifying accounting and billing considerably. Other schemes require extensive measurements.

Support for congestion control or *traffic management* indicates the applicability of the pricing scheme for controlling network congestion by exploiting the price-sensitivity of users, thus inducing them to transmit less or during less congested periods.

The *degree of network efficiency* evaluates the expected utilisation levels in networks. While high utilisation levels are desirable for the network operators, they are not necessarily desirable for the user. Low utilisation levels imply availability of service. A highly utilised network, on the other hand, may have to deny service to some customers. To take the user's point of view into account, the *economic efficiency* of the pricing scheme has also been evaluated. Low utilisation rates do not necessarily indicate inefficiency. For example, even though household phone lines are only utilised on average 4% of the time, many households prefer to pay for additional lines for fax/Internet services rather than miss important calls.

Network congestion occurs in short time frames, so the pricing time frame reflects on the scheme's flexibility to respond to congestion periods. If a pricing scheme supports short time frames, (for example Smart-market pricing discussed in 4.7.3), the network operator can change prices in response to varying network conditions, thus, controlling network congestion. Short-term fluctuations of prices may however, not be desirable from the user's point of view. These criteria used to evaluate the various charging/pricing schemes is not intended to provide an absolute ranking of the charging schemes, but to summarise the main features, allowing the reader to establish a final evaluation based on his/her preferences. In particular, network operators and users may have totally opposing views on the value of some criteria. As a result, their preferred charging schemes will not necessarily coincide.

4.7.1 Duration based

As we discovered in chapter 1, and as we all are used to, in GSM and PSTN as well as all other Circuit-Switched communication technologies, charging is based primarily on the **duration** of the call. There are obviously other factors that affect the final invoice seen by the end customer, some of which are the time of the day/day of the week the call was made, the kind of network the terminating terminal is attached to, the distance between the calling parties etc. These other factors however, are dependent on the network operator and issues like interconnect charges also depend on whose network resources have been used during the duration of the call. For example, if a BT customer makes a call to a Telia customer in Sweden, then Telia charges BT for the use of their resources to make the connection possible. All these issues were discussed in section 1.8 and the reader can have a quick glance to remind himself of the important issues concerned with charging within GSM. The primary factor affecting GSM charging is the **call duration.** We are however, about to see the end of billing **solely** by units of time.

Within GPRS/UMTS, or indeed any Packet-Switched based service, the duration of the call, although still an important record, is no longer the primary parameter of interest. GPRS permits the GSM network to be directly connected to LANs, WANs and the Internet. In the same manner that a LAN user, ADSL users or cable modem users log on to the network, and are "always connected, always on-line", the GPRS/UMTS user will be "constantly" logged on to the network, but may or may not be actively using it at any given time. As a result, the concept of initiating a call in the telephony sense will be lost. The GPRS/UMTS user is always connected, always on-line. As a result, duration-based billing invented to fit traditional voice-telephony Circuit-Switched paradigm does not fit well into the GPRS/UMTS Packet-Switched concept. Charging by the duration of connection time would encourage users to log off after sending data thereby stifling usage, which is undesirable for both service providers and consumers. The goal for network operators and service providers is to encourage usage especially having spent billions of pounds on UMTS licenses alone not to mention network infrastructure and maintenance. Average Revenue Per User/customer (ARPU) has to at least double if network operators are to survive these huge investments. This can only be achieved by implementing new creative billing mechanisms discussed in the subsections below. These charging schemes, provide unique opportunities to enhance the profitability of the service, and to attract and retain customers. Although, duration–based charging has its pros, - it is easiest to bill and support and is a familiar concept for consumer markets, it has its cons. The point of the "always on, always connected" functionality is missed, control of network congestion cannot be accurately executed, and most importantly, network operators miss out on several other opportunities to maximise revenue on service provision.

Although, the user is "always connected", the duration of specific applications may be of interest to the operator. For example, within a PDP context lasting for 8 hours, a user may only use a voice application (make a voice call) for a fraction of the time, say 5 minutes.

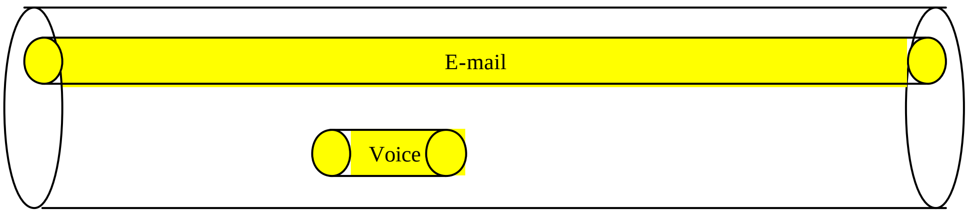


Figure 4.6 Duration of various applications within the same PDP context.

Since users are used to the duration-based concept of charging for voice coupled with the fact that the transfer of voice data is more or less continuous, the operator may wish to charge this particular application on a duration basis. To accomplish this, record producing nodes within the GPRS/UMTS network need to be able to capture the duration of such an application. This is achieved either by identifying the APN used, which may be different for different applications (configurable) or by using the Traffic Data Volume (TDV) containers produced in the SGSN and GGSN, if the same APN is used for several applications. Hence, a user connecting to the corporate network for e-mail and other office applications may connect via APN 1, while APN 2 is used if he wishes to make a voice call. The figure below illustrates this concept. 1a and 2a below represent "Activate PDP context Request Messages" while 1b and 2b are virtual connection paths to the GGSN after DNS resolution has taken place.

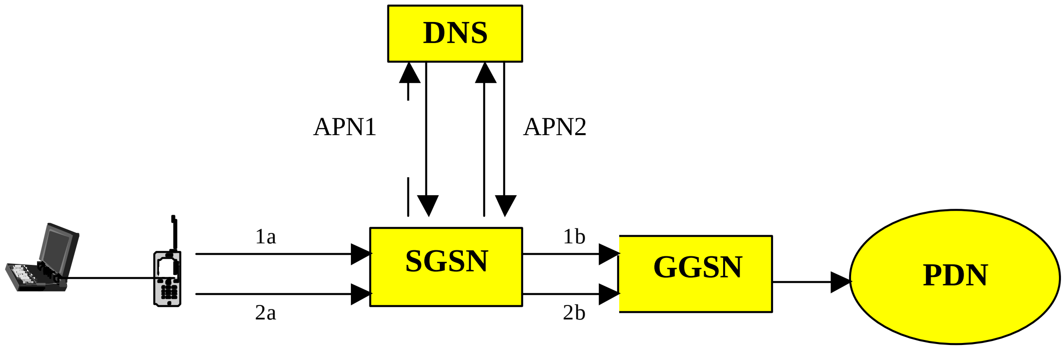


Figure 4.7 APN utilised is used to identify the application running.

If however, the same APN is used for several applications, as described in section 4.5.5.6 above, then the CDRs produced at the SGSN and GGSN have parameters known as Traffic Data Volumes (TDV). These are in turn are made up of several parameters, one of which is the Quality of Service profile. Hence, a user currently running an e-mail application may decide to run a voice application. To do this, the QoS profile[10] in the TDV field will change. The **point in time** during which these changes occur is recorded. The operator may then use these records to identify the duration of various applications as shown in the figure below.

[10] The UMTS QoS class changes from background to conversational and the delay and throughput requirements both change too

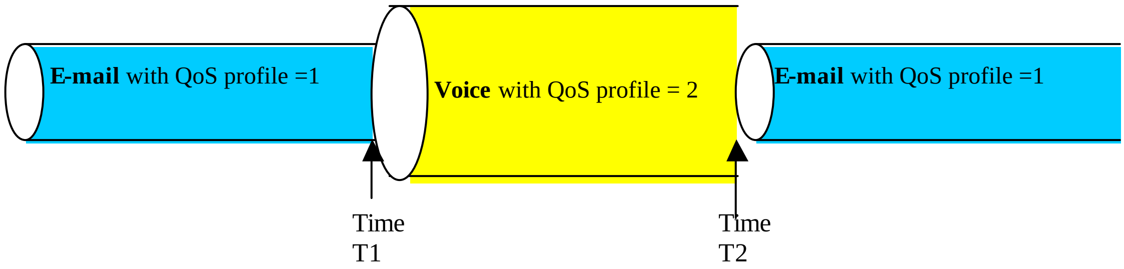


Figure 4.8 QoS profile within the Traffic Data Volume field of the CDR is used to identify the duration of various applications.

It should be noted that the QoS profile might change even during a single application. In this case, the operator can most likely still recognise that the same application is being run and may alter the tariffs based on the QoS provisioned.

Duration based Charging	
Advantages	**Disadvantages**
Easy to bill and support	Network Operators miss out on other opportunities to maximise revenue
Compliant with existing Billing Systems	No means of implementing Network congestion control
Most familiar concept to consumer markets.	"Always on, always connected" functionality is missed, i.e. users are discouraged from being constantly logged on.
	May be meaningless as user may be logged on without necessarily utilising network resources.

Table 4.5 Advantages and Disadvantages of Duration based charging

4.7.2 Subscription-based

Some services may be charged by subscription only. Subscription-based charging is also known as flat rate charging. This will imply that the CDRs need not be rated, and may not even need to be collected. This charging mechanism is familiar to many of us. It is the favoured charging mechanism employed by some existing ISPs and most digital television providers and is independent of the duration of connection or the amount of data transfer. Examples of the pricing structure from various packet data services around the world may be helpful in giving some key pointers towards strategies for GPRS/UMTS subscription-based pricing.

ADSL[11] is an example of such a subscription-based packet data service already in existence in some parts of the world. The interested reader may refer to Appendix 4E for a brief summary of ADSL. In the US, telecommunication companies offering ADSL, have set three flat-rate monthly tariffs as shown in the table below, depending on the QoS (in particular the maximum throughput) subscribed for by the user.

	Subscriber class		
Cost	$30	$60	$80
QoS subscribed for	256Kbps	512Kbps	768Kbps

Table 4.6 User classification based on the QoS subscribed for

[11] ADSL connections are point-to-point-dedicated circuits meaning that its users are always connected.

This phenomenon of classifying users based on the QoS subscribed for is strongly supported by many data service providers, and is most likely going to play an important role in charging for UMTS services. DSL data traffic is billed at a flat rate monthly subscription, but questions still remain about how Voice over DSL (VoDSL) will be billed. Will it be treated like traditional telephony calls based on duration or will it be treated like DSL data traffic. Before DSL services are actually billed, data must be collected from the appropriate network elements. These CDRs will most likely be collected on the class 5 switches just as they are for traditional PSTN traffic. ADSL is fast becoming popular in Europe too and France telecom has adopted a similar charging strategy to that of the US, i.e. flat-rate monthly subscription fee based on the throughput subscribed for by the user. France telecom has however specified only two different classes of subscribers (500Kbps/128Kbps downstream/upstream and 1Mbps/256Kbps downstream/upstream).

This charging mechanism, however has a couple of **drawbacks**:

- Customers may be treated differently depending on the applications used.
- Network congestion control or traffic management cannot be implemented.
- Operators may not maximise their revenue possibility and

Flat-rate subscription charging takes no consideration of the amount of data being transferred and hence ignores the demands on network resources. For example, a consumer browsing the web is charged equally as another customer using the same service for megabyte-hungry applications, assuming they have subscribed for the same QoS profile. A simple way round this unfairness is to tie the flat-rate monthly charges to a maximum megabyte allowance. A subscriber who uses up this monthly allowance pays an additional charge per megabyte transferred. This is described in more detail in section 4.7.4 below.

Also, flat rate pricing does not allow the network to influence users' transmission behaviour, as there are no incentives for users to alter their transmission behaviour to support the network operation. It is therefore unsuitable for congestion control or traffic management. Since flat rate pricing fails to exchange the appropriate signals between users and network operators/service providers, it may lead to congestion and insufficient investment. On the other hand, the network may be running at very low utilisation levels, without the network operator being aware of this. This unawareness may lead to equally bad future network dimensioning. Flat rate pricing does not explicitly support individual QoS guarantees to the user, unless various flat rate prices are imposed on various customer classes as described in the ADSL examples above. Operators may not maximise their revenue possibility, but analysis of existing flat rate pricing shows that below-average users subsidise above-average users, thus allowing network operators to obtain sufficient revenue.

With respect to the evaluation criteria mentioned above however, flat-rate subscription this charging has several desirable advantages. It is simple and convenient, and makes no assumptions about the underlying network technology. Since charges are unrelated to usage, no measurements are required for billing and accounting purposes. It is also socially fair in the sense that no distinction is made between rich and poor provided that the flat-price is affordable by all.

Flat rate subscription based Charging	
Advantages	**Disadvantages**
Simple and convenient	May be unfair depending on applications used
No measurements required making it compliant with existing technology	No means of implementing Network congestion control if no records are collected
Familiar concept to consumer markets especially among Fixed line and Digital TV customers	Network Operators miss out on other opportunities to maximise revenue

Table 4.7 Advantages and Disadvantages of Flat-rate subscription charging

4.7.3 Volume Based

Wireless network traffic (most notably GSM) has grown exponentially in recent years, and there seems to be no dispute that this trend will continue in the future, especially with the introduction of UMTS, enabling wireless Remote LAN Access, essential for corporations wishing to compete effectively in the global marketplace. The roll out of applications with ever increasing resource requirements (e.g. Napster for download of MP3 files), the recent shift toward real-time applications (e.g. Video conferencing) requiring strict QoS guarantees from underlying network infrastructure, and the constant increase in wireless user base all suggest a scarcity of network resources. **Usage-based** charging can thus be used as a means to resolve the allocation of these scarce resources. Network resources are a key factor to deciding pricing. In the mid 90s, a UK mobile operator in an effort to attract customers before Christmas when sales are at a peak, was offering free calls on Christmas day, resulting in network traffic over and above its capacity. As a result, the network crashed and its marketing director was fired. Also, from the social-welfare point of view, high prices would prevent the growth of user using the services, and thus make network services a luxury, in extreme cases. On the other hand, providing free/cheap network services leads to “tragedy of the commons”. Hence, care needs to be taken to ensure that the right prices are set. However the actual process of pricing is outside the scope of this project.

The “Always on, Always connected” functionality of a Packet-Switched world, results in enormous connection duration. This coupled with the fact that network resources are used only when data is being transmitted/received and not necessarily during the entire duration of the connection strongly suggests that duration-based charging is meaningless. Besides, the geographic destination of a particular Packet-Switched connection may be unknown. Hence, the parameters that affect the charge of a Circuit-Switched voice/data call are potentially irrelevant on the Packet-Switched domain. A “simple” charging mechanism that uses the data volumes as the main charging parameter has thus been suggested. This charging mechanism poses a number of challenges on network operators. These challenges include an increased number and level of complexity of CDRs as opposed to the CDRs generated for CS calls. This has already been discussed in section 4.2.1. These CDRs need to include a number of new parameters, which are non-existent on the CS domain. An example of such a parameter is the QoS, which is itself composed of a number of attributes. Refer to 4.5.5 for QoS. The network requirements needed to enable recording and processing of these new CDRs are numerous, but most of these requirements have already been dealt with by standards bodies, network infrastructure vendors, and network operators and have been discussed in sections 2.4 to 2.6 above. The benefits resulting from the ability to base charging of GPRS/UMTS services on the volume of data traffic include

- Ease of establishing profitable GPRS/UMTS business
- Management of the network bandwidths simplification of scarce resource-allocation.
- Charging the customer for what he actually uses.

A problem associated with volume based charging however is that consumer users may find it hard to relate to data volume. What does 10 MB mean? Is it sufficient to view a web page or 30 pages? Can I download an MP3 audio file or 20 audio files if I’m limited to 10MB? Is the price I’m currently paying for Internet access at home comparable to what I’ll be paying for a similar amount of web surfing on my GPRS/UMTS handset or PDA? Besides, subscribers are often unable to predict or measure their usage. Some web pages have heavier graphics and are therefore megabyte-hungrier than others.

<table>
<tr><th colspan="2">Volume based Charging</th></tr>
<tr><th>Advantages</th><th>Disadvantages</th></tr>
<tr><td>Possibility to maximise revenue</td><td rowspan="3">Prior unawareness of web page and multimedia file size</td></tr>
<tr><td>Charges the customer for what they actually use</td></tr>
<tr><td>Easily resolves scarce resource allocation and good for management of network bandwidth</td></tr>
<tr><td>A new mechanism but nevertheless compliant with existing technology. Nodes responsible for data volume count have been standardised</td><td>Alien to consumer markets</td></tr>
</table>

Table 4.8 Advantages and Disadvantages of Volume based charging

Three subclasses of volume-based charging are briefly discussed below.

With ***priority pricing*,** users are forced to indicate the value of their traffic by selecting a priority level, an important QoS parameter. During periods of congestion, the network can then carry traffic depending on the priority level indicated. Priority pricing requires a priority field in every packet header. Such a field is provided in IPv6 (and possibly IPv4?), thus ensuring compatibility with such networks. Priority pricing requires extensive measurements for billing and accounting to keep track of the priority level of each transmitted packet for each user. Individual QoS guarantees cannot be given and in the worst case scenario, low-priority traffic may even be dropped. Obviously, high-priority packet owners are charged higher than owners of lower-priority packets are. Since prices are a function of the relative importance of each packet, priority pricing raises the economic efficiency of the network. Any dropped packets will have a lower priority level and thus a lower value to the user. With a large number of priority levels and/or high prices for high priority levels, traffic from poor users (in a relative sense) may be starved. Under flat pricing, users are not faced with monetary incentives and thus will demand the highest priority for both loss and delay, while priority pricing instigates price sensitivity and thus, improves network performance. It is then easier for the network to set prices such that the overall user satisfaction is higher under priority pricing than subscription/flat rate pricing. Priority pricing assumes that resources will be scarce or that networks will be running at high levels of utilisation.

***Smart-market pricing*,** also a usage-based charging mechanism, is slightly different from priority pricing, in the sense that it focuses on the issues of capacity expansion and the social cost imposed on other users. It introduces an extra usage-charge when the network is congested. This is similar to the paradigm of peak and off-peak rates in present days voice telephony with the major difference being that network congestion can occur anytime of the day in the world of data communications. Hence peak rates can occur anytime of the day, depending on the current situation of the network. Here, the charge is determined through an auction. The user associates a price with each packet, carried in the packet's header, communicating the user's willingness to pay for transmission. Typically, such a price would be derived from default values, with the option to the user of overriding the default value for special transmissions. The network collects and sorts all the bids, determines a threshold value and transmits all the packets whose bid exceeds the threshold value. The threshold value is determined by the network's capacity and represents the marginal cost of congestion. Each transmitted packet is then charged this marginal congestion cost, not the value of the bid. The introduction of an auctioning mechanism leads to non-compliance with existing technologies. Smart-market pricing requires significant technical changes to protocols and networking hardware, which is a considerable cost. For each transmitted packet the user's billing records need to be updated. This also implies considerable overhead, even if the auctioning mechanism is only applied during congested periods. Smart-market pricing is probably a pricing scheme that may be considered in mature markets to gain a competitive edge and will most likely not be implemented in initial charging mechanisms due to its complexity and cost. The scheme also does not provide the users with service guarantees or even a guarantee of transmission. Like priority pricing, it only ensures that the packets are transmitted according to their relative priority, determined by the bid. Prices may deter poor users from using the service, thus creating a case for government regulation. On the other hand, smart-market pricing encourages both network and economic efficiency. Allowing users to re-negotiate the tariff periodically is important in cases where user traffic is highly variable over time. A periodic re-negotiation of the tariff would reveal to the network the anticipated load over a short-term interval, and thus improve the network's resource use. However, such re-negotiations would only make sense if the short-term interval exceeds the round-trip delay. To account for the additional complexity arising from the re-negotiation, the service provider/network operator may decide to bill the user for a set-up charge for every re-negotiation. This would encourage the user to re-negotiate the tariff only if the expected savings exceed the set-up cost. Resources are again assumed to be scarce, at least over short time intervals, due to the introduction of bandwidth-intense applications. The bids carried in each packet could also be used to facilitate routing decisions. Packets with high bids could be routed over shorter paths, whereas packets with low bids may be routed through longer paths.

Yet, another subclass of usage-based charging is ***responsive pricing*,** which is in many respects similar to smart-market pricing, particularly due to the fact that it only comes into operation during periods of congestion. This pricing scheme is based on the fact that users are adaptive and respond to price signals. No auctioning mechanism is introduced. The network simply increases the price for network resources in the case of high network utilisation. Adaptive users then, reduce the traffic offered to the network. Similarly, in case of low network utilisation, the network decreases the price and the community of adaptive users increases their offered traffic. In this way, adaptive users do not just increase the network efficiency, but also economic efficiency. This is the pricing mechanism employed today by *"easy everything"*, UK's market leading Internet café. This mechanism requires network control nodes to keep track of the current utilisation level of the network, and thus the need to keep detailed billing and accounting records, therefore incurring significant overhead. Also an IN subsystem, capable of mediating the current tariff per packet to the user may be required. This mechanism may however, not be a realistic approach, as users prefer a long-term assurance of the intended tariff to be applied on their data packets. It may also introduce instability, as experienced by traditional congestion-control functions. Upon declining network utilisation, the network would reduce prices. This encourages adaptive users to transmit, thereby causing congestion.

4.7.4 Combination of Subscription and Volume-based or duration-based

More often than not, volume-based charging will be used in conjunction with other charging mechanisms, particularly subscription-based charging. Hence, rather than charging the subscriber for each KB or MB of data transferred (with QoS, time of day, etc playing a roll on the tariff), the user will be charged a monthly fee, which entitles them to a particular amount of data as discussed below. The monthly fee paid by the subscriber will depend on the **QoS** requested for by the user. Users can then be classified as either, Gold, Silver, or Bronze users depending on the QoS profile they've subscribed for, similar to the ADSL example in 4.7.2. This QoS profile will include all the QoS attributes but the most notable to the customer would be the **throughput** or **bit rate**. An example of such a subscriber classification is shown below.

	Subscriber Class		
	Gold	**Silver**	**Bronze**
Subscription fee (£)	45	36	18
Maximum Throughput subscribed for	768	512	256

Table 4.9 Subscription fee based on QoS profile

Every user may then on a per-service or per-session basis request a particular QoS profile. The QoS profile subscribed for is stored in the HLR and is obtained by the SGSN at GPRS Attach. A user is then granted the QoS profile requested for, as long as this does NOT exceed the QoS profile subscribed for. Hence, a Bronze user may request for a throughput of 256kbps for a particular application. If however, he requested for 512kbps, he will not be granted this QoS as 256kbps is his maximum allowed throughput. On the other hand, a Gold user wishing to be charged less for a particular session may request for a lower QoS profile than that subscribed for. This particular scenario may however be implemented as follows. Users belong to no particular class and choose on a per-session basis which class they require for that particular session. Hence, the same user may request a Gold service for multimedia file download and a Bronze service for e-mail application. That is, all subscribers pay a set monthly subscription and are not differentiated and thus, belong to no particular subscriber class.

	Gold	**Silver**	**Bronze**
Guaranteed Throughput	512	256	128
Maximum Throughput	768	512	256

Table 4.10 Guaranteed Throughput based on QoS profile

Subscribers are then only charged for the volume of data they transfer, and the QoS negotiated. The QoS negotiated depends on that requested for by the user for that particular session. Hence, every user is given the freedom to choose on a per-session basis what QoS profile he requires. There is no notion of a “Silver subscriber” or a “Gold subscriber”. This option however, involves more CDR recording and processing than the other. For ease and convenience, only three classes of QoS profiles have been defined. A network operator may however choose to have as many profiles as he wishes to with increasing discounts as the QoS steps down from high to low. The number of profiles possible is however limited by standard definitions and network operators may further limit this to suit their customer needs and reduce complications.

The initial users of GPRS services are anticipated to be corporate employees. Typical applications to be used by these are **Remote LAN Access (RLA)** and **Wireless Internet**. These services are likely to be charged, on a subscription and data volume basis. With more than 137 million business users world-wide estimated to be involved in some form of remote work by 2003, remote access is no longer an option but a necessity for corporations wishing to compete effectively in the global marketplace, says Concert, BT/AT&T global venture. [31]

A survey done by Northstream [19] suggested that GPRS Remote LAN Access and Wireless Internet should be priced somewhere in between the cost of use (from the consumer’s point of view) for current Circuit-Switched data services (GSM 9.6kbps & 14.4kbps and fixed-line ISDN 128kbps). 14.4kbps data services were used as a benchmark on the high side and the fixed-line ISDN on the low side. The table below shows the suggested price for the different services.

RLA

Connection fee	125
Subscription fee	40
Allowance in MB	50
Overage usage fee/MB	0.5

Wireless Internet

Connection fee	0
Subscription fee	25
Allowance in MB	30
Overage usage fee/MB	1

	Charge (USD)	
Monthly Usage (MB)	**RLA**	**Wireless Internet**
20	40	25
40	40	35
100	65	95
200	115	195

Table 4.11 Monthly revenue from RLA and Wireless Intemet users for 40, 100, and 200 MB limits

BT Cellnet has also developed the Remote LAN Access (RLA) Service for corporate users wishing to access their company Intranets and a similar charging scheme has been defined, i.e. subscription and volume based charging. This RLA service just as in the case above incurs a connection charge for installation of Host Servers at the corporate site, end-user software etc. The connection between the user’s data network and the BT Cellnet GPRS network could either be via leased circuits providing a dedicated 2Mbit/s circuit, able to carry a mixture of voice, data and SMS services or indirect access via the Internet. BT Cellnet has also proposed a tariff structure intended for other initial GPRS services, which is again based on a combination of monthly subscription and data volume.

	Cost (£)
Charge per user per month	17.63
Connection charge	Nil
Usage Allowance	50 MB per user per month. See Note 1
Overage usage fee per MB	1.50

Table 4.12 BT Cellnet’s proposed tariff structure intended for initial GPRS services.

"Note 1: The usage limit of 50 Megabytes per user per month has been set at a high level to allow maximum use of this new technology during the introductory period following this launch only. This will enable Cellnet to obtain the best feedback on likely usage patterns to allow us to design longer-term tariff propositions. The usage limit should not be taken as an indication of the likely level of permitted use in those longer-term tariffs."

The charging scenarios described above are quite acceptable for the corporate market. Corporate employees are used to dealing with megabytes and kilobits and are familiar with the concept of 'always-on' packet-based services. This concept is however alien to consumers who are used to paying for services by the minute or in the case of SMS, by the event. For this reason, operators may wish to charge on a subscription and duration basis as opposed to subscription and volume basis.

Hence, Deutsche Telekom charge a fixed monthly subscription charge with a limited connection time offered to the customers (for example 50 hours) for their ADSL services. This is probably based on the assumption that within this specified duration, only a limited amount of traffic can be transferred. Hence, the customers are in essence, charged on the basis of "volume of traffic transmitted".

The charging mechanism adopted by Deutsche Telekom may thus, be the preferred method initially, especially for consumer users as it could be hard for them to relate to traffic data volume. Hence, subscription based charging coupled with either volume-based or duration-based, depending on the user category is probably going to be the most widely adopted charging mechanism for GPRS/UMTS data services like RLA and Wireless Internet. That is, services that have nothing to do with third party content providers.

<table>
<tr><th colspan="2">Combination of Subscription and Volume or duration based Charging</th></tr>
<tr><th>Advantages</th><th>Disadvantages</th></tr>
<tr><td>Allows flexibility and provides possibility to maximise revenue</td><td rowspan="4">Volume based charging is alien to consumer markets</td></tr>
<tr><td>Charges the customer for what they actually use or what they prefer to be charged for</td></tr>
<tr><td>Management of network bandwidth is possible</td></tr>
<tr><td>Easily incorporates QoS profile into charge</td></tr>
</table>

Table 4.9 Advantages and Disadvantages of Volume based charging

4.7.5 Event-based

Some services may be charged on an "event" basis. This requires the CDRs to have a fixed charge applied, based on the service. The Short Message Service (SMS) is currently based on this charging mechanism and is likely to continue to be charged in a similar fashion in future. This is due to the fact that SMS messages are transmitted and/or received over signalling channels and not traffic channels. Hence, the limitations in the number of octets that may be transmitted. Neither the data volume nor the Quality of Service is of any importance as no traffic resources are used. As mentioned in chapter 1, SMS messages are transmitted during quiet periods i.e. when little or no signalling is taking place and as such, its arrival time at the destination is inherently unpredictable. SMS messages are delivered on a best-effort basis. Due to all of the above-mentioned factors, this is the charging mechanism most appropriate for this service. In future, it may be that those urgent messages with a required guaranteed arrival time and/or greater character capacity may be transmitted over data traffic channels and as such may cost more than normal SMS messages. As GPRS/UMTS develop, more and more services are likely to be charged on an event-basis. For example, the CarMaker service example described in section 4.3.4 may be charged on an event basis.

Let us have a closer look at this example. The reader may refer to 4.3.4 to remind himself of the service scenario.

This deal gives CarMaker a competitive advantage and will therefore be happy to install this UMTS handset into their new vehicles. For Fab Telecom, the deal gives increased network traffic and hence increased revenue and may therefore not charge CarMaker for the service. We can further assume that the garage is owned by CarMaker and that any successful bookings will bring forth additional revenue. For the user, the system manages his car servicing system automatically and he will therefore be willing to pay for the service. Since the messages transferred are small in volume and duration, the operator is most likely to maximise his revenue by charging on an event basis. Hence, each time a successful booking is made, the user is charged for the service. The Lifestyle Management System and diary providers would probably receive a settlement charge from the operator, otherwise known as revenue sharing. This charge is for using their system and services to enable the booking system to function properly.

4.7.6 Content value based

The traditional use of monthly subscription flat rate charges in telephony billing is a model that operators feel comfortable with. How tempting, therefore, to apply that same model when introducing new 3G services – after all, it works today and would save further development costs. This is however a short-term view that could seriously damage the longer-term profitability of many network operators, reduce innovation and restrict customer choice [8]. Furthermore, charging mechanisms, which charge users by data volume, duration, and/or subscription almost invariably fail to reflect the inherent value of the information, i.e., the content of the data packets. The particular content of a data packet may be extremely important to one user, but less so to another user. The value of a data packet to the same user may also vary with time. Content-value based charging aims to take advantage of this to further improve revenue from services provisioned.

"Billing for content is going to be in every operator's interest. Right now, fixed line operators are facing this challenge with the current Internet model. Revenues from access are steadily decreasing, with more and more free services being provided. So the critical question for operators is: Will the increase in traffic be enough to compensate for the loss of margin?" Those operators not prepared to take this gamble will have to find a way of identifying what is passing over their networks, so they can maximise the opportunities to base billing on the value of the product or service provided and not just basic access provision. The ability to identify these packet contents is essential if Service Providers wish to charge content partners - banks, restaurants, cinemas etc - in order to maximise their revenues from GPRS and UMTS services. Without knowing what is crossing their network, an operator will have no chance of adding revenues from the content of products and services, which is where the real value of future services lies. Thus, the need for Network Operators, to be able to identify the 'content' of packets passing across their networks, and being delivered to their customers, is paramount. The mediator's ability to handle Internet Data Records, otherwise known as IPDRs [12] is of utmost importance if the operator wishes to be able to charge for content, acquire commissions on accomplished transactions, and identify applications. Application identification may be used to implement differential charging. Refer to Appendix 4D and/or [29] for more information on IPDR and IPDR.org

Since the main focus of GPRS/UMTS operators is to give mobile users access to IP networks and associated IP-based services, data records pertaining to IP data usage needs to be collected and merged with GPRS/UMTS usage data for transfer to the Billing System. However, data records related to external IP network usage is unfortunately not produced at the SGSN or the GGSN. The SGSN and GGSN are also limited by their inability to capture and record the **contents** of the data packets transferred through them. Another significant limitation of the GSNs is their inability to identify the applications used (i.e. whether it is voice, data or video) and their incapability to record data related to transactions accomplished (e.g. financial transaction). Hence, the important issue related to content billing is to first of all **identify where to collect the data from,** and then to **ensure that the data collected can be correlated** using some kind of Transaction Identification. This is part of ongoing work initiated by IPDR.org and is discussed briefly in Appendix 4D. As we discovered earlier, in section 4.6.2, diverse IP usage data is produced at the various network elements and hence can be collected from these elements. In the remote corporate user example illustrated above, the router, RADIUS server, RSVP policy manager, network management node, and routing tables each produce

[12] IPDR is the IP equivalent of CDRs generated by telephone switches (MSC in GSM).

various usage records, hat may help in accomplishing the shortcomings (e.g. application identification) of the records generated at the SGSN and GGSN.

Content providers will be looking to sell their content as opposed to the airtime required to deliver the content. Content providers may be selling their content either to the operator or directly to the end user. Hence, content as well as network costs should be a component of mobile data pricing. However Content based pricing has so far been uncertain and only few operators have been able to build content into its mobile data charging mechanism. Questions such as “Who should pay?” “Are consumers interested in paying for content?” etc have been raised. If the content provider is to be charged, the mobile network operator may have no contract or mechanism for tracking the sender. On the other hand if the receiver is charged for the service, how is the receipt of unsolicited messages controlled, i.e. how do network operators control abuse of push services?

A probable solution for this may be for network operators to introduce some kind of Application Alliance Partners scheme so that content/push service providers can be tracked easily. This scheme allows members of the Alliance partners only, to provide push services to end-users. Application providers register with the network operator and are provided with a code/password to enable the delivery of these push services. This is in accordance with the example scenario discussed in section 4.3.2, whereby the network operator charges the content provider for the delivery of push services. Consumer users are not charged for it. Advertising companies rely on the consumer to buy their product/service just as any other form of advertisement. Similarly banks and other similar service providers may register with the mobile network operator so that the operator’s subscribers can access their services via their GPRS/UMTS handsets. These content partners, banks etc pay a monthly charge to the network operator for this service or may be charged on a per-use basis depending on contractual terms agreed on between the concerned parties. This charging architecture is similar in concept to that adopted by NTT DoCoMo in Japan for its I-mode[13] services. Refer to Appendix 4F for a brief summary of I-mode. NTT DoCoMo has set an example of what is possible with content. DoCoMo, having redesigned a handset on an ordinary cellular network to accept a version of HTML, is now delivering a variety of new services – I-mode services - and is charging content providers both on a volume basis and on a content basis. The charging strategy used for I-mode services is a mix of basic service access and usage-based charges. Usage-based here implies both **volume- and content-based**. Applications offered include a daily cartoon sent to subscribers for $1 a month, with the opportunity to SMS (Short Message Service) a particular cartoon to a friend for a small additional sum. The cartoon subscriber base is 750,000 and growing rapidly. Other examples of applications and their cost can also be found in Appendix 4F. For this particular example (cartoon example), where the content provider generates revenue for the operator, the operator rather than charging the content provider, shares their revenue with the content provider concerned. Hence, we have identified two classes of content providers.

- Those who push advertisements and rely on consumers responding to these advertisements and
- Those who provide services ‘beneficial’ to the consumer e.g. cartoon providers or game providers

The first class of content providers are charged by the mobile network operators, while the latter case is set to increase in popularity among application providers, who rather than being charged by the operator, charge the operator for their services. For example, ‘Picofun’ with its fantasy multi-player picofootball game has struck several deals with European WAP service providers. These games help in generating revenue to the operator as the game vastly increases airtime. These WAP cellular operators pay an up-front license fee to Picofun and revenue generated from these WAP calls are shared between the operator and Picofun.

According to our evaluation criteria, content billing is semi-compliant with existing technologies, as on the one hand, it may require additional nodes for handling of push services and advertisements. It also requires an extensive data collection facility from numerous network nodes and platforms as well as building alliances with content partners, banks restaurants, theatres etc, which may involve a bit of record storage and all the additional costs involved with building partnerships. On the other hand however, the technology required for this data collection already exists and many mediation and Billing System vendors are aware of the diverse data requirements needed for content billing and are developing mediation systems to cater for this requirement.

[13] NTT DoCoMo, Japans market leading mobile network operator introduced in February last year, a new service known as I-mode, which enables its subscribers to access a wide variety of data services.

4.8 Summary of Packet-Switched Charging Models

Just as various charging mechanisms are employed for various GSM services, - duration based for voice and event based for Short Message Service, so will the wide variety of GPRS/UMTS services be charged differently. In other words, not all services will be charged using the same charging mechanism. Various mechanisms will be employed for various services. For example, the transfer of voice and video data is more or less continuous, thus requiring network resources more or less for its entire duration. Consequently, we can expect voice and video to be charged on a duration basis, whereas other services, such as, file download or fax/email applications may be charged on a volume basis, i.e. charged per MB of data transferred.

Moreover, some applications like voice and video applications require bi-directional latency and bandwidth guarantees, whereas file download and fax/email applications do not have the latency constraint, but may require significant bandwidth depending on the resource (file) being accessed. Hence, QoS plays a significant role in most new charging mechanisms.

A problem operators may face if a combination of charging mechanisms is used is that they may have to reduce their tariff on certain services for these services to be charged at comparable prices if another charging mechanism had been used. An example would clarify my point.

According to the charging mechanisms described for Wireless Internet, described in section 4.7.4, 100MB of data traffic costs $95.00, which is equivalent to approximately £0.62 for 1MB (using an exchange rate of £0.65 to $1.00). Mobile Network Operators in the U.K today charge £0.05 - £0.10 for a 160-character SMS message. One character is one byte, which means that the price of transferring 1MB would be £312.50, which is astronomical compared to the £0.62/MB suggested above.

The table below illustrates which charging schemes in my opinion will be employed for various UMTS services.

	Charging Mechanisms for GPRS/UMTS Packet Switched Services					
Services	**Subscription**	**Event**	**Volume**	**Duration**	**Content**	**QoS**
Normal Voice Telephony	–	✗	✗	–	✗	✗
ISP Access	–	✗	–	✗	✗	–
Corporate Network Access	–	✗	–	✗	✗	–
Mail Server	–	✗	–	✗	✗	✗
Broadcast Service (News, weather, Stock exchange etc)	✗	–	✗	✗	–	–
Video on Demand	–	–	–*	–*	–	–
Telemetry	–	–	✗	✗	✗	✗
Games	✗	✗	✗	–	–	–
Location-Based Services	–	–	✗	✗	–	✗
Advertisement	✗	–	✗	✗	–	✗

* Volume-based and Duration based charging are most likely not going to be used simultaneously. In this case, charges will be based either on the size of the video clip or on its duration.

_ Will be employed

✕ Will NOT be employed

5 Conclusion

Charging GPRS/UMTS packet based services by the duration of connection time would encourage users to log off after sending data thereby stifling usage, which is undesirable for both service providers and consumers. The goal for network operators and service providers is to encourage usage especially having spent billions of pounds on UMTS licenses alone not to mention network infrastructure and maintenance. Average Revenue Per User/customer (ARPU) has to at least double if network operators are to survive these huge investments. This can only be achieved by implementing new creative billing mechanisms. These include

- Flat rate subscription based charging
- Volume based charging
- Combination of subscription based and volume based or duration based.
- Event based Charging
- Content based Charging

Each of these various charging schemes have been evaluated on a number of criteria including

- Compliance with existing technologies
- Measurements required for billing and accounting and
- Congestion Control or Traffic Management

For example, flat rate subscription charges are independent of data volume usage and hence, no measurements are required for billing and accounting purposes. This may seem ideal for the network operator as it is easiest to implement but there are several disadvantages associated with it. Since flat rate pricing fails to exchange the appropriate signals between users and network operators/service providers, it may lead to network traffic congestion and insufficient investment. On the other hand, the network may be running at very low utilisation levels, without the network operator being aware of this. This unawareness may lead to equally bad future network dimensioning. Worst of all, the operator may not maximise their profit if flat rate subscription is implemented.

On the other hand, with priority pricing,[14] users are forced to indicate the value of their traffic by selecting a priority level, an important QoS parameter. This enables the network to make the right choice of which packets to carry during periods of congestion. Priority pricing requires a priority field in every packet header. Such a field is provided in IPv6, thus ensuring compatibility with such network technology. It however requires extensive measurements for billing and accounting to keep track of the priority level of each transmitted packet for each user.

Just as various charging mechanisms are employed for various GSM services, - duration based for voice and event based for Short Message Service, so will the wide variety of GPRS/UMTS services be charged differently. In other words, not all services will be charged using the same charging mechanism. Various mechanisms will be employed for various services. For example, the transfer of voice and video data is more or less continuous, thus requiring network resources more or less for its entire duration. Consequently, we can expect voice and video to be charged on a duration basis, whereas other services, such as, file download or fax/email applications may be charged on a volume basis, i.e. charged per MB of data transferred. Moreover, some applications like voice and video applications require bi-directional latency and bandwidth guarantees, whereas file download and fax/email applications do not have the latency constraint, but may require significant bandwidth depending on the resource (file) being accessed. Hence, QoS plays a significant role in most new charging mechanisms.

Incumbent GSM operators have long lived on revenue generated by being a ‘carrier’ of voice and data. Mobile Network Operators of Packet-Based services can however not rely on this as their sole source of revenue. This coupled with the fact that increased data services and applications associated with

[14] Priority pricing is an important subclass of volume based or duration based charging. It is however important to note that it can only be employed for packet based services.

UMTS open up new potential revenue possibilities for mobile operators, have resulted in operators looking out for additional sources of revenue, some of which are listed below.

- Portal-type service provider to enable the acquisition of financial m-commerce transactions and
- Provider of personalised and location-based push services

Provision of portal-type services enables revenue from purchases made by subscribers, while the provision of personalised and location-based services generate revenue from business partners either via commission or advertisement.

As personalised and location-based push services are speculated to be one of the UMTS killer applications, so does the importance of recognising the content of packets transferred over the UMTS network increase. The value of the packet content may vary from one user to the other and may also vary with time. Content value based charging aims to take advantage of this to further improve revenue from services provisioned. The task of identifying the packet contents is not exactly straightforward and several network nodes (including non-UMTS or external data network nodes) need to be involved. The records produced by each of these nodes need to be standardised[15] so that they can easily be correlated for further processing. This standardisation process is important and is under development by the IPDR.org in conjunction with several major players in the telecommunications and data communications sector. It is important that Mobile Network Operators and Content Providers are involved in the development of IPDR. The IPDR record format identifies the information required to characterise an IP service.

The volume and complexity of CDRs produced by GPRS/UMTS network nodes is far higher than those produced for voice and Circuit-Switched data services in GSM. Besides, several other external nodes[16] are involved in Data Record production, especially if content based charging is to be employed. As a result, a device capable of collecting the various records (UMTS CDRs and non-UMTS IPDRs) from disparate network nodes, processing the records and forwarding them onto Business Support Systems like the Billing System is no longer an advantage but a necessity. This device is known as a Mediation device and is key to the successful implementation of packet based charging models. The mediation should also be flexible and scalable to match the potential market uptake of GPRS/UMTS services.

[15] 3GPP have standardised Call Detail Records produced at the SGSN and GGSN. Other record-producing nodes outside of the GPRS/UMTS network however produce records (IPDRs), which are of importance for some packet based charging mechanisms. These are the records that need a standard format.

[16] These external nodes may be Cisco routers, RADIUS servers, RSVP policy managers etc, and each of them produce various records that may be of interest depending on the charging mechanism employed.

6 Appendices

Appendix 1A

A Description of Some GSM Services

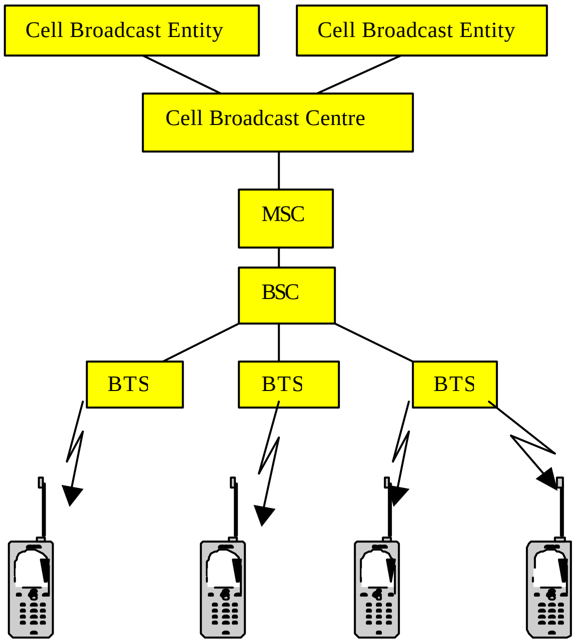
Short Message Service

Due to the recent explosion in the growth of SMS usage among GSM consumers, this subsection is designated to give a brief description of SMS technology. It allows the transfer of short messages between a Short Message Service Centre (SM-SC) and a GSM mobile. SMS however offers more than a simple message transport. It provides a store-and-forward functionality retaining messages that cannot be delivered and making repeated delivery attempts until successful delivery or time limit expiry. This is the nearest to a guaranteed delivery that a radio network can provide. It also provides acknowledged delivery, so the sender knows if and when the delivery was made. SMS can be either point-to-point (i.e. mobile to mobile otherwise known as point-to-point) or cell broadcast. The SM-SC is connected to the GSM network via an SMS gateway. GSM however, does not specify the protocol, which connect the SM-SC to the network. SMS messages can be up to 140 octets in length, (160 characters). They can be originated or terminated on either idle or dedicated Mobile Stations (MS). A mobile is said to be in Dedicated Mode, if it is involved in a call, thus, requiring a two-way communication link between itself and a Base Station (BS) as opposed to Down-link Information Only when it is in Idle Mode. This down-link information includes system information and paging messages, transferred from the BS to the MS when the MS is switched on, and has made all its initial selection decisions, and is monitoring the paging channels attempting to maintain the state of readiness to make/receive calls. For dedicated mobiles, the duplex link has to be maintained throughout the duration while the MS is in this mode. This may require that the traffic and radio connection is passed to another BS, a process termed Handover, which could be either intra- or inter-BSC. Handover may be due to Poor Signal Quality/Level, Link Distance, Power Budget, or Interference.

For service to or from an idle mobile, the Short Message must be assigned a dedicated channel over which the message transfer takes place. For a mobile in dedicated mode on a traffic channel, the SMS transfer takes place over the Slow Associated Control Channel (SACCH), which has an impact on transfer times. Whether in idle or dedicated mode, Short Messages are transferred over the Common Control Channels (CCCH). SMS has been extremely successful and has seen remarkable growth over the last few years, with 1billion messages/month in March 1999 to 2billion messages/month in October 1999. [15]

Cell Broadcast Message

The Cell Broadcast Service CBS permits the network to broadcast short data messages to all subscriptions within a particular cell. It is similar in concept to a Teletex service. Individual CBS messages may be allocated one or more CBS areas. The network requires a cell broadcast entity (Network operator or outside organisation) which collects the CBS messages and passes it on to a central cell broadcast centre which in turn passes it on via network nodes to the MS as shown in fig 1 below. Data services include Message Handling, Videotex, Teletex, GPRS, and Facsimile. The range of Tele services offered by GSM grows continuously as the needs of network users grow.



Supplementary Services

The Call Hold service

This allows a served mobile subscriber to interrupt communication on an existing call and then subsequently, if desired, to re-establish communication. The traffic channel remains assigned to the mobile subscriber after communication is interrupted, to allow the origination, or possible termination, of other calls.

Call waiting

This simply allows the subscriber to be notified of an incoming call whilst their mobile station is in the busy state.

The Multi-party Service

This enables a subscriber to establish multi-party conversation. They can whilst active on a call, hold that call and make other calls, switch from one to another (providing privacy between the calls) or join the calls together (Audio conferencing).

The Closed User Group

This provides the possibility for a group of subscribers to communicate "only" among themselves. One or more subscribers may be provided incoming/outgoing access to users outside the group.

Advice of Charge Information

This allows the mobile to calculate, in real time, the charge for a call and to display this information.

Appendix 1B

GSM Call Record Types

The following is a list of Call Record Types (also referred to as Call Data Records CDRs) produced from most network operators.

Mobile Originated	MO
Mobile Terminated	MT
Transit	TR
Call Forwarding	CF
Roaming Call Forwarding	RCF
Mobile Originating SMS in MSC	MOSM
Mobile Terminating SMS in MSC	MTSM
Supplementary Service Subscriber Procedure	SSSU
Advice Of Charge	AOC
Supplementary Service invocation Event module	SSEV
Services Switch Event module	SSWEV

Each of the CDRs above contains parameters that give details of it.

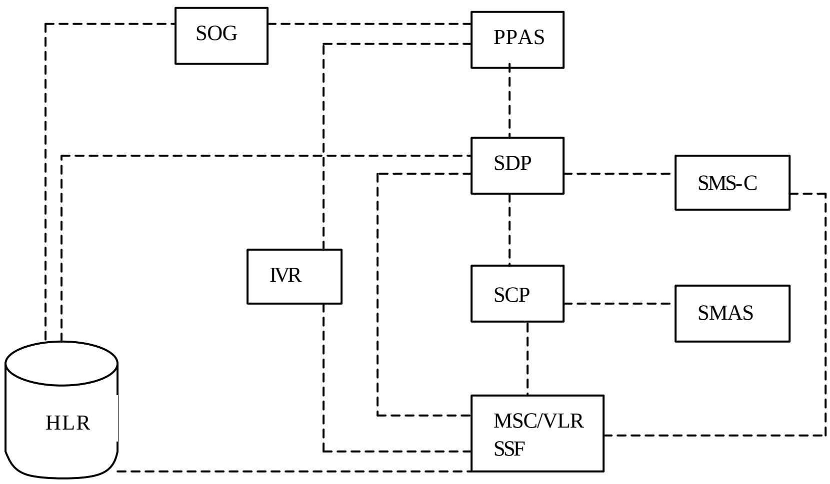
An example is the MT CDR, which comprises of the following parameters. The MO CDR comprises of the same parameters.

Traffic Activity Code	Call Identification Number
Record Sequence Number	Calling Party Number
Called Party Number	Called Subscriber IMSI
Called Subscriber IMEI	Original Called Number
Redirecting Number	Date for Start of Charge
Time For Start of Charge	Chargeable Duration
GSM Tele Service Code	GSM Bearer Service Code
Transparency Indicator	MSC Identification
Outgoing Route	Incoming Route
Call Position	Internal Cause and Location
Last Partial Output	Partial Output Record Number
Related Call Number	ICI Ordered
Cell Identity of the First Cell Used by the Called Subscriber	Supplementary Service Code
RF Power Capability Class Called Subscriber	First Radio Channel Used
Acceptance of Call Waiting	Terminating Location Number
GSM Call Reference Number	MSC Address
Time from Register seizure to Start of Charging	Fault Code
Cell ID for last Cell Called	Network Call Reference

Appendix 1C

The Pre-Paid Service Intelligent Network Subsystem

Basically, the IN system is involved in an awful lot of signalling to monitor and maintain the subscriber details and account data as discussed briefly below.



All of the lines shown above between the disparate nodes is some kind of signalling using one of the many protocols used for general GSM signalling, including ISUP, MAP SMS and MML etc. These signalling messages will not be discussed here.

A short description and function of each node follows.

PPAS (Administrative System for PPS)

This hosts all PPS specific administrative functions such as :

- Voucher management
- Refill and inquiries
- Subscriber provisioning
- Administration of account expiration dates
- Service Class data

The PPAS provides the IVR with the necessary data for refill and inquiries.

SDP (Service Data Point)

This consists of one or several physical servers. It holds the following data

- Subscriber Data
- Account data
- Service Class data
- Tariff and charging analysis data

It provides the SCP with relevant data for performing charging of PPS calls. It also sends SMs, initiates SMS call back calls on demand and charges MO-SM. It also notifies PPAS of change of PPS account status (for example low account value) for evaluation of Service Scenario.

HLR

The HLR stores PPS subscriptions just like ordinary subscriptions. This record contains a flag to indicate whether a subscription is PPS or not. All PPS calls are routed to the SSF. It may also be used to bar terminating calls to a PPS account if the Service Scenario of the account justifies this. Barring in the HLR rather than in the SCP reduces network load.

IVR (Interactive Voice Response)

Used to guide a PPS subscriber that has called the Service Provider's service number. The IVR helps the caller to interact with the PPAS system with account balance inquiries, expiration date and refill procedures.

MSC

The MSC routes PPS calls to the SSF, which is usually integrated in the MSC. The SSF may also be a stand-alone node.

SCP (Service Control Point)

This node executes the service. The SDP or SSF triggers Service execution. It controls the call by interacting with the SSF and interacts with the SDP for data retrieval/update.

SMAS (Service Management Application System)

Used for Service provisioning and management

SOG (Service Order Gateway)

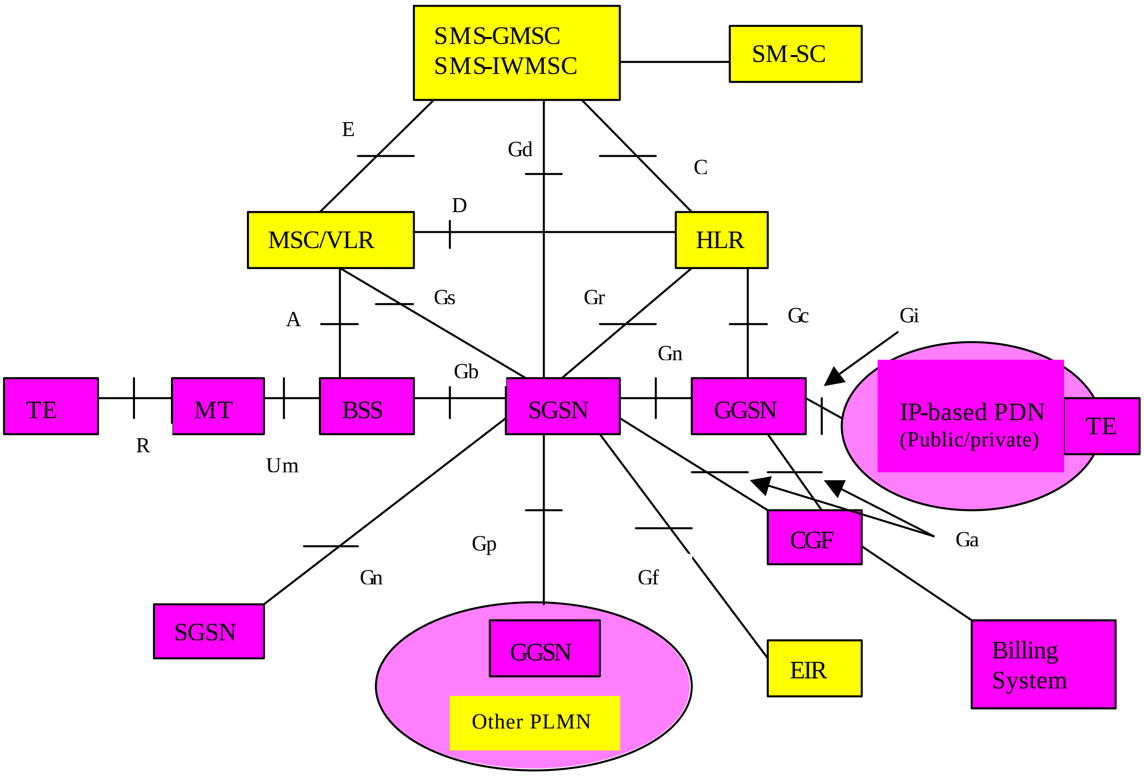
Used for uniform interface interconnection between different network elements.

SSF (Service Switching Function)

This invokes service execution in an SCP. It interacts with the SCP for information on how to handle a call. It reports call duration, supervises calls and orders the AST-DR to play announcements.

Appendix 2A

GPRS Logical Network Architecture



GPRS Interfaces

All GPRS Interfaces are open and defined by 3GPP.

- Ga Interface between a CDR transmitting unit (e.g. GGSN or SGSN) and a CDR receiving functionality (CGF).
- Gb Interface between an SGSN and a BSS.
- Gc Interface between a GGSN and an HLR.
- Gd Interface between an SMS-GMSC or SMS-IWMSC and an SGSN.
- Gf Interface between an SGSN and an EIR.
- Gi Reference points between GPRS and an external PDN.
- Gn Interface between two GSNs within the same PLMN. This is an internal IP ‘backbone’ network
- Gp Interface between two GSNs in different PLMNs.
- Gr Interface between an SGSN and an HLR.
- Gs Interface between an SGSN and an MSC/VLR.
- R Reference point between a non-ISDN compatible TE and MT.
- Um Interface between the mobile station and the GPRS fixed network part. This is the GPRS network interface for providing packet data services over the radio to the MS.

Appendix 2B

GPRS GGSN PDP context data record.

Field		Description
Record Type	M	GPRS GGSN PDP context record.
Network initiated PDP context	C	Present if this is a network initiated PDP context.
Anonymous Access Indicator Served IMSI	C M	Set to true to indicate anonymous access (and that the Served IMSI is not supplied). IMSI of the served party (if Anonymous Access Indicator is FALSE or not supplied).
GGSN Address	M	The IP address of the GGSN used.
Charging ID	M	PDP context identifier used to identify this PDP context in different records created by GSNs
SGSN Address	M	List of SGSN addresses used during this record.
APN Network Identifier	M	The logical name of the access point to the external PDN (network identifier part of APN).
APN Selection Mode	O	An index indicating how the APN was selected.
PDP Type	M	PDP type, e.g. X.25, IP, PPP, or IHOSS:OSP
Served PDP Address	M	PDP address of UE, e.g. an IPv4, IPv6 or X.121.
Remote PDP Address	O	List of PDP addresses of the remote host or DTE e.g. an IPv4, IPv6, or X.121 (Included if the PDP type is X.25)
Dynamic Address Flag	C	Indicates whether served PDP address is dynamic, i.e. if it is allocated during PDP context activation. Field is missing if address is static.
List of Traffic Data Volumes	M	A list of changes in charging conditions for this PDP context, each time stamped. Charging conditions are used to categorise traffic volumes, such as per –tariff period. Initial and subsequently changed QoS and corresponding data values are listed. Data volumes are in octets above the GTP layer and are separated for up-link and downlink traffic.
Record Opening Time	M	Time stamp when this record was opened.
Duration	M	Duration of this record in the GGSN.
Cause for Record Closing	M	The reason for the release of record from this GGSN.
Diagnostics	O	A more detailed reason for the release of the connection.
Record Sequence Number Node ID	C O	Partial record sequence number, only present in case of partial records. Name of the recording entity.
Record Extensions	O	A set of network/ manufacturer specific extensions to the record.
Local Record Sequence Number	O	Consecutive record number created by this node. The number is allocated sequentially including all CDR types.

In the table above, M stands for Mandatory, C for Conditional and O for Optional.

Appendix 4A

UMTS QoS Technical Requirements

UMTS QoS is defined with a set of parameters, which should meet the following criteria. Only some of these criteria have been mentioned below.

1. The UMTS QoS control mechanisms shall provide a mapping between application requirements and UMTS services
2. The UMTS QoS control mechanisms shall be able to efficiently interwork with current QoS schemes. Further, the QoS concept should be capable of providing different levels of QoS by using UMTS specific control mechanisms (not related to QoS mechanisms in the external networks)
3. The UMTS shall provide a finite set of QoS definitions
4. The QoS parameters are needed to support asymmetric bearers
5. Applications (or special software in UE or 3G gateway node) should be able to indicate QoS values for their data transmissions
6. Multiple QoS streams per address
7. QoS behaviour should be dynamic, i.e., it shall be possible to modify QoS parameters during an active session
8. User QoS requirements shall be satisfied by the system, including when change of SGSN within the Core Network occurs, etc.

Others can be found in 3GPP Standards Document TS 23.107 V3.2.0 (2000-03), QoS Concept and Architecture.

Appendix 4B

UMTS QoS Classes and parameters defined in R99

Traffic class Parameter	Conversational class	Streaming class	Interactive class	Background class
Maximum bit rate	X	X	X	X
Delivery order	X	X	X	X
Maximum SDU size	X	X	X	X
SDU format information	X	X		
SDU error ratio	X	X	X	X
Residual bit error ratio	X	X	X	X
Delivery of erroneous SDUs	X	X	X	X
Transfer delay	X	X		
Guaranteed bit rate	X	X		
Traffic handling priority			X	
Allocation/Retention priority	X	X	X	X

A description of some of the parameters follows below.

Delivery order

This indicates whether the UMTS bearer shall provide in-sequence SDU delivery or not, and specifies if out-of-sequence SDUs are acceptable or not. Whether out-of-sequence SDUs are dropped or re-ordered depends on the specified reliability.

Delivery of erroneous SDUs (y/n/-)

'yes' implies that error detection is employed and that erroneous SDUs are delivered together with an error indication, 'no' implies that error detection is employed and that erroneous SDUs are discarded, and '-' implies that SDUs are delivered without considering error detection.

Transfer delay (ms)

Transfer delay of an arbitrary SDU is not meaningful for a bursty source, since the last SDUs of a burst may have long delay due to queuing. The meaningful response delay perceived by the user is the delay of the first SDU of the burst.

Allocation/Retention Priority

Specifies the relative importance compared to other UMTS bearers for allocation and retention of the UMTS bearer. The Allocation/Retention Priority attribute is a subscription parameter, which is not negotiated from the mobile terminal. Priority is used for differentiating between bearers when performing allocation and retention of a bearer. In situations where resources are scarce, the relevant network elements can use the Allocation/Retention Priority to prioritise bearers with a high Allocation/Retention Priority over bearers with a low Allocation/Retention Priority when performing admission control.

Appendix 4C

Network Elements Involved in CDR production in "Corporate User Example"

Netflow accounting record: Routers are network devices used to forward IP packets across the ISP network. Cisco's routing equipment has the capability to produce per-connection usage Records on a per-connection, unidirectional flow aggregate. These types of records provide the basic raw material from which billing information is built. Some of the useful parameters given in a NetFlow record include:

Source and destination IP Address	Initiating and terminating IP addresses
Destination port	Terminating IP port (in some cases representing the application used)
Protocol Sent and received bytes/packets	TCP or UDP Number of bytes
Start/end timestamps	Time that the flow began and ended

RADIUS (Remote Authentication Dial-In User Service): When a user dials into a remote access server, they are transparently assigned a dynamic IP address as part of the login procedure. The IP address is valid until the user log-out, when it is assigned to the next user logging in. RADIUS servers authenticate remote dial-in access, log authentication request and results, as well as account for login and logout events. Mediations systems can use RADIUS records, given below, to associate network transactions with customers.

IP Address	Dynamic IP address that was given to the user at login time
User name	Name of user
Start (login), End (logout)	Times of login and log-out

RSVP (Resource reSerVation Protocol): An RSVP policy manager, such as Class-Data, accounts for QoS request events. By correlating this data with the Netflow records, it is possible to determine the connection's requested and negotiated QoS

Source and destination IP address	Initiating and terminating IP addresses
Start and end	Timestamps of the QoS session
Requested QoS	Level of QoS

Network management records: IP Network management stations such as HP-Openview, among other purposes, map network (IP) addresses to machine locations and user groups. By accessing the network management system's database, it is possible to associate IP addresses to billing entities and geographic locations. The following information can be obtained:

IP group	Groups of IP addresses and network masks
Real-world entity	Name
Geography	Location associated with entity

Routing tables: Routers use dynamic IP routing schemes to balance network load among several paths, and to allow for redundancy and fault tolerance. In some cases, to find the path by which network transaction passed and identify external billing entities, it is necessary to collect the following information from routing tables:

IP group	Groups of IP addresses and network masks
Path (IP)	Identifier of the path to be taken by IP packets destined to IP group

Appendix 4D

Internet Protocol Data Record (IPDR)

If a network operator wishes to charge content partners - banks, restaurants, cinemas etc on a content-basis, then the operator requires network nodes and devices capable of recording and collecting IP usage data. These IP-related data records are known as **Internet Protocol Data Record (IPDR)** and the IPDR standard format is still under development by The Internet Protocol Detail Record Organisation (IPDR.org). The Internet Protocol Detail Record Organisation (IPDR.org) has taken the initiative to define the essential elements of data exchange between IP network elements, and OSS/BSS. I.e. identify the information required to characterise an IP service (including e-commerce transactions) to the level of granularity desired by a BSS.

The IPDR.org will also provide the foundation for development of open, carrier-grade IP support systems that enable next-generation IP networks to operate efficiently and cost effectively. IPDR.org specific goals include:

- Define an open, flexible record format (the IPDR record) for exchanging usage information.
- Define essential parameters for any IP transaction.
- Provide an extension mechanism, so network elements and support systems can exchange optional usage metrics for a particular service.
- Provide a repository for defined IPDRs.

IPDR.org concluded that it was much better to develop the basic framework for the specification, allowing companies to develop working code against the framework, supporting interoperability and the usefulness of the specification. Following analysis of the results, the IPDR.org will identify additional elements for the specification that are more service-specific. This iterative approach will best meet the needs of the industry, and facilitate the achievement of IPDR's goals according to IPDR.org.

The IPDR.org Working Group is on the verge of approving and launching the IPDR compliant source code and has just announced the release for public comment of the organisation's first approved technical specification - Network Data Management - Usage (NDM-U) for IP-Based Services, version 1.0. The NDM function collects data from devices and service elements in a provider's network. Usage refers to the type of data. IPDR.org has also extended their reach throughout the Telecommunications industry by building relationships with global associations, such as the GBA, to ensure that their work is aligned with the priorities of the entire industry. GBA's next steps are to validate interoperability against the IPDR specification and address additional issues such as provisioning, record transport mechanisms, real-time APIs, security etc. The GBA is actively producing pricing and billing models to illustrate the scenarios where an operator might maximise revenues.

Appendix 4E

Asymmetric Digital Subscriber Line (ADSL)

ADSL is one of the many different versions of DSL technologies, which dramatically improve the bandwidth of existing analogue phone systems. DSL connections are point-to-point-dedicated circuits, meaning that they are always connected. DSL providers can offer up to 16 voice channels plus data over one line. There is no dial-up, and no switching. The line is a direct connection into the carrier's frame relay, ATM, or Internet Connect System. A DSL modem is required at the customer site.

ADSL is an asymmetric technology meaning that the downstream data rate is much higher than the upstream data rate, which suits Internet connections, as more information is usually downloaded from web servers than is uploaded. Downstream rates vary with the distance from the Local Exchange and range from 1.544Mbps to 8.448Mbps. Upstream speeds range from 16Kbps to 640Kbps. ADSL operates in a frequency range that is above the frequency range of voice services, so the two systems can operate over the same cable. One of the major additions to ADSL, and one that some call the killer application, is the voice service, Voice over DSL (VoDSL). Whilst it may sound ironic, that soon, many ADSL subscribers will get voice services over the same copper loop that supplies traditional telephone service, VoDSL is much more than just a dial tone. VoDSL providers need to install Integrated Access Devices (IADs) at their customer premises. The IAD takes the voice and data traffic from the customer, turns the traffic into packets, and sends it over the DSL pipe to the Local Exchange (LE). In most ADSL installations, ATM is the transport technology running over the last-mile-connection from the LE to the customer. ATM enables providers to deliver high QoS.

Appendix 4F

Brief Summary of I-mode, its services, and charging schemes used for these services.

I-mode is based on the underlying technology – Personal Digital Communications (PDC), which is similar to GPRS in that it uses packet-switched voice and data channels. It however differs from GSM/GPRS Radio Access Network in that it uses very small cells (300 – 500 metres), which is well suited for the Japanese geography, where 80% of the population occupies 20% of the land. I-mode today has over 10 million subscribers and makes up a substantial fraction of DoCoMo's subscriber base as opposed to the low percentage of European operators' subscribers making use of wireless data applications. The I-mode service has almost 300 banks included, over 600 Application Alliance partners and over 13000 voluntary web sites paying DoCoMo resulting in official content partners taking 70% (Entertainment 52 and transactions 21) of all traffic. The huge success of I-mode is due to a number of factors including:

- The fact that it complements fixed-line Internet, providing different services rather than being just another access solution for the Internet.
- Its graphic capability
- Subsidised handsets and
- Not least the charging and pricing structure of I-mode's services.

I-mode uses compact HTML (cHTML) to present Web information on the mobile phone while WAP uses Wireless Mark-up Language (WML). There are three level levels I-mode services.

I-mode Site Connection Service: which enables people to access I-mode service sites to retrieve information, check their bank balances, transfer money, reserve concert or airline tickets and connect to the Internet under restricted mode (i.e. text information and compact graphics)

I-mode Push Message Service: This enables users to receive desired information automatically once they have registered with the specific site. Typically this includes news flashes, weather forecasts etc.

I-mode Mail Service: The cellular phone numbers are used for e-mail addressing using the format 090(MSISDN)@docomo.ne.jp.

The charging strategy used for I-mode services is a mix of basic service access and usage-based charges. Usage-based here implies both **volume- and content-based**

- The basic charge for service access is 300¥ (approximately $2.76) per month.
- Transmission of data packets is charged at 0.3¥/packet (approximately $0.002), where one packet is 128 bytes long.
- I-mode Site Connection Service is charged at a flat rate per month.
- Mobile banking service (balance checking and money transfer), weather forecasts, airline reservations restaurant recommendations etc are all charged per transaction and typically cost between 10¥ and 40¥ depending on the service.
- Sending and receiving e-mails are charged similarly, i.e. on a per-transaction basis.

NTT DoCoMo take a cut of all I-mode communication services while the application service providers take the rest.

7 References

1. Allen Rick (Logica): ' A brave new world of billing', The operator Guide to GPRS, pp 69-77 (2000).

2. Blomgren Olle: 'Description of the PPS 3.0 System' (1998)

3. Channing Ian (Technology Editor Mobile Communications International): 'GPRS-The bridge to 3G', The operator Guide to GPRS, pp 5-9 (2000).

4. Chorleywood Consulting: 'Billing Plus', volume 2 number 8 (August 2000)

5. Dann Gavin: 'BT Cellnet GSM Charging Architecture' (2000)

6. Dann Gavin: 'CME Call Records and Parameters' (1999)

7. Darnbrough John (Nortel Networks): 'Wings of Light, Delivering Profitable GPRS Applications', Presentation slides (2000)

8. Global Billing Association: 'GBA Workshop Billing for Content II', (2000-06-08), (http://www.xacct.com/library/index.htm)

9. Grobel Ian: 'Billing and Usage Collection Strategies for General Packet Radio Services' (2000-09-27), (http://www.xacct.com/library/index.htm)

10. Holma Harri and Toskala Anti: ' WCDMA for UMTS Radio Access For Third Generation Mobile Communications'

11. IPDR.org, Inc.: 'Network Data Management – Usage (NDM-U) For IP-Based Services' (v1.1 2000-06-26)

12. Karvé Anita: 'Billing for Voice over DSL', (2000), (http://www.billingworld.com/content_frames/full.asp?id=1594&action=article)

13. KSCL: 'Jupiter Billing - Jupiter Convergent Billing, Administration and Customer Care', (http://www.kscl.com)

14. KSCL: 'Jupiter Services Management Portfolio – Jupiter Mediation, Jupiter Interconnect', (http://www.kscl.com)

15. Locke-Wheaton John (Mobile Lifestreams): 'Using Voice and Data services together', The operator Guide to GPRS, pp 51-55 (2000).

16. Lucas Matthew and Schweitzer Limor: 'Mediation in a Multi-Service IP Network', (2000-09-27), (http://www.xacct.com/news/art100198.html)

17. Motorola: 'General Packet Radio Service Billing Mediator Functionality - Proposal to Cellnet from Motorola ECID' (1999)

18. Newman Mark (Editor 3G Mobile): 'Getting the price right: data's biggest challenge?', The operator Guide to GPRS, pp 37-39 (2000).

19. Northstream: 'GPRS business modelling', The operator Guide to GPRS, pp 29-34 (2000).

20. Potterill Brian (Pricewaterhouse Coopers): 'GPRS & EDGE', Presentation slides (2000-07-07).

21. Sheldon Tom: ‘LAN Times Encyclopaedia of Networking’ (1998).

22. Walke Bernhard H: ‘Mobile Radio Networks, Networking and protocols’, (1999)

23. Wray Castle Telecoms training: ‘GSM System Overview’, Sections 1,2,3,4 and 7 (1998).

24. Wray Castle Telecoms Training: ‘Advanced telecoms training for the forward thinking’, Sections 1,8 and 9 (1998).

25. 3GPP Standards Document TS 32.015 v3.1.1: ‘GSM Call and event data for the packet switched domain’ (2000-03 Release 1999)

26. 3GPP Standards Document TS 32.015 v3.2.0: ‘Charging and Billing; GSM Call and event data for the packet switched domain’ (2000-07 Release 1999)

27. 3GPP Standards Document TS 23.107 v3.2.0: ‘QoS Concept and Architecture’ (2000-03 Release 1999).

28. http://www.bluetooth.com

29. http://www.ipdr.org

30. http://www.ipv6.org

31. http://today.intra.bt.com/today 19/07/2000