Optimal Radio Acess Bearer Configuration for Voice over IP in 3G UMTS networks

Xavier Pérez-Costa[a,b], Albert Banchs[c1], Juan Noguera[a] and Sebastià Sallent-Ribes[b]

[a]**NEC Network Laboratories, Heidelberg, Germany**
[b]**Universitat Politècnica de Catalunya, Barcelona, Spain**
[c]**Universidad Carlos III, Madrid, Spain**
e-mail: {perez,noguera}@ccrle.nec.de, banchs@it.uc3m.es and sallent@mat.upc.es

***Abstract*: The Radio Access Bearer (RAB) is the entity responsible for transporting radio frames of an application over the access network in UMTS. The parameters of a RAB, namely the maximum bandwidth and the allowed frame sizes, can be configured according to the requirements of the application using it. In this paper we take up the issue of identifying the optimal RAB configuration for VoIP applications. For this purpose, we have developed our own simulator, which models accurately all the aspects that have an impact onto the RAB configuration, and have evaluated the voice quality resulting from different RAB configurations. Based on the results obtained, we propose two new RAB configurations for VoIP, one when Robust Header Compression (RoHC) is used and another when it is not used.**

1. Introduction

The Universal Mobile Telecommunications System (UMTS) is a third generation mobile system developed by the 3rd Generation Partnership Project (3GPP) [1]. It aims at providing mobile users with multimedia services (voice, video and data) with assured QoS and up to a dedicated capacity of 2 Mbps per user.

UMTS originally defined two different domains, Circuit Switched (CS) for voice and transparent/non-transparent CS data, and Packet Switched (PS) for packet data. Recently, 3GPP has defined the IP Multimedia Sub-system (IMS), which uses the PS domain to provide IP multimedia services. IMS sets a very flexible framework for service creation, and enables the migration of CS services to the PS domain, with consequent reduction on network deployment and operation cost. For speech services, this migration means the replacement of traditional CS speech service by VoIP provided via IMS and the PS domain.

3GPP has defined the concept of Radio Access Bearer (RAB) as a user plane connection provided by the UMTS Terrestrial Radio Access Network (UTRAN) between a User Equipment (UE) and the Core Network. The general characteristics of a RAB (data rates, QoS, etc) are normally set by the Core Network (CN) based on subscription and/or requirements of the media or set of medias using the RAB. The actual configuration for a RAB is decided by UTRAN based on the RAB information received from the CN.

The RAB configuration has a direct impact on network resource usage. The more suited the RAB configuration is to the actual pattern of the data being transferred, the more efficient the RAB is in terms of usage of network resources. Choosing proper RAB configurations is key to UTRAN, given the high cost of last-mile transport (Iub interface) and the rather limited radio resources.

Between the RNC and the UE data is always transferred inside frames which length is within a set of allowed frame sizes. The set of allowed frame sizes is configured when the RAB is setup but, for complexity reasons, the size of the set has to be small. In general, no PS reference RAB has a set size larger than three. This is, the RAB has three allowed sizes (e.g. 200 bits, 400 bits and 600 bits). In simplified terms, when data is to be transferred, an appropriate frame size is selected and padding bits are added, if needed, to fill the remaining bits of the frame.

The RAB bandwidth determines the QoS received by the application, and the set of allowed frame sizes for the RAB determines the amount of bandwidth wasted for padding. Given a certain application, it is crucial to define its RAB well adjusted to its requirements; too small bandwidth will result in a bad quality, while too large bandwidth or improper frame sizes will result in a waste of resources.

From the above it follows that a RAB design optimized for VoIP is an important issue in UMTS networks. The fact that VoIP traffic sends at a variable bit rate, may include control traffic like e.g. RTCP and, if Robust Header Compression (RoHC) is used, headers have a variable size, makes the definition of the optimal RAB a challenging task. While 3GPP has already defined a reference RAB for VoIP support [2], this RAB provides no optimized handling of VoIP traffic.

The focus of the present paper is on the search for an optimized RAB for VoIP. The outline is as follows. In Section 2. we describe the RAB concept in the UMTS architecture and the reference RAB currently defined for VoIP. In Section 3. we describe the simulator that we have developed for evaluating the resulting voice quality from a certain RAB configuration. In Section 4. we present the results obtained for sweeping along the configuration space; from these results we propose two optimal RAB configuration for VoIP, one when RoHC is used and one when it is not. Section 5. concludes the paper with some final remarks.

2. Radio Access Bearer

The transmission of data within a RAB in UMTS works as follows. Data (namely IP packets) generated by an application at the UE is stored in an internal buffer. This buffer is emptied periodically, every Transmission Time Interval (TTI), when a radio frame is created with

[1]This work was carried on while working at NEC

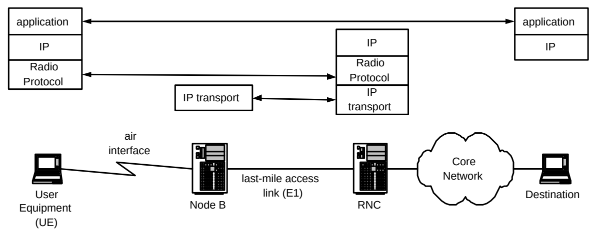

Figure 1: UMTS Architecture (air interface and radio access network).

the data stored at the buffer up to a certain maximum frame size (MFS). The RAB bandwidth corresponds to MFS/TTI. In this paper we take TTI equal to 20 ms which is a commonly used value in current implementations.

In case the amount of data in the buffer is less than MFS, a frame of size smaller than MFS may be created. However, only a few frame sizes are allowed, so that we need to fill the frame with padding up to the next frame size allowed.

Once the frame has been created as described above, it is transported through the air interface to the Node B, where an IP packet containing the frame is created[2]. The IP packet is then transported through the Radio Access Network (RAN) to the RNC. The last-mile link in the RAN is commonly E1 on leased lines or microwave links.

The RNC terminates the radio protocol; it extracts the radio frames from the IP transport packets, and the data from these frames, discarding the padding, and transmits the resulting data (which are IP packets) further into the Core Network (CN). The UMTS architecture and protocol stack are illustrated in Fig. 1.

A reference RAB for VoIP is defined by 3GPP in [2], which defines a RAB bandwidth of 46 Kbps and frame sizes of 920, 304 and 96 bits. As we will see later on with our simulation results, this RAB is not well optimized for VoIP.

3. Simulator Description

In order to evaluate performance of VoIP depending on the RAB configuration we have developed our own simulator. Ours is an event-driven simulator, written in the C++ programming language, that closely simulates the frame creation, buffering and transmission at the UE, and all the aspects involved with it, that have an impact on the appropriate RAB configuration. We focus in the upstream part because it is where the packets to be transmitted over the RAB are created. The configuration for the upstream case will also apply to the downstream case since the packets to be sent over the RAB are the ones received from the upstream part. Fig. 2 depicts the modules of which the simulator consists; in the following we describe them in detail.

[2]According to 3GPP specifications, the Radio Access Network (RAN) may be based on IP or ATM transport [3]; throughout this paper we focus in IP transport since we argue that it is better suited for supporting a mix of traffic types, allows for economy of scale and that having an All-IP network saves management and operational costs.

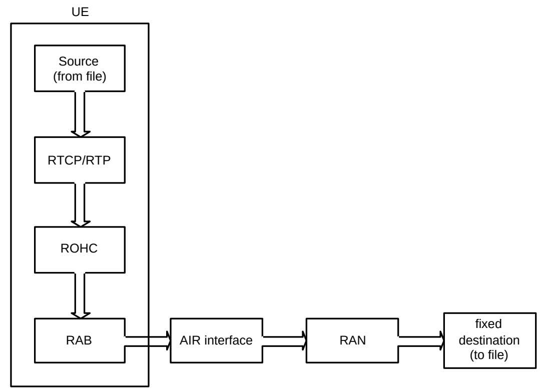

Figure 2: Simulator modules.

3.1. Source Model

Voice sources are modeled according to real-life audio clips obtained from the 3GPP web site [4]. We extract the frame arrival times and sizes from these clips, which are written into a file and then used as input into our simulator. The specific audio clip that we have used in this paper corresponds to a 12.2 Kbps AMR codec audio clip of about 20 seconds with active and idle periods. Each speech frame is encapsulated into an RTP/UDP/IPv6 packet. Overhead includes the IPv6 header (40 bytes), the UDP header (8 bytes), the RTP header (12 bytes), the profile (1 byte) and the PDCP (1 byte).

3.2. RTCP Model

In addition to the speech frames, Real-Time Control Protocol (RTCP) packets are also sent periodically by the UE. In the simulation results we present in this paper, we have considered two different scenarios for RTCP, average and worst-case, depending on the size and inter-arrival times of RTCP packets.

3.2..1 Average RTCP

With average RTCP we consider a payload size of 60 bytes, which roughly matches the size of an RTCP packet with a useful CNAME tag and a Receiver Report[3].

[6] specifies that the maximum rate at which RTCP packets may be sent is of one packet every 5 seconds. This is the RTCP inter-arrival time that we have taken in our average RTCP scenario.

3.2..2 Worst-case RTCP

From [6] we have that the bandwidth used by RTCP should be limited to 2.5% of the total bandwidth. Assuming an AMR RTP payload of 32 bytes [7], which produces a total sending rate of 36900 bps, and an RTCP inter-arrival time of 5 seconds, this gives the following RTCP packet size:

$$S_{RTCP} = 5 \cdot 0.025 \cdot 36900 \approx 4600 \text{ bits} \quad (1)$$

[3]For example, in [5] (section 3.4), it is mentioned that a typical RTCP packet size is 90 bytes, which taking out the IPv4/UDP header leaves us with this 60 byte payload, approximately.

Header Name	Header Function	Header Size
IR	Initialization/Refresh	60
IR_DYN	Partial context update	21
UOR2-ID	Compressed header	6
UO1-ID	Compressed header	5
UO0	Compressed header	4

Table 1: RoHC header sizes (Bytes).

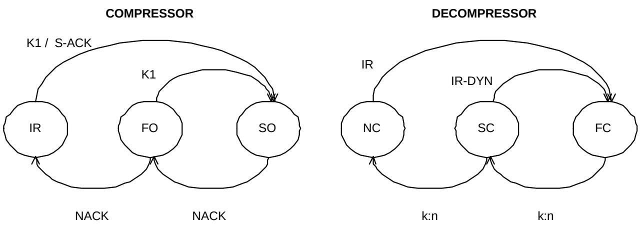

Figure 3: Compressor and decompressor state-machine.

We have taken the above in our worst-case RTCP scenario[4].

3.3. RoHC Model

In our simulator we assume the Bidirectional Optimistic mode of RoHC [9] and emulate its behavior by means of a simplified state machine. We consider the compressed header sizes shown in Table 1 for a large CID of one byte; optional ACKs are not sent, in order to avoid additional overhead. Fig. 3 illustrates the compressor and decompressor state-machine, which we detail in the following.

3.3..1 Compressor

The compressor adds the IR header size to the payload while in the IR (Initialization/Refresh) state, the IR-DYN size while in the FO (First Order) state and the average size of the UOR2-ID, UO1-ID and UO0 while in the SO (Second Order) state.

The transition from the IR to the SO state is performed when a fixed amount of IR headers, K1 (which we take equal to 3), has been sent, such that we can safely assume that the decompressor is in the FC (Full Context) state or when a S-ACK has been received.

The transition from the FO to the SO state is performed when a fixed amount of IR-DYN headers, also K1, has been sent, such that we can safely assume that the decompressor is in the FC.

The transition from the SO to the FO state or from the FO to the IR state is performed when a NACK from the decompressor is received.

3.3..2 Decompressor

The transition from the NC (No Context) to the FC state is performed when an IR header is received and produces a S-ACK feedback message to the compressor.

[4]Note that [8] gives as maximum payload size 130 bytes, but states that this may change with the RTCP packet size.

The transition from the SC (Static Context) to the FC state is performed when an IR-DYN header is received; no feedback message is sent.

The transition from the FC to the SC state or from the SC to the NC state is performed when k packets out of n (we take 3:6) are received with errors, resulting in a NACK feedback message.

We do not consider the impact of packet drops onto the state machine in our model, since the probability of a burst of drops large enough to cause a sequence number wraparound is negligible.

Errors introduced by the air interface into the (possibly compressed) header result, considering a 3 bits CRC, in a detected error with probability 7/8, which leads to an increase in the k counter, and in an undetected error with probability 1/8. As explained above, after 3 detected errors, a NACK is sent.

When errors are detected we assume that RoHC will be able to repair k out of n, i.e., the audio quality will not be harmed. On the other hand, when the errors are not detected they will not be repaired resulting in a packet drop, since a packet with an incorrect header will not reach its destination.

3.3..3 Feedback

The RoHC feedback messages experience the same delay as the data packets from the UE to the UPS, i.e., no immediate feedback is assumed. A medium free of errors is assumed for the feedback path.

3.4. RAB

The RAB module in our simulator works according to the explanation given in Section 2.. The various RAB parameters are configurable in the simulator, so that we can sweep along the parameter space in the search for the optimal configuration.

3.5. Air interface

The ARROWS IST project [10] shows that there is no correlation between errors in different TTI's [11], as

- For vehicle speeds above 13 km/h the coherence time is lower than 10 ms, so there is no correlation between consecutive TTI's.
- Below 13 km/h the power control, updated 1500 times/s, will cancel fast fading and shadowing fading, so the Eb/N0 will remain approximately constant.

The above leads to independence among errors at different TTI's; thus errors can be assumed Markovian. Based on [12], we consider packet error rates of 10^{-2} and 10^{-3}. We consider that an erroneous packet with header of size H and payload of size S has its error in the header with probability $\frac{H}{S+H}$ and in the payload with probability $\frac{S}{S+H}$. In the former case, the error impacts the RoHC, while in the latter it impacts the resulting audio quality.

3.6. RAN

In order to model the RAN behavior, we have used the OPNET simulator [13], which includes a UMTS mod-

Scenario	FER AI	RAN cong.	RTCP	playback time
Ideal cond.	10^{-3}	no	no	100 ms
Ideal cond. + avg. RTCP	10^{-3}	no	avg.	100 ms
Bad cond. + avg. RTCP	10^{-2}	yes	avg.	100 ms
Bad cond. + w-c RTCP	10^{-2}	yes	w-c	100 ms
Bad cond. + w-c RTCP + PB=50 ms	10^{-2}	yes	w-c	50 ms

Table 2: Simulation scenarios.

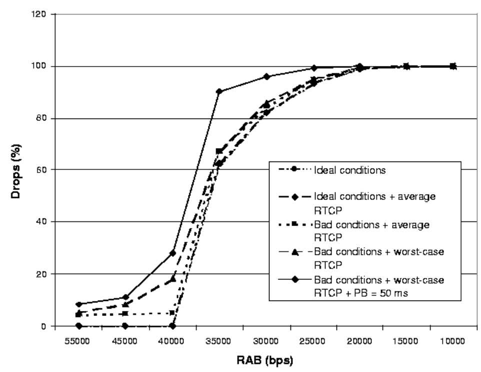

Figure 4: No RoHC: Packet drops vs. bandwidth.

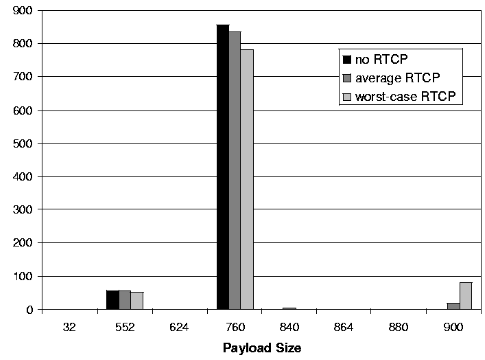

Figure 5: No RoHC: Packet distribution.

ule, to simulate two scenarios where the RNC is connected to a Node B through an E1 link. The first scenario consists of a few UEs and therefore no congestion, while the second one increases the number of UEs to create congestion. From each simulation we have taken the delays and drops suffered by a sequence of packets. Then, we have treated the packets in our simulator according to this sequence, i.e., we drop or introduce a certain delay to packet i in our simulator according to the behavior experienced by packet i of the recorded sequence. The average drop rate for the congested sequence is of about 4%; for the uncongested sequence there are no drops.

3.7. Destination Playback time

We assume audio and video playback applications, which are intolerant to packets arriving later than their playback time [14]. Therefore, we drop packets that arrive later than this time. The playback time represents the maximum delay that can be supported between the creation of a frame in its source and the moment when it is played in the destination, and is a configurable value.

4. Discussion of the Results

In order to identify the optimal RAB configuration, we conducted a series of simulation experiments sweeping along the configuration parameters. In this section we describe the results obtained for the no RoHC and RoHC cases. In each case, we consider the scenarios shown in Table 2 where 'AI' stands for Air Interface, 'cong.' for congestion, 'cond.' for conditions, 'avg.' for average and 'w-c' for worst-case.

4.1. No RoHC

Fig. 4 shows the drop rate experienced by the VoIP application in the no RoHC case as a function of the RAB bandwidth for the five scenarios described above. By drop rate we count the frames dropped at the IP-RAN, the frames resulting with errors in the air interface and the frames that arrive at the destination later than the playback time.

From the figure it can be seen how the drop rate decreases with an increase in the RAB bandwidth. As long as RTCP is not worst-case, the smallest possible drop rate is already achieved with a RAB of 40.000 bps, and a larger bandwidth does not further decrease the resulting drop rate.

Based on the above, we propose a RAB of 45.000 bps, i.e., the 40.000 bps plus some additional safeguard. We argue that provisioning the RAB bandwidth taking into account the worst-case RTCP would require a very large bandwidth and lead to an inefficient use of this bandwidth in most cases.

One additional recommendation from our results is that special care must be taken by the application developers in the design of RTCP. In fact, our results show that the worst-case RTCP allowed by the standards produces a quite large additional packet drops with our recommended bandwidth. This drop rate is specially harmful for the perceived QoS, as drops occur in bursts at the instants when long RTCP messages are sent.

The remaining configurable parameters are the allowed frame sizes. Frame sizes adjusted to the payloads will result in smaller frames, and therefore smaller IP packets in the RAN, so that less bandwidth will be used by them in the expensive last-mile links.

In order to understand the payload distribution resulting from the different scenarios, we study the histogram of payloads in the no RTCP, average RTCP and worst-case RTCP[5] cases, when the RAB bandwidth is of 40.000 bps. This is illustrated by Fig. 5 (even though not

[5] In our scenarios, the payload size distribution turned out to be

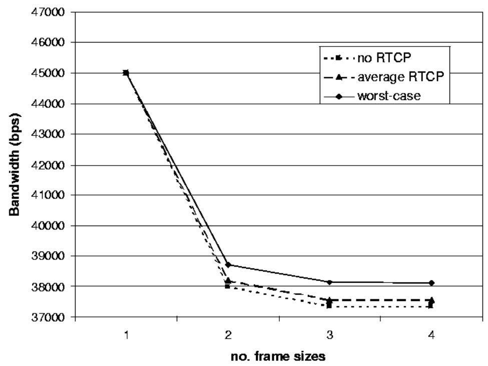

Figure 6: No RoHC: RAN bandwidth consumption.

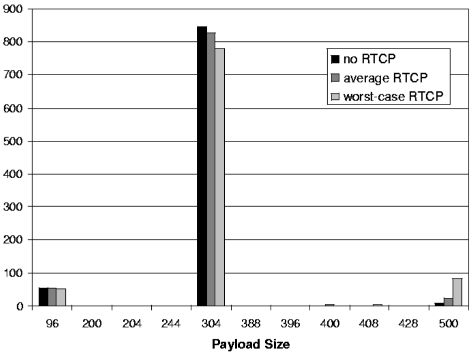

Figure 8: RoHC: Packet distribution.

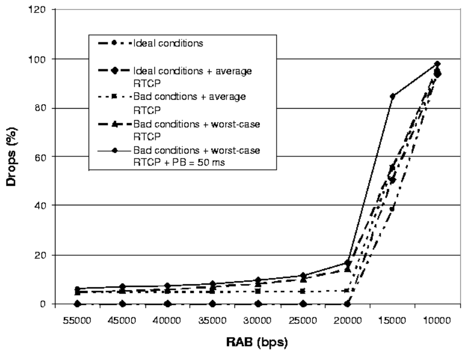

Figure 7: RoHC: Packet drops vs. bandwidth.

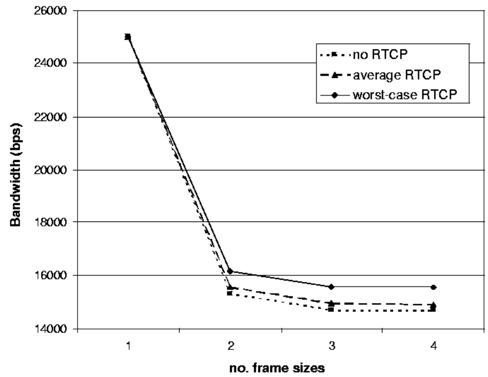

Figure 9: RoHC: RAN bandwidth consumption.

appreciated in the graph due to scale problems, all sizes that appear in the y-axis contain at least one frame for one of the cases).

From the figure, it can be seen that the most frequent payloads are 900, 760 and 552, the last two corresponding to the size of an active and an idle frame, respectively. The maximum payload size (900) is used when the data to be transmitted does not fit into one frame, due to some additional load caused e.g., by an RTCP. In such case one, or more maximum payload frames may occur, plus possibly one transition frame of intermediate size.

In Fig. 6 we study the bandwidth consumed in the RAN as a result of using one frame size (900), two (900, 760), three (900, 760 and 552) and four (900, 760, 552 and 880). It can be observed that significant savings are achieved when using up to three frame sizes, and that using a fourth one produces no further gain. As a consequence, our proposal for the allowed frame sizes in the optimal RAB for RoHC is of 900, 760 and 552 bits.

4.2. RoHC

We repeated the above experiments when RoHC is used. Fig. 7 illustrates the drop rate as a function of the bandwidth; following the same reasoning as in the previous case, we propose a RAB bandwidth of 25.000 bps for the RoHC case.

Fig. 8 depicts the payload size histogram when the RAB bandwidth is set equal to 25.000 bps. It can be observed from the distribution that the more frequent payload sizes are 500, 304 and 96 bits.

Fig. 9 illustrates the bandwidth consumption at the RAN when one frame size is allowed (500 bits), two (500 and 304), three (500, 304 and 96) and four (500, 304, 96 and 428). Results show that there is no gain when using more than three frame sizes; therefore, our proposal for the allowed frame sizes is 500, 304 and 96 bits.

Finally, we note that the RAN bandwidth consumption in the RoHC case is of about half of the bandwidth consumption when no RoHC is used. We conclude that the bandwidth savings when using RoHC is of about 50%, both in the air interface and in the RAN.

4.3. Comparison with the existing reference RAB

The existing reference RAB for VoIP proposes the same frame sizes for the intermediate and smaller frame as the ones we propose for the RoHC case, but a much larger maximum frame size. Among the possible reasons for the difference between our proposal and the reference, we conject the following:

- Our simulation allows us to derive a more adjusted value of the required bandwidth; a less detailed

mainly dependent of RTCP.

study would require a larger bandwidth as a safeguard.

- We consider play-back applications that can adapt to a certain delay/jitter without seeing their quality seriously harmed; less adaptive applications would require a larger bandwidth.

5. Conclusions

VoIP is targeted as one of the key applications for the present and future UMTS. One of the objectives for supporting this traffic type is to minimize the bandwidth consumption at the air interface and last-mile link, which commonly are the most expensive parts of the network for operators, while ensuring a sufficient QoS.

In order to achieve the above goal, it is fundamental to define optimized RABs for VoIP support. In this paper we take up this issue: we develop a simulator and simulate a range of RAB configuration; as a result, we propose two RAB configurations for VoIP (one with RoHC and one without).

The main merit of the present work is the elaboration of thorough models for the various aspects that impact the choice of the RAB configuration, against which we validate the proposed RAB configurations. These aspects could not be reliably accounted for otherwise, so any RAB configuration validated by less accurate methods will cause an uncertain behavior.

For the no RoHC case, we propose a maximum RAB bandwidth of 45.000 bps and frame sizes of 900, 760 and 552 bits. For the RoHC case, we propose a maximum RAB bandwidth of 25.000 bps and frame sizes of 500, 304 and 96 bits. The resulting bandwidth consumption at the RAN is of about 30.000 and 15.000 bps, respectively.

It follows from the above that the gain of using RoHC is of about 50%, both in the air interface and the RAN. The gain of our RoHC RAB is also about 50% with respect to the current reference VoIP RAB by 3GPP, which proposes a RAB bandwidth of 46.000 bps.

The bandwidth consumption at the air interface depends on the number of Channel Elements required, which in turn is a function of the RAB bandwidth. As the VoIP data channel studied here is multiplexed together with the SIP and signaling channels in the same RAB, the optimization of these two additional channels should be analyzed before the total gain at the air interface can be derived. Our work in this paper is a first step towards this objective.

Our simulator allows us to input real-life audio clips. In the future we plan to use the simulator to evaluate the resulting audio quality from the various RAB configurations, using e.g. the Mean Opinion Score as voice quality measure.

Acknowledgments

We would like to thank Lindsay Frost for fruitful discussions and helpful comments.

REFERENCES

[1] "3GPP Partnership Project Web page," `http://www.3gpp.org`.

[2] 3GPP TR 34.128, "Common test environments for User Equipment (UE) conformance testing," version 4.7.0, June 2003.

[3] 3GPP TR 25.933, "IP Transport in the UTRAN Work Task Technical Report," version 5.3.0, June 2003.

[4] 3GPP TS 26.074, "AMR Speech codec speech processing functions, AMR Speech Codec Test Sequences," version 5.0.0, June 2002.

[5] S. Casner and V. Jabobson, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links," RFC 2508, February 1999.

[6] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," RFC 1889, January 1996.

[7] 3GPP TR 26.236, "Packet switched conversational multimedia applications; Transport protocols," version 5.3.0, June 2003.

[8] 3GPP TR 26.937, " Transparent end-to-end packet switched streaming service (PSS); Realtime Transport Protocol (RTP) usage model," version 1.5.0, June 2003.

[9] C. Bormann et al, "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed," RFC 3095, July 2001.

[10] "Arrows Project Web page," `http://www.arrows-ist.upc.es`.

[11] J. J. Olmos and J. Perez-Romero, "Simulation tools presentation," `http://www.arrows-ist.upc.es/publications/workshop/UTRA-FDDSimulators.pdf`.

[12] 3GPP TR 25.942, "RF system scenarios," version 6.0.0, January 2003.

[13] "OPNET Simulator," `http://www.opnet.com`.

[14] S. Shenker, C. Partridge, and R. Guerin, "Specification of Guaranteed Quality of Service," RFC 2212, September 1997.