

---

## **New fair QoS-based charging solution for mobile multimedia streams**

---

Zsolt Butyka\*, Tamás Jursonovics  
and Sándor Imre

Department of Telecommunications,  
Budapest University of Technology and Economics,  
Magyar Tudósok körútja 2., 1117 Budapest, Hungary  
Fax: +36 14633263                      E-mail: buzso@mcl.hu  
E-mail: jursonovicst@gmail.com      E-mail: imre@hit.bme.hu  
\*Corresponding author

**Abstract:** In this paper, we present a new streaming proxy based charging architecture for QoS differentiated charging in 3G. We give a short overview of charging concepts in next generation mobile networks and we describe the most important network attributes on which the quality depends. Our goal is to point out the importance of bursty packet loss which determines the quality of the multimedia streaming in contrast with the average loss rate. We propose several methods for QoS differentiated charging which are in compliance with the recommendations of 3GPP. We describe three models which can predict the QoS of a stream. The algorithms are evaluated on a real GPRS streaming flow.

**Keywords:** 3G mobile network; streaming; multimedia; QoS; real-time; charging.

**Reference** to this paper should be made as follows: Butyka, Z., Jursonovics, T. and Imre, S. (2008) 'New fair QoS-based charging solution for mobile multimedia streams', *Int. J. Virtual Technology and Multimedia*, Vol. 1, No. 1, pp.3–22.

**Biographical notes:** Zsolt Butyka received his MSc Degree in Telecommunication Engineering from the Budapest University of Technology and Economics in 1998. He is currently a third year PhD student at the Department of Telecommunications at BUTE and he also works at PKI Telecommunications Development Institute (Magyar Telekom Plc.). His current research focuses on charging and billing in heterogeneous mobile networks, streaming multimedia over wireless networks and NGN networks.

Tamás Jursonovics is a PhD student from Budapest University of Technology and Economics. He is currently a Research Engineer in the Service Development of T-Mobile Hungary, Hungary. His current research interest include QoS-based billing solutions for 3G networks, Streaming media transport in next generation mobile networks.

Sándor Imre received his MSc Degree in Telecommunication Engineering from the Budapest University of Technology and Economics in 1993. In 1999, he received his PhD Degree in Electrical Engineering. He is an Associate Professor at the Budapest University of Technology and Economics, and the R&D Director of the Mobile Innovation Center, Hungary. His research areas include IP mobility, reliability; wireless access technologies; software defined radio; quantum computing.

## 1 Introduction

The evolution of mobile telecommunication technology has accelerated in the last five years. Users can access the internet faster with the newly developed, IP based network technologies (i.e., EDGE and UMTS). The existing GPRS system allows only a limited range of services, like e-mail or WAP based internet surfing, but the 3G offers users the option to connect to a large number of Value Added Service Providers (VASPs). The VASPs can deploy their services (like real time multimedia streaming, value added, location based services or audio/video conferencing) through a 3G network provider (Olariu et al., 2005); therefore, the 3G systems require new methods for charging.

Many billing solutions are developed for 3G. Baug (2002) looks into the key challenges of GPRS billing. He seeks to show how solving these challenges will help operators succeed in the not so distance future of UMTS. Koutsopoulou et al. (2001) introduce a novel billing scheme for UMTS networks which is capable of handling efficiently charging, accounting and billing for value added services that are provided either by the network operator or third trusted parties. Tewari and Omahony (2003) present new AAA methods for Real Time Payments of Mobile IP.

But the majority of solutions are not able to consider the special charging requirements of streaming. The visual quality of a streaming media depends on various network conditions (for example delay, delay jitter, loss distribution, etc ...); therefore, the 3G needs new methods for the QoS differentiated charging. Moreover, if the customers connect to a third party streaming provider through a network operator then the service which is used by the customer is on the side of the third party, therefore only the 3rd party can charge this service usage, which must be shared with the NO, which must trust this information. Summing up, the 3G requires new methods for charging and billing.

This paper is organised as follows: Section 1 reviews the charging requirements in 3G networks. Sections 2–4 present a new architecture fitting our business vision point of view. Sections 2 and 3 give a brief description of charging concepts in a mobile environment. In Section 4 the business models are presented. Section 5 explains the difference between online and offline charging. Section 6 introduces the streaming protocols. Section 7 describes the streaming attributes from mobile operator point of view. Section 8.1 reviews the charging requirements in 3G networks. Sections 8.2–8.4 present a new architecture fitting to our business vision point of view. Section 8.5 explains the mechanisms for QoS differentiated charging; Sections 8.6 and 8.7 describe and evaluate these new models, which can predict the QoS of a streaming flow. Finally, Section 9 concludes the paper.

## 2 Charging difference between internet and mobile networks

The charging, billing and accounting schemes used in the internet have been quite simple until now. Users have been mainly billed at a flat rate, based on their subscription and/or the duration of their connection for accessing the internet. In mobile telecommunication networks, users have been mainly billed on their subscription and the call duration, as well as a number of other parameters (e.g., type of communication, location and destination, etc ...). In the near future (Barry et al., 2005) these schemes are expected to

receive modifications as a consequence of recent technological advances combined with the emerging dominance of the internet protocol.

Although IP is the glue that will tie together the internet and mobile networks (Barachi et al., 2005), the business model and the related charging frameworks considered by the two communities are diverse in view of the placement and management of the charging functionality. Thus, the internet community considers a business model that requires direct agreement between the user and independent provider, while the telecommunications world insists on the operator centric model. These communities have been working separately for many years and, therefore, there are many differences even in the respective terminology (Koutsopoulou et al., 2004).

The internet research community has focused more on the protocols used for accounting data exchange, while the mobile world paid more attention to the specification of the network entities that should generate, process and collect charging information. Since both worlds are converging and new dynamic links are now possible with the application/service providers, it is crucial that a minimum compatibility is achieved between these systems. Thus, efforts should be made in order to align the accounting protocols and even new advanced functionalities, such as content and location based charging.

### **3 Billing requirements in mobile networks**

To bypass a complicated charging architecture, a multi level charging architectural approach structured in several levels is proposed in 3GPP. The management and processing of the relevant information should be made separately for each level. Furthermore, different charging models should be applicable on each charging level. In such a charging architecture, subscribers require the provision of 'one stop billing'. Users would like to receive a single itemised bill for using voice and data services offered by network operators and independent application or service providers. This requirement implies that the network operator would be responsible for collecting charging data from all players and billing the users. Another requirement is that the charging information should be in a form easily understood by the average user. Also, the users should be constantly aware of the charges to be levied for each chargeable event.

On the other side, the mobile operators require a flexible charging architecture that accommodates various pricing models (e.g., time based, volume based, QoS based, etc ...) in order to fulfil not only the traditional business models but also innovative models. In addition, the selection of a specific pricing model could possibly be based on the user and the service profile parameters. Another important requirement, imposed by the mobile operator is the support of both pre-paid and post-paid charging mechanisms.

From the independent application or service providers' perspective, the evolving requirement is that each authorised player should be able to dynamically apply the desired pricing policy for its services usage. The independent providers should be able to add or modify tariffs for the service and content portion. This dynamic modification should be made in a standardised way in order to update whichever entity will handle the charging, accounting and billing functionality.

#### 4 General business models in next generation networks

First of all, let us describe how a general business model is built up. The first player in this model is the user, which in business parlance, means the customer. Then let us define three entities: the network operator, the content aggregator and the content provider.

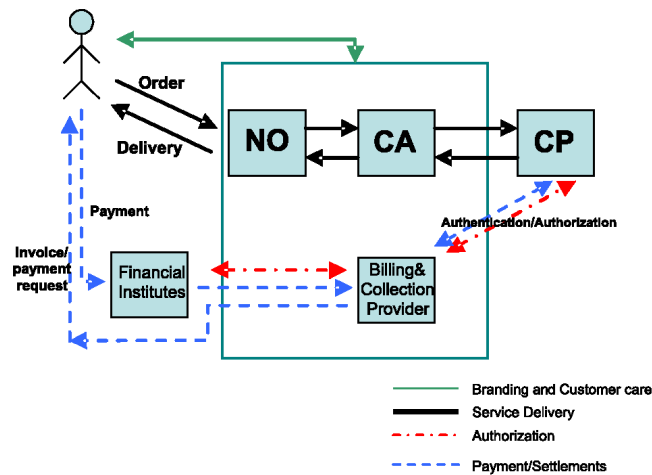
The network operator will probably be the 3G license holder and its key function is to provide access and transport services to the customers. The content aggregator's function is similar to the well known mobile portal and its key function is to package and offer services from content providers. The third entity is the content provider whose main function is to provide services that add value to access and transport services (UMTS Forum Report, 2002). Before a service is offered, there should be an agreement between the customer and these third parties. With an agreement, the customer may start using services offered by a service provider.

The next step is the authentication and authorisation. Authentication is needed to prove that the customer is a rightful customer, and authorisation is necessary to verify that the customer is allowed to use the particular service. The main reason for using this process is to limit credit risk exposure for the involved parties.

##### 4.1 Network operator centric business model

In the network operator centric business model (Figure 1) the customer has a direct relationship with the network operator. The network operator sets the prices of the services and handles the payments.

**Figure 1** Network operator centric business model



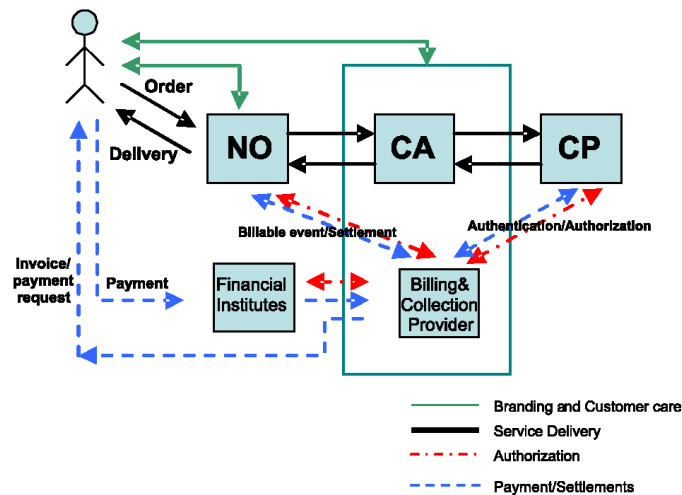
The content is normally acquired wholesale from content providers or home made by the network operator itself. The network operator, therefore, manages its own content aggregator role. Services are, in many cases, offered as bundled packages as part of subscriptions. Besides 'traditional' event charges, new charging techniques may develop that handle the same basic process but in real time. Network operators will use this model to increase ARPU and retain their customers. External parties involved may be content providers and financial institutions.

#### 4.2 Content aggregator centric business model

The content aggregator ('m-portal') model is not limited to providing physical access to services through a mobile portal, but rather includes a range of value added services. Added value that might be offered on top of access and transport services could include authentication, security, simplicity, and payment aggregation.

In the content aggregator centric business model (Figure 2) the customer has an agreement with the content aggregator, but may still have a relationship with the network operator. The content aggregator determines the price of its content, while the customer may pay access charges to the network operator separately – this can be arranged in different ways depending on agreements made between the parties. It is also indicated the possibility that the content aggregator settles access and transport charges with the network operator.

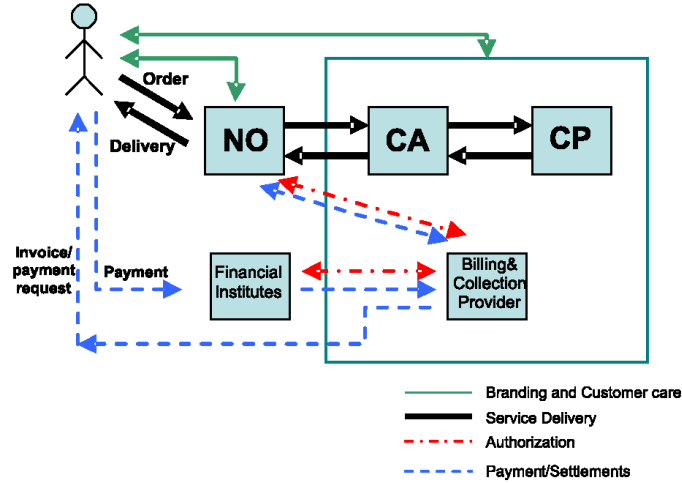
**Figure 2** Content aggregator centric business model



#### 4.3 Content provider centric business model

At first approach, the content provider centric business model (Figure 3) is similar to the content aggregator centric model. The main difference is that the content provider has a considerable content portfolio and wants to align itself with a network operator and also take up the content aggregator role. In the previous case the content aggregator will most probably sign up agreements with a number of content providers. The customer may have relationships with many content providers in this model and the network operator will only gain access and airtime charges. Content providers may settle access and transport charges with the network operator to offer a complete price for delivering a service.

The main disadvantage of this model is that content providers have to arrange billing and payment on their own. Also, the customer has to deal with each content provider individually – for example, by paying with a credit card. In this model, the likelihood of one time buyer-seller transactions is higher compared with the other models. The diversity of service offerings is likely to be very high, while the number of transactions per buyer-seller combination is, probably, rather low.

**Figure 3** Content provider centric business model

## 5 Offline and online charging terminology

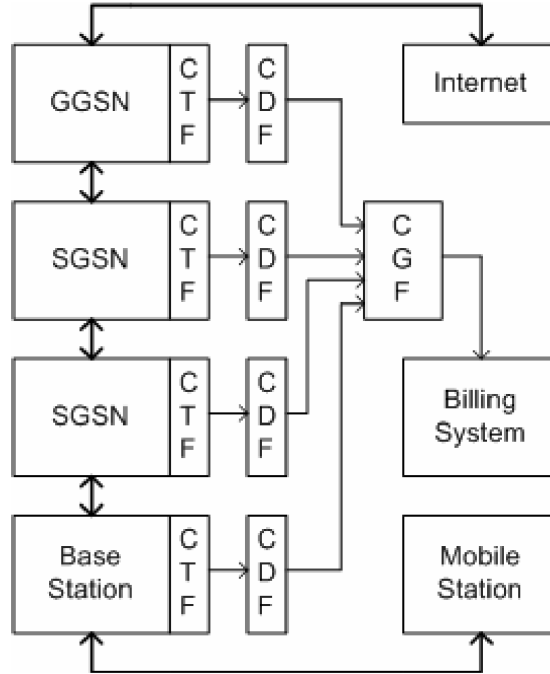
The standards define two charging methods: online and offline charging. The difference between these types is that the online charging is needed in case of prepaid users, while for the post paid users the offline charging method is sufficient because they do not need to be charged in real-time (3GPP TS 32.240 V6.3.0, 2005).

### 5.1 Offline charging

In this case the gateway (GGSN) and the inner-nodes (SGSN) are sending the charging information to the Billing System (BS). This charging information must be in a standardised format, called Charging Data Record (CDR). The Charging Trigger Function (CTF) of the network elements generates charging events based on the observation of network resource usage. The Charging Data Function (CDF) receives charging events from the CTF, and then uses the information contained in the charging events to build CDRs. These records are sent to the Charging Gateway Function (CGF), which acts like a storage buffer, cleans, and pre-processes the CDRs. Finally, the CGF sends these processed CDRs to the BS (Figure 4). Because these charging records carry all information about the services required, the functionality of the CDRs extends beyond charging. With CDRs it is possible to analyse service-utilisation, and gain statistical information about the services and content. By archiving the CDRs, the user-complaints can also be easily settled.

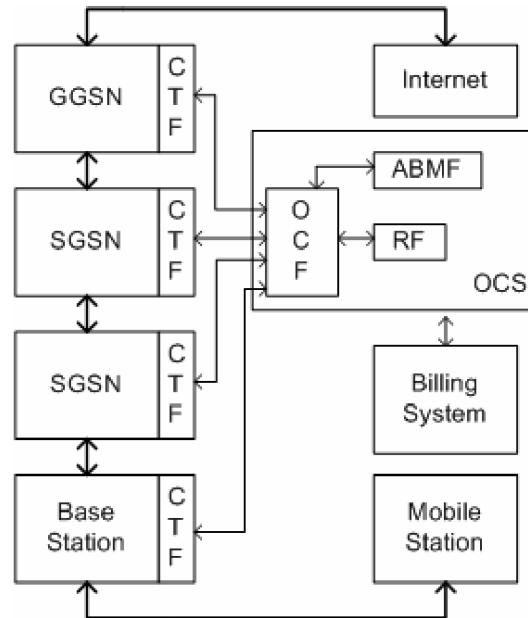
In the offline charging mechanism, the user's account will be decreased after a few minutes from the service usage. The network overhead will be smaller because of transmitting less number of charging records over the entire network (Ary and Imre, 2005).

Figure 4 Offline charging method



### 5.2 Online charging

In online mode (Figure 5) a new entity, the Online Charging System (OCS) is responsible for the proper charging. The main task of this entity is to realise real-time charging by continuously delegating certain amounts of credit to the serving network elements. The user's account will be deducted by so-called unit reservation. If the service terminates before all credits are consumed, the network elements are re-transferring the remaining credits to the OCS. To assure continuous service delivery, if the users do not terminate the service, a new amount of granted credit should be sent to the serving network element before the previous one runs out. This unit-granting function is represented by the Online Charging Function (OCF) inside the OCS. The CTF generates charging events for the OCF as well, but this communication is bidirectional, as the OCF has to grant credits for the service. The OCS also includes the Account Balance Management Function (ABMF) and the Rating Function (RF). The Account Balance Management Function is the location of the subscribers account balance within the OCS, and the Rating Function is used to determinate the value of the network resource usage.

**Figure 5** Online charging method

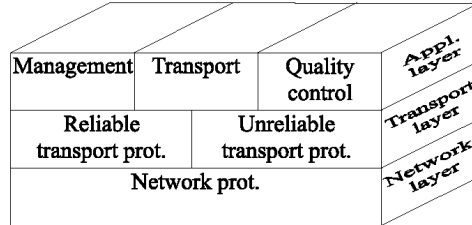
## 6 Streaming protocols

Streaming is a network technology to real time broadcast live or on demand multimedia contents (like audio and/or video) from one centralised streaming server to many clients on a wide network. This content delivery is a one way communication from the server to the clients (on the other hand, the video conferencing is a bidirectional communication among the users). This delivery is a unicast or multicast transport based on IP protocol. Certainly, the streaming uses bidirectional communication for the management between the clients and the server. By means of streaming, users can listen to the radio, watch TV or a movie trailer on their mobile.

At present there are many streaming solutions such as: RealNetworks Helix, Windows Media Streaming, Quick Time Streaming and NextReaming. These use different protocols and formats which impeded the proliferation of the technology because there was no universally accepted streaming method. The 3GPP realised this problem and it tried to guide the manufacturers with the aim of creating a wide and well uniform streaming technology for 3G. So, in this paper, we follow the way of 3GPP; therefore, our measurements and examinations are based on a 3GPP compliant technology.

Streaming has three, separate planes based on the same Network layer (Figure 6) (in most cases, on IP): management, transport and quality control plane (Kim and Park, 2005).



**Figure 6** Streaming protocols

The Management plane allows the users to control the streaming, such as: connection opening and closing, getting information about the media, preparing the server for the broadcast and starting/pausing/stopping the delivery. The transport plane makes the real time data delivery; it reserves the greatest part of available bandwidth. It must adjust oneself to the variable network parameters, so it cooperates with the quality control plane to avoid jitter and traffic congestion.

The quality control plane measures the transport actual quality, and generates QoS reports for the both side.

## 7 Streaming attributes

In this paper we are going to describe new quality based charging methods but first we describe some important IP network attributes which determine and indirectly describe the quality of the streaming. Video streaming over the internet is difficult because the internet only offers best effort service (Montelius and Larsson, 2001). That is, it provides no guarantees on bandwidth, delay jitter, or loss rate. Specifically, these characteristics are unknown and dynamic (Apostolopoulos, 2002).

### 7.1 Bandwidth

The available bandwidth between two points in the internet is generally unknown and time varying. If the sender transmits faster than the available bandwidth then congestion occurs, packets are lost at the IP routers and that causes deterioration in video quality. If the sender transmits slower than the available bandwidth then the receiver produces sub-optimal video quality. The streaming can use two methods to avoid this effect. If the variation of the actual bandwidth is small, then the client can lower estimate the bandwidth and it can send the stream on Constant Bit Rate (CBR). Otherwise, if the above mentioned effect is high then the client must estimate the average bandwidth and it must send the stream with Variable Bit Rate (VBR), which can adjust itself to the actual bandwidth.

### 7.2 Delay (and jitter)

The end-to-end delay that a packet experiences may fluctuate from packet to packet. This variation in end-to-end delay is referred to as the delay jitter. Delay jitter is a problem because the receiver must receive/decode/display frames at the original sequence, and any late frames resulting from the delay jitter can produce problems in the

reconstructed video, e.g., jerks in the video. This problem is typically addressed by including a play out buffer at the client. While the play out buffer can compensate for the delay jitter, it also introduces additional delay.

### 7.3 *Packet losses*

The third fundamental problem is losses. A number of different types of losses may occur, depending on the particular network under consideration. For example, wired packet networks such as the internet are afflicted by packet loss, where an entire packet is erased (lost). On the other hand, wireless channels are typically afflicted by bit errors or burst errors. Losses can have a very destructive effect on the reconstructed video quality. To combat the effect of losses, a video streaming system is designed with error control. Approaches for error control can be roughly grouped into four classes: Forward Error Correction (FEC), retransmissions, error concealment, and error-resilient video coding.

## **8 The proposed method for charging of streaming**

In the previous chapters we have given a short overview of the main backgrounds of existing charging mechanisms and streaming attributes. Starting now, we will focus only on our new charging approach.

In today's mobile networks there are no universal, scalable, uniform methods for charging of streaming. Most operators build their own charging procedures on the existing network usage measurement, so the streaming bills cannot be separated from other IP services bills. In this case, the streaming plays only a data volume generators role and the operators cannot make any extra profit out of these services.

The GSM IP payment system (like NDS) is developed for WAP services, so it can handle the multimedia download, but the mobile equipment memory capacity limits the maximal size of the downloadable content. This solution is unsuitable for real-time delivery, and applies only to event based charging methods.

Streaming manufacturers carried out their own charging solutions (like Real SDS, see at "Helix Service Delivery Suite"), but these are very platform dependent and, in most cases, not centralised.

Solving the above problems, in this chapter, we present our proposal for CAB model in today's mobile networks, with the aim of creating a platform-independent, centralised, well scalable architecture. So, in the following subsections, we offer a brief survey of charging requirements of streaming in 3G, and we describe our new charging architecture and mechanisms.

### 8.1 *Charging requirements*

3GPP recommends (3GPP TS 32.200 V5.9.0, 2005) three methods for charging of streaming in its technical specification:

- charging by duration of session
- one off set up charge
- charging by volume of data, optionally QoS differentiated.

The great advantage of duration based charging is that, it can be easy implemented, but it cannot adjust itself to the temporary network parameters, like available bandwidth or packet loss. Two users with different access speeds pay for the same multimedia stream equally, but they have different visual experiences. Whereas the volume based charging can adapt itself to the diverse network bandwidth, the users pay for what they get. This is enough for HTTP or FTP services, which do not need delay sensitive protocols, but the streaming or real time transfer requires permanent delay and reliable transport. If these conditions are not fulfilled, the quality of streaming decreases, so QoS differentiation must be used to charge these services.

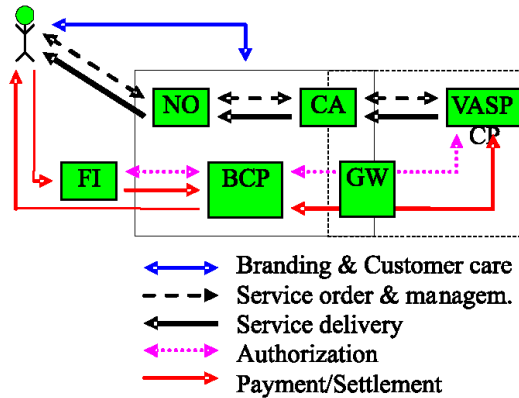
The 3GPP describes in its report (see above) additional requirements, which we accentuate out of three:

- for network operators and 3rd parties to charge each other for the use of their resources
- to charge for different level of QoS applied for and/or allocated during a session for each type of medium or service used
- to charge using pre pay, post pay charging techniques.

### 8.2 Proposed business model

In the immediate future, the implementation of the 3G will based on the 2G system; thus, in the early age of 3G, the main concept of billing scheme will correspond with the present GSM concept, which is a very mobile operator centred. So, we have chosen the NO model (Figure 7).

Figure 7 The proposed NO model with network entities



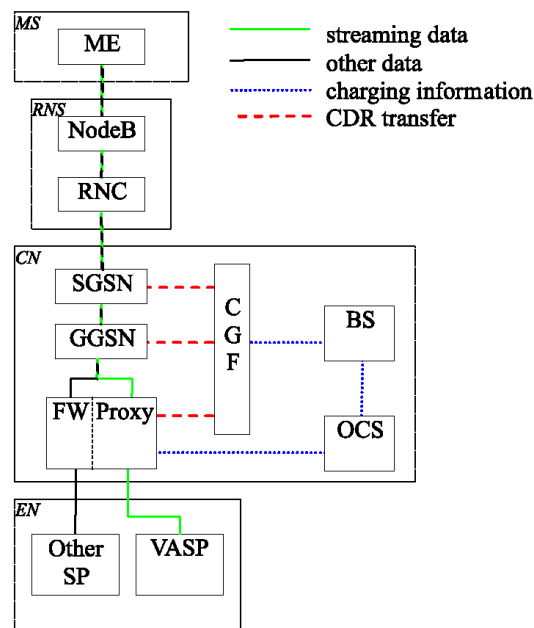
External parties involved may be VASP and Financial Institutions (FI). The NO has its own billing, collections and customer care infrastructure (BCP). The customers have a direct relationship with the NO, which sets the prices of the services, and handles the payments. The VASPs can receive the charging and authentication information of their services from the NO through a charging gateway (GW). This model has the following key characteristics:

- price of services defined by the network operator
- billing and payment arranged by the NO
- NO revenues gained as airtime, data volume, message, subscription, advertising and transaction/event.

### 8.3 The new architecture

Over a classic 3G data session, the customers' communication begins with a PDP context establishment through the RNS, SGSN and GGSN. The GSNs are able to separate the users' flows from one another with their PDP context, but they cannot look inside the IP payload (clearly speaking, they do not decode the encapsulated protocol (UDP, TCP, RTP, etc ...)). So, the GSNs are unsuitable for a charging method, which is based on properties of these protocols. Therefore we introduce a new network element, which makes the existing network structure complete (Figure 8).

**Figure 8** Proxy based charging architecture



The customers cannot connect directly to the VASPs after the PDP context activation, because we put a streaming proxy and a firewall near the GGSN, on its internet side. The proxy monopolises the management of streaming and it enables the users to connect through itself only to streaming providers (VASPs). Other services are accessible through the firewall, so its key function is only the access restriction (the proxy can be bypassed without it).

Offline charging information (for example: usage of radio interface, usage duration, destination and source) are collected in the Packet Switched domain network for each MS by the GSNs, and they are transferred in a CDR to the BS via the CGF. The generation of CDR depends on the charging characteristic profile; the GSNs are able to define separate

trigger conditions, like data volume limit, time or maximum number of charging conditions. In this architecture the proxy can measure the streaming flow due to its central position, so it has its own CDR generation function for offline charging.

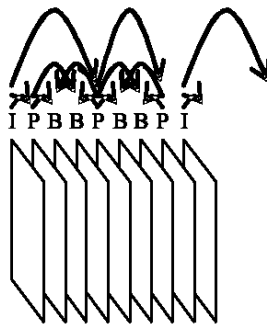
The pre-paid account is limited. If the amount of money runs out while a customer uses a service, the access must be denied to the customer. The offline charging mechanism cannot provide this condition therefore the proxy uses account reservation for online charging. Firstly, it reserves a piece of the customers' account. If the customers use up this amount of money for a value added service, then the proxy tries to reserve another piece of the account. If the allocation of money is unavailable, the proxy restricts the access of service.

#### 8.4 Related mechanisms

The radio interface between the RNC and MS is a really unreliable medium. It is possible to lose some data during a call session; so, to protect from inaccurate charging, the 3G SGSN will always instruct the RNC at RAB set up to count the unsent downlink data toward the MS (Waluyo et al., 2005). At RAB release the RNC reports the unsent data to the SGSN, which sends this value to the BS in the RNC Unsent Downlink Volume file of a CDR.

In case of a streaming flow, a single frame loss (mostly, one frame represents one network packet) is probably imperceptible to the user, because the streaming server uses MPEG 4 (or similar) standard to encode the media stream. The available bandwidth in a mobile network is very limited and these algorithms archive high compression rate. Typically, the encoded frames have three different types (Figure 9): 'I frames' are intra coded images, coded independently of other frames. These are reference frames. 'P frames' are coded predicatively from the closest previous reference frame, and 'B frames' are coded bi-directionally from the preceding and succeeding reference frames.

**Figure 9** Frame dependencies



The decoding of a multimedia stream depends on the successful receipt of the reference frames ('I frames'). If there is a packet loss with an I frame, then the decoding of dependent frames fails. This causes skewing and degrades the quality. On the other hand, if there is a packet loss with a non I frame, it causes a small, mostly imperceptible pixelisation on a small part of the display. If the loss distribution is bursty, then the longer the loss burst, the higher the probability of loss of an I frame. Moreover, the streaming

technology can use a FEC algorithm, which can prevent only the non bursty packet loss. So, one can realise that it is the loss distribution – and not the average loss rate – that determines the quality of a multimedia stream.

Today's charging mechanisms do not handle this problem, they measure only the summarised volume of unsent data (see above), and they do not use QoS differentiation. So, it is possible during two streaming calls that ten single packet losses do not cause, and ten neighbouring errors cause, quality deterioration; however, the two payments (the two numbers of successfully received packets) are equal.

We recognised this problem, and developed new mechanisms which are presented in Subsections 8.5 and 8.6. These methods can determine the momentary loss distribution (e.g., the numbers of various length loss bursts), which is in proportion to the QoS of the stream. These values should be sent from time to time to the BS in CDRs, which would be able the quantification of the cash equivalent of the actual loss distribution. We give some advice in this decision in Subsection 8.8.

### *8.5 Measuring the loss distribution*

Measuring bursty packets at the RNC: if the RNC measured not only the cumulative number of unsent data but also the occurrence of error runlengths, then it would send this extra information in a CDR to the BS, which would determine the actual QoS class.

This is an exact method, but it measures only the QoS of the radio bearer (Papadimitriou and Tsaoussidis, 2005). It cannot consider the other side of the network, so it is usable, when the transmission between the RNC and VASP is 'very reliable', namely the connection supports QoS (ATM, leased lines, etc ...). In this case, the VASP must have a direct connection with the NO, so the 3rd party can not be charged with QoS differentiation on the internet.

Measuring bursty packets at the proxy: most streaming protocols (like RTP-RTCP) have a built in QoS measurement procedure: the receiver can report back periodically the actual QoS of the stream, for example: the cumulative number of the lost packets, the interval jitter, etc ... The exact number of loss runlengths could not be calculated from these values, because the client does not send an acknowledgement after each packet, therefore these are time averages values. The RTP/RTCP standard enables definition of new, profile dependent report elements, so if the receiver measured and stored the loss runlengths, it would be sent to the sender in a new report. Of course the proxy should understand these new types of reports, which should be converted to a CDR for the BS.

This method is available to measure the loss bursts, when the mobile operator trusts the software of clients. But it is possible that if a hacker modified the media player, and if he (or she) sent back higher error numbers than the real value, then he (she) would decrease the payment. To prevent cheating, the NO has to use client authentication, but only those users can be served whose media players support this authentication method. This limits the number of marketable mobile equipments.

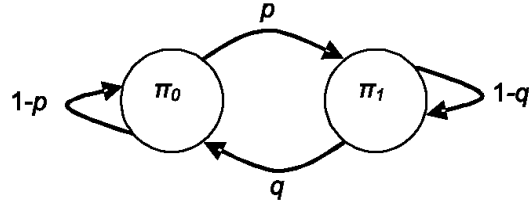
### *8.6 Estimating the loss distribution*

Usually the QoS is specified in few classes, so, mostly, the exact number of lost packets is not needed to be known at charging. It is enough if a few parameters of the streaming are given from which the loss distribution can be deduced or predicted. So, we suggest that the continuous measuring of the packet loss is pointless, it is enough to take samples

of the packet transfer from time to time, and the loss runlengths should be estimated between two sampling periods from these samples. Clearly speaking, the client (or the proxy) should take 1000 samples from the stream; it should use some parameters from these from which the loss distribution can be calculated for the next 10,000 packets. To determine these parameters, we present three models for packet loss metrics.

Sanneck and Carle (2000) use two Markov models in their paper to predict temporal loss dependency. Firstly, we describe the two state Markov model (known as Gilbert model) (see Figure 10).

**Figure 10** Gilbert model



The Gilbert model has only two states with the following probabilities:  $\pi_0$  represents the event that a packet is received successfully while  $\pi_1$  refers to the opposite.  $p$  denotes the probability the packet is lost, provided that the next is not lost,  $q$  is the opposite.  $1 - q$  is called conditional loss probability (clp). If  $p + q = 1$  the Gilbert model reduced to the Bernoulli model.

The state probabilities can be calculated in the following way:

$$p_0 = \frac{q}{p+q}, \quad p_1 = \frac{p}{p+q} \quad (1)$$

$p$  and  $q$  can be computed from the streaming data flow, using the loss length distribution statistics:

$$p = \frac{\sum_{i=0} m_i}{m_0}, \quad 1 - q = \frac{\sum_{i=0} m_i i - 1}{m_i} \quad (2)$$

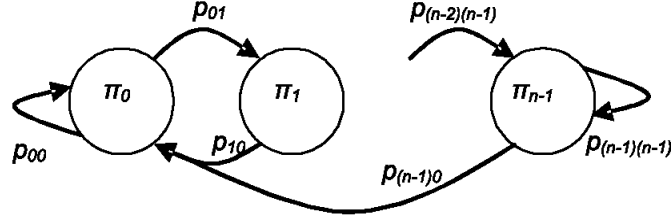
where  $m_i$ ,  $i = 1, 2, \dots, n - 1$  denotes the number of loss burst having length  $i$ , and  $n - 1$  is the maximal length of bursts.

The probability distribution of loss runs ( $p_k$ ) with length  $k$  has geometric distribution

$$p_k = (1 - q)^{k-1} q \quad (3)$$

where  $p$  (mean loss rate) and  $1 - q$  (conditional loss probability) can be used to describe the QoS.

The third model is the extended Gilbert model. It remembers not only the last one but the last  $n - 1$  events; therefore, it needs  $n$  states to describe these  $n$  events (Figure 11), because only the past  $n$  consecutive loss events will affect the future, if the correlation between two loss bursts is small.

**Figure 11** Extended Gilbert model

The transition matrix of the extended Gilbert model is the following

$$\begin{pmatrix} 1-p_{01} & 1-p_{02} & 1-p_{03} & \dots & 1-p_{(n-2)(n-1)} & 1 \\ p_{01} & 0 & 0 & \dots & 0 & 0 \\ 0 & p_{12} & 0 & \dots & 0 & 0 \\ 0 & 0 & p_{23} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_{(n-2)(n-1)} & 0 \end{pmatrix}.$$

We can compute the state probabilities like this

$$\begin{pmatrix} 1-p_{01} & 1-p_{02} & 1-p_{03} & \dots & 1-p_{(n-2)(n-1)} & 1 \\ p_{01} & 0 & 0 & \dots & 0 & 0 \\ 0 & p_{12} & 0 & \dots & 0 & 0 \\ 0 & 0 & p_{23} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_{(n-2)(n-1)} & 0 \end{pmatrix}.$$

The probability distribution of loss runs ( $p_k$ )

$$p_{01} = \sum_{i=0}^{n-1} \frac{m_i}{m_0}, \quad p_{(k-1)(k)} = \sum_{i=1}^{n-1} \frac{m_i}{\sum_{i=k-2}^{n-1} m_i}. \quad (4)$$

The state probabilities ( $\pi_k, \dots, \pi_{n-1}$ ) can determine the QoS class.

These methods enable us to decrease the load of the measuring node and to give the capability to predict the fluctuation of the customers account.

### 8.7 Model evaluation

To evaluate and compare the above described models, we captured a real streaming flow over GPRS calls with the following properties (Table 1). We could not measure over 3G, because it was inaccessible for us.

#### MPEG

We compared how to fit the three probability distributions of loss runs to the occurrence of real loss bursts. So, we computed the state probabilities and state transitions out the first 1000 packets with the equations (1), (2), (4) and (5), and we computed the



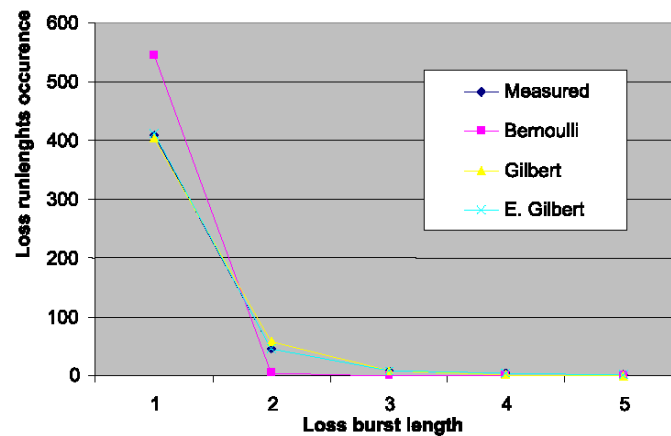
predictions (3), (6) for the next 10000 packets (the estimated occurrences of loss runlengths can be seen in Table 1).

**Table 1** Streaming properties for evaluation streaming flow

<i>Properties</i>	<i>Value</i>
Encoder	MPEG 4
Bit rate control	CBR
Bit rate	23 Kbps
Key frame period	5000 s
Error resilience	None
Size	QCIF
Max. packet size	400 byte

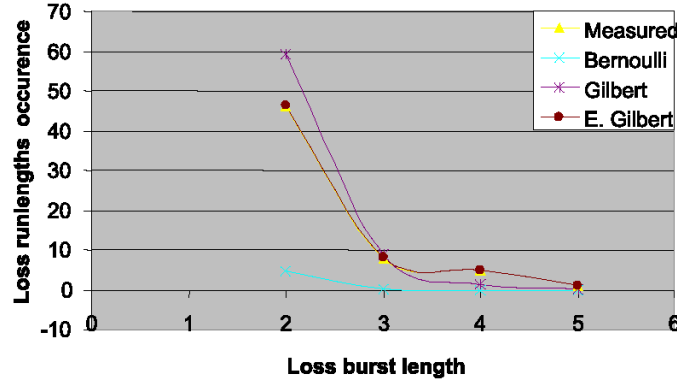
The Bernoulli model is a stateless model, so it cannot predict the time relations between errors; it overestimates the occurrence of single losses, and underestimates the multiple losses (Figure 12).

**Figure 12** Model evaluation



The Gilbert model approaches the real occurrence in a good way, when the maximal length of the error burst is smaller than 4, but it underestimates the multiple losses, which causes the quality deterioration of streaming (it has only two states for prediction). So this model is not usable for charging, because it does not predict well the significant error bursts (Figure 13, Table 2).

The extended Gilbert model makes the closest prediction, but it requires as many states as the maximal loss runlengths is. Its state probability can make the base of the QoS differentiation for charging (Figure 13, Table 2).

**Figure 13** Mode detailed evaluation of our model**Table 2** Lengths of the loss bursts

Model	Lossburstlength				
	1	2	3	4	5
Measured	408	46	8	5	1
Bernouli	544.40	4.56	0.04	0.00	0.00
Gilbert	402.32	591.36	8.76	1.29	0.19
E. Gilbert	411.45	46.39	8.07	5.04	1.01

### 8.8 Using the loss distribution for charging

We determined the loss distribution in the above, but afterwards, an algorithm is needed which can classify the QoS from the probability of loss bursts. This can be based on limits or on a special function, which depends on subjective decisions (for example: on the marketing strategy of the network operator, on the content of the stream, etc ...); therefore, in this paper we do not discuss these methods.

## 9 Conclusion

We have presented a new, streaming proxy based charging architecture for QoS differentiated charging in 3G networks. The recommended solution is available for online and offline charging. We have pointed out the importance of bursty packet loss, which determines the quality of streaming, in contrast with the average loss rate. We have proposed several methods for QoS differentiated charging, which are in compliance with the recommendations of 3GPP. We have described three models, which could predict the QoS of a streaming flow. The algorithms were evaluated them on a real GPRS streaming flow.

In the future we would implement our charging method in a simulation environment and we plan make an experimental charging system, which can be used to evaluate the developed methods.

## Acknowledgements

This work is supported by ETIK and OTKA F042590.

## References

- 3GPP TS 32.240 V6.3.0 (2005) *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication Management; Charging Management; Charging Architecture and Principles (Release6)*, Technical Specification, 3GPP.
- 3GPP TS 32.200 V5.9.0 (2005) *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication Management, Charging Management; Charging Principles (Release5)*, Technical Specification, 3GPP.
- Apostolopoulos, J.G. (2002) *Video Streaming: Concepts, Algorithms, and Systems*, Technical Report, HPL-2002-260, Mobile and Media Systems Laboratory, HP Laboratories Palo Alto, September.
- Ary, B. and Imre, S. (2005) 'Real-time charging in mobile environment', *Hungarian Telecommunication Magazine (Selected Papers)*, Vol. LX, No. 6, pp.54–59.
- Barachi, M.E., Glitho, R. and Dssouli, R. (2005) 'Developing applications for internet telephony: a case study on the use of web services for conferencing in SIP networks', *International Journal of Web Information Systems*, Troubador Publishing, UK, Vol. 1, No. 3, pp.147–160.
- Baug, L. (2002) *GPRS Billing: Getting Ready for UMTS*, SITICOM Group, [http://www.siticom.com/press/pdf\\_hi2001/billingnews050301.pdf](http://www.siticom.com/press/pdf_hi2001/billingnews050301.pdf).
- Barry, M., Gutknecht, J., Kulka, I., Lukowicz, P. and Stricker, T. (2005) 'From motion to emotion: a wearable system for the multimedia enrichment of a Butoh dace performance', *Journal of Mobile Multimedia*, Rinton Press, Princeton, New Jersey, Vol. 1, No. 2, pp.112–132.
- Kim, J. and Park, S. (2005) 'Periodic streaming data reduction using flexible adjustment of time section size', *International Journal of Data Warehousing and Mining*, Sogang University, Korea, Idea Group Inc., Vol. 1, No. 1, pp.37–56.
- Koutsopoulou, M., Gazis, E. and Kaloxylos, A. (2001) 'A novel billing scheme for UMTS networks', *International Symposium on 3rd Generation Infrastructure and Services*, Athens, Greece.
- Koutsopoulou, M., Kaloxylos, A., Alonistioti, A. and Merakos, L. (2004) 'Charging, accounting and billing management schemes in mobile telecommunication networks and the internet', *IEEE Communications Surveys*, Vol. 6, No. 1, First Quarter, pp.50–58.
- Montelius, F. and Larsson, O. (2001) *Streaming Video in Wireless Networks*, MSc Thesis No. LiTH-ISY-EX-3227, Linköping Department of Electrical Engineering, Linköping University, Sweden.
- Olariu, S., Maly, K., Foudriat, E.C., Overstreet, C.M., Yamany, S.M. and Luckenbach, T. (2005) 'Telemedicine for disaster relief: a novel architecture', *Journal of Mobile Multimedia*, Rinton Press, Princeton, New Jersey, Vol. 1, No. 4, pp.285–306.
- Papadimitriou, P. and Tsaoussidis, V. (2005) 'On transport layer mechanisms for realTime QoS', *Journal of Mobile Multimedia*, Rinton Press, Princeton, New Jersey, Vol. 1, No. 4, pp.342–363.
- Sanneck, H. and Carle, G. (2000) 'A framework model for packet loss metrics based on loss runlengths', *SPIE/ACM SIGMM Multimedia Computing and Networking Conference*, San Jose, CA, pp.2–8.

Tewari, H. and Omahony, D. (2003) 'Real time payments for mobile IP, management of next generation wireless networks and services', *IEEE Communications Magazine*, February, pp.126–136.

UMTS Forum (2002) *Charging, Billing and Payment Views on 3G Business Models*, Report 21 Version 4.4a, [www.umts-forum.org](http://www.umts-forum.org).

Waluyo, A.B., Srinivasan, B. and Taniar, D. (2005) 'Research in mobile database query optimization and processing', *Mobile Information Systems*, Vol. 1, No. 4, pp.225–252.