University of Essex

Department of Electronic System Engineering

M.Sc. in Telecommunication & Information Systems

Thesis on:

UMTS Core Packet-Switched Network

By

Evangelos Vlachogiannis

17/09/2001

ABSTRACT

The deployment of 3rd Generation mobile system, UMTS, will provide a plethora of new services. These new services require certain quality of service, depending on a wide range of parameters. The purpose of this thesis is to try to analyse this quality of service parameters so that an operator can configure his network, depending of his needs. It has to be noted that this thesis is concentrated on the Core network.

In the beginning, there is a brief introduction on the architecture and general aspects of UMTS procedures. The next part gives a description of the quality of service on both UMTS and ATM, which is the common layer 2 protocol. Then, this thesis makes an attempt to analyse and map the attributes of the several UMTS traffic classes to the appropriate ATM traffic classes.

AUTHOR'S DECLARATION

The work described in this thesis was carried out by the author at the University of Essex in the Department of Electronic Systems Engineering in cooperation with BT Cellnet between June and September 2001. All work and ideas in this thesis are original unless otherwise acknowledged in text or by reference. All work presented was undertaken by the author without collaboration.

Evangelos Vlachogiannis

17 September 2001

DISCLAIMER

Any views or opinions expressed in this thesis are purely the personal of author's, and do not necessarily present the views or opinion of the University of Essex or its employees.

To my parents

Table of Contents

List of Figures

List of Tables

Abbreviations

2nd Generation
3rd Generation
ATM Adaptation Layer type 5
ATM Adaptation Layer type 2
Asynchronous Transfer Mode
Authentication Centre
Business to Consumers
Border Gateway
Base Station Controller
Base Station Subsystem
Base Transceiver Station
Bandwidth
Constant Bit Rate
Cell Delay Variation Tolerance
Cell Loss Probability
Cell Loss Rate
Core Network
Common Part Convergence Sublayer
Cell Transfer Delay
Dynamic Host Configuration Protocol
Differentiated Services
Domain Name System
Effective Bandwidth
Equipment Identity Register
Frequency Division Duplex
File Transfer Protocol
Generic Cell Rate Algorithm
Guaranteed Frame Rate
Gateway GPRS Support Node
Gateway Mobile Switching Centre
General Packet Radio Service
Global System for Mobile Communication
GPRS Support Node
GPRS Tunnelling Protocol
GTP User Plane
Home Location Register
Home Public Land Mobile Network
Independent Identically Distribution
Integrated Services

Internet Protocol
Internet Protocol version 4
Internet Protocol version 6
Integrated Services Digital Network
Layer 1
Layer 2
Location Area
Medium Access Control
Maximum Bit Rate
Maximum Burst Size
Minimum Cell Rate
Maximum Frame Size
Multiprotocol Label Switching
Mobile Station
Mobile Switching Centre
Mobile Terminal
Non-real-time VBR
Octet Stream Protocol
Peak Cell Rate
Packet Data Network
Packet Data Protocol, e.g., IP
Protocol Data Unit
Public Land Mobile Network
Point-to-Point Protocol
Public Switched Telephone Network
Permanent Virtual Circuit
Quality of Service
Radio Link Control
Radio Network Controller
Radio Network Subsystem
Resource Reservation Protocol
Real-time VBR
Service Access Point
Segmentation And Reassembly
Service Data Unit
Short Message Service
Time Division Multiple Access
Terminal Equipment
Unspecified Bit Rate
User Datagram Protocol
Universal Mobile Telecommunication System
Usage Parameter Control
UMTS Terrestrial Radio Access Network
Variable Bit Rate

Visited Location Register
Visited Public Land Mobile Network
Wireless Application Protocol
Wideband Code Division Multiple Access
World Wide Web

Symbols

Ga	Charging data collection interface between a CDR transmitting unit (e.g., an SGSN or a GGSN) and a CDR receiving functionality (a CGF).
Gb	Interface between an SGSN and a BSS.
Gc	Interface between a GGSN and an HLR.
Gd	Interface between a SMS-GMSC and an SGSN, and between a SMS-IWMSC and an SGSN.
Gf	Interface between an SGSN and an EIR.
Gi	Reference point between GPRS and an external packet data network.
Gn	Interface between two GSNs within the same PLMN.
Gp	Interface between two GSNs in different PLMNs. The Gp interface allows support of GPRS network services across areas served by the co-operating GPRS PLMNs.
Gr	Interface between an SGSN and an HLR.
Gs	Interface between an SGSN and an MSC/VLR.
Iu	Interface between the RNS and the core network. It is also considered as a reference point.
R	Reference point between a non-ISDN compatible TE and MT. Typically this reference point supports a standard serial interface.
Um	Interface between the mobile station (MS) and the GSM fixed network part. The Um interface is the GSM network interface for providing GPRS services over the radio to the MS. The MT part of the MS is used to access the GPRS services in GSM through this interface.
Uu	Interface between the mobile station (MS) and the UMTS fixed network part. The Uu interface is the UMTS network interface for providing GPRS services over the radio to the MS. The MT part of the MS is used to access the GPRS services in UMTS through this interface.

1

Introduction

1.1 General overview [1]

The third-generation UMTS system was designed essentially to meet the new requirements associated with mobile data services. This service is going to enable much greater bit rates, which, in its turn, is going to enable a wide number of new services. The main service categories addressed are:

- Location-based services
- Edutainment and Infotainment
- B2C Services
- Office Extension
- Telemedicine
- Telematics/Telemetry/Monitoring

Some of the enabling applications are:

- Multimedia
- Mobile commerce
- Voice over IP
- Interactive broadcasting
- Positioning

1.2 Thesis aim

The aim of this thesis is to give a brief introduction to the Universal Mobile Telecommunications Service network and then concentrate on the quality of service aspects of its core network.

UMTS core network is based on an IP backbone over which the mobile user's IP data is transferred transparently by means of tunnelling mechanisms. Every user depending on the application he uses, and of course the money he pays, requires a certain QoS. This QoS is indicated using parameters assigned by the application running on the mobile station. Then, the operator needs to have mechanisms to map all these parameters to all protocols used by the tunnelling mechanism (end-to-end). The basic aim of this thesis is to map those UMTS parameters to ATM parameters, so that an operator can configure his network PVCs.

1.3 Thesis's structure

An outline of this thesis is indicated below in a Chapter sense:

1. Introduction

2. UMTS architecture

This section gives a brief description of the UMTS network. In order to do so there is firstly an overview of GSM concepts. A step further is the introduction of GPRS, and the final step is the UMTS network. As, the subject of consideration of this thesis, is the core network, there are more details on this and just a brief description on the rest.

3. Accessing UMTS - PDP Context

At this part of the thesis, a brief description of the way a mobile station is registered (attached) to the network is given. This is important, as during the attachment procedure QoS parameters are set.

4. Tunnelling mechanism

The purpose of this chapter is basically to explain what is the reason for wireless network complexity and how this can be overcome.

5. User transport plane (UMTS only)

This section consists of the protocol stack used to transport data end-to-end (mobile to mobile). This is necessary so that headers and parameters of each protocol can be considered. Of course, also in this case, most important is the transport plane in the core network.

6. QoS in UMTS

This is a very important section. It describes analytically the classes of service used in UMTS by defining end-to-end bearer services. Then, there is a detailed description of the traffic attributes, which are necessary to be mapped to ATM traffic attributes. Finally, there is a brief description of the basic functions in order to understand how QoS is applied.

7. ATM QoS

A review of ATM traffic aspect is necessary, so that UMTS attributes can be mapped to ATM attributes. So, a description of ATM QoS classes and ATM attributes is given, but only those that are used for the mapping, given by a UMTS-to-ATM class mapping.

8. Mapping

This chapter contains the analysis and the mapping of the attributes of both UMTS and ATM, so that an operator would be able to configure his network.

9. Conclusion

This chapter contains a summary of the results obtained from the mapping process. There is also a description of aspects that future work can be based on (MPLS and OPNET).

2

UMTS Architecture

This chapter is used to provide a description of UMTS architecture. There is a brief description of the complete UMTS architecture, but this chapter is focused on the core network architecture, which is important for the purpose of this thesis. In order to understand UMTS system architecture, it is necessary to review basic aspects of the current GSM network.

2.1 GSM architecture [2]

A brief public land mobile network (PLMN) picture can be shown in Figure 1. A GSM mobile station is denoted as MS. In each cell there is a base transceiver station (BTS), which is responsible for the radio coverage of that area. Several BTS together are controlled by one base station controller (BSC). The BTS and BSC together form the base station subsystem (BSS). Traffic coming from the MSs through the above is routed through a switch called the mobile switching centre (MSC). In case of a call coming from or going to fixed network, traffic is then routed through a gateway mobile switching centre (GMSC).

GSM networks are structured hierarchically. For at least one location area (LA), is assigned at least one administrative region, MSC. Each LA consists of several cell groups assigned to BSCs. For call control and network management, there are several databases available: the home location register (HLR), the visited location register (VLR), the authentication centre (AUC), and the equipment identity register (EIR). HLR stores permanent and temporary data for each user, and is the first queried in case of a call, in order to determine user's location. A VLR is responsible for a group of LAs and stores the data of users being in its area of responsibility. The AUC is responsible for generating and storing security-related data such keys used for authentication and encryption. Finally, EIR registers equipment data rather then subscriber data.

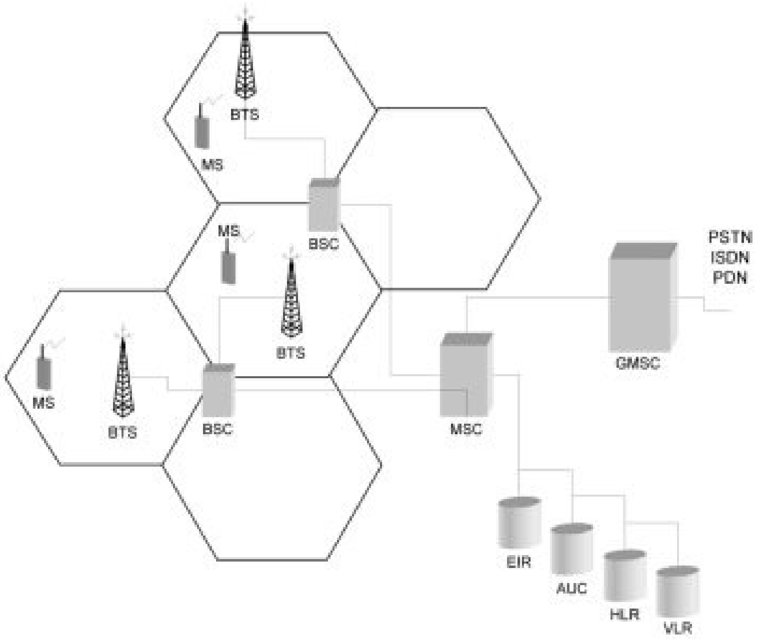

Figure 1:GSM architecture

2.2 Intermediate solution - GPRS [2]

The classic GSM network does not provide sufficient capabilities for routing packet data. This is the reason for which an extension of the current network was necessary. By the end of 2000, the General Packet Radio Service (GPRS) provided packet switched data services (up to 171 kbits/s). The packet switching mode, as opposed to the circuit-switching mode in current use, occupies the GSM network's transmission resources only when there is information to be transmitted. The user pays only for the actual amount of information sent, rather than the entire time that the communication remains active. This permits permanent connections and all its efforts. At that time, the Wireless Application Protocol (WAP) makes it possible to adapt information from the Internet to special application platforms (MS) and Short Message Service (SMS) is updated (graphics e.t.c.)

The main problem that had to be solved for the GPRS, is associated with the compatibility between an environment which is heavily oriented towards circuit switching (used in the GSM network) and the requirements of packet switching. The solution was to introduce a new class of network nodes; called GPRS (Figure 2) support nodes (GSN). GSNs are responsible for the delivery and routing of data packets between the mobile stations and the external packet data networks (PDN).

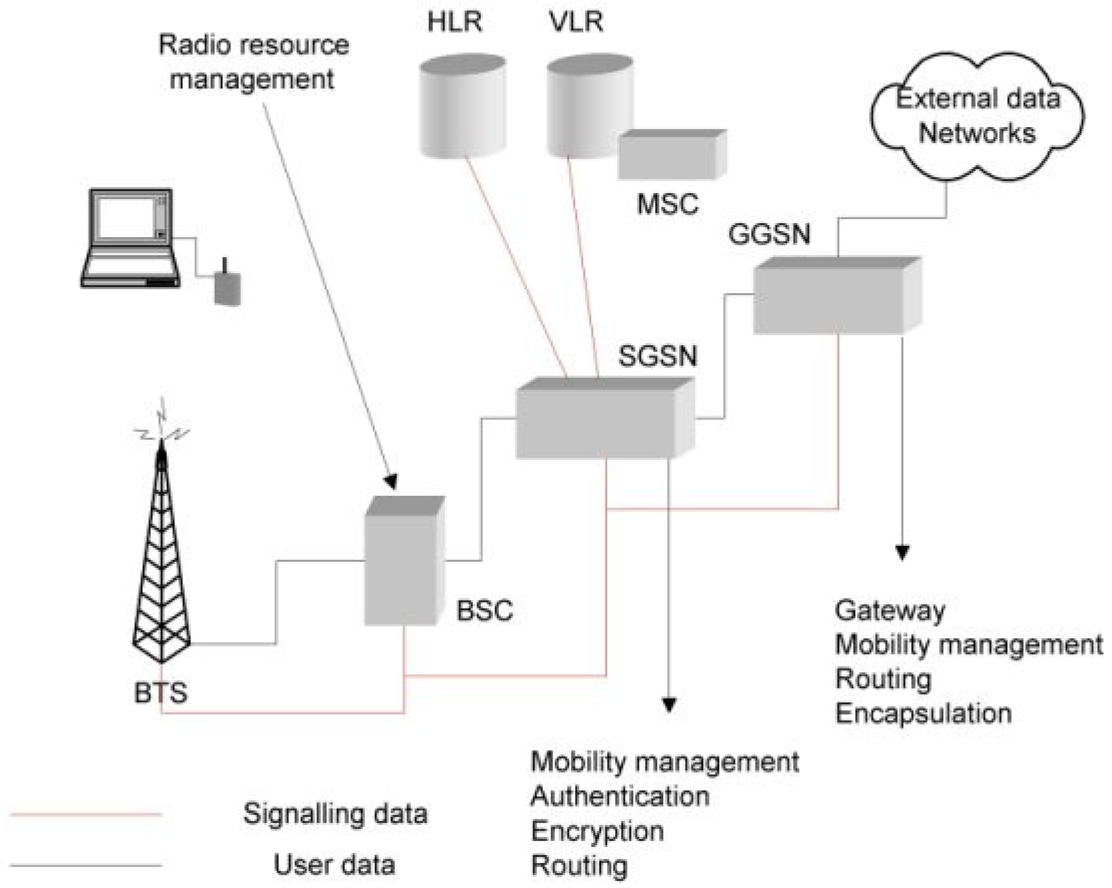

Figure 2: GPRS system

A GSN contains the functionality required to support GPRS functionality for GSM and/or UMTS (see later). There are two types of GSNs: The serving GPRS Support Node (SGSN) and the Gateway GPRS Support Node (GGSN):

- SGSN can be viewed as a "packet-switched MSC". It delivers packets to mobile stations within its service area. SGSNs send queries to home location registers (HLRs) to obtain profile data of GPRS subscribers. SGSNs detect new GPRS MSs in a given service area, process registration of new mobile subscribers, and keep a record of their location inside a given area. Therefore, the SGSN performs mobility management functions such as mobile subscriber attach/detach and location management.

- A GGSN is used as an interface between the GPRS backbone network and the external packet data networks. It converts the GPRS packets coming from the SGSN into the appropriate packet data protocol (PDP) format (e.g., IP or X.25) and sends them to the corresponding packet data network. In the other direction, PDP addresses of incoming data packets are converted to the GSM address of the destination user. The readdressed packets are sent to the responsible SGSN. For this purpose, the GGSN stores the current SGSN address of the user and his or her profile in its location register. The GGSN also performs authentication and charging functions. One (or more) GGSNs may be provided to support multiple SGSNs.

Within the network (Figure 3), the GGSN is connected to the SGSN nodes by means of an all-IP based GPRS backbone network. Within this backbone, the GSNs encapsulate the PDN packets and transmit (tunnel) them using the GPRS Tunnelling Protocol GTP (see later). There are two kinds of GPRS backbones:

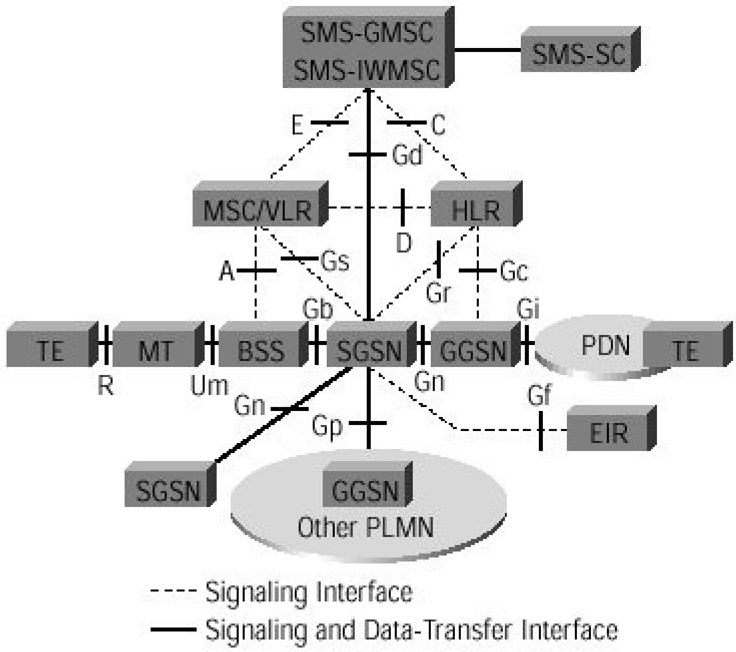

Figure 3: Packet Domain Logical Architecture [5]

- Intra-PLMN backbone networks connect GSNs of the same PLMN and are therefore private IP-based network of the GPRS network provider.
- Inter-PLMN backbone networks connect GSNs of different PLMNs. A roaming agreement between two GPRS network providers is necessary to install such a backbone.

[5]To be able to view all this in practice, let's consider the examples shown in Figure 4. There are three different routing schemes: mobile-originated message (path 1), network-initiated message when the MS is in its home network (path 2), and network-initiated message when the MS has roamed to another GPRS operator's network (path 3). In these cases, the operator's GPRS network consists of multiple GSNs (with a gateway and serving functionality) and an intra-operator backbone network. It has to be noted that the border gateway (BG) is not considered here.

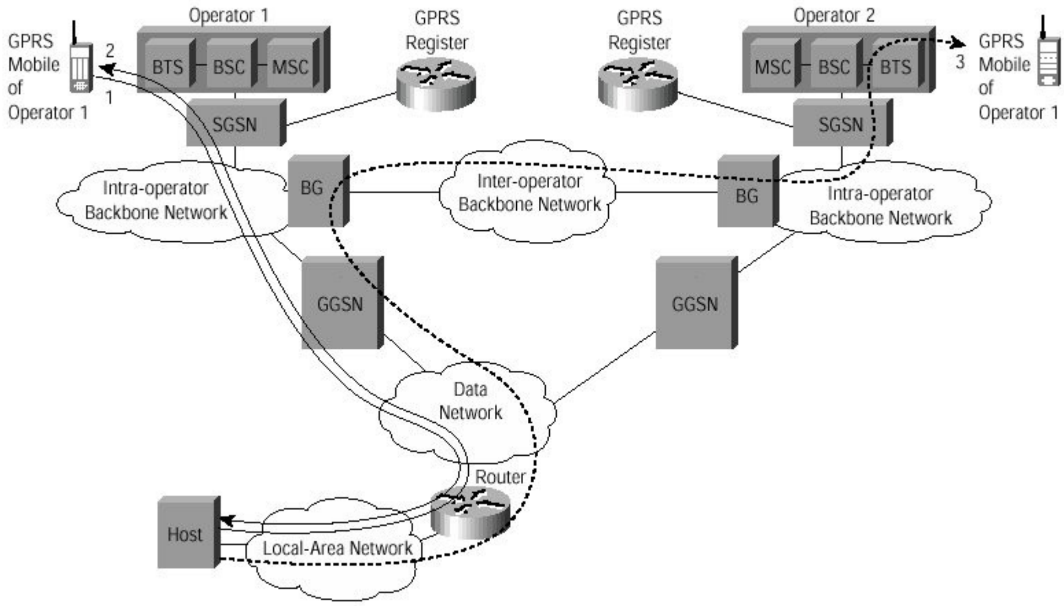

Figure 4: GPRS Routing example [5]

2.3 UMTS architecture

The next step will be UMTS. This new system combines new technologies with past techniques.

2.3.1 UMTS access network architecture [6,3]

A completely new access network architecture was developed, which took advantages of the high-speed switching capabilities of ATM, the valuable support for both WCDMA and TDMA, as well as delivering standard open interfaces within the radio network. Figure 5 gives an overview of the UMTS Terrestrial Radio Access Network (UTRAN) architecture.

Basically, UTRAN consists of nodes B and radio network controllers (RNC). The node B could be said to be equivalent to the GSM BTS in connecting the antenna site to the network. The most important requirement of the node B is to minimize the cost/functionality as the network could comprise a large number of B 'nodes'.

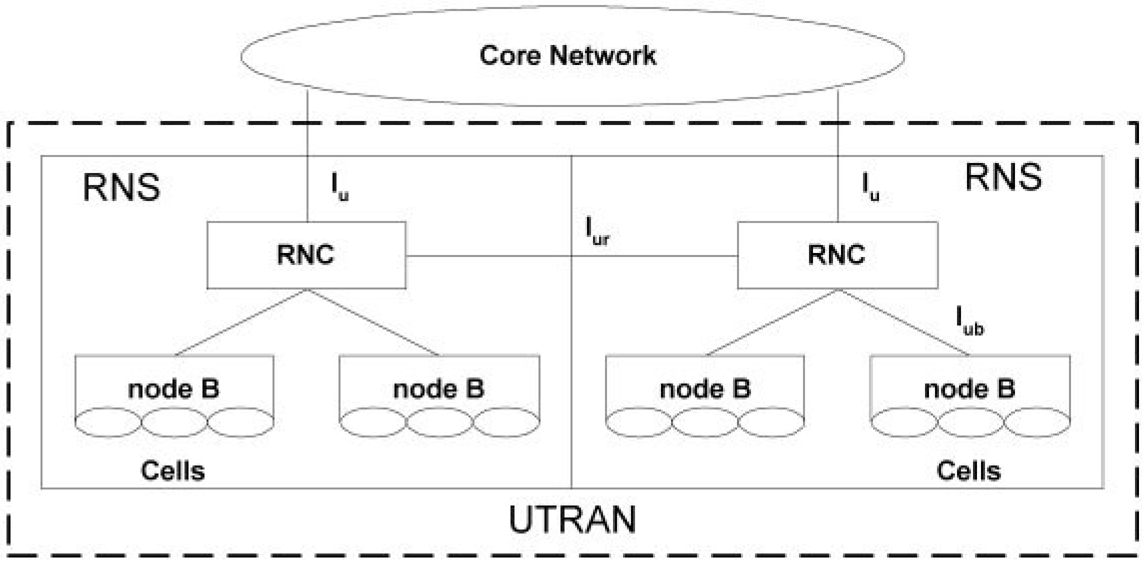

Figure 5: UTRAN architecture

The RNC could be said to be roughly equivalent at a peer level to the GSM BSC. Its responsibility is controlling the resources associated with a number of nodes B, and negotiating with the core network for aspects such as bearers and quality of service (QoS).

The RNCs and nodes B are connected to each other and to the core network with three interfaces as shown in Figure 5.

2.3.2 UMTS network architecture [3,4]

UMTS network consists of two network domains: the Circuit switched domain and Packet switched domain (Figure 6). These two domains rely on two separate and parallel backbones. The first backbone is based on ISDN (Integrated Service digital Network) technologies and is responsible for carrying voice traffic, whereas the second one based on IP technologies is used to carry data traffic. The access network connects the two domains. It has to be noted that the circuit switched UMTS backbone is derived directly from the classic GSM network infrastructure, while the packet switched one, from the infrastructure used to introduce GPRS in the GSM network. While the UMTS access network is entirely new and separate from that used for GSM, the core network infrastructure is a direct evolution of the GSM infrastructure. In the early stages of UMTS, GSM operator will be able to share the network infrastructure between 2G and 3G access networks.

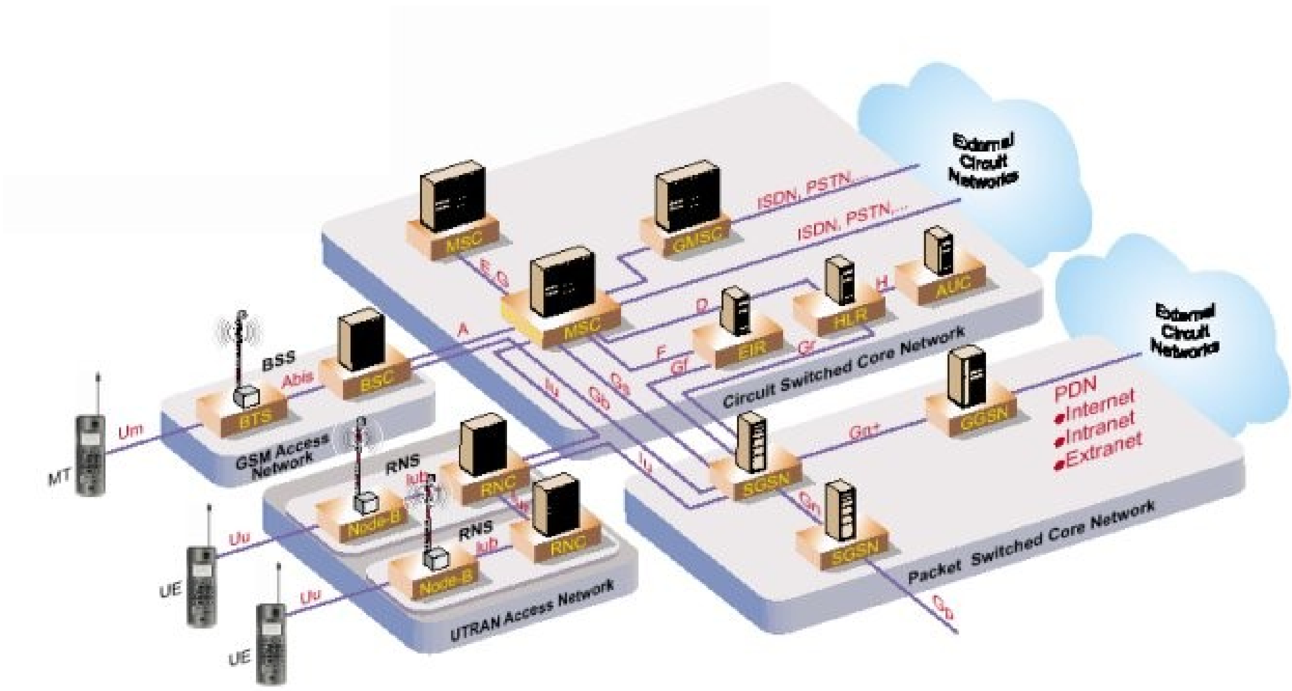

Figure 6: UMTS architecture [7]

3

Accessing UMTS – PDP Context

[3,4,6] Before a mobile terminal can access GPRS services, it must inform the network of its presence. Performing a GPRS Attach procedure (Figure 7). When to the SGSN node does this. The Attach procedure contains: updating location information in the HLR, transferring information from the old SGSN (where the mobile terminal was formerly registered), and the new SGSN.

When a MS wants to transmit or receive data, it has to activate a PDP context activating a PDP context, it informs the reference GGSN of the presence of the mobile terminal and makes it possible to transfer data packets to and from the corresponding user. The PDP context describes the characteristics of the link with the external data networks; type of networks, destination address, the address of GGSN to be used and the quality of service parameters (see later). QoS is defined through:

- Precedence: indicates the way the service is prioritised among users
- Delay: end-to-end delay in transmitting a packet from the origin to destination
- Reliability: depends on its correction capacity and fault tolerance
- Throughput: throughput per user

For each PDP context, the mobile terminal can be assigned a static address established at the time of subscription, or a dynamic address allocated at the time the GGSN activates the PDP context by the operator of the user's home network (HPLMN - Home Public Land Mobile Network) or of the visited network (VPLMN - Visited Public Land Mobile Network). Dynamic addresses only allow data transfers originated by the mobile terminal.

Each terminal can request one or more PDP context activations. These could correspond to different applications and can request different QoS. Such applications can be a file transfer (FTP) while the user talks to phone using UMTS.

When the detach procedure is requested (by the network or by the mobile terminal), all PDP contexts for a given terminal are deactivated. The detach procedure can also be originated implicitly when a predetermined time expires during the period in which there is no mobile terminal activity (e.g., no data sent or received).

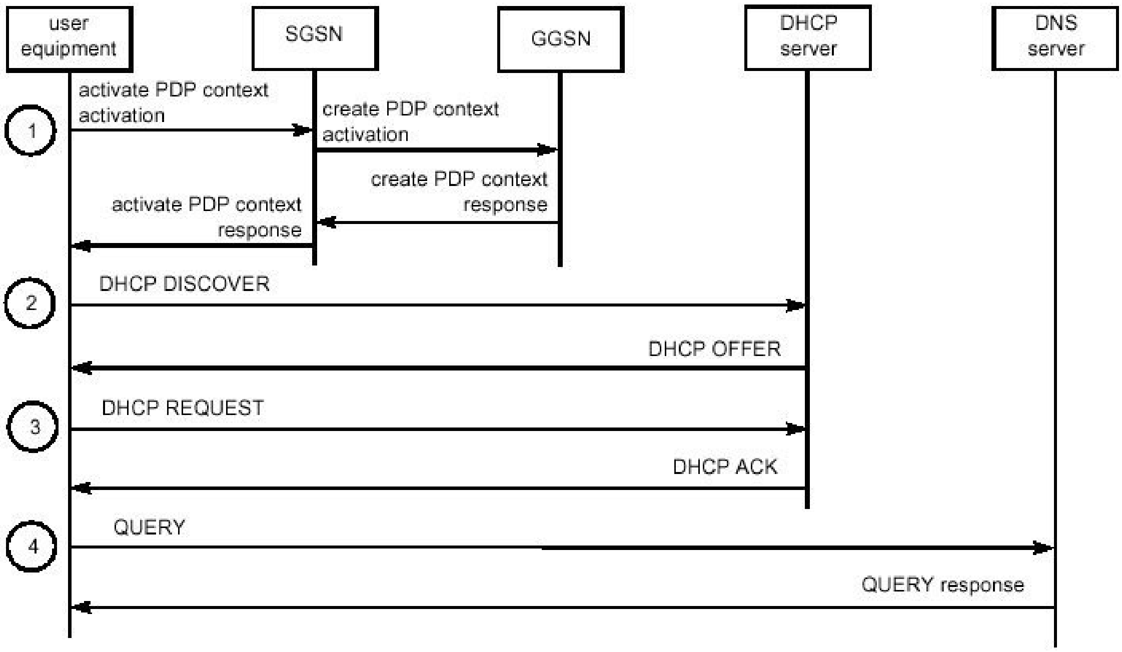

Figure 7: Discovery message sequence [6]

4

Tunnelling Mechanism

[3,30] The most commonly used protocol in the Internet that guarantees that packets are routed between the networks nodes to their final destination is Internet Protocol (IP). This involves each node to be given an IP address. Every IP address contains three fields named class network and host number. IP address is used by the routers to route packets to their destination. They first look packet's network address, route it to its network and then looks for the host. This is more efficient way than using complete IP addresses, in which case the routing tables would be too large.

Now, obviously, if a host is moved from its place/network it will not receive any packets anymore. This is not a problem with wired network but this is a vital problem with wireless networks. Specifically, in the GPRS network packets coming from external networks need a mechanism to be routed to the appropriate host, which can usually change position.

As already mentioned it would be a disaster for the internet the usage of complete IP address. In order to solve that problem the GGSN 'encapsulates' each IP packet containing the address of the SGSN node controlling the mobile station at the time the packet enters the network. In this way, packets addressed for users controlled by a certain SGSN are transferred to that node, creating 'tunnels' between GGSN and SGSN. When a user changes location, this means that it comes under the control of a new SGSN. Then, GGSN changes the encapsulation address. This causes redirection of the opened tunnel of a user to a new destination.

All those problems will be solved by the use of IPv6 in future.

5

User transport plane (UMTS only)

[4] The user plane consists of a layered protocol structure, which provides user information transfer, carrying associated control procedure at the same time. Such procedures can be: flow control, error detection, error connection and error recovery. User plane used in UMTS can be shown in Figure 8.

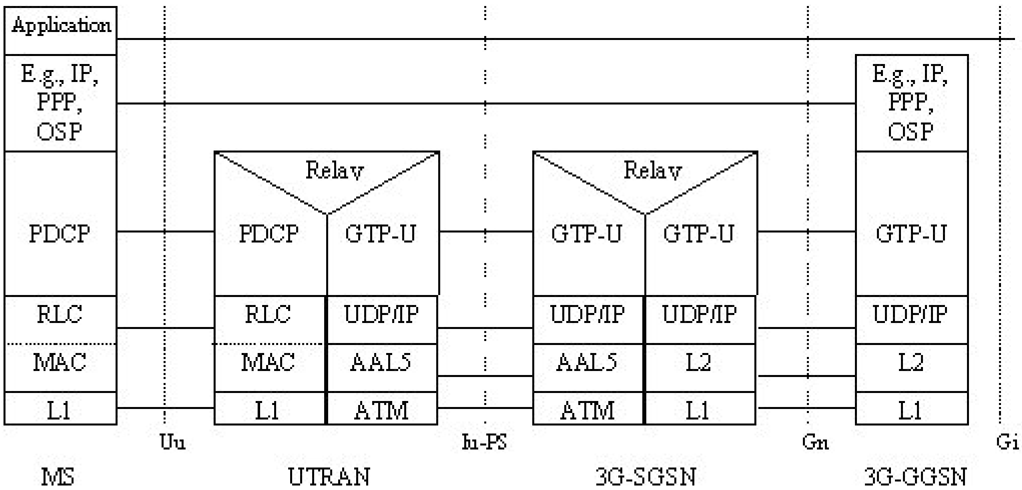

Figure 8: User Transport Plane

For the purpose of this thesis the only interesting part of this user plane is between SGSN and GGSN. For that part there is a protocol stack shown in Table 1.

GTP-U
UDP
IP
AAL-5
ATM

Table 1: SGSN-GGSN User plane

- **GTP-U**: This is the GPRS Tunnelling protocol for the user plane. This protocol is responsible for tunnelling user data between GGSNs. All PDP PDUs shall be encapsulated by the GTP [11].
- **UDP**: This carries GTP PDUs for the protocols that do not need a reliable data link (eg. IP). It also provides protection against corrupted GTP PDUs [11,12].
- **IP**: The well-known IP is used as the backbone network protocol used for routing user data and control signalling. By this time IP v4 is used but IP v6 is on the way [13].
- **ATM**: Even if UMTS does *not* specify layer 2 protocol, the most commonly used is ATM.the information to be transmitted is divided into fixed size cells (53 octets), multiplexed and transmitted. For voice traffic AAL-2 is used. For data traffic AAL-5 provides support for variable-bit rate connection-oriented or connectionless data services [16].

Headers of these protocols can be found in Appendix I. Encapsulating, SAR-PDU is shown in Figure 9:

Stack (octets) = Data + GTP + UDP + IP + AAL-5 + ATM
= Data + 12 + 8 + 20 + AAL-5 + ATM
= Data + 40 + AAL-5 + ATM

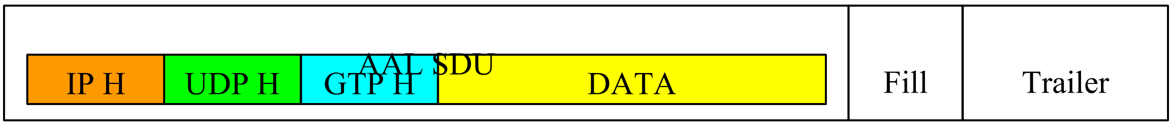

Figure 9: SAR-PDU

6

QoS in UMTS

6.1 QoS Concepts [21]

Quality of Service (QoS) is the ability of a network element (e.g. an application, host or router) to have some level of assurance that its traffic and service requirements can be satisfied. To enable this QoS, it is necessary that all network layers from top-to-bottom and every network element from end-to-end cooperate successfully.

QoS does not create bandwidth. The network is not able to give what it does not have, so availability of bandwidth is the starting point. QoS only manages bandwidth according to application demands and network management settings. Hence, QoS with a guaranteed service level requires resource allocation to individual data streams. The bandwidth allocated to an application in a "resource reservation" is no longer available for use by "best-effort" applications. Considering that bandwidth is a finite resource, a priority for QoS designers has been to ensure that best-effort traffic is not starved after reservations are made. High-priority applications must not disable the low-priority Internet applications. The worst case should be that low-priority applications simply have a slower service, but still function.

There are essentially two types of QoS available:

- ***Resource reservation*** (integrated services): network resources are apportioned according to an application's QoS request, and subject to bandwidth management policy. RSVP provides the mechanisms to do this [IntServ].

- ***Prioritization*** (differentiated services): network traffic is classified and apportioned network resources according to bandwidth management policy criteria. To enable QoS, classifications give preferential treatment to applications identified as having more demanding requirements. [DiffServ] provides this service.

6.2 End to End UMTS QoS [9]

Traffic has to pass different bearer services of the networks on its way from the source terminal equipment (TE) to the destination TE. End-to-end network services are characterised by a certain Quality of service (QoS). This is provided to the user who has requested and paid for this. In order to provide a certain QoS it is necessary to establish a bearer service, with specified characteristics and capabilities, from the source to the destination. End-to-End Service used by the mobile terminal will be realized through several different bearer services: a TE/MT Local Bearer Service, a UMTS Bearer Service and an External Bearer Service.

Given the error characteristics of the radio interface it seems reasonable not to define complex mechanisms like those for fixed networks. These mechanisms must be robust and, at the same time, capable of guaranteeing reasonable resolution.

This part is going to define an end-to-end UMTS QoS, the Classes of services and their parameters. A UMTS bearer service layered architecture is illustrated in Figure 10. Each bearer service on a specific layer offers its individual services provided by the layers.

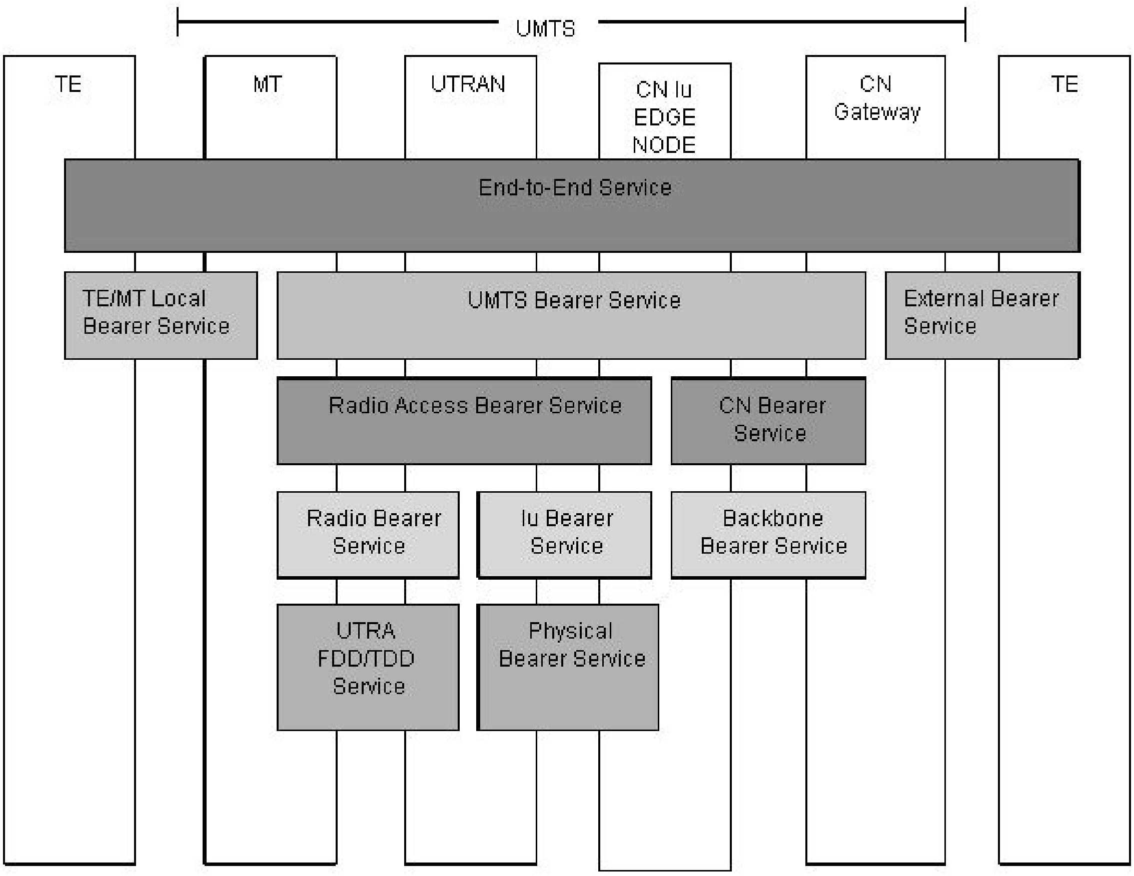

Figure 10: UMTS QoS Architecture [9]

6.3 UMTS bearer service [9]

UMTS bearer service defines the UMTS QoS. It consists of two parts: the Radio Access Bearer Service and the Core Network Bearer Service.

- **Radio Access bearer Service**: Provides transport of user data and signalling between MT and CN Iu Edge Node with the QoS adequate to the negotiated

UMTS Bearer Service or with the default QoS for signalling. This serviced is based on the characteristics of the radio interface and is maintained for a moving MT.

- **Core Network Bearer Service**: Connects the UMTS CN Iu Edge Node with the CN Gateway to the external network. It is responsible for efficiently controlling and utilising the backbone network in order to provide the contracted UMTS bearer service. It has to be noted that the UMTS packet core network shall support different backbone bearer services for different kinds of QoS.

6.4 UMTS QoS Management Functions [9]

The QoS is applied using QoS management function both in user and control plane. A detailed description of these functions is out of the scope of this thesis. However, a simple description of the basic function will give a brief idea how QoS is applied, and how the following analytical work may be realized.

The basic functions that provide end-to-end service is provided by translation and mapping with UMTS external services.

- **Translation function**: This function is necessary for interfacing between UMTS and external networks (ex. IP). It translates the UMTS bearer service attributes to QoS parameters of the external networks.

- **Mapping function**: It is this function that marks each data unit properly, so that it receives the appropriate QoS at the transfer by a bearer service. I other words, application parameters have to be mapped to UMTS Bearer Service parameters, and then, these have to be mapped to the parameters of the underlying bearers (Radio Access Bearer and the Core Network Bearer).

6.5 Classes of Service [9,3,10]

The following QoS classes are proposed for the UMTS system:

- Conversational class
- Streaming class
- Interactive class
- Background class

The main distinguishing factor between these classes is delay-sensitivity. The conversational class is for very delay-sensitive traffic, while the background class is the most delay-insensitive. Conversational and Streaming classes are mainly intended to be used to carry real-time traffic flows. Interactive class and Background are mainly meant to be used by traditional Internet applications like WWW, Email, Telnet, FTP and News. The main difference between Interactive and Background class is that Interactive class is mainly used by interactive applications, e.g. interactive Email or interactive Web browsing, while Background class is meant for background traffic, e.g. background download of Emails or background file downloading. 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 low compared to fixed networks.

6.5.1 Conversational Class

This class is employed for real-time services like telephony speech (e.g. GSM), voice over IP and videoconferencing. These kinds of services are always performed between peers of live end-users (human). For this reason, it is required characteristics given by human perception. The end-to-end delay has to be low with low delay variation, as well. Otherwise the quality would be unacceptable.

6.5.2 Streaming Class

This class employe when the user is looking at (listening to) real time video (audio). This technique of transferring data is called multimedia streaming and allows processing data as a steady and continuous stream. This becomes more and more important with the growth of the Internet, as most of the users have slow connections.

In this case, comparing with Conversational one, only the destination is live (human), so this is a uni-directional transport. It is characterised by that the time variation between information entities (samples, packets) within the flow shall be constant. However, there is not any limitation on low transfer delay. So, as the stream is time aligned at the receiver, the delay variation limit is determined by the capability of the time aligner. Thus, acceptable delay variation is much greater than the delay variation given by the limits of human perception.

6.5.3 Interactive Class

This scheme applies when the end user, that can either be a machine or a human, is on line requesting data from the remote equipment. The destination expects for a response within a certain time. Such cases can be web browsing, database retrieval, tele-machines etc. Therefore, the main requirements for this class concern round-trip delay and data integrity, i.e. guaranteed low bit error rate.

6.5.4 Background Class

This scheme applies when the end-user sends and receives data files in a background process, which is thus secondary in comparison with higher-priority processes.

Examples are background delivery of E-mails, SMS, download of databases and reception of measurement records. Background traffic is characterised by that the destination is not expecting the data within a certain time, i.e. more or less delivery time insensitive. However, low bit error rate has to be preserved.

6.6 QoS parameters [9,3,10]

This section describes the parameters used by UMTS network to manage its resources. The allocation of these is sown in Table 2.

6.6.1 Traffic class

This attribute indicates the type of application (traffic class) for which the Radio Access Bearer service is optimised. In this way, UTRAN is able to make assumptions about the traffic source and optimise the transport for the predetermine traffic class. This is very useful to buffer allocation.

6.6.2 Maximum bit rate (MBR)

Maximum UMTS bits transfer rate at a Service Access Point (SAP). The purpose of this attribute is to make code reservations in the radio interface and to limit the delivered bit rate to applications or external networks.

6.6.3 Guaranteed bit rate (kbps)

The purpose of this attribute is to inform the network about the guaranteed number of bits delivered by UMTS at a SAP within a period of time, divided by that time period. This attribute is necessary for admission control based on available resources and for resource allocation within UMTS.

6.6.4 Delivery order (y/n)

In order UMTS networks to specify if out-of-sequence SDUs are acceptable or not, this attribute indicates whether UMTS bearer shall provide in-sequence SDU delivery or not.

6.6.5 Maximum SDU size (octets)

Obviously, this parameter determines the maximum SDU size allowed by the network. This is used for admission control and policing.

6.6.6 SDU format information (bits)

This attributes gives a list of possible exact sizes of SDUs. Knowing that bearer is less expensive.

6.6.7 SDU Error Ratio

Fraction of erroneous detected or lost SDUs. Defined only for conforming traffic. Its performance is independent of loading conditions in case of reserving resources. Otherwise (Interactive & Background classes), it is used as target value. The purpose of this attribute is to configure the protocols, algorithms and error detection schemes, primarily within UTRAN.

6.6.8 Residual bit error ratio

Undetected bit error ratio in the delivered SDUs. Alternatively, it indicates the bit error ratio in the delivered SDUs, in case no error detection is requested. Its purpose is to be used for the configuration of radio interface protocols, algorithms and error detection coding.

6.6.9 Delivery of erroneous SDUs (y/n/-)

‘yes’ indicates that error detection is employed and that erroneous SDUs are delivered together with an error indication. ‘no’ indicates that error detection is employed but the erroneous SDUs are discarded. Finally, ‘-‘ indicates that SDUs are delivered without considering error detection. This is used to decide whether error detection is needed and whether frames with detected errors shall be forwarded or not.

6.6.10Transfer delay (ms)

This attribute indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service. Delay for an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP. This is necessary for specifying the UTRAN part of the total transfer delay for the UMTS bearer.

6.6.11Traffic handling priority

Specifies the relative importance for handling of traffic of all SDUs belonging to the UMTS bearer compared to the SDUs’ traffic of others. The purpose of this attribute is to differentiate between bearer qualities within the interactive class. It has to be noted that, by definition, priority is an alternative to absolute guarantees, and thus these two attributes cannot be used together for a single bearer.

6.6.12 Allocation/Retention Priority

It specifies the relative importance compared to other Radio access bearers for allocation and retention of the Radio access bearer. This attribute is a subscription parameter, which is not negotiated from the mobile terminal and is used for differentiating between bearers when performing allocation and retention of a bearer.

6.6.13 Source statistics descriptor ('speech'/'unknown')

This final attribute is used in order to specify characteristics of the source of submitted SDUs (ex. speech source).

Traffic Class	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 rate	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

Table 2 : UMTS bearer attributes defined for each bearer class [9]

7

ATM QoS

[17,18] Configured UMTS attributes have to be mapped to ATM, which is the common layer 2 protocol so that operators' ATM PVCs can be configured and the operators can manage their networks. For this reason, there is a need to review ATM QoS and its attributes. This part is going to give a description of the ATM QoS classes (Table 3) that are necessary for the purpose of this thesis; classes that UMTS QoS classes are mapped to.

7.1 Unspecified Bit Rate (UBR)

UBR is made for non-real-time, bursty applications that are tolerant of delay and loss, like ftp. UBR service does not specify service guarantees and is sometimes referred to as the "best effort" service. No commitment on Cell Loss Ratio (CLR) and Cell Transfer Delay (CTD) is made; the sharing is not necessarily fair and there is no specific traffic contract, there is not even a commitment on transmitting data at all. It is the traffic with the least Quality of Service support and can be compared to the traditional best-effort traffic.

The parameters Peak Cell Rate (PCR) and Cell Delay Variation Tolerance (CDVT) are specified:

7.1.1 Peak Cell Rate (PCR)

The Peak Cell Rate (PCR) traffic parameter specifies an upper bound on the rate at which traffic can submitted on an ATM connection. Enforcement of this bound by the Usage Parameter Control (UPC) allows the network to allocate sufficient resources to ensure that the network performance objectives (e.g., for Cell Loss Ratio) can be achieved. PCR is specified as cells per second.

7.1.2 Cell Delay Variation Tolerance (CDVT)

ATM layer functions (e.g., cell multiplexing) may alter the traffic characteristics of connections by introducing Cell Delay Variation. When cells from two or more connections are multiplexed, cells of a given connection may be delayed while cells of another connection are being inserted at the output of the multiplexer. The upper bound on this measure is the Cell Delay Variation Tolerance (CDVT).

7.2 nrt-VBR

Non-Real-Time (nrt-VBR) Service Category. The non-real-time VBR service category is intended for bursty non-real-time traffic applications whose cells are delivered with priority so that the connections with critical response time can go over this service.

VBR-nrt connections are characterized in terms of a PCR, SCR, and MBS. Cell Loss Ratio (CLR) is guaranteed if the sender does not exceed the agreed parameters. CLR is the only QoS parameter that is specified. Typical applications for this service are data transfers for transaction-processing applications such as airline reservation, banking transactions, and process monitoring. Frame relay traffic can also use nrt-VBR service.

7.2.1 Sustainable Cell Rate (SCR)

The SCR (cells/sec) traffic parameter is defined as an upper bound on the average rate of the conforming cells of an ATM connection. This is defined over time scales that are long relative to those for which PCR is defined. Enforcement of this bound by the UPC could allow the network to allocate sufficient resources, but less than those based on the PCR, and still ensure that the performance objectives (e.g., for Cell Loss Ratio) can be achieved.

7.2.2 Maximum Burst Size (MBS)

The MBS traffic parameter specifies the burst size that is allowed in services that are explicitly supporting bursts (rt-VBR and nrt-VBR). This parameter is important to allocate the buffers size and it’s also needed for the Generic Cell Rate Algorithm (GCRA) that decides whether the cells are conformant, marked out-of profile or dropped.

7.2.3 Cell Loss Ratio (CLR)

The CLR QoS parameter for a connection is defined as:

$$CLR = Lost\ Cells\ /\ Total\ Transmitted\ Cells$$

Lost and transmitted cells counted in severely eroded cell blocks should be excluded from the cell population in computing cell loss ratio. The CLR parameter is the value of CLR that the network agrees to offer as an objective over the lifetime of the connection.

ATM parameters for the corresponding ATM Traffic classes can be shown in Table 3:

7.3 rt-VBR

The real-time VBR service category is made for bursty real time links with tightly constrained delay and delay variation. Sources are expected to transmit at a rate that varies with time. Equivalently the source can be described as “bursty”. Voice and video would be appropriate applications for this category. Comparing with nrt-VBR class, the additional attributes that are being specified for this class of service are: Peak-to-peak CDV and maxCTD. Cells that are delayed beyond the value specified by maxCTD are assumed to be of significantly reduced value to the application. Finally, real-time VBR service may support statistical multiplexing of real-time sources.

7.3.1 Peak-to-peak Cell Delay Variation (CDV)

Peak-to-peak CDV is the difference between the fixed delay (best case) and the delay (worst case) determined by the probability, whether cells arrive within a certain time or whether they are regarded as late or lost, according to the traffic agreement (Figure 11). This, worst case, is equal to a value likely to be exceeded with probability no greater than a (The cells that arrive late cannot expect further priority treatment and can be dropped or given a lower priority). Thus, the peak-to-peak CDV is equal to the difference between the (1-a) and the fixed delay. Assuming that the fixed delay is the reference delay for the two point CDV, the range of the distribution of the two-point (cell transfer delay between two measurement points) CDV is the same as the peak-to-peak CDV.

7.3.2 Maximum Cell Transfer Delay (MaxCTD)

MaxCTD is defined as the sum of the fixed delay, given through physical parameters and switching time over components, and the peak-to-peak Cell Delay Variation, which has a probability of 1- a that the cells arrive in time.

maxCTD = Fixed Delay + Peak-to-peak CDV

CTD is measured at two points in different ATM switches, at the exit point of the first ATM switch and at the entry point of the next ATM switch and then compared with the agreed maxCTD parameter:

CTD = Cell entry event at Point 2 - Cell exit event at Point 1

Early cells are queued for later transmission and nonconforming cells can be queued or dropped.

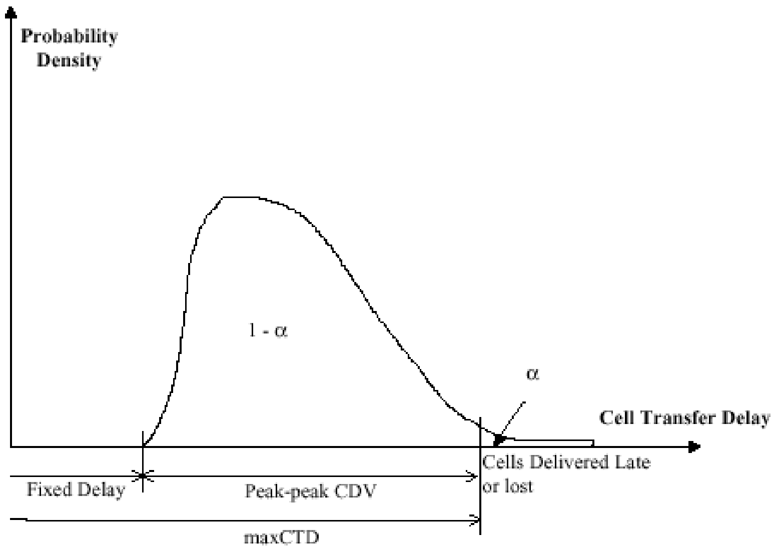

Figure 11: Cell transfer delay probability density model (for real-time service categories)

7.4 CBR

The constant bit rate (CBR) service category is made for connections that require a constant bandwidth, which is continuously available during the connection. CBR service is intended for real-time applications requiring tightly constrained delay variation (e.g., voice, video, circuit emulation) but is not restricted to these applications. The source may emit cells at, or below the negotiated Peak Cell Rate (and may also even be silent), for periods of time. Cells which are delayed beyond the value specified by the maximum cell transfer delay (maxCTD) are considered less important and can be discarded at any time. Traffic marking, policing and shaping is important for real-time traffic support. Comparing with rt-VBR class, SCR and MBS attributes are not available.

	ATM Layer Service Category					
Attribute	CBR	rt-VBR	nrt-VBR	UBR	ABR	GFR
Traffic Parameters						
PCR and CDVT	Specified			Specified	Specified	Specified
SCR, MBS, CDVT	n/a	Specified		n\a		
MCR	n\a				Specified	n\a
MCR, MBS, MFS, CDVT	n\a					Specified
QoS Parameters						
Peak-to-peak CDV	Specified		Unspecified			
MaxCTD	Specified		Unspecified			
CLR	Specified			Unspecified	See Note 1	See Note 7
Other Attributes						
Feedback	Unspecified				Specified	Unspecified

Table 3: ATM attributes per traffic class [17]

The current (might be changed) standards for the mapping from UMTS to ATM traffic classes are illustrated in Table 4:

UMTS class	Conversational	Streaming	Interactive	Background
ATM class	CBR	rt-VBR	nrt-VBR	UBR

Table 4: UMTS-ATM class mapping

8

Mapping

In this section, there is an attempt to map UMTS attributes of the UMTS QoS classes to ATM attributes of the appropriate ATM traffic classes, determined by Table 4. Probably, the best way doing that is by starting from the simplest class (background) as there are the least attributes to bear in mind and then build up towards more complex classes.

8.1 Generally

Each MS can request one, or more PDP contexts. Each one might request for a different QoS. An example could be simultaneous appearance of a voice service and an ftp application, which are a Conversational (or Circuit Switched) and a Background class application, respectively. Between two given ATM network node, there will be (in theory) 5 types of PVCs for traffic (Figure 12): one set for circuit-switched voice and then one set for each of the QoS classes. The number of PVCs for voice depends on the number of calls that can be supported. This is due to the way in which AAL-2 multiplexes the call into PVCs. For data, in fact, one PVC per QoS is enough, and each of them will transport the traffic for all the PDP contexts of a particular class.

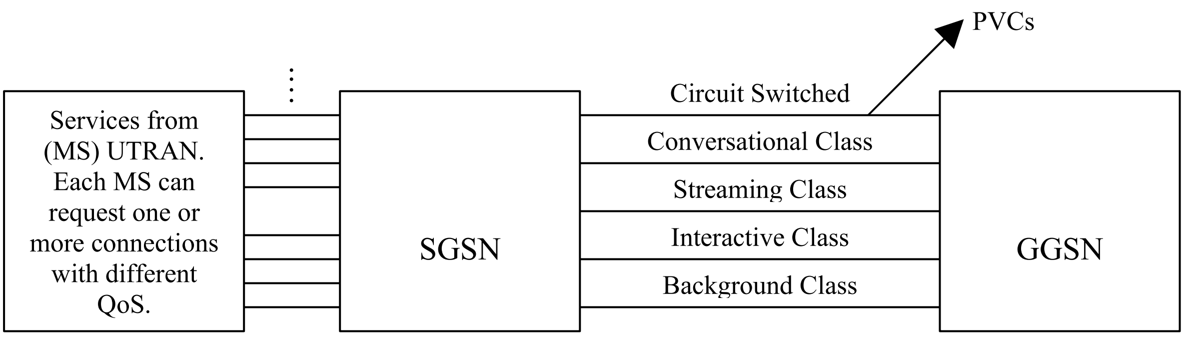

Figure 12: From PDPs to PVCs

8.2 Background to UBR

For the background class, the only parameter that has to be mapped to ATM is the **Maximum bit rate**. Even if UMTS needs to specify **SDU error ratio**, **Residual bit error ratio** and **Delivery of erroneous SDUs**, there is no need to map these to ATM as all these are used to guarantee bit integrity, which is already guaranteed by ATM. Additionally, ATM does not provide any error control mechanisms. The higher layers can only do this.

Assuming an instant that there are m PDP contexts (users), each one using a different SDU (Figure 13). This means that there are up to m different SDU sizes. SDU varies and depends on the type of application (email, ftp, e.t.c.). These SDUs are multiplexed by AAL-5 and the ATM cells are transmitted through the background PVC.

[29] ”AAL-SDU is received at the AAL-SAP (length =1 to 65535 octets). Data passes to CPCS to become CPCS-SDU. An 8-octet trailer and a pad are added to form a CPCS-PDU with a length of multiplies of 48 octets. The trailer has fields for the CPCS-SDU length and error protection for the whole PDU. The CPCS-PDU is passed to the SAR and becomes an SAR-SDU. The SAR segments the SAR-SDU into 48-octet SAR-PDUs without any additional header or trailer information. The SAR passes the SAR-PDUs to the ATM layer. The last one from an SAR_SDU is signified to the ATM layer. The ATM layer builds an ATM header and includes the end of message (EOM) flag as a 1 in bit 2 of the payload type field of the cell containing the SAR PDU from the frame”.

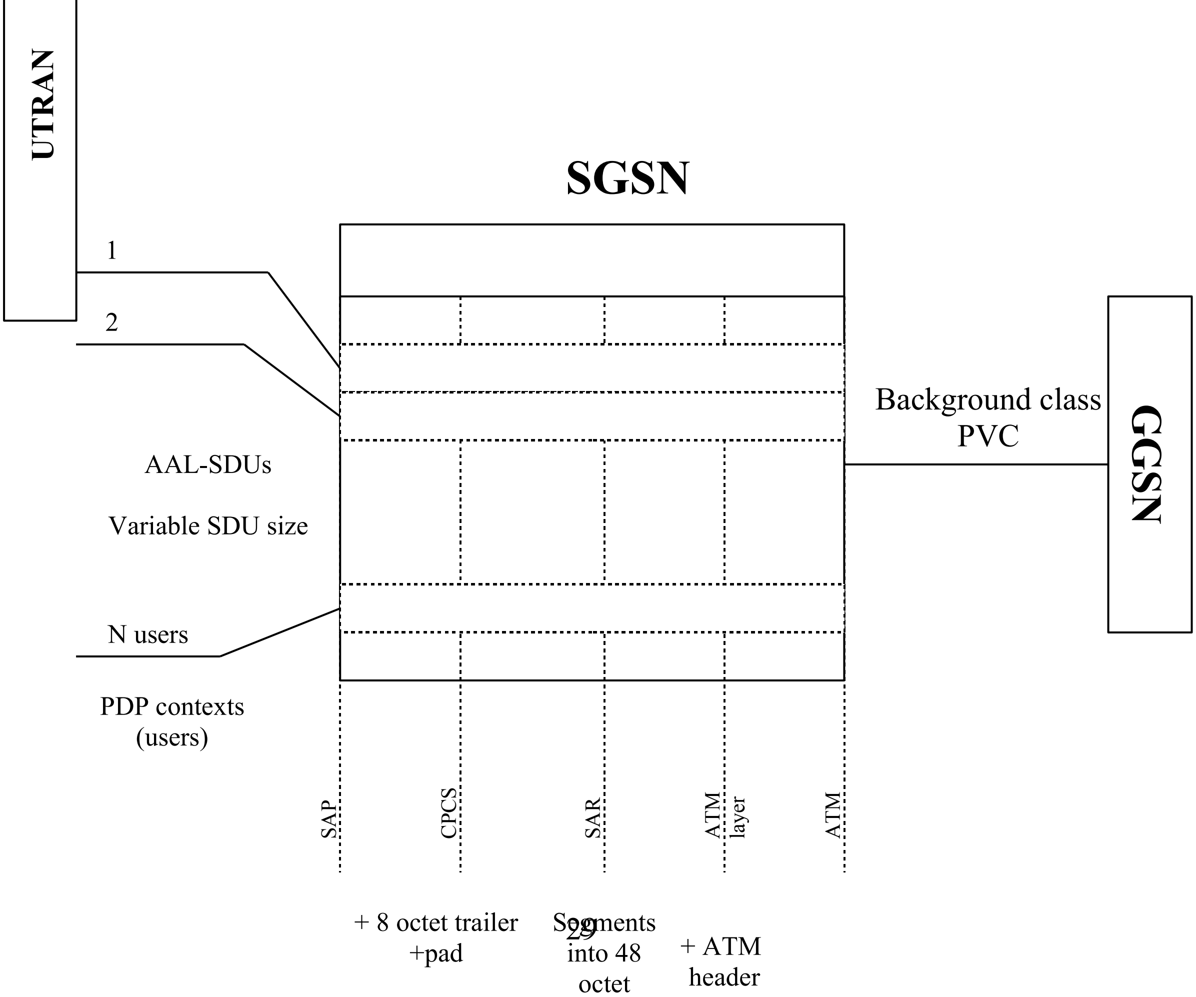

Figure 13: PDP Contexts through AAL-5

Given the maximum bit rate (MBR), the Peak cell rate of the ATM PVC, that is going to carry the background class PDP contexts, is going to be computed.

The easier way doing that is by start considering a single SDU, arriving at the SAP-AAL. In this case, the number of cells for a given single SDU can be calculated:

$$\textit{No of cells for given single SDU} = (SDU + 8 + pad) / 48$$

Or,

$$\textit{max no. of cells per SDU} = (\textit{max } SDU + 8 + pad) / 48$$

Note: Median might be used in order to avoid peaks.

Then, knowing the MBR for each PDP context (MBR_i) the rate of SDUs from a single PDP context can be computed:

$$\textit{SDU rate for } i^{th} \textit{ PDP context} = (MBR_i/8)/ \textit{ mean } SDU$$

But, for computing the PCR, it is necessary to have the maximum SDU rate. The smaller the SDU size the biggest the SDU rate. So:

$$\textit{max SDU rate for } i^{th} \textit{ PDP} = (MBR_i/8)/ \textit{ min } SDU$$

The next step is to compute the number of cells for a single PDP context:

$$\textit{No. of cells for } i^{th} \textit{ PDP} = \textit{max SDU rate for } i^{th} \textit{ PDP} * \textit{max no. of cells per SDU}$$

$$\textit{No. of cells for } i^{th} \textit{ PDP} = ((MBR_i/8)/ \textit{ min } SDU)* (\textit{max } SDU + 8 + pad) / 48$$

So worst case PCR for i^{th} PDP with $MBR_i=MBR_{max}$ (PDP context with maximum MBR) context is:

$$PCR_{max} = ((MBR_{max}/8)/ \textit{ min } SDU)* \textit{ ceil } ((\textit{max } SDU + 8) / 48)$$

where *ceil* is a function that rounds to the infinity.

The final stage is to compute PCR for all PDP contexts (assume m) that are about to be carried by the background PVC.

Suppose there are m PDP contexts supplying packets at a maximum rate PCR_{max}. Then, $\lambda/m = PCR_{max}$, according to Appendix II. This system can be modelled as m M/M/1 systems operating in parallel, each with a packet arrival rate of, λ/m and a service rate of μ/m. Then,

$$PCR = m * \textit{average no of packets in any these M/M/1 systems}$$

$$\boldsymbol{PCR = m*\lambda/(\mu-\lambda)}$$

where:

m: number of PDP contexts
μ: transmission rate
λ: PCR for i^{th} PDP (arrival rate)
(see appendix III, function *PCR,* for Matlab implementation)

It would also be really useful if the throughput could be computed. Each SDU also contains Stack header bits (40), coming from the upper layer protocols headers. Then, throughput for single PDP would be:

$$\textit{Throughput for } \text{single PDP} = ((MBR_i\text{-}40)/8) / (53 * \textit{PCR for } \text{single PDP})$$

And a good estimation of the overall throughput would be:

$$\textit{Throughput for m PDPs} = (m*(MBR_{max}\text{-}40)/8) / (53 * PCR)$$

8.3 Interactive to nrt-VBR

Additional attributes, to take into account:

- UMTS: Traffic handling priority.
- ATM: SCR, MBS, CLR.

An acceptable CLR has to be requested (typically 10^{-9}). It would be really useful if the effective bandwidth could be found for each case. By saying effective bandwidth is meant the necessary bandwidth for one or more variable bit rate applications for a specific QoS without wasting the channel.

When traffic parameters are known, effective bandwidth can be calculated and may be used to obtain a circuit-switched style call acceptance and routing for ATM networks [26].

[27] ATM can be modelled as an independent identically distribution source.

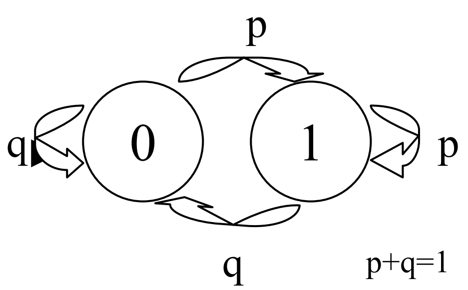

Figure 14: IID Marcov Model

For this simple source effective bandwidth is:

$$EBW = \frac{\ln(pe^{\theta} + q)}{\theta} \text{ when, } CLP = \exp(-\theta B)$$

where:
EBW: Effective Bandwidth
B: Buffer size (cells)
CLP: Cell loss probability

Assuming that a time unit is the time it takes one cell to go to the link, the link bandwidth BW =1. Then the number of sources that the given channel can carry (with the specific QoS/CLP) can be easily calculated by dividing BW by the EBW or in other words No of Sources = 1/EBW.

This can be implemented using Matlab (function EBW in Appendix III).

Now being able to calculate the effective bandwidth it is easier to configure ATM PVCs parameters so that they guarantee the required QoS.

Effective BW is always somewhere between mean BW (least expensive) and peak BW (most expensive). Considering the case of the interactive UMTS class, SCR parameter could be considered as the worst case of the mean BW, as this is the maximum average cell rate. So by specifying SCR, mean BW is specified, as well.

The probability of an arriving cell p = Mean Bandwidth / Link Bandwidth.

Then, p=SCR/BW and ***SCR=p*BW***

Given the number of sources (users or applications requests of same user) and the capacity of the buffer (cells), SCR can be calculated.
Analytically,

$$CLP = \exp(-\theta * B) \Leftrightarrow$$
$$\ln(CLP) = -\theta * B \Leftrightarrow$$
$$\theta = \frac{\ln(1/CLP)}{B}$$

also, EBW=1/Number of Sources

Thus,

$$EBW = \frac{\ln(pe^{\theta} + q)}{\theta} \Leftrightarrow$$
$$\theta * EBW = \ln(p * e^{\theta} + 1 - p) \Leftrightarrow$$
$$e^{\theta * EBW} = p * (e^{\theta} - 1) + 1 \Leftrightarrow$$
$$p = \frac{e^{\theta * EBW} - 1}{e^{\theta} - 1}$$

This can be implemented using Matlab (function SCR in Appendix III).

The next task is to configure MBS parameter. This parameter has to do with the buffer size. The bigger is the MBS the more is the buffer capacity requirement, assuming the same departure rate (ex. processing speed of the router). In case that the buffer is too big so that that MBS can be big as well, the delay in the buffer is increased and consequently congestion is increased, as well. Big buffers are more expensive anyway. On the other hand, reducing MBS causes bit rate reduction (less speed). So a compromise has to be done. For this reason it would be useful to have a graph showing the CLP due to buffer overflowing against MBS.

[28] Analytically, the probability that the buffer can overflow can be found using queuing theory. In particular, this can be modelled as a finite Source and Server Marcov model (Figure 15):

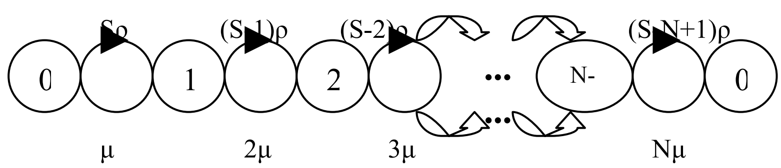

Figure 15: Finite Source and Server Marcov Model

The probability that the buffer blocks at a state r is an engset distribution given by:

$$p_r = \frac{(\rho * h)^r \binom{S}{r}}{\sum_{i=0}^{N} (\rho * h)^i \binom{S}{i}}$$

Then, the probability that the buffer overflows, or in other words that the buffer blocks at the last state is an engset distribution given by:

$$p_N = \frac{(\rho * h)^N \binom{S}{N}}{\sum_{i=0}^{N} (\rho * h)^i \binom{S}{i}}$$

This can be implemented using Matlab (function *engset*, *comb*, in Appendix III).

Concluding, a single program that will be able to help configuring a PVC of interactive class would be useful (function *inter* in Appendix III).

Running the program…

```
Give the link capacity (bits/s) ex.=5000000000 : 550000000
Give the maximum MBR between all PDPs (bits/s) ex.=2000000 :
2000000
Minimum number of sources should support (worst case) = 275
So Minimum BW for each source = 2000000.0000 bits/s
Give the minimum SDU (1 to 65535 octets) : 1000
Give the maximum SDU (1 to 65535 octets) : 20000
Give the transmission rate ex.=2000000000 : 200000
Give number of PDP contexts (typ. x*1000) : 1000
Give the Buffer Size (cells) ex.=600 : 750
Give the Cell loss probability (10^-x) ex.=9 : 9
Give the required number of sources ex.=100 : 150
Effective Bandwidth = 0.00667
So PVC should be configured with : PCR = 1088.77 cells/s , SCR =
3616576.16 cells/s
MBS can be estimated from the plot
```

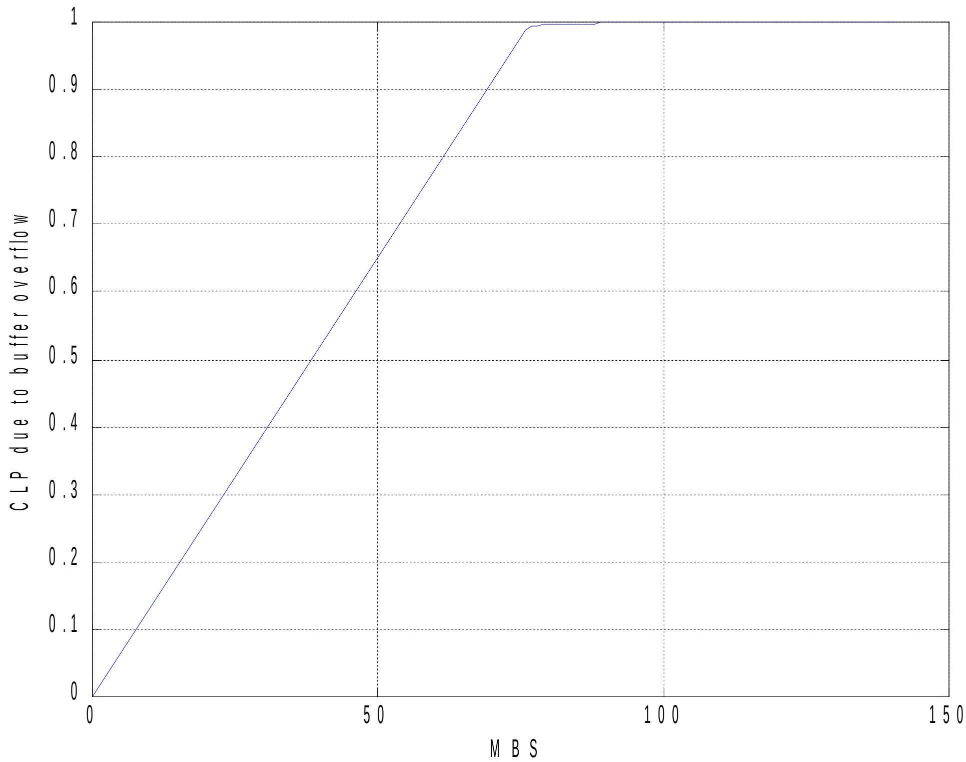

Figure 16: Graph from *inter* function

8.4 Streaming to rt-VBR

Additional attributes to take into account:

- UMTS: Guaranteed bitrate, Transfer delay.
- ATM: Peak-to-peak CDV, MaxCTD .

Assume m PDP contexts' data arriving at a common "concentrator" so that they can be statistically multiplexed to a single data stream, routed and transmitted through the appropriate ATM PVC. This system can be considered as an M/M/1 system (Appendix II). Then, PDPs' SDUs can be taken as variable length packets. In case of this traffic class, both maximum SDU size and minimum SDU size is given by Maximum SDU size and SDU format information, respectively. The maximum SDU rate can now be calculated: (MBR/8)/minSDU size, which is the maximum λ according to M/M/1 theory. When λ has its maximum value, this is the worst case of delay, as there are "long queues". So, worst case average delay = $1/(\mu-\lambda)$ where $1/\mu$ is the mean transmission time. But average number of packets at the concentrator = $\lambda/(\mu-\lambda)$. Thus,

$$\textit{overall average delay} = [1/(\mu-\lambda)] * [\lambda/(\mu-\lambda)] = \lambda/(\mu-\lambda)^2$$

where λ=(MBR/8) / minSDU size.

In order to be able to relate that with Peak-to-peak CDV ATM attribute, it is necessary to find the average cells per packet, which can be computed from the above SDU attributes, as well. Assuming average SDU size = SDU_a, then,

$$\textit{average cells per SDU} = \mathrm{ceil}\,(SDU_a/48)$$

$$\begin{aligned}\textit{Average delay per cell} &= \textit{Average delay per SDU /average no of cells}\\ &= [1/(\mu-\lambda)] / \mathrm{ceil}\,(SDU_a/48)\\ &= 1 / [(\mu-\lambda) / \mathrm{ceil}\,(SDU_a/48)]\end{aligned}$$

Transfer delay $D_{transfer}$ indicates maximum delay for 95% of the distribution delay of all delivered SDUs during a connection.

Then,

$$\textit{max. delay per cell} = \textit{max delay per SDU /max number of cells per max SDU}$$

but,

$$\textit{max SDU } (SDU_{max}) \textit{ is given,}$$

so,

$$\textit{number of cells per maxSDU} = \mathrm{ceil}(SDU_{max} / 48)$$

Thus,

$$\textit{max delay per cell} = D_{transfer} / \mathrm{ceil}(SDU_{max} / 48)$$

Summing all the PDP contexts' maximum delays, for 95% of the distribution delay of all delivered SDUs during a connection, this max delay per cell, would be near that (95%) percentage of cells, arriving within a certain connection.

Also,

$$\textit{Peak-to-peak CDV} = \textit{fixed delay} - \textit{max delay per cell}$$

Obviously,

$$\textit{maxCTD} = \textit{Fixed Delay} + \textit{Peak-to-peak CDV}$$

Guaranteed bit rate provides the minimum guaranteed bandwidth required by a certain connection. The sum of all the connections (PDP contexts) for instance, has to be equal, or less than the ATM PVC capacity. So, this UMTS attributes can be used for easily computing the number of connections that an ATM PVC can provide.

8.5 Conversational to CBR

UMTS traffic attributes, for this class, are identical to the Streaming class one. The only difference on mapping from UMTS Conversational class to CBR ATM class, comparing mapping process between Streaming to rt-VBR, is that in this case, SCR and MBS ATM attributes are not available. This is reason, that such connections are characterised as circuit-switched-like links.

9

Conclusions

9.1 Summary of the results

The work done for mapping UMTS to ATM attributes has shown that even for the simplest traffic classes, this task can be very complex. This is getting worse in case radio bearer has to be taken into account. In that case, things become much more complex and it is necessary to do many “better world” assumption, which reduces approximation’s accuracy.

In this thesis, there were obtained some formulas. These might be simplistic but at least they give an idea. Anyway, there was not available time for doing something more complex, keeping in mind supervisions problems.

It was an attempt to map most important attributes in a hierarchical manner.

Starting from the simplest, background, class the first attribute that has to be configured is PCR. The formulas obtained (Section 8.2), make use of some characteristics that in real life might or might not be possible to obtain (like min. SDU). Even though, they can be very useful.

Then, is Interactive class’s turn. SCR, MBS and CLR are the new attributes that have to be configured. Using theory for finding the effective bandwidth and queuing theory, SCR and MBS can be configured. SCR is related to CLR (CLP) so a compromise has to be done.

Streaming class requires configuring Peak-to-peak CDV and MaxCTD. For this purpose maximum delay per cell is required which is computed using queuing theory. Then, it is easy to calculate both Peak-to-peak CDV and MaxCTD.

The final class is conversational class for which, theoretically, there are no any additional parameters to configure. But, SCR and MBS are not available which require consideration of all the parameters and their combination. Unfortunately there was not time doing that.

9.2 Further Work

It would be really interesting if instead of mapping UMTS parameters to ATM, trying to map these to MPLS, which is probably the next layer 2 protocol. An interesting task could also be to do some modelling using software packages that make life better and definitely better. This section aims to introduce MPLS and its basic features, and also considering the idea of using software-modelling packages.

9.2.1 MPLS

Although QoS support is available in conventional IP, Multiprotocol Label Switching (MPLS) gives the ability to scale IP QoS to core network without the complicated process of mapping onto alternative transport QoS mechanisms, as with ATM [22].

MPLS is a label swapping (mapping) and forwarding technology, but it integrates label swapping with network layer routing. Layer swapping/mapping means the changing of the label value in the packet header as the packet moves from one node to another [23].

MPLS is probably the solution to address the problems faced by today-networks; speed, scalability, QoS, traffic engineering. MPLS has emerged as an elegant solution to meet the bandwidth management and service requirements for next generation IP based backbone networks. This can exist over existing ATM (and frame relay) networks [24].

9.2.2 Optimising Networked Applications

[20] "Highly complex and influenced by so many factors, network applications are difficult to build and deploy successfully in real world. Traffic is hard to decipher; protocols interact in ways that are difficult to follow, and the new network application languages add unexpected overheads. Secondly, the demand of these new applications is so great, that neither a step-by-step implementation nor a test-bed set-up to examine performance and operations is viable anymore. Finally, the development lab environment is far different than the production environment of a working corporate network.

To avoid the loss and damage caused by poorly performing applications, developers need to follow a methodology for application development, testing, and deployment and use a software suite that is cost-effective, easy to use, accurate, and flexible enough to handle the most complex of network scenarios. This approach must enable both the developer and IT professional to check the performance of an application before it is deployed and to adjust both the network and the application characteristics, in accordance with observed performance and traffic loading." Such a useful commercial application is OPNET (Figure 15).

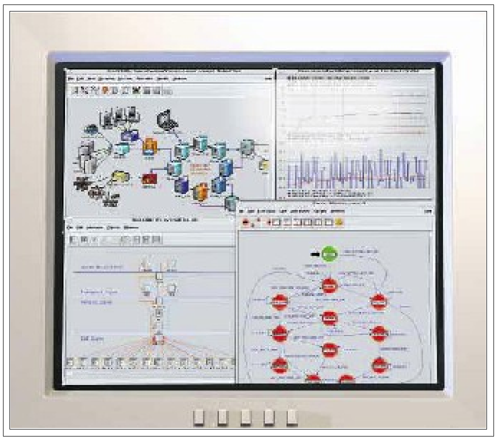

Figure 17: OPNET Modeller [19]

ACKNOWLEDGEMENTS

This thesis, and generally my studies, would have not been possible without both the continuous encouragement and the financial support of my parents, who I would like to express my gratitude and love.

There are also some people who provided me valuable help and I would like to thank.

Dr. Sean Monaghan, my university supervisor, for his advices and his help whenever I needed that.

Mr. Luis Alcantara Garcia, BT Cellnet Supervisor (not any more), for his help even if the conditions were not the best.

Dr. Ali Salman, BT Cellnet line manager, for his interest on allocating to me a project with BT Cellnet.

My brother Georgios Vlachogiannis and my friend Maria Levanti, for their encouragement and help during my project.

Tim Berners-Lee, for his undoubted useful invention named World Wide Web. And of course all the forums and organisations that provide such useful information on the WWW that everyone can access free of charge.

References

[1] No. 11 Report from the UMTS Forum

[2] GSM PHASE 2+ GENERAL PACKET RADIO SERVICE GPRS: ARCHITECTURE, PROTOCOLS, AND AIR INTERFACE. - IEEE Communications Surveys - Third Quarter 1999, vol. 2 no. 3 http://www.comsoc.org/pubs/surveys

[3] UMTS Mobile Communications for the future, by Flavio Muratore.

[4] 3G TS 23.060 General Packet Radio Service (GPRS); Service description; http://www.3gpp.org/

[5] GPRS White Paper - Cisco Systems, http://www.cisco.com

[6] BT Technol J Vol 19 No 1 January 2001
GSM to UMTS — architecture evolution to support multimedia
http://www.bt.com/bttj/

[7] DS-UMTS Revision A - UMTS protocol support
http://www.protocols.com

[8] BT Technol J Vol 19 No 1 January 2001
Voice and Internet multimedia in UMTS networks
http://www.bt.com/bttj/

[9] 3G TS 23.107 Technical Specification Group Services and System Aspects; QoS Concept and Architecture; http://www.3gpp.org/

[10] Master's Thesis Report Dan Ouchterlony Måns Molin - New service opportunities in 3G: A relative system resource usage model for UMTS QoS classes

[11] 3G TS 29.060 Technical Specification Group Core Network; Technical Specification Group Core Network; General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp Interface; http://www.3gpp.org/

[12] IETF RFC 768 (8/1980): "User Datagram Protocol".

[13] IETF RFC 791 (9/1981): "Internet Protocol".

[14] IETF RFC 793 (9/1981): "TCP, Transmission Control Protocol".

[15] 3G TS 25.414 Technical Specification Group Radio Access Network; UTRAN Iu Interface Data Transport and Transport Signalling; http://www.3gpp.org/

[16] ITU-T Recommendation I.363.5 (8/1996): "B-ISDN ATM Adaptation Layer Type 5 Specification".

[17] ATM FORUM Traffic Management Specification Version 4.1 AF-TM-0121.000 March 1999 – http://www.atmforum.com

[18] University of Bern Computer Networks and Distributed Systems Prof. T. Braun Project: Differentiated Services in ATM-Networks

[19] OPNET Technologies, Inc. – http://www.opnet.com

[20] Optimizing and Deploying Networked Applications Technology Guide – http://www.techguide.com

[21] White Paper –The Need for QoS - http://www.qosforum.com

[22] University of Essex - IP and Optical Layer MPLS by M. J. Reed, 2001

[23] MPLS and Label Switching Networks, by Uyless Black

[24] MPLS - Web ProForum Tutorials – http://www.iec.org

[25] Class Notes - Santosh S. Venkatesh 1997 TCOM 501: Networking - Theory and Fundamentals University of Pennsylvania -http ://www.seas.upenn.edu:8080/~tcom501/On-line_Book/Queueing_Systems/queueing_systems.html

[26] Effective Bandwidths for Multiclass Marvov fluids and other ATM sources, by George Kesidis, Jean Walrand and Cheng-Shang Chang. IEEE/ACM TRANCACTIONS ON NETWORKING, VOL1, NO. 4, AUGUST 1993

[27] University of Essex – Broadband Networks: Infrastructures and Technologies. Effective Bandwidth by Sean Monaghan, 2001.

[28] University of Essex – Network Engineering. Traffic Problems with Simple Markov Model by Mark C. Sinclair, 2001.

[29] University of Essex – Broadband Networks: Infrastructures and Technologies. ATM by M. J. Reed, 2001

[30] Computer Networks by Andew S. Tanebaum – 3rd Edition

Appendix I

Headers

I.1 GTP Header

[11] "The GTP header is a variable length header used for both the GTP-C and the GTP-U protocols. The minimum length of the GTP header is 8 bytes. There are three flags that are used to signal the presence of additional optional fields: the PN flag, the S flag and the E flag. The PN flag is used to signal the presence of N-PDU Numbers. The S flag is used to signal the presence of the GTP Sequence Number field. The E flag is used to signal the presence of the Extension Header field, used to enable future extensions of the GTP header defined in this document, without the need to use another version number. If any of these three flags are set, the length of the header is at least 12 octets and the fields corresponding to the flags that are set shall be evaluated by the receiver. The sender shall set all the bits of the unused fields to zero. The receiver shall not evaluate the unused fields.
The GTP-C and the GTP-U use some of the fields in the GTP header differently. The different use of such fields is described in the sections related to GTP-C and to GTP-U.

Always present fields:

- Version field: This field is used to determine the version of the GTP protocol. For the treatment of other versions, see subclause 11.1.1, "Different GTP versions". The version number shall be set to '1'.

- Protocol Type (PT): This bit is used as a protocol discriminator between GTP (when PT is '1') and GTP' (when PT is '0'). GTP is described in this document and the GTP' protocol in GSM 12.15. Note that the interpretation of the header fields may be different in GTP' than in GTP.

- Extension Header flag (E): This flag indicates the presence of the Next Extension Header field when it is set to '1'. When it is set to' 0', the Next Extension Header field either is not present or, if present, must not be interpreted.

- Sequence number flag (S): This flag indicates the presence of the Sequence Number field when it is set to '1'. When it is set to '0', the Sequence Number field either is not present or, if present, must not be interpreted. The S flag shall be set to '1' in GTP-C messages and in GTP-U/GTP signalling type of messages.

- N-PDU Number flag (PN): This flag indicates the presence of the N-PDU Number field when it is set to '1'. When it is set to '0', the N-PDU Number field either is not present, or, if present, must not be interpreted. This flag is significant only for GTP-U. As such, this flag is unused by GTP-C and it shall be ignored by a GTP-C receiving entity.
- Message Type: This field indicates the type of GTP message. The valid values of the message type are defined in subclause 7.1 for both GTP-C and GTP-U.
- Length: This field indicates the length in octets of the payload, i.e. the rest of the packet following the mandatory part of the GTP header (that is the first 8 octets). The Sequence Number, the N-PDU Number or any Extension headers shall be considered to be part of the payload, i.e. included in the length count.
- Tunnel Endpoint Identifier (TEID): This field unambiguously identifies a tunnel endpoint in the receiving GTP-U or GTP-C protocol entity. The receiving end side of a GTP tunnel locally assigns the TEID value the transmitting side has to use. The TEID values are exchanged between tunnel endpoints using GTP-C (or RANAP, over the Iu) messages.

Optional fields:

- Sequence Number: This field is an optional field in G -PDUs. It is used as a transaction identity for signalling messages having a response message defined for a request message, that is the Sequence Number value is copied from the request to the response message header. In the user plane, an increasing sequence number for T-PDUs is transmitted via GTP-U tunnels, when transmission order must be preserved.
- N-PDU Number: This field is used at the Inter SGSN Routeing Area Update procedure and some inter-system handover procedures (e.g. between 2G and 3G radio access networks). This field is used to co-ordinate the data transmission for acknowledged mode of communication between the MS and the SGSN. The exact meaning of this field depends upon the scenario. (For example, for GSM/GPRS to GSM/GPRS, the SNDCP N-PDU number is present in this field).
- Next Extension Header Type: This field defines the type of Extension Header that follows this field in the GTP-PDU.

Octets	Bits 8	7	6	5	4	3	2	1
1	Version			PT	(*)	E	S	PN
2	Message Type							
3	Length (1st Octet)							
4	Length (2nd Octet)							
5	Tunnel Endpoint Identifier (1st Octet)							
6	Tunnel Endpoint Identifier (2nd Octet)							
7	Tunnel Endpoint Identifier (3rd Octet)							
8	Tunnel Endpoint Identifier (4th Octet)							
9	Sequence Number (1st Octet)1) 4)							
10	Sequence Number (2nd Octet)1) 4)							
11	N-PDU Number2) 4)							
12	Next Extension Header Type3) 4)							

(*) This bit is a spare bit. It shall be sent as ‘0’. The receiver shall not evaluate this bit.
1) This field shall only be evaluated when indicated by the S flag.
2) This field shall only be evaluated when indicated by the PN flag.
3) This field shall only be evaluated when indicated by the E flag.
4) This field shall be present when any one or more of the S, PN and E flags are set.“

I.2 UDP Header

[12]

Octets	Bits 8	7	6	5	4	3	2	1
1	Source Port							
2	Source Port							
3	Destination Port							
4	Destination Port							
5	UDP Length							
6	UDP Length							
7	Checksum							
8	Checksum							

I.3 IP Header

[13]

Octets	Bits 8	7	6	5	4	3	2	1
1	Version				IP Header Length			
2	Type of Service							
3	Total Length							
4	Total Length							
5	Identification							
6	Identification							
7		DF	MF	Fragment Offset				
8	Fragment Offset							
9	Time to Live							
10	Protocol							
11	Header Checksum							
12	Header Checksum							
13	Source address							
14	Source address							
15	Source address							
16	Source address							
17	Destination address							
18	Destination address							
19	Destination address							
20	Destination address							
Up to 60	Word Options							

Appendix II

M/M/1Queing Theory

[25] "Suppose there are m independent Poisson data streams, each supplying packets at rate λ/m, arriving at a common "concentrator" where they are mixed into a single data stream of combined rate λ. Suppose that packet lengths are independent and exponentially distributed so that packet transmission times are exponentially distributed with mean transmission time, say, 1/μ. The concentrator then forms an M/M/1 system, which statistically multiplexes the independent data streams into a single data stream. The average number of packets at the concentrator is hence given by $N_{SM} = \lambda/(\mu-\lambda)$ while the average delay per packet is $T_{SM} = 1/(\mu-\lambda)$.

If transmission capacity is instead spread evenly over the m data streams so that each data stream effectively sees a dedicated line with service rate μ/m (equivalently, mean packet transmission time m/μ), then the system can be modeled as m M/M/1 systems operating in parallel, each with a packet arrival rate of ¸λ/m and a service rate of μ/m. This is the situation in time-division multiplexing (TDM) and in frequency-division multiplexing (FDM). The average number of packets in any of these individual M/M/1 systems is hence $N_{TDM/FDM} = \frac{\lambda}{m}/(\frac{\mu}{m} - \frac{\lambda}{m}) = N_{SM}$, while the average delay per packet is $T_{TDM/FDM} = 1/(\frac{\mu}{m} - \frac{\lambda}{m}) = mT_{SM}$.

Thus, statistical multiplexing reduces the average packet delay over frequency- or time-division multiplexing m-fold! The reason is simply that in time- or frequency-division multiplexing a dedicated portion of the resources is allocated *a priori* to each data stream. In low duty cycle traffic streams this leads inevitably to "dead time" when resources allocated to a given data stream are idle; if, simultaneously, there is congestion in a different data stream, overall delay would be reduced by making the idle resources available to help ameliorate congestion at a currently backed up data stream. Indeed, statistical multiplexing deploys resources where needed on demand resulting in an efficient usage of resources and an over all lower delay per packet. Thus statistical multiplexing is the method of choice in data networks serving many low duty cycle traffic streams.

Mean packet delay, however, is not the only factor affecting the quality of service in a given application. In telephony, for instance, traffic streams are voice conversations in which packets arrive regularly with no packet arriving while one is still in transmission. This regular character of packet arrivals allows a dedicated distribution of resources across the various streams via TDM or FDM in such a way that there is no waiting in queue. Thus, the delay in transmission of each packet is fixed and there is no variability in this quantity. This is critical in telephony as speech recognition is particularly sensitive to variance in delay. Statistical multiplexing on the other hand, while lowering the overall mean packet delay, introduces variability in the delay on a packet-to-packet basis thus resulting in a positive variance of delay. This results in a potentially severe quality degradation, which far outweighs the savings in mean packet delay. Thus it is that time- and frequency-division multiplexing systems are still widely deployed in telephony."

Appendix III

Matlab Code

```
% function Z = PCR(MBRmax,minSDU,maxSDU,mu,m)
% For given maximum MBR (between PDPs) calculates
% Peak cell rate (PCR)
% MBRmax : max maximum bit rate (bits)
% minSDU: minimum SDU (length =1 to 65535 octets)
% maxSDU: max SDU
% mu: transmission rate
% m: number of PDP contexts

function Z = PCR(MBRmax,minSDU,maxSDU,mu,m)
PCRmax = ((MBRmax/8)/ minSDU)* ceil((maxSDU + 8)/48);
lamda= m * PCRmax;
mu=m*mu;
Z = m*lamda/(mu-lamda);
```

```
% function Z = EBW
% Calculates the Effective Bandwidth, given:
% CLP: Cell loss probability
% B: Buffer size
% p: Probability of receiveing 1 cell

function Z = EBW
CLP=input('Cell loss probability(10^x): ');
B=input('Buffer size (cells): ');
p=input('Probability of receiving 1 cell (or unity BW): ');
CLP=power(10,CLP);
q=1-p;
thita = (log10(1/CLP)/log10(exp(1)))/B;
EBW = (log10(p*exp(thita)+q)/log10(exp(1)))/thita;
disp(sprintf('EBW = %.5f',EBW));
disp(sprintf('Number of ATM sources = %d',fix(1/EBW)));
```

```
% function Z = SCR(BW,B,CLP,NofS)
% Calculates the Sustainable Cell Rate (SCR) for given:
% BW: link capacity (bits/s)
% B: Buffer size (cells)
% CLP: Cell loss probability in form
% 10^-x  %...ex=9
```

```
% NofS: Number of sources needed with specific CLP

function Z = SCR(BW,B,CLP,NofS)
CLP=power(10,-CLP);
thita = (log10(1/CLP)/log10(exp(1)))/B;
EBW = 1/NofS;
p = (exp(thita*EBW)-1)/(exp(thita)-1);
Z = p * BW;
```

```
% function engset(p,h,r,s,N)
% Calculates the engset distribution given:
% p : arrival rate : SCR
% h : 1/mu = 1/departure rate : Tpr
% N : no of servers/buffer places : N=B
% r : where to block (end of buffer) : r=N
% s : no of sources/incoming cells : s=MBS

function z=engset(p,h,r,s,N)
a=0;
for i=0:1:N
   a = a + power((p*h),i) * comb(s,i);
end
b = power((p*h),r) * comb(s,r);
z=b/a;
```

```
% function z=comb(n,k)
% Calculates combinations of n taken by k

function z=comb(n,k)
if((n-k)>n/2)%less overflows for big numbers
   a=0;
      for i=0:1:k-1
           if(i==0) a=n;
           else
                if(a==0) error('overflow during factorial
process ==> Too low probability')
                end
           a = a * (100-i);
           end
      end
   z=a/factorial(k);
else
   z = factorial(n)/(factorial(k) * factorial(n-k));
end
```

```
% function Z = MBS(B,SCR,MBS)
% Plot a graph so that the user is able to choose best MBS
% B: Buffer size (cells)
% SCR: Sustainable Cell Rate (SCR) ... see SCR.m
% Maximum Burst Size (MBS)

function Z = MBS(B,SCR)
MBSp=0;
```

```
xp=0;
for MBS=B/10:1:(2*B)/10
   x=engset(SCR/10,1/100,B/10,MBS,B/10);
   plot([MBS MBSp],[x xp])
   MBSp=MBS;
   xp=x;
   hold on
end
grid on
hold off
grid on
xlabel('MBS');
ylabel('CLP due to buffer overflow');
disp('MBS can be estimated from the plot');
```

```
% function Z = inter
% Helps to configure Interactive Class PVC
% Calculates: theoritical number of sources, min BW for
% each source, Effective BW, PCR and SCR given:
% BW: the link capacity (bits/s)
% MBRmax: the maximum MBR between all PDPs (bits/s)
% minSDU: minimum SDU
% maxSDU: maximum SDU
% mu: the transmission (departure) rate
% m: number of PDP contexts
% B: Buffer Size (cells)
% CLP: Cell loss probability
% NofS: the required number of sources

function Z = inter
BW=input('Give the link capacity (bits/s) ex.=5000000000 : ');
MBRmax=input('Give the maximum MBR between all PDPs (bits/s)
ex.=2000000 : ');
disp(sprintf('Minimum number of sources should support (worst
case) = %d',fix(BW/MBRmax)));
disp(sprintf('So Minimum BW for each source = %.4f
bits/s',BW/fix(BW/MBRmax)));
minSDU=input('Give the minimum SDU (1 to 65535 octets) : ');
maxSDU=input('Give the maximum SDU (1 to 65535 octets) : ');
mu=input('Give the transmission rate ex.=2000000000 : ');
m=input('Give number of PDP contexts (typ. x*1000) : ');
B=input('Give the Buffer Size (cells) ex.=600 : ');
CLP=input('Give the Cell loss probability (10^-x) ex.=9 : ');
NofS=input('Give the required number of sources ex.=100 : ');
EBW = 1/NofS;
disp(sprintf('Effective Bandwidth = %.5f',EBW));
SCRv=SCR(BW,B,CLP,NofS);
PCRv = PCR(MBRmax,minSDU,maxSDU,mu,m);
disp(sprintf('So PVC should be configured with : PCR = %.2f
cells/s , SCR = %.2f cells/s',PCRv,SCRv));
MBS(B,SCRv);
```