IP Multimedia Services Improvements in the GSM/EDGE Radio Access Network

Benoist Sébire
Nokia Research Center
Itämerenkatu 11-13
FIN-00180 Helsinki
email: benoist.sebire@nokia.com

Tommy Bysted
Nokia Mobile Phones
Frederikskaj
DK-1790 Copenhagen V
email: tommy.bysted@nokia.com

Kent Pedersen
Nokia Mobile Phones
Frederikskaj
DK-1790 Copenhagen V
email: kent.pedersen@nokia.com

***Abstract* - In the Release 6 of the 3GPP standards, a new type of physical layer is being designed for the GSM/EDGE Radio Access Network (GERAN): the Flexible Layer One (FLO). The main advantage of FLO is that configuration of the physical layer is done by the network at call setup. Consequently support of new services such as IP Multimedia Subsystem (IMS) services can be handled smoothly without having to specify new coding schemes in each release of the 3GPP specifications. Through several enhancements, FLO significantly improves the link level performance of real-time IMS services compared to GERAN Release 5. Furthermore, the architecture of FLO brings GERAN even closer to UTRAN and together with Iu alignment, it enables seamless provision of 3G services over the two radio access technologies.**

I. INTRODUCTION

During the past years, a major topic within the 3GPP standardization has been the introduction of the IP Multimedia Subsystem. The aim of the IMS is to provide IP based multimedia services to the user through the UMTS network. A key issue for the user of IMS Services is the Quality of Service (QoS). Since each element of the network contributes to the total QoS, the Radio Access Network be it UTRAN or GERAN, has great importance when considering the total QoS. The radio bearers of UTRAN are designed in a way that is flexible enough to cope with the introduction of future IMS services such as real time multimedia. However, the radio bearers of GERAN have until now solely been either dedicated bearers specialized for given services, such as AMR speech services or they have been generic data bearers having fixed payload sizes, such as the (E)GPRS coding schemes. The fact that the nature of future IMS services to a large extend is unknown, has initiated work to improve the physical layer of GERAN enabling fast introduction of new services and higher spectral efficiency of these services. This paper describes the new flexible layer one (FLO) and is organized as follows. Section I explains how the quality of service is managed in UMTS. Section III and IV provide the motivation for and principles behind FLO. Section V describes the architecture in details, section VI validates the architecture and section 0 shows how FLO can improve the performance of IMS services. Finally a conclusion is drawn in section VIII.

II. QUALITY OF SERVICE IN UMTS

Before explaining why FLO is needed it is important to understand what a radio bearer is and how the QoS is managed in UMTS.

From the mobile user perspective, the UMTS network is a network to access multimedia services (see Figure 1). Its ability to satisfy the user is more important than the technology itself. The quality of service (QoS) is the collective effect of service performances, which determine the degree of satisfaction of a user for a given service. It is characterized by the combined aspects of performance factors and is measured by the degree of satisfaction of the users, and by its spectral efficiency (or degree of satisfaction of the operator). In order to provide satisfactory services, end-to-end QoS requirements are set for each service.

To ensure that the QoS requirements are met from one end to another, a QoS architecture is defined (see Figure 2) [1]. By introducing a hierarchy of services, the architecture allows the QoS to be controlled at different levels, and within different elements along the transmission chain. Since each element of the network contributes to the total QoS, the radio access network be it UTRAN or GERAN, has great importance when considering the end-to-end QoS.

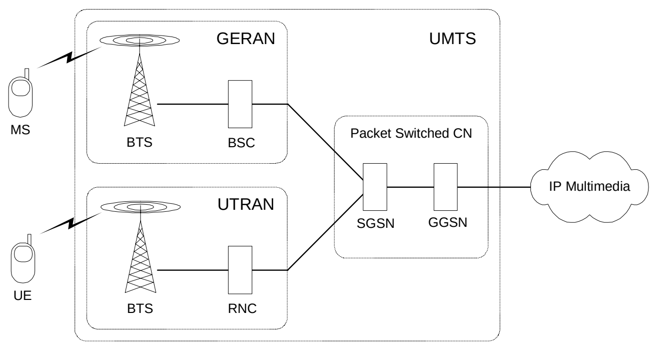

Figure 1. UMTS Architecture

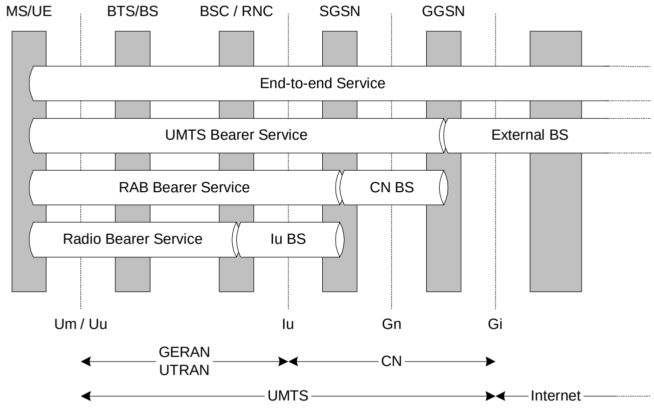

Figure 2. UMTS QoS Architecture

The end-to-end QoS is ensured by two services: the external bearer service, and the UMTS bearer service. A bearer is a service providing QoS between two defined points. The external bearer service manages the QoS within external networks, while the UMTS bearer service contains mechanisms to manage QoS over the UMTS network. It is equally important for both services to fulfil the QoS requirements in order to guarantee the end-to-end QoS.

The UMTS acts as an infrastructure allowing services to be provided, and maintained while the Mobile Station (MS) moves and hides these moves from the IP multimedia subsystem. The UMTS bearer service can be split into: radio access bearer (RAB) service, and core network bearer service (CN BS). The RAB service can in turn be split into: radio bearer (RB) service, and Iu bearer service. Such a split allows the core network (CN) to be independent from the radio access technology used by the RAN. The main task of the RAN is to create and maintain RAB for communication between MS and CN. The RAB gives the CN the illusion of a fixed communication, relieving the CN of radio related aspects. The RAN and the CN map the end-to-end QoS requirements over the Iu interface (Iu bearer services), while the RAN only takes care of satisfying the QoS requirements over the radio path (radio bearer service). This is where FLO takes effect.

Having defined radio bearers and how the QoS is managed in UMTS, the following will deal with the requirements that are set on the RAN for the support of IMS services.

III. Motivation

As mentioned, the need for a flexible layer one in the Release 6 of GERAN is driven by the introduction of IMS services. For an efficient support of IMS services, the following requirements are set on the radio bearer service of the RAN [1] [2] [3] [4]:

- the radio bearers should be flexible enough to efficiently deploy any IP multimedia application;
- the radio bearers should allow the transport of several flows in parallel (e.g. text and video);
- the radio bearers should satisfy the user in a spectral efficient manner;
- the radio bearer should support the UMTS QoS concept and architecture.

In order to fulfil these requirements in an efficient manner, a flexible layer one similar to the UTRAN one is needed. The introduction of such a flexible layer one enables optimised support of real time IMS services in GERAN.

IV. Overview

A. *General*

In GERAN Release 5, the MAC sublayer is responsible for the mapping between the logical channels (traffic or control channels) and the basic physical subchannels [5]. The logical channels are the channels the physical layer offers to the MAC sublayer. Until now these logical channels and the mapping to the basic physical subchannel have been fully specified in [5]. In UTRAN a different approach has been taken, where instead of providing a fixed set of logical channels the physical layer offers transport channels (TrCH), which can be configured at call setup by the network [6]. A transport channel is used to transmit one data flow with a given QoS over the radio interface. A number of transport channels can be active at the same time and multiplexed at the physical layer. With FLO, the concept of transport channels is introduced in GERAN.

B. *Principles*

When introducing FLO, the physical layer of GERAN offers one or several transport channels to the MAC. The configuration of a transport channel (number of input bits, channel coding, interleaving…) is denoted the transport format (TF). As in UTRAN, a number of different transport formats can be associated to one transport channel. The transport formats of each TrCH are defined by a number of attributes, which can be either dynamic or semi-static. The semi-static attributes are common among all transport formats associated to a given TrCH (e.g. CRC size). The dynamic attributes can however be different for each transport format (e.g. number of bits for a transport format).

The transport formats are completely controlled by the RAN and signalled to the MS at call setup or when the channel is reconfigured. In both the MS and the BTS, the transport formats are used to configure the encoder and decoder units. When configuring a transport format, the RAN can choose between a number of predefined CRC lengths, interleaving depths, rate matching attributes etc.

On transport channels, transport blocks (TB) are exchanged between the MAC sublayer and the physical layer on a transport time interval (TTI) basis. Only a limited number of combinations of the transport formats on the different TrCHs are allowed. A valid combination is called a transport format combination (TFC). The set of valid TFCs on a basic physical subchannel is called the transport format combination set (TFCS).

In order to decode the received sequence the receiver needs to know the active TFC for each radio packet. This information is transmitted in a layer 1 header: the transport format combination indicator (TFCI). Each of the TFC within a TFCS are assigned a unique TFCI value and when a radio packet is received this is the first to be decoded. From the decoded TFCI value the transport formats for the different transport channels are known and the decoding can start.

V. ARCHITECTURE

The architecture of FLO in GERAN is depicted on Figure 3. In the following each building block is explained.

A. CRC Attachment

Error Detection is provided on each transport block through a CRC. The size of the CRC to be used is fixed on each TrCH and configured by layer 3 (semi-static attribute of the transport format). The entire transport block is used to calculate the parity bits. Code blocks are output from the CRC attachment. CRC sizes of 0 (no error detection), 6, 12 and 18 bits are available [7] [8].

B. Channel Coding

After CRC attachment, the code blocks are processed through channel coding, producing encoded blocks. Only one code is used, the same 1/3 non-recursive non-systematic convolutional code of constraint length 7 as in EGPRS defined in [7]. Using only one code rate simplifies the architecture, and eases the design of upper layers since no code selection is to be performed.

C. Rate Matching

The rate matching is the core of FLO. In rate matching, bits of an encoded block on a TrCH are repeated or punctured. Since the transport block size is a dynamic attribute, the number of bits on a TrCH can vary between different transmission times. When this happens, bits are repeated or punctured to ensure that the total bit rate after TrCH multiplexing is identical to the total channel bit rate of the allocated basic physical subchannel. The rate-matching attribute (RMA) is semi-static and used when the number of bits to be repeated or punctured for each transport channel is calculated. Higher layers assign the RMA for each transport channel. The RMAs define priorities between the coded bits of different transport channels: the higher the RMA is, the more important the coded bits are. By setting different RMAs to different transport channels, the coding rate of the transport channels can be adjusted.

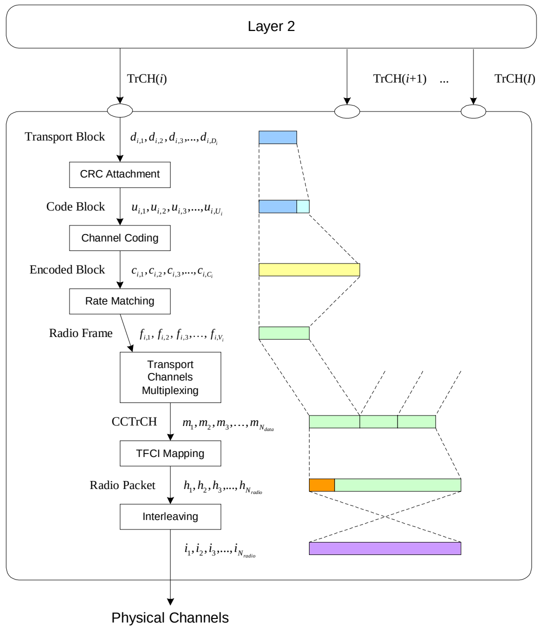

Figure 3. FLO Architecture

Outputs from the rate matching are called radio frames. For every radio packet the rate matching produces one radio frame per encoded block, i.e. per TrCH.

The rate matching algorithm for GERAN is based on the algorithm defined for UTRAN [6]. In GERAN a few simplifications have been made since there is no spreading factor, nor compressed mode, nor special cases such as turbo codes and therefore many parameters of the UTRAN algorithm can be fixed either to 0 or 1 [8].

D. Transport Channel Multiplexing

For every radio packet to be transmitted, one radio frame from each active TrCH is delivered to the TrCH multiplexing. These radio frames are serially multiplexed into a Coded Composite Transport Channel (CCTrCH).

E. TFCI Mapping

For every radio packet, an indicator for the active TFC is transmitted in the TFCI. In GERAN the TFCI is limited to 5 bits, allowing a maximum of 32 different transport format combinations on the same basic physical subchannel. During a connection these transport format combinations can be changed by a reconfiguration.

F. Interleaving

The coded TFCI and the CCTrCH are interleaved together on bursts. The interleaving can be either block diagonal or block rectangular and is configured at call set-up [8].

Having defined the architecture, the next section will assess the performance.

VI. AMR OVER FLO

Since FLO can be configured to support any service, a natural way of evaluating the performance is to compare with the performance of an existing well defined service, such as the Adaptive Multi-Rate (AMR) speech service, as standardised in Release 98 of the 3GPP specifications (TCH/AFS of [7]). That is, the aim of the comparison is to evaluate how well FLO can implement an AMR service compared to the existing optimised channel coding (TCH/AFS).

A. FLO Configuration

In order to support AMR over FLO, two transport channels are set-up: one for the class 1a bits and another one for the class 1b bits. In the following, configurations of each FLO building block are given.

CRC Attachment: A CRC is used for error detection on the transport channel of the class 1a bits while there is none on the transport channel of the class 1b bits. The same 6 bits CRC as TCH/AFS is used.

Channel Coding: The same mother code is used for the two transport channels: 1/3 non-recursive non-systematic convolutional code of constraint length 7.

Rate Matching: Proper rate matching attributes must be selected for each AMR mode so as to get as close as possible to the existing coding rates. Table I lists the best rate matching attributes for every bit class of every mode. When these rate matching attributes are used, the coding rate difference between what is specified in [7] and what FLO gives, is below 0.001 (at equivalent bandwidth).

TFCI Mapping: a TFCI of 3 bits is considered. It allows up to 8 different TFCs such as for instance: 4 AMR modes, FACCH, SID_UPDATE and SID_FIRST.

Table I. Rate Matching Attributes for AMR over FLO

AMR Mode	Bit Class	Coding Rate	Best RMA
12.2 kbit/s	1a (81 bits) 1b (163 bits)	0.5000 0.5949	6 5
10.2 kbit/s	1a (65 bits) 1b (139 bits)	0.4465 0.4810	13 12
7.95 kbit/s	1a (75 bits) 1b (84 bits)	0.3476 0.3907	10 9
7.4 kbit/s	1a (61 bits) 1b (87 bits)	0.3350 0.3508	1 1
6.7 kbit/s	1a (55 bits) 1b (79 bits)	0.2748 0.3496	5 4
5.9 kbit/s	1a (55 bits) 1b (63 bits)	0.2735 0.2800	1 1
5.15 kbit/s	1a (49 bits) 1b (54 bits)	0.2292 0.2596	7 6
4.75 kbit/s	1a (39 bits) 1b (56 bits)	0.2284 0.2231	1 1

Interleaving: As for TCH/AFS an 8 bursts diagonal interleaving is used. The result of the interleaving is a distribution of the reordered 464 bits over 8 bursts using the even numbered bits of the first 4 bursts and odd numbered bits of the last 4 bursts [8].

B. Link Level Results

Simulations were run for TU3iFH at 900Mhz over 20000 frames. Decoding errors of the TFCI and typical MS impairments were included. Link level results are summarized in Table II below. The loss is very little compared to the existing AMR schemes. In average there is a 0.15dB loss at 1% of FER and 0.04dB loss at 0.1% of RBER. This demonstrates that even though FLO is generic, the FLO architecture enables design of channel coding schemes having similar performance as coding schemes optimised for a specific service. Note that the constraint length of the convolutional code always is 7 when using FLO, while the existing AMR schemes use constraint length 5 for modes 5.15, 6.7, 7.4, 10.2 and 12.2. Hence the observed gains in some cases.

Table II. AMR over FLO vs. TCH/AFS
(C/Ico in TU3iFH at 900Mhz)

AMR Mode	1% FER	0.1% Rber1b
4.75 kbit/s	+0.4 dB	+0.3 dB
5.15 kbit/s	0.0 dB	+0.3 dB
5.9 kbit/s	+0.4 dB	0.0 dB
6.7 kbit/s	-0.2 dB	-0.1 dB
7.4 kbit/s	+0.1 dB	-0.2 dB
7.95 kbit/s	+0.4 dB	-0.1 dB
10.2 kbit/s	+0.1 dB	+0.6 dB
12.2 kbit/s	0.0 dB	-0.5 dB

VII. EXAMPLE OF RT IMS SERVICE

One key aspect in the support of IMS services is that the RAN should be flexible enough to accommodate any service (see section III). For this example, no assumption is therefore made on the content of the RTP packets. It could be MPEG4 audio, MPEG4 video or anything else. Only the three following QoS requirements are set in order to set up the radio bearers:

Source rate:

- 160 bits RTP/UDP/IP packet every 20ms (8kbit/s)
- 240 bits RTP/UDP/IP packet every 20ms (12kbits/s)
- 320 bits RTP/UDP/IP packet every 20ms (16kbit/s)

Residual Bit Error Rate (RBER) that requires a 12 bits CRC to be used.

Delay that allows a 40ms interleaving depth to be used but does not allow the RLC to be acknowledged, i.e. link level improvement by retransmission is not possible.

A. *Radio Bearers*

In Release 5 the radio bearers that can be used for RT multimedia services are based on the (E)GPRS coding schemes. However, these coding schemes are not optimised for the support of unacknowledged RT multimedia services in general and for the support of the given RT multimedia service example in particular. Firstly, although a 40ms interleaving depth is allowed according to the delay requirements, it is not possible to apply it when using the (E)GPRS coding schemes. Then since the payload to be carried does not exactly fit into one RLC packet some padding has to be used. The resulting release 5 radio bearers are therefore: for the 8kbit/s source rate, MCS-1 is used with 16 bits of padding; for the 12kbit/s source rate, MCS-3 is used with 56 bits of padding; for the 16 kbit/s source rate, MCS-4 is used with 32 bits of padding.

With FLO in Release 6, it is possible to optimise the physical layer according to the QoS requirements at radio bearer setup. For the RT multimedia service example, it means that: the coding scheme can be adjusted at a bit level and no padding bits are used; the RLC/MAC header can be optimised for unacknowledged mode; 12 bits CRC is used as requested by the RBER requirement; 40ms interleaving depth is used as allowed by the delay requirement.

B. *Link Level Performance*

The link level performance of the radio bearers of Release 5 and the ones that can be obtained with FLO in Release 6 are here compared for the support of the RT multimedia service example. The RLC was unacknowledged, the MAC mode dedicated, 20 000 frames were simulated in TU3iFH, and typical MS impairments were included. For FLO, a 3 bits TFCI was assumed. Simulation results are presented on Figure 3 and summarized in Table III.

For an 8kbit/s RT multimedia service, the simulations show that FLO can bring a 3.6dB gain. At this rate, only a few padding bits are used in the Release 5 radio bearers. Consequently, the gain of FLO mainly comes from the reduced overhead and the longer interleaving depth. For a 12kbit/s RT multimedia service, FLO is shown to bring 8.7dB gain and for a 16kbit/s RT multimedia service, the gain is 9.7dB. The increased gains compared to 8kbit/s come from the padding bits. Clearly by introducing FLO the performance of RT multimedia services will be improved.

Table III. Link Level improvement with FLO
(C/Ico at 1% of FER in TU3iFH at 900MHz)

RT multimedia service Source Rate	Release 5	Release 6 with FLO	Gain from FLO
8 kbit/s	11.0 dB	7.4 dB	3.6 dB
12 kbit/s	19.5 dB	11.3 dB	8.2 dB
16 kbit/s	25.1 dB	15.4 dB	9.7 dB

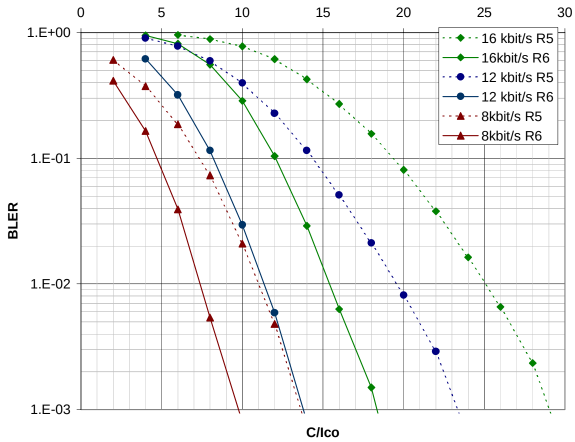

Figure 4. Link Level Performance (TU3iFH - 900MHz)

Note that it is important to remember that it does not mean that the performance of EGPRS is poor, but simply that EGPRS was not optimised for the support of unacknowledged RT services on dedicated channels as it was designed for acknowledged traffic on shared channels.

VIII. CONCLUSION

For the support of new RT services, new coding schemes have been standardised and implemented in GSM/EDGE for more than a decade. Although such an approach is optimised from a link level performance point of view it presents a few drawbacks: the channel coding is not flexible; it is a burden for implementation as memory consumption and testing time increase for each new service; fast introduction of new RT services is difficult. In order to efficiently support IMS services a flexible layer one is therefore needed and being designed for GERAN. The main advantage of FLO is that configuration of the physical layer is specified at call setup allowing smooth introduction of new services and alleviating implementation in the long term. Performance simulations have shown that that the produced channel coding by FLO is as good as optimised coding schemes and that FLO improves the performance of RT IMS services compared to Release 5. The gains originate from enhanced granularity, reduced overhead and flexible interleaving. Last but not least, the architecture of FLO brings GERAN even closer to UTRAN and together with Iu alignment, it enables seamless provision of same 3G services over the two radio access technologies.

REFERENCES

[1] 3GPP TS 23.107, “QoS Concept and Architecture”
[2] 3GPP TS 22.101, “Service principles”
[3] 3GPP TS 22.228, “IP multimedia subsystem - Stage 1”
[4] 3GPP TS 23.228, “IP multimedia subsystem - Stage 2”
[5] 3GPP TS 45.002, “Multiplexing and multiple access on the radio path”
[6] 3GPP TS 25.201, “UTRAN Physical Layer General Description”
[7] 3GPP TS 45.003, “Channel Coding”
[8] B.Sébire, T.Bysted, K.Pedersen, “Flexible Layer One for the GSM/EDGE Radio Access Network”, *ICT’2003*, Feb. 2003