End-2-End QoS Provisioning in UMTS networks

Haibo Wang    Devendra Prasad

January 12, 2005
Contents

1 Overview and Problem Definition
  1.1 Introduction .............................................. 4
  1.2 Quality of service architecture ....................... 5
    1.2.1 Applications point of view ....................... 5
    1.2.2 IP Point of view .................................. 6
    1.2.3 UMTS Point of view ............................... 7
  1.3 Problem Definition ..................................... 9
  1.4 Project Time Plan ...................................... 10

2 Background .................................................. 12
  2.1 UMTS QoS .................................................. 12
    2.1.1 UMTS Network Architecture ....................... 12
    2.1.2 UMTS protocol stack ............................... 15
    2.1.3 UMTS QoS Management ............................. 18
    2.1.4 Main QoS mechanism in UMTS ....................... 20
  2.2 IP QoS ..................................................... 21
    2.2.1 DiffServ Policy .................................... 21
    2.2.2 IntServ/RSVP Policy ............................... 21
    2.2.3 Comparison between IntServ and DiffServ ............ 21
  2.3 IP QoS Versus UMTS QoS ................................. 21

3 Delimitation .................................................. 24
  3.1 State-of-the-Art ......................................... 24
    3.1.1 ARROWS project .................................... 24
    3.1.2 SAMU project ....................................... 25
    3.1.3 SEACORN project ................................... 25
    3.1.4 Other related works ................................ 26
    3.1.5 Conclusion ......................................... 26
  3.2 Problem Analysis ........................................ 26
    3.2.1 From E2E system level ............................. 27
    3.2.2 Play Scenarios ...................................... 29
3.3 Network Simulator 2 (NS-2) review ........................................... 32
  3.3.1 IP Features ......................................................... 32
  3.3.2 Differentiated services ............................................ 32
  3.3.3 UMTS extension .................................................. 34
  3.3.4 Limitation of NS-2 and its UMTS extension ................. 35
3.4 Conclusion ............................................................... 36

Bibliography ................................................................. 37
Chapter 1

Overview and Problem Definition

1.1 Introduction

There is a growing interest in the end-to-end Quality of Service (E2E QoS) provision in both the standards for 3G mobile radio systems (in 3GPP/3GPP2) and for IP infrastructure (in IETF), which reflects the trend that as the mobile Internet-like services get more and more popular, the market requires the real-time and quality-assured services provision by the mobile communication world which is going to be a convergence of all-IP backbone [3GPP R5] and Universal Mobile Telecommunications System (UMTS) radio access network.

UMTS has been designed to support integrated services no matter voice, data, video, etc. Among them broadband multi-media services is regarded as the key to UMTS success. Some most important services require UMTS to provide bandwidth on demand, mixed traffic types, efficient network transport and/or guaranteed QoS features, and those services may be roughly divided into realtime and non-realtime with quite different demand to the UMTS network.

In the convergence of IP and UMTS environment, two possible key research topics concern QoS and mobility management (Mobile IP). This project focus on the QoS issue especially for realtime services in IP/UMTS combined domain from a End to End point of view. Related concepts are reviewed, well-known approaches in both Internet and UMTS system to provide QoS are described, and a new E2E QoS provision mechanism is proposed.
1.2 Quality of service architecture

Generally, Quality of Service is a set of requirements to be met so that a service or application can be delivered to the end-user in a quantitative and qualitative service level. The QoS level can be quantified by packet loss probability, guaranteed bandwidth, end-to-end (E2E) delay and jitter, reflect how the traffic flow through a network. Put it in another way, QoS can be seen as the degree of satisfaction of an end-user for an delivered service.

The concept above leads to the basic idea of QoS, to distinct traffic into different types, corresponding to their different features and different demands to the networks, and to be delivered to the customers on different charges.

Consequently, certain QoS mechanisms must be implemented to provide/ensure the E2E QoS features of applications matching their traffic type. We distinguish two main categories of mechanisms: QoS provision mechanisms and QoS control mechanisms [2]:

- QoS provision mechanisms include parameters mapping, admission and resource reservations schemes.
- QoS control mechanisms consist of traffic shaping, scheduling, policing and control mechanisms.

QoS can be offered by network operators with Service Level Agreements (SLAs). A SLA is a contract to specify the transit of services through network domains (more accurate definition and reference needed).

1.2.1 Applications point of view

In packet-switched networks, while discussing QoS, the applications are viewed in two different way with respect to network. One way in IP world (Proposed by IETF) are termed as real-time and non-real-time (elastic). Those applications deliver time-sensitive information are real-time, where the data blocks must be displayed consecutively at predetermined time intervals, thus require specific delay, jitter and error parameters. On the other hand, the applications don’t include time-sensitive information are non-real-time, which may be much more tolerable to delay and jitter but more sensitive to error parameters. Another way was proposed by 3GPP in UMTS network, where applications are classified into four types based on the generated traffic: Conversational, Streaming, Interactive and background.
Before stepping into more detailed discussion on IP QoS viewpoint and UMTS QoS viewpoint, it is necessary to first describe some well-known QoS metrics for presenting E2E service/application requirements, those are delay, jitter, loss rate and throughput.

**Delay**  
It is the elapsed time for a packet to traverse the network from the source to the destination. At the network layer, the end-to-end packet latency is the sum of processing delay, transmission delay, queuing delay and propagation delay.

**Jitter**  
It is defined as the variation in delay encountered by similar packets following the same route through the network. The jitter requirement only affects real-time streaming applications because this QoS requirement arises from the continuous traffic characteristics of this class of applications. Jitter is generally included as a performance parameter since it is very important at the transport layer in packetized data systems, due to the inherent variability in arrival times of individual packets. Services intolerant of delay variation will usually try to reduce the delay variation by means of buffering. However, late data arrivals make data useless, resulting in receiver buffer underflow, and early arrival can lead to receiver buffer overflow.

**Loss Rate**  
Loss rate refers to the percentage of data loss among all the delivered data in a given transmission time interval, which can be evaluated in frame level or packet level. Loss rate requirements apply to all classes of applications. In general, real-time applications might tolerate a limited amount of data lost, depending on the error resiliency of the decoder, and the type of application. On the other hand, non-real-time applications typically have much more strict requirement on data loss.

**Throughput**  
It is defined as the rate at which packets are transmitted in a network. It can be expressed as a maximum rate or an average rate.

### 1.2.2 IP Point of view

The IP QoS point of view can be seen in a scenario, where the transport of IP packet through a set of network and one of them is UMTS network, as represented in Fig.1.1. The known approaches for dealing with IP QoS are Integrated Services (IntServ) and Differentiated Services (DiffServ).

DiffServ is an attempt to design a simple architectural framework for QoS that can provide a variety of scalable end-to-end services across multiple separately
administered domains without requiring complex behaviors in forwarding equipment [2]. This approach is interesting in sense that it overcomes many issues such as scalability, inter-operation and administration without using high signaling. Actually, unlike IntServ, DiffServ minimizes signaling by using Per-Hop-Behaviors (PHBs).

Integrated Service provides application requirements to the network, which have to maintain QoS mechanisms to insure the promised QoS. The Resource reservation protocol (usually Resource Reservation Protocol (RSVP)), transports the QoS requirements along the path from the sender to the receiver in order to make resource reservation. It sets states in router per flow and is supposed to gain a better E2E performance in case the networks on along the path is in congestion since there are enough transmission resource reserved for all the existing data flows.

1.2.3 UMTS Point of view

3GPP defines UMTS QoS classes as:

1. Conversational class;
2. streaming class;
3. interactive class;
4. background class.

The main distinguishing factor between these QoS classes is how delay sensitive the traffic is: Conversational class is meant for traffic which is very delay sensitive while Background class is the most delay insensitive traffic class [3]. Table 1.1 shows the application associated with those types of traffic and their main characteristics:

![Figure 1.1: Reference Architecture](image-url)
To provide these QoS requirements, 3GPP proposed the concept named Bearer Service.

"To realise a certain network QoS a Bearer Service with clearly defined characteristics and functionality is to be set up from the source to the destination of a service.

A bearer service includes all aspects to enable the provision of a contracted QoS. These aspects are among others the control signalling, user plane transport and QoS management functionality. A UMTS bearer service layered architecture is depicted in figure 1.2, each bearer service on a specific layer offers it’s individual services using services provided by the layers below." [3] The Bearer Service (BS) concept is important in that it allowed us to divide the E2E QoS problem into

<table>
<thead>
<tr>
<th>Traffic Class</th>
<th>Examples of Application</th>
<th>Fundamental characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversational(RT)</td>
<td>Voice/Video Telephony</td>
<td>Low E2E delay, low Jitter, two-way</td>
</tr>
<tr>
<td>Streaming(RT)</td>
<td>streaming video</td>
<td>Low jitter, one-way</td>
</tr>
<tr>
<td>Interactive(BE)</td>
<td>Web Browsing</td>
<td>Two-way, Low loss rate/error</td>
</tr>
<tr>
<td>Background(BE)</td>
<td>Email, Background download</td>
<td>One-way, Low loss rate/error</td>
</tr>
</tbody>
</table>

Table 1.1: UMTS QoS classes (RT: real-time, BE: best-effort)
a set of sub-problems concerning the QoS provision inside a specific Bearer Service and the QoS profile mapping among them (between 2 BSs in the same plane or between an upper layer and a lower layer).

1.3 Problem Definition

When considering an End-2-End QoS provisioning via a UMTS network, there could be a long chain (e.g., from a MS user to a real-time application server located in Internet), as depicted in figure 1.3. In terms of BS concept, the E2E QoS or application level QoS is provided by two main domain: UMTS bearer service domain and IP bearer service domain.

As mentioned in section 1.2, the first problem in E2E QoS study is the mapping mechanism, which includes:

1. Application/service type mapping to UMTS BS QoS attributes;
2. UMTS QoS BS attributes mapping to IP BS QoS attributes, in terms of DiffServ or InterServ;
3. UMTS QoS BS attributes mapping to RAB Service attributes and to CN Service attributes; the further mapping to even lower BS attributes beyond the scope of our study.

Within the application level mapping, there could be two cases in the UE side, since IP level BS manager is an optional function in UE [4]. If

1. the IP BS manager exist in UE
   The IP bearer services manager communicates with the UMTS bearer services manager through a translation function, and the interaction between UMTS bearer services and IP bearer services shall only occur at a translation function in the UE and GGSN.

2. there is no IP BS manager in UE
   The application QoS profile has to be translated directly into UMTS QoS profile and there is no IP level QoS between UE and GGSN.

No matter in which case, IP QoS interaction still occur between GGSN and the application server in external IP network. And once the mapping is fixed, (i.e, UMTS BS attributes are mapped into DiffServ), admission and resource reservations schemes will be decided according to DiffServ in IP domain.
UMTS BSs are provided between the UE and the GGSN, where UMTS QoS negotiation is provided by the PDP context, and the UMTS QoS profile is used to specify the QoS parameters in a PDP context. Hence PDP context is important and needed to be studied also.

Secondly, QoS control mechanisms including traffic shaping, scheduling, policing and control mechanisms, have to be decided according to the mapping mechanism both on UMTS domain and IP domain.

Figure 1.3: Layered Service architecture of an E2E QoS chain in UMTS

1.4 Project Time Plan

1. Overview: Sept. 8th - Oct. 1st
   (a) Overview Paper study
   (b) More relevant study in order to rectify the existing problem.
   (c) Problem definition(preliminary)

2. Play-Scenario decision: Sept. 25th - Oct. 29th
   (a) Study on UMTS PDP context and QoS.
(b) Study on IP QoS.
(c) Problem delimitation
(d) Simulation study and installation.

3. Simulation: Oct 22th - Dec 23th
   (a) Make the first end-2-end simulation model(Best Effort).
   (b) Propose a new strategy for End-2-End QoS provision

4. 9-Semester Presentation: Jan 25th, 2005

5. Simulation Analysis and Measurement: Jan 26th - Feb 28th
   (a) Simulation results Analysis with possible analytical model.
   (b) Design measurement setup for proposed E2E QoS strategy.
   (c) Measurement with emulator

6. Further output study: March 1st - April 1st Analysis
   (a) Measurement results analysis together with simulation result.
   (b) upgrade simulator based on output analysis if necessary.

   (a) Build or polish analytic model based on simulation and measurement output.
   (b) Finish all report/Thesis documents.

8. Thesis Defence Preparation: May - June
   (a) Thesis hand in
   (b) Preparation for defence.
Chapter 2

Background

In this chapter, QoS mechanism in both UMTS domain and IP domain is studied.

2.1 UMTS QoS

Before step into UMTS QoS issues, it is necessary to review the source of UMTS concept and its basic network structures.

3G Systems are intended to provide a global mobility with wide range of services including telephony, paging, messaging, Internet and broadband data. International Telecommunication Union (ITU) started the process of defining the standard for third generation systems, referred to as International Mobile Telecommunications 2000 (IMT-2000). In Europe European Telecommunications Standards Institute (ETSI) was responsible of Universal Mobile Telecommunications System (UMTS) standardization process. In 1998 Third Generation Partnership Project (3GPP) was formed to continue the technical specification work.

2.1.1 UMTS Network Architecture

A network architecture is defined by 1) functional groups and 2) reference points. Functional groups are defined by a set of functions, which may be performed by one or more physical piece of equipment; reference points are conceptual points separating functional groups, representing a physical interface between pieces of equipment.

A UMTS network consist of three interacting domains; Core Network (CN), UMTS Terrestrial Radio Access Network (UTRAN) and User Equipment (UE) [5]. The main function of the core network is to provide switching, routing and
transit for user traffic. 

Core network also contains the databases and network management functions. The UTRAN provides the air interface access method for User Equipment. Base Station is referred as Node-B and control equipment for Node-B’s is called Radio Network Controller (RNC).

Figure 2.1: UMTS Basic Network Architecture

- **Core Network (CN)**

  The basic Core Network architecture for UMTS is based on GSM network with GPRS. All equipment has to be modified for UMTS operation and services. The Core Network is divided in circuit switched and packet switched domains. Some of the circuit switched elements are Mobile services Switching Centre (MSC), Visitor location register (VLR) and Gateway MSC. Packet switched elements are Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). Some network elements, like EIR, HLR, VLR and AUC are shared by both domains.

  According to the latest 3GPP R5 specification, IP will be the main protocol for transporting user packets inside the PS core network. However, user packets are not directly routed via IP routing protocol but based on GPRS Tunneling Protocol (GTP).

- **UTRAN**
Wide band CDMA (WCDMA) technology was selected for UTRAN air interface, which supports a transmission rate theoretically up to 2Mbit/s (realistic up to about 300kb/s). UMTS WCDMA is a Direct Sequence CDMA system where user data is multiplied with quasi-random bits derived from WCDMA Spreading codes. In UMTS, in addition to channelisation, Codes are used for synchronisation and scrambling. WCDMA has two basic modes of operation: Frequency Division Duplex (FDD) and Time Division Duplex (TDD). A specific structure of UTRAN is depicted in figure 2.3.

As in figure 2.3, the UTRAN physical entities are made up of Node-B and RNC and their functions include:

- **Node-B**: Air interface Transmission/Reception, Modulation/Demodulation, CDMA Physical Channel coding, Error Handing, Closed loop power control, etc.
- **RNC**: Radio Resource Control, Admission Control, Channel Allocation, Handover Control, Broadcast Signalling, Open Loop Power Control, etc.

- **User Equipment** User equipment is the equipment used by the user to access UMTS services.

### 2.1.2 UMTS protocol stack

The protocols over Uu and Iu interfaces are divided into two structures [6]:

1. **User plane protocols**: These are the protocols implementing the actual radio access bearer service, which carrying user data through the UMTS network.

2. **Control plane protocols**: These are the protocols for controlling the radio access bearers and the connection between the UE and the network from different aspects (including requesting the service, controlling different transmission resources, handover, etc.).

In the following, we briefly explain these two protocol structures.

#### UMTS user plane

Figure 2.5 depicts the user plane of the UMTS network. The UE and UTRAN include L1, MAC, RLC and PDCP layers. The definition of these layers is as below:

![UMTS Protocol Stack: User Plane](image-url)
• **Packet Data Convergence Protocol (PDCP):** This transmission functionality maps higher-level characteristics onto the characteristics of the underlying radio-interface protocols. PDCP provides protocol transparency for higher-layer protocols. PDCP supports e.g., IPv4, PPP and IPv6. Introduction of new higher-layer protocols shall be possible without any changes to the radio interface protocols. PDCP provides protocol control information compression. User data compression is not supported in UMTS, because the data compression efficiency depends on the type of user data, and because many applications compress data before transmission. It is difficult to check the type of data in the PDCP layer, and compressing all user data requires too much processing.

• **Radio Link Control (RLC):** The RLC protocol provides logical link control over the radio interface. There may be several simultaneous RLC links per UE. Each link is identified by a Bearer Id.

• **Medium Access Control (MAC) [6]:** The MAC protocol controls the access signalling (request and grant) procedures for the radio channel.

• **Physical layer (PHY or L1) [7]:** The physical layer offers information transfer services to MAC and higher layers.

UTRAN, SGSN and GGSN include the following layers:

• **GPRS Tunnelling Protocol for the user plane (GTP-U) [8]:** This protocol tunnels user data between UTRAN and the SGSN, and between the SGSN and the GGSN in the core network. GTP encapsulates all PDP Protocol Data Units (PDU).

• **User Datagram Protocol/Internet Protocol (UDP/IP) [9][10]:** These are the backbone network protocols used for routing user data and control signalling.

• **ATM Adaptation Layer 5 (AAL5) [11]:** This adaptation layer protocol provides support for variablebit rate connection-oriented or connectionless data services.

• **Asynchronous Transfer Mode (ATM) [12]:** ATM is a cell-based switching and multiplexing technology designed to be a general purpose connection-oriented transfer mode for a wide range of services. The information to be transmitted is divided into fixed-size cells (53 octets), multiplexed, and transmitted.
UMTS control plane

The UMTS control plane is illustrated in Figure 3. The control plane consists of the following layers:

- **UMTS Mobility Management and Session Management (GMM/SM) [3]**: GMM supports mobility management functionality such as attach, detach, security, and routing area update and SM supports PDP context activation and PDP context deactivation.

- **Short Message Service (SMS) [13]** supports the mobile-originated and mobile-terminated short message service.

- **Radio Resource Control (RRC) [14]**: This a signalling protocol for the control and configuration of the radio interface.

- **The RLC protocol offers logical link control over the radio interface for the transmission of higher layer-signalling messages and SMS.**

- **The MAC protocol controls the access signalling (request and grant) procedures for the radio channel.**

- **The physical layer offers information transfer services to MAC and higher layers.**

- **Radio Access Network Application Protocol (RANAP) [15]**: This protocol encapsulates and carries higher-layer signalling, handles signalling between the SGSN and UTRAN, and manages the GTP connections on the Iu interface. The layers below RANAP are defined in [16].
2.1.3 UMTS QoS Management

The concept of Bearer Service (BS) has been introduced in Chapter 1 as in figure 2.6. QoS in each layer is expressed by a series of QoS attributes specific to the layer. These attributes are used in order to request the appropriate service from the lower layers.

It is the UMTS Bearer Service that provides the UMTS QoS to the UMTS users. The UMTS bearer service consists of two parts, the Radio Access Bearer (RAB) Service and the Core Network (CN) Bearer Service. The radio access bearer service provides confidential transport of signalling and user data between MT and its corresponding SGSN with the QoS adequate to the negotiated UMTS bearer service or with the default QoS for signalling. This service is based on the characteristics of the radio interface and is maintained for a moving MT. The radio access bearer service is realised by a Radio Bearer Service and an Iu Bearer Service. The role of the radio bearer service is to cover all the aspects of the radio interface transport. This bearer service uses the UTRA FDD/TDD mechanism. The Iu bearer service provides the transport between UTRAN and CN. The Iu bearer services for packet traffic shall provide different bearer services for a variety of QoS.

The core network bearer service of the UMTS core network connects the SGSN with the GGSN. The role of this service is to efficiently control and utilise the core network in order to provide the contracted UMTS bearer service. The UMTS
packet core network shall support different bearer services for a variety of QoS mechanisms. The core network bearer service uses a generic Layer1/Layer2 functionality in order to fulfil the QoS requirements of the core network bearer service. These layers are not specific to UMTS but may reuse an existing standard and is selected according to operator’s choice [17].

QoS management functions for UMTS bearer service in the control plane

![UMTS QoS Management functions in the control plane](image)

**Figure 2.7: UMTS QoS Management functions in the control plane**

QoS management functions for the UMTS bearer service in the user plane

![UMTS QoS Management functions in the user plane](image)

**Figure 2.8: UMTS QoS Management functions in the user plane**
2.1.4 Main QoS mechanism in UMTS

UMTS provides QoS to its user through the PDP context mechanism. Once attached to the UMTS network, a UE must activate a PDP context in order to send or receive data. PDP context procedure consists of a series of signalling allowing a UMTS user to establish a virtual connection with the GGSN and to express its QoS requirements for this connection. Moreover a user has also the possibility to define rules that can identify the packets to which this connection and these QoS parameters may apply. Two major parts of the PDP signalling procedure are QoS profile and Traffic Flow Template (TFT).

UMTS QoS mechanism tries to overcome the limitations existing in the General Packet Radio Service (GPRS) QoS framework. The GPRS also uses the PDP context signalling procedure in order to establish a PDP context. However for a given PDP address, only one QoS profile can be used. This means that all the application flows sharing the same PDP context are forced to experience the same QoS profile defined for the PDP context and no per-flow prioritisation is possible [19] [20] [21]. In UMTS, the PDP context mechanism is improved to support multiple application flows and to provide a more flexible QoS negotiation and set-up.

QoS profile

The QoS requirements of a user are grouped into a QoS profile. This QoS profile is then used to establish an appropriate UMTS bearer service. A QoS profile is associated with each PDP context. During the QoS profile negotiation, it is possible for the UE to request a value for each of the QoS attributes of a QoS profile, including its subscribed default values. The network may negotiate each attribute to a level that is in accordance with the available UMTS resources. The network always attempts to provide adequate resources to support the negotiated QoS profiles.

Traffic Flow Template

A TFT is a series of rules allowing the UMTS network to detect a flow requiring a certain QoS. It consists of one up to eight packet filters, each identified by a unique packet filter identifier. A packet filter also has an evaluation precedence index that is unique within all TFTs associated with the PDP contexts that share the same PDP address. This evaluation precedence index is in the range of 255 (lowest evaluation precedence) down to 0 (highest evaluation precedence). The UE manages packet filter identifiers and their evaluation precedence indexes, and
creates the packet filter contents. A PDP context can never have more than one TFT associated with it.

A valid packet filter contains a unique identifier within a given TFT, an evaluation precedence index that is unique within all TFTs for one PDP address, and at least one of the following attributes:

*To be extended according to the problem delimitation*

PDP context

### 2.2 IP QoS

#### 2.2.1 DiffServ Policy

Per Hop Behaviors (PHB) and Differentiated Services Code Point (DSCP)

Traffic classification and conditioning

Per Hop Behaviors

1. EF
2. AF
3. BE

#### 2.2.2 IntServ/RSVP Policy

#### 2.2.3 Comparison between IntServ and DiffServ

### 2.3 IP QoS Versus UMTS QoS

Let us first consider the differentiated services architecture in the Internet. In order for a customer to receive differentiated services from its Internet Service Provider (ISP), it must have a service level agreement with its ISP. An SLA basically specifies the service classes supported and the amount of traffic allowed in each class. An SLA can be static or dynamic. Static SLAs are normally negotiated on a regular basis such as monthly or yearly. For a dynamic SLA, a signalling protocol such as RSVP is required to request services on demand from an ISP.
Customer domain should decide how its hosts share the services specified by the SLA. There are basically two choices:

- Each host makes its own decision and marks an appropriate DS code according to its desired service.
- A resource controller called the Bandwidth Broker (BB) makes decision for all hosts.

If the host has not the ability to mark the packets by itself, then it has to ask the BB by the use of a signalling protocol such as RSVP. Once the BB receives the host’s demand for a service, it has must also use a protocol such as RSVP or Lightweight Directory Access Protocol (LDAP) [30] to set the classification, marking and shaping rules at the leaf router directly connected to the sender. As a result, the leaf router knows how to mark the packets coming from the host.

For static SLAs, boundary routers of the ISP can be manually configured with the classification, policing and shaping rules so that they know how to handle the incoming traffic. If the SLA between a customer and its ISP is dynamic, the BB in the customer domain must again use a signalling protocol such as RSVP to request resources on demand from the ISP.

In case of UMTS, assuming that the UE can not mark its traffic, then the GGSN plays the role of BB for the UE. It marks the traffic coming from the UE with a code point according to the QoS requirements of the traffic. In Internet, a host that can not mark its traffic by itself must use a signalling procedure such as RSVP in order to express its QoS requirements to the BB. While this option is feasible by the communication between the IP BS manager entities in the UE and the GGSN, we think that it may be redundant. In fact, in UMTS, we have the PDP context mechanism which enables the UE to communicate its QoS requirements with the GGSN. The QoS profile and the TFT elements of the PDP signalling protocol can perfectly enable the GGSN to set its classification and marking rules. If the host can mark its traffic by its own, then there is no need for BB functionality in the GGSN.

Now if the SLA between the UMTS and its ISP is static, then the traffic can be immediately transmitted to the ISP. If the SLA is dynamic, then the GGSN must communicate with its ISP by RSVP or another signalling protocol in order to request resources and to configure the boundary routers with the corresponding classification, policing and shaping rules.
In case of integrated services, the reservations along the path between the sender and the receiver are always triggered by RSVP messages.
Chapter 3

Delimitation

3.1 State-of-the-Art

3.1.1 ARROWS project

Advanced Radio Resource Management for Wireless Services
IST project
Duration: Jan 2001 – Dec 2002
This project aims at providing advanced Radio Resource Management (RRM) and Quality of Service (QoS) management solutions, for both UTRA-TDD and UTRA-FDD modes, to support integrated voice and data services. It includes packet access, asymmetrical traffic and high bit rate (2 Mbit/s) services for multimedia IP based applications. An intelligent management of the radio resources is essential in order to fulfil the required QoS. A multimedia test-bed was developed to validate the proposed concepts, supports four basic applications: 1) audio-video telephony, based on VIC (video) and RAT (audio) conferencing tools, 2) video streaming, based on the applications developed by the MPEG4IP group, 3) Web, based on Mozilla client and Apache server and 4) mail, based on Mozilla client and sendmail/pop3d/imapd servers for STMP/POP3/IMAP services, respectively.

Although ARROWS concentrates on the QoS aspects of UTRA, a global QoS framework is proposed for two reasons. First, from the applications point of view, QoS is an end-to-end issue and, second, a mapping between UMTS and end-to-end QoS parameters is required. And the QoS-related contributions include packet scheduling, radio network congestion control, Admission Control-MAC Algorithms and so on.

http://www.arrows-ist.upc.es
3.1.2 SAMU project
Advanced Services by Mastering UMTS (SAMU)
Duration: July 1999 – September 2002
SAMU project is a collaborative project approved by the French RNRT (Telecommunication Research National Network) and aims at improving the services offered to the users in an UMTS mobile system. For that purpose, the following topics were under investigation:

1. Defining smart resource allocation strategies for a CDMA mobile system. (RRM)
2. Defining an efficient way to map IP over UMTS, to provide the user with an end-to-end QoS. (QoS)
3. Complementary to the 2 previous ones. It aims at using agent software to make the system autonomous. (Agent)
4. Demonstrating the feasibility of IP services over UMTS for users in their cars.

The innovative work in QoS from SAMU includes UMTS/IP QoS mapping architecture and UMTS link layer optimization for TCP.

http://samu.crm-paris.com/Edocuments.htm

SAMU project brings together 3 industrial research centres: Motorola Labs (which is the leader of the project), SFR and Renault and 2 academic research laboratories: Eurecom Institute and PriSM laboratory.

3.1.3 SEACORN project
Simulation of Enhanced UMTS Access and Core Networks (SEACORN)
Duration: March, 2002 – March, 2004
SEACORN aimed at Enhanced UMTS networks research areas including:

1. Identification and characterization of the new services;
2. radio link techniques required to increase the bit rate and the capacity gain;
3. new network protocols for the access network and TCP/IP for the core network,
4. dynamic simulation build-up.
In area 3, the contribution of SEACORN is

- Development, and implementation of resource management algorithms enabling QoS provisioning and differentiation while optimising resource efficiency, in a multi-service UMTS dynamic network simulator considering both up- and downlink communications;

- Development and implementation of resource management mechanisms for IP-based core networks in a dynamic network simulator.

And a system level simulator was developed based on Network Simulator 2 (NS-2).

http://seacorn.ptinovacao.pt/

3.1.4 Other related works

"A FRAMEWORK FOR DYNAMIC SLA-BASED QOS CONTROL FOR UMTS". etc.

3.1.5 Conclusion

Most of the work still stay on proposing all types of E2E frameworks and QoS attributes but few solid specific simulation has been done to evaluate the performance of the proposed frameworks.

3.2 Problem Analysis

As described by TS 23.207, the main challenges that the UMTS QoS architecture has to overcome are (more, we synthesized, merge bullets):

- Translation parameters and mechanisms - Service differentiation based on a set of traffic classes needs a simple and reliable translation mechanism between the different domains involved.

- UMTS QoS Management - The network should be monitored and managed to assure the implementation of the user agreements. Negotiation and modification of the QoS available from the network should be possible.
3.2.1 From E2E system level

1.2.1.1 QoS Handling in the UMTS Core Network

There are three basic timescales for traffic management in UMTS networks:

- Capacity planning and network dimensioning. Not considering.

- UMTS Call Admission Control
  Call admission control is performed in every multiplexing point involved (MSC, SGSN, GGSN, border gateway, media gateway, etc). CAC answers two questions. Can the node accept this new call? Will the node be able to meet the QoS requirements of the new call and of the established calls?

UMTS CAC is implemented using the simple and flexible concept of equivalent bandwidth. The purpose is to estimate the network resources required to provide the requested QoS and to determine whether these resources are available. If they are available, the necessary resources are reserved. However, if they are not available, a mechanism is initiated to downgrade the QoS. For example, the user could be offered a smaller guaranteed bit rate or lower traffic handling priority. equivalent bandwidth for the duration of the call, which is typically minutes for voice calls or hours for Internet Protocol (IP) sessions.

But without PDP context process implemented in simulator, as a result of that we can play in a very simple scenario which will be discussed later in the chapter.

- QoS Differentiation between User Equipment
  Several mechanisms are available to isolate traffic coming from every UMTS user equipment: queuing differentiation and weighted fair buffer allocation within the SGSN and GGSN; policing and traffic flow templates within the GGSN.

1. Queuing differentiation
  Scheduling techniques, such as WFQ or Weighted Round Robin (WRR), are used to resolve any contentious access to a resource. These techniques guarantee a minimum allocated bandwidth for every PDP context. Fair queuing mechanisms also provide an implicit policing function since bandwidth is allocated in proportion to the weights. It is known as implicit policing as it only takes any action when a network element is congested, always dividing the bandwidth up in a fair manner. Implicit policing is mandatory to support a QoS commitment:
user behavior has to be constrained to that specified in the contract if resources are short so that a misbehaving or malicious user does not adversely affect the QoS delivered to other users.

2. **Weighted fair buffer allocation**

Weighted Fair Buffer Allocation (WFBA), also known as Weighted Random Early Discard (WRED), is used to discard packets in an intelligent and proactive way. It is particularly useful in allowing the Transmission Control Protocol (TCP) to react quickly and efficiently to congestion. The basic principles are as follows. The buffer is divided fairly among the active user equipment. The fraction of the buffer allocated to a user equipment depends on its UMTS bearer service QoS attributes, as defined in Table 2. Buffer overbooking per PDP context is allowed, depending on the buffer occupancy. A new packet is accepted only if the user equipment is not already using too much space in the buffer.

3. **GGSN policing**

The GGSN is the edge node in the UMTS network, so it receives the incoming traffic and checks that the downlink user data traffic conforms with the QoS attributes of the corresponding UMTS bearer service. The GGSN includes a traffic conditioner, which guarantees traffic conformance. Another role of the GGSN is to filter the incoming traffic according to the Traffic Flow Templates (TFT). These are used to distinguish between different user packets going to the same PDP address which have different QoS requirements identified by different PDP contexts.

- **Use of IP Differentiated Services at Layer 3** QoS differentiation is necessary to ensure that UMTS traffic is suitably handled within the network. On the UMTS backbone, Internet Engineering Task Force (IETF) Differentiated Services (DiffServ or DS) are used at layer 3, the IP transport layer.

### 1.2.1.2 Between UMTS and External IP network

- **IP and UMTS QoS Attribute Mapping on GGSN**
  UMTS QoS classes mapping to DiffServ classes.

- **Service Level Agreements**
  Taking the UMTS PLMN network is a DiffServ domaine, each packet data network can also be a DiffServ domain. The PHB policy is uniform within a DiffServ domain. To guarantee the end-to-end QoS across external networks, Service Level Agreements (SLA) are required between the UMTS
network and each external network. SLAs can be signed between the UMTS and every Packet Data Network (PDN) operator, such as a corporate intranet/extranet or Internet Service Provider (ISP), and between UMTS operators and inter-PLMN backbone providers or other PLMN UMTS/GPRS operators.

In a UMTS R5 architecture, which are the common parameters can connect the external Internet and UMTS network (IP backbone)?

In Differentiated services architecture we have no hard guarantees in the sense that there are no fixed delay or loss rate values related to each QoS class. The SLA only includes the traffic profile (maximum burst size and the arrival rate of the traffic) as absolute values. In UMTS, on the other hand, we have several parameters for each QoS class. Some of these parameters have an equivalent in the DiffServ world. These are the guaranteed bit rate and maximum SDU size of the UMTS QoS parameters. They can be easily mapped to the arrival rate and the maximum burst size in the traffic profile of the SLA. Other attributes are specific to the UMTS network and they do not have any equivalent in the fixed network. These attributes are residual bit error ratio and the delivery of the erroneous SDUs.

### 3.2.2 Play Scenarios

The IP BS managers in the UE and GGSN provide the set of capabilities for the IP bearer level as shown in Table 3.1. According to 3GPP TS 23.207, Provision of the IP BS Manager is optional in the UE, and required in the GGSN. These options give us different scenarios for an end-to-end QoS architecture.

**Scenario 1**

This scenario assumes that the GGSN supports DiffServ edge functions. The UE does not provide an IP BS Manager. As a result, the end-to-end IP QoS bearer service towards the remote terminal is controlled from the GGSN. The remote network as well as the backbone IP network are DiffServ enabled.

<table>
<thead>
<tr>
<th>Capability</th>
<th>UE</th>
<th>GGSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiffServ Edge Function</td>
<td>Optional</td>
<td>Required</td>
</tr>
<tr>
<td>RSVP/IntServ</td>
<td>Optional</td>
<td>Optional</td>
</tr>
</tbody>
</table>

Table 3.1: IP BS Manager Capability in the UE and GGSN)
The end-to-end QoS is provided by the PDP context over the UMTS access network, DiffServ through the backbone IP network, and DiffServ in the remote access network in the scenario shown in the figure below.

**Scenario 2**
This scenario assumes that the UE and GGSN support DiffServ edge functions, and that the backbone IP network as well as the remote network are DiffServ enabled.

The end-to-end QoS is provided by the PDP context over the UMTS access network, DiffServ through the backbone IP network, and DiffServ in the remote access network in the scenario shown in Figure below.

**Scenario 3**
This scenario assumes that the UE and GGSN support DiffServ edge functions, and that the backbone IP network is DiffServ enabled. In addition, the UE supports RSVP signalling which interworks within the UE to control the DiffServ and to enable end-to-end QoS using IP layer signalling towards the remote end. There is no IP layer signalling between the IP BS managers in the UE and the GGSN and the GGSN is not RSVP aware. In this scenario, the terminal supports
signalling via the RSVP protocol to control the QoS at the local and remote accesses, and DiffServ to control the IP QoS through the backbone IP network. The RSVP signalling protocol may be used for different services.

The QoS for the wireless access is provided by the PDP context. The UE may control the wireless QoS through signalling for the PDP context. The characteristics for the PDP context may be derived from the RSVP signalling information as shown in Figure below and Figure below. Alternatively, subscription data accessed by the SGSN may override the QoS requested via signalling from the UE.

**Scenario 4**

This scenario assumes that the UE and GGSN support RSVP signalling which may control the QoS directly, or interwork with DiffServ. The backbone IP network is RSVP and/or DiffServ enabled. The UE relies on the GGSN in order to provide the end-to-end QoS in the sense that it is the GGSN that maps the RSVP QoS parameters into the DS parameters to be used in the backbone network.

This scenario assumes that the UE and GGSN support RSVP signalling which may control the QoS directly, or interwork with DiffServ. The backbone IP network is RSVP and/or DiffServ enabled. The UE relies on the GGSN in order to provide the end-to-end QoS in the sense that it is the GGSN that maps the RSVP QoS parameters into the DS parameters to be used in the backbone network.

The QoS for the wireless access is provided by the PDP context. The UE may control the wireless QoS through signalling for the PDP context. The characteristics for the PDP context may be derived from the RSVP signalling information as shown in Figure ?? and Figure ?? . Alternatively, subscription data accessed by the SGSN may override the QoS requested via signalling from the UE.

**Conclusion**

We decide to choose scenario 1 for further investigation. The reason includes the limitation on the PDP Context activation and the consideration of downlink data packets instead of uplink. Since in the simulator we don't have the PDP context as well as the signalling implemented from the UE to the GGSN. Hence it put a limitation in our assumptions. Therefore, the DiffServ will be used in order to map the IP QoS profile attributes with the UMTS CN QoS profile attributes. Another consideration will be the predefined negotiation algorithm.
3.3 Network Simulator 2 (NS-2) review

3.3.1 IP Features

3.3.2 Differentiated services

DiffServ uses codepoints (DSCP) attached to a packet’s IP-header to distinguish traffic with different PHBs (Per Hop Behaviour). A PHB defines a forwarding treatment of a single packet in a router. They do not offer any guarantees regarding gained bandwidth, packet delay or jitter. It is only a means of defining better treatment for some class of traffic than to another.

Within a router, traffic is divided in different queues based on DSCPs. Two things have to be considered:

1. How to manage packets inside a single queue (or buffer)?

2. How to control scheduling between multiple queues, i.e. how link access is shared between queues?

Buffer management

Buffer size - usually defined in number of packets - is the first thing that has to be considered. Since buffers have a tendency of filling up something has to be done to access packets arriving in a queue. Two most used buffer types are drop-tail and RED.

1. Drop-tail
   This is the most simple type of buffer. Dropping occurs only to packets arriving in a full buffer. This is a simple and light implementation but usually causes problems with multiple constant packet flows and so called global synchronization of TCP connections.

2. RED
   RED (Random Early Detection) drops packets at an increasing probability as the buffer begins filling up. Minimum and maximum thresholds of buffer occupation are defined and between them the probability of dropping a packet is increased linearly. Below the minimum threshold packets are not dropped at all and above the maximum threshold all arriving packets are dropped.
An average number of packets in the buffer is maintained. This value, rather than the instantaneous value of the buffer length, is compared to the dropping thresholds when a packet arrives.

Queue management

When dealing with multiple buffers a scheduler must be included for distributing link access rights. Two different schedulers are used in this simulation: PQ (Priority Queueing) and WRR (Weighted Round Robin).

1. PQ
   Priority queueing sets different priorities to different buffers. Link access is always granted to the buffer with the highest priority if it has packets to send. This type of scheduling has the advantage of offering very small packet delay and jitter. On the other hand high priority traffic can easily exhaust traffic with smaller priorities unless bandwidth limits are configured properly.

2. WRR
   Weighted round robin shares link access circularly, one buffer at a time. Each buffer has a weight which assigns a relative share of link access time compared to other buffers. With proper configuration WRR can offer similar type of preferential treatment to high priority traffic than PQ with the advantage of not completely exhausting classes with lower forwarding treatments.

Core Routers

Core routers are routers, which are only connected to other routers - not hosts (Servers or PC or UE). They only have the functionality of performing PHB on incoming packets. For this, core routers maintain multiple physical queues and each have virtual queues or precedences. An incoming packet is placed in a virtual queue within a physical queue based on the DSCP-value in the packet’s IP-header.

A physical queue is implemented as a RED-queue in NS2. Virtual queues within the same physical queue all use the same buffer to temporarily store packets. However, all virtual queues can be configured with their own RED-parameters (minimum and maximum thresholds and dropping probability).

The following parameters must be defined for a core router:

- Number of physical and virtual queues
- All DSCPs needed
- Mapping of each DSCP to a physical and a virtual queue
- RED-parameters (min and max thresholds and dropping probability) for each virtual queue

**Edge routers**

Edge routers, or access routers, are used between the hosts and the core network. They perform the same functionality as core routers (PHB), and in addition manage the following functions:

- policing
- metering
- packet marking
- translation

When configuring edge routers the same parameters must be defined as for the core routers. In addition, the criteria of marking the packet with a DSCP must be included. The decision is two-fold:

1. Distinguish the packet of belonging to a certain traffic aggregate. The decision is based on the following IP-header values: source address, destination address and traffic type (or any combination of these).

2. Measure transmission rates of the given traffic aggregate and the conformance of it to pre-defined values

### 3.3.3 UMTS extension

It is the intention of this document to provide a description of the installation and configuration requirements for the Enhanced UMTS Radio Access Network extensions to ns-2 (version 2.26), developed within the SEACORN project [1] for Ericsson Telecommunicatie B.V.

http://www.ti-wmc.nl/eurane

The Enhanced UMTS extensions for ns-2 comprise of an additional three nodes, namely the Radio Network Controller (RNC), Basestation (BS) and the User Equipment (UE), whose functionality allow for the support of the following transport channels:
• FACH
• RACH
• DCH
• HS-DSCH

The main functionality additions to ns-2 come in the form of the RLC Acknowledged Mode (AM), Unacknowledged Mode (UM) MAC-d/-c/sh support for RACH/FACH and DCH, and MAC-hs support for HS-DSCH, i.e. HSDPA.

RLC
At the RNC, two implementations of Acknowledged Mode (AM) are available for RLC. The type of RLC (AM or AM-HS) to use is dependent on the transport channel. Two implementations of Unacknowledged Mode, UM and UM-HS, are also available, and are basically a functional sub-set of the AM and AM-HS, respectively.

MAC
As mentioned previously, there are two possible MAC architectures to choose from. The basic MAC (Mac/Umts) used for the DCH and common channels (RACH and FACH), and the more complicated MAC-hs (Mac/Hsdpa) used for the HS-DSCH.

RNC
It is a UMTS specific node as well as BS and UE nodes.

SGSN and GGSN
They are nothing but normal NS-2 nodes, or in another word, a router.

3.3.4 Limitation of NS-2 and its UMTS extension
No PDP context has been implemented.
3.4 Conclusion

Based on ongoing research as well as the functionality of the simulator, we choose to go for a Scenario1 with the implementation of new algorithm. That algorithm takes care of Scheduling, queueing, traffic shaping and allocation of available channels.

The incoming IP packets will be mapped with the UMTS CN QoS profile attributes if it meets the threshold parameter value based on GGSN. If the threshold parameter is above the limit then the packet will be routed to the UE with the predefined route and that provides a simple way of PDP Context activation.
Bibliography


[6] 3G TS 25.401: "UTRAN Overall Description".

[7] Hans Peter Schwefel; Lecture notes: Wireless Networks II Aalborg University, 2004