Voice over IP Realized for the 3GPP Long Term Evolution

Fredrik Persson
Ericsson Research
Ericsson AB, SE-164 80 Stockholm, Sweden
fredrik.f.persson@ericsson.com

***Abstract*—The paper outlines voice over IP for the 3GPP long term evolution and investigates the capacity of the service. Some important components affecting capacity are antenna diversity and an efficient radio realization, including concatenation, scheduling, and link adaptation. In the paper, high capacity is achieved by using dynamic scheduling together with receiver diversity and MIMO. However, the capacity could be even higher if not limited by the physical downlink control channel in combination with large and frequent signaling messages. Persistent scheduling in its simplest form cannot match the dynamic scheduling approach, in spite of the lower protocol overhead. This is mainly due to the lack of active link adaptation. Finally, by utilizing concatenation of voice packets over the radio substantial capacity gains are possible derived from both decreased protocol overhead and fewer transmitted signaling messages.**

Keywords- Long term evolution, voice over IP.

I. Introduction

Telephony is, and has been for some time, by far the most important of all personal mobile communication services. Voice is a very efficient way of communicating and for most cellular operators, telephony constitutes an important part of the revenues. Consequently, high capacity is crucial when also launching the corresponding cellular telephony service based on the Internet protocol (IP), here referred to as voice over IP (VoIP). There is good support for IP-based services already in the high speed packet access (HSPA) of the 3rd generation partnership project (3GPP), but the importance of such services will be even higher for the long term evolution (LTE), which is purely packet based. However, challenges are found especially for real-time services like VoIP, not least when it comes to important performance metrics like capacity and coverage.

The paper takes a closer look at service requirements, service mapping and how the VoIP service could be realized for the LTE radio access. The expected voice capacity performance is thereafter shown for some possible implementations, including scheduling, advanced antennas, and potential improvements in the radio protocol area.

II. Service Quality Requirements

The service quality requirements for packet switched telephony are to large extent set by the circuit switch telephony service of today. For example, this means an end-to-end delay of no more than 220 ms in a mobile-to-mobile call, low delay jitter, and a low call drop rate. An operator may also put requirements on coverage and capacity, but these should be evaluated rather from a network deployment point of view. In 3GPP, the adaptive multi rate (AMR) wideband and narrowband codecs are used, with bit rates ranging from 4.75 kbps up to 23.85 kbps. It has further been agreed in 3GPP to allow for a certain grade of redundancy coding [1]. AMR can tolerate a frame loss rate in the range of 1% per radio link (considering a mobile-to-mobile call). Used with the discontinuous transmission (DTX) feature, both comfortable background noise and good capacity are achieved.

III. Service and QoS Mapping

A new quality of service (QoS) concept is being agreed within 3GPP [2]. The session QoS is first negotiated between the peer entities using the session initiation and session description protocols (SIP/SDP) [3][4]. For VoIP this means, among others, agreements on codec specific parameters like AMR codec modes and number of media frames per real-time transport protocol (RTP) packet. In the QoS concept each service data flow is assigned a Label, which is a pointer defining the behavior and forwarding policy in the radio network. Some Labels and Label characteristics will be standardized for interoperability reasons, and VoIP should be mapped to such a Label (even though it still will be an issue for each operator to map services to Labels). VoIP can be mapped to either a guaranteed bit rate (GBR) bearer or a Non-GBR bearer – which to choose to secure the service quality and retainability, as proposed in the previous section, depends on network dimensioning and is ultimately the operator's decision. To allow for efficient rate adaptation, the GBR parameter could be set to the lowest acceptable codec rate, and the maximum bit rate (MBR) parameter to the highest allowed bit rate.

IV. IP Transport

According to [1], the bandwidth efficient AMR payload format is preferred over the octet aligned format, typically adding 10 bits to the RTP payload plus possible padding bits. Headers for RTP, the user datagram protocol (UDP) and IP further increase the packet sizes with 40 bytes, or 60 bytes in case of IPv6. The use of the RTP control protocol (RTCP) is not recommended in [1] for a point-to-point telephony call, and is therefore not considered here.

V. LTE Radio Realization

A. Packet Data Convergence Protocol

At top of the radio protocol stack is the packet data convergence protocol (PDCP) layer [5]. For the user plane PDCP does robust header compression (ROHC) [6] and ciphering. ROHC is especially important for VoIP to make efficient use of the radio resources, but several traffic profiles are defined making it possible to compress practically all user plane flows. The most common and smallest regular compressed header size, at least for continuous flows like voice, is 3 bytes. The outcome of the compressor depends on the traffic behavior and loss patterns, giving varying compressed packet sizes. Ciphering, which prevents unauthorized acquisition of data, is done after the header compression. The protocol overhead to add to each IP packet at PDCP is 8 or 16 bits (16 bits assumed here).

B. Radio Link Control

The radio link control (RLC) protocol layer [7] provides the medium access control (MAC) protocol layer with one packet data unit (PDU) per transport block (and layer, in case of layered transmission). At RLC the following important functions are supported: retransmissions with automatic repeat request (ARQ), segmentation, re-segmentation, and concatenation. The ARQ and the hybrid ARQ (HARQ) at the MAC layer have the same termination points and similar round trip times (RTT), which makes a tight interaction between them possible. Therefore it is interesting to consider ARQ retransmissions even for real-time services like VoIP. Segmentation, re-segmentation (which may be used in case of ARQ retransmissions), and concatenation are important tools to ensure high resource efficiency. Concatenation, which helps reducing the total number of physical layer transmissions, can be utilized to decrease not only the protocol overhead but also the scheduling related signaling load.

C. Medium Access Control

The most important functions of the MAC protocol layer [8] are multiplexing of RLC PDUs, HARQ, and scheduling. RLC PDUs from different radio bearers of the same user can be multiplexed to the same transport block. This can increase the resource efficiency for VoIP, where the small AMR packets do not always fill out the allocated resource blocks entirely. A similar effect can also be obtained with the RLC concatenation. HARQ, in combination with soft combining and also possibly ARQ, is an important and powerful tool to achieve high power utilization. The transmission time interval (TTI) is fixed to 1 ms, giving a HARQ RTT of down to 5 ms. This makes it possible to distribute the power over multiple transmissions without affecting the service end-to-end delay significantly. The short RTT also enables the options to minimize the end-to-end delay or to maximize the time for making channel dependent scheduling decisions.

For the scheduling, two extremes are considered here. For the first extreme, resources are reserved based on the instant need and channel conditions. This is sometimes referred to as *dynamic scheduling*, and is very efficient from a resource utilization point of view. For VoIP, delay can also be utilized in the scheduling decisions to increase both quality and capacity. However, dynamic scheduling requires large amounts of signaling that limit the potential capacity. For the other extreme a fixed resource allocation is done for the entire lifetime of a session, giving similarities to a dedicated channel. This is sometimes referred to as *persistent scheduling*, and solves the signaling load issue. However, the possibility to adapt to changing channel conditions is lost and resources are possibly wasted since they must also be reserved for estimated retransmissions. Further, since no channel adaptation is possible anymore, the same coding and modulation scheme must be selected for transmissions at all possible locations, including the cell edge. Other possible scheduling strategies should end up in between these two extremes.

The total RLC and MAC protocol overhead is not finally defined in 3GPP, but several options have been discussed. In total 40 bits are considered here for the case of one single voice packet per RLC PDU and octet alignment of the total layer 2 protocol overhead. The overhead is slightly different if also considering RLC concatenation or MAC multiplexing.

D. Physical layer

In LTE, the downlink is based on orthogonal frequency division multiplexing (OFDM), and the uplink on single-carrier frequency division multiple access (SC-FDMA). The resources are organized in a time-frequency structure according to [9]. In case of 5 MHz bandwidth, there are 25 resource blocks of 12 sub-carriers each in frequency, and 2 resource blocks of 7 OFDM/SC-FDMA symbols each in time per TTI. The smallest resource to allocate to a user is two resource blocks (one resource block can carry 84 resource elements in total, carrying physical layer signals, control information, and data). The resource granularity limits in some cases the potential resource efficiency and capacity for VoIP, since the voice packets can be very small. The physical layer also provides link adaptation, which is tightly integrated with scheduling and HARQ. It adapts the modulation and coding to the channel conditions and the target number of HARQ retransmissions. The modulation and channel coding rate decides the number of information bits a resource block can carry and consequently, also the number of resource blocks required for transmission of a certain amount of data. This makes link adaptation crucial for achieving high resource efficiency.

In this paper, receiver diversity is obtained in both the uplink and the downlink by antenna diversity with single stream transmission and interference rejection combining (IRC). A multiple-input multiple-output (MIMO) antenna configuration is achieved in the downlink by two antennas and two streams per antenna rate control (PARC). There are also other important functions at the physical layer, such as power control, which are not mentioned further here. Finally, the generic frame structure and normal cyclic prefix are considered and a cyclic redundancy check (CRC) of 24 bits is added to each transport block.

E. Physical signals and out of band control signaling

Several physical control signals are defined in [9]. Downlink reference signals are used for channel estimation and cell identification and consist of known reference symbols (one reference signal transmitted per downlink antenna port). Uplink reference signals are used for channel estimation for coherent demodulation. There are also synchronization signals and the common control physical channel (CCPCH) used for transmission of broadcast information in the downlink, and a random access preamble in the uplink. The overhead assumptions are presented in Table II.

Out of band control signaling is sent in both the uplink and the downlink. In the uplink, the physical uplink control channel (PUCCH) carries HARQ feedback information (positive and negative acknowledgements), channel quality indicator (CQI) reports and scheduling requests (which here are assumed to be dedicated). Two resource blocks in frequency are considered for this type of signaling in case of dynamic scheduling, and one in case of persistent scheduling (no scheduling request needed). In the downlink, the physical downlink control channel (PDCCH) carries downlink resource assignments, uplink scheduling grants, and HARQ feedback information. The resource assignments and scheduling grants are quite large – the sizes are not decided in 3GPP yet but are here considered to be of 46 and 38 bits, respectively. According to 3GPP, the PDCCH is allowed to occupy up to the 3 first out of 14 OFDM symbols in each sub-frame. In this paper 3 symbols are used in case of dynamic scheduling, and only 1 in case of persistent scheduling (basically only used for HARQ feedback).

VI. CAPACITY EVALUATION METHOD AND ASSUMPTIONS

This section presents the evaluation method and assumptions used to evaluate the performance of some of the protocol issues mentioned in the previous section, namely the scheduling, concatenation, and the RLC/MAC protocol overhead.

A static simulation-based evaluation methodology presented in [10] has been used here. In each iteration of the simulation, terminals are randomly positioned in the system area, and the radio channel between each base station and terminal antenna pair is calculated according to the propagation and fading models. A single user is selected independently of channel quality, which models channel-independent time domain scheduling, like e.g. round robin. Further, ideal link adaptation has been assumed. Based on the channel realizations and the active interferers, a signal-to-interference-and-noise ratio (SINR) is calculated for each receive antenna. The SINR values are then finally mapped to active radio link bit rates for each active user. Statistics are collected over a large number of iterations. The method assumes that users are scheduled an equal amount of time. The mean and the 5th percentile of the active radio link bit rates are used as measures of average and cell-edge user quality, respectively (giving 95% coverage). The relation between cell size and capacity is visualized by considering different inter-site distances (ISD). The traffic model assumptions are basically the same as in [10], and the most important ones are summarized in Table I. VoIP traffic and protocol overhead assumptions are according to Table II. Session set-up signaling (SIP/SDP) and other related traffic has been disregarded.

Finally, it should be noted that the results obtained in this paper are strongly dependent on the used models, and that other models may give different results.

TABLE I. SUMMARY OF MODELS AND ASSUMPTIONS

Traffic Model	
User distribution	Uniform
Terminal speed	0 km/h
Data generation	100% system load, see Table II for per user traffic model
Radio Network Models	
Distance attenuation	L = 35.3+37.6·log(d), d = distance in meters
Shadow fading	Log-normal, 8dB standard deviation
Multipath fading	Spatial Channel Model (SCM), Suburban macro
Cell layout	Hexagonal grid, 3-sector sites, 57 sectors in total
Cell radius	167, 333, 577 m (500, 1000, 1732 m ISD)
System Models	
Spectrum allocation	5 MHz for DL and 5 MHz for UL (FDD)
Base UE output power	250 mW into antenna (no minimum power)
Max antenna gain	15dBi
Modulation and coding schemes	QPSK, 16QAM and 64QAM (downlink only), turbo coding according to WCDMA Rel-6
Transmission scheme	DL: Single stream (1×2), 2 stream PARC (2×2) UL: Single stream
Receiver	MMSE with interference rejection combining

TABLE II. VOICE TRAFFIC AND OVERHEAD ASSUMPTIONS

Voice traffic	
Codec rate	12.2 kbps
Voice frames per RTP packet	1 (reference) or 2
DTX – voice activity	50% (SID frames considered)
ROHC compressed RTP/UDP/IP size	24 bits (AMR), 48 bits (SID)
Total PDCP payload size	280 bits (AMR), 104 bits (SID)
Layer 2 protocol overhead	
PDCP	16 bits
RLC/MAC	40 (reference) or 24 bits
Physical layer protocol overhead	
CRC	24 bits
Uplink physical channels and signals	
Reference signals	1 of 7 SC-FDMA symbols
Random access preamble	72 sub-carriers over 1 ms per frame
L1/L2 out of band signaling (PUCCH)	2 resource blocks in frequency
Downlink physical channels and signals	
Reference signals	1/21 (1×2) or 2/21 (2×2) of all resource elements
Synchronization signals	72 sub-carriers over 2 OFDM symbols, twice per frame
CCPCH	72 sub-carriers over 1 slot per frame
L1/L2 out of band signaling (PDCCH)	1 or 3 OFDM symbols

VII. CAPACITY RESULTS

A. Scheduling comparison without antenna diversity

First of all the dynamic and persistent scheduling approaches are compared for a 1×1 antenna configuration. Persistent scheduling requires here 53% less control overhead in the downlink and 15% in the uplink, compared to dynamic scheduling. However, since coverage must be secured without considering link adaptation, more attention must be given to users at the cell edge when using persistent scheduling. The capacity is therefore lower than with dynamic scheduling in both directions, as shown in Figure 1.

B. Scheduling comparison with receiver diversity and MIMO

The capacity results including receiver diversity and MIMO are shown in Figure 2. As expected, receiver diversity and MIMO give a clear advantage capacity-wise. However, at least two important observations can be made here.

Firstly, considering a code rate of 1/3 for the signaling and code rate 1/6 for the HARQ feedback, about 12 resource assignments or 15 scheduling grants can be carried by the PDCCH in total. With 50% voice activity and considering symmetric distribution between active uplink and downlink users, the total average number of simultaneously transmitting voice users is limited to about 250-260 users per link. From Figure 2 it is thus obvious that the maximum PDCCH resource has been overridden in some cases. However, even considering

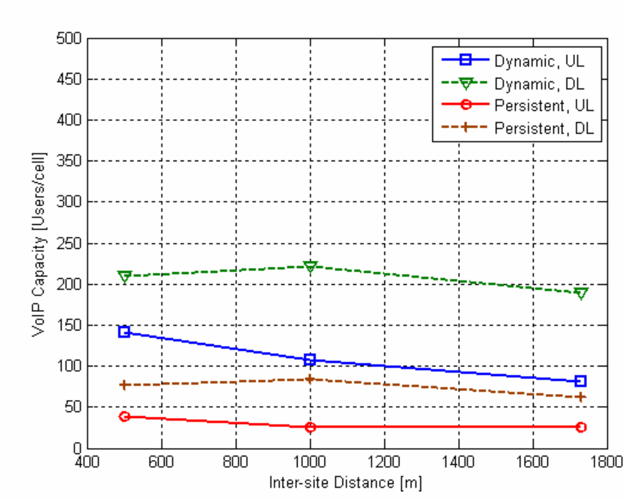

Figure 1. Comparing uplink and downlink capacity for dynamic and persistent scheduling, without antenna diversity.

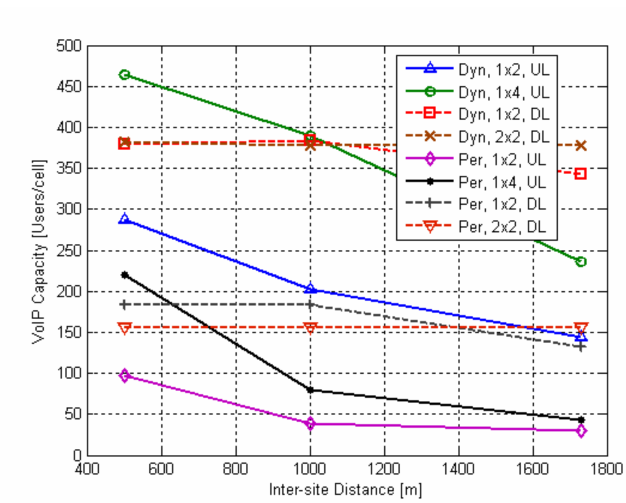

Figure 2. Comparing uplink and downlink capacity for dynamic and persistent scheduling, with receiver diversity and MIMO.

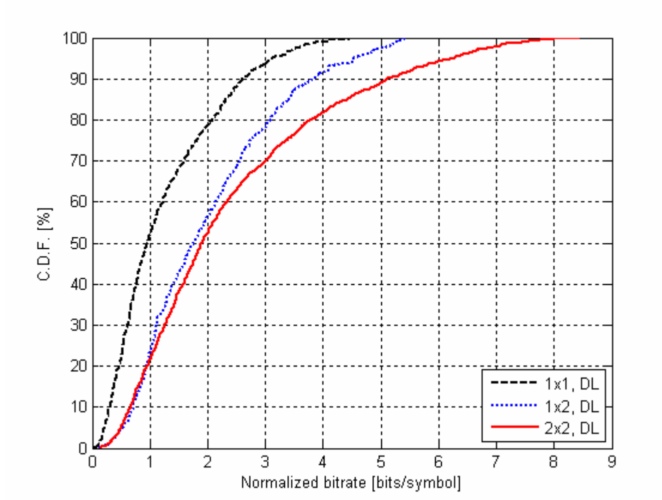

Figure 3. The CDF of the normalized bit rate in the downlink, considering dynamic scheduling and an ISD of 500 m.

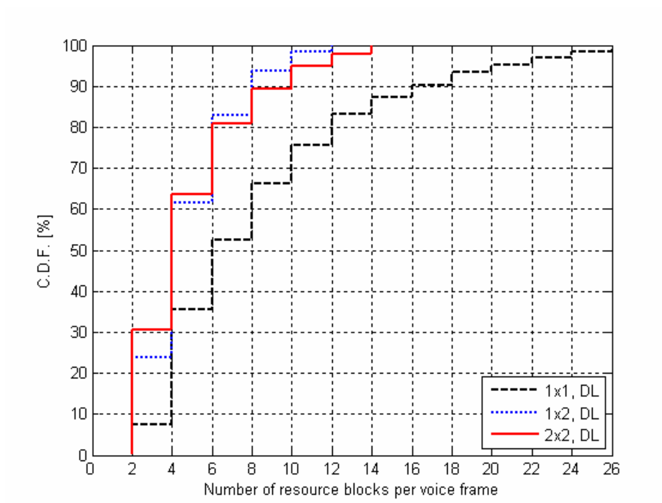

Figure 4. The CDF of number of resource blocks per voice packet in the downlink, considering dynamic scheduling and an ISD of 500 m.

the PDCCH limitation, the capacity will still be higher with dynamic scheduling than with persistent scheduling. The conclusion one can draw is that the gain with antenna diversity is limited with the given assumptions on PDCCH and signaling message sizes, using dynamic scheduling. If not increasing the space for the PDCCH, it is important to limit the sizes of the control messages and/or make sure that an efficient scheduling framework is being agreed on in 3GPP, considering not only low overhead but also the possibility to make channel dependent scheduling decisions and apply link adaptation.

Secondly, when adding MIMO, no additional gain is achieved in the downlink besides receiver diversity. To understand the reasons for this one should consider the normalized bit rate and resource block allocation distributions depicted in Figure 3 and Figure 4, respectively.

In Figure 3 it can be seen that MIMO, as expected, improves the throughput for high SINRs where the normalized bit rate is already high, and also that the gain is decreasing towards zero for lower SINRs. However, looking at the corresponding re source block allocation per voice frame in Figure 4, it can be noted that there is only a small gain for MIMO at the lower percentiles, which correspond to higher SINRs. The explanation for that is found in the total fraction of resource blocks sent at such high SINRs, which is relatively small and therefore does not affect the capacity significantly. The users situated close to the cell edge, on the other hand, occupy a much higher number of resource blocks, which therefore have a decisive effect on the capacity. Another reason to the shown behavior is that the small voice packets already without MIMO occupy the minimum, or close to the minimum, resource allocation per TTI (two resource blocks) at high SINRs. A third reason is that the reference signal overhead is doubled in case of MIMO, compared to the 1×2 case. Note also that even though only one MIMO scheme has been evaluated here, the result should hold in general also for other MIMO schemes.

C. *Overhead reductions with dynamic scheduling*

Two possible ways of reducing the protocol overhead have been evaluated here:

1. Sending two voice packets in one RLC PDU (concatenation). By doing so both the total RLC/MAC overhead and the number of transmissions will be almost halved, which also will decrease the downlink signalling load.
2. Reducing the RLC/MAC protocol headers further. It has here been assumed that the RLC/MAC headers have been reduced from 40 bits down to 24 bits. The improvement could be realized by, e.g., decreasing the size of the length field and by padding gains.

In Figure 5, the capacities for the two options are shown together with the reference case with one voice packet per RLC PDU and 40 bits RLC/MAC overhead, considering dynamic scheduling and receiver diversity.

The largest gains are, as expected, achieved by concatenation. To be able to judge the actual gain it should be noted that the capacity limit related to PDCCH is roughly being doubled in case of two voice packets per RLC PDU, since each user now gives rise to half the number of physical transmissions. With the assumptions given in this paper the actual gain is about 28% in the uplink and as much as 70% in the downlink. If one can manage to avoid the DPCCH limitation, the capacity will increase in general, and the gain of concatenation would be less (about 16-17% in both directions). The potential gain in improving the RLC/MAC overhead by two octets is about 5% in the uplink and 2% in the downlink.

VIII. CONCLUSIONS

In this paper, the realization of VoIP has been outlined for LTE, looking into service requirements, QoS mapping, radio realizations, and a couple of performance evaluations related to scheduling and advanced antennas.

The LTE defines a short TTI of 1 ms, which gives a low RTT. The low RTT enables a low end-to-end delay, an increased scheduling time, or the possibility to utilize HARQ and/or ARQ retransmissions for higher power efficiency. It further opens up for high data throughput and voice user capacity. However, the short TTI also gives challenges in the area of control and protocol overhead.

It can be concluded that scheduling has a central role also in LTE, as was seen already for HSPA. It has been shown that the basic persistent scheduling cannot match the dynamic scheduling approach for VoIP, in spite of the lower protocol overhead. This is mainly due to lack of active link adaptation (giving margins for preserving coverage). It has also been shown that high capacity can be achieved by using dynamic scheduling together with receiver diversity and MIMO, even though the voice capacity was limited with the present assumptions on PDCCH and signaling message sizes. If it is not possible to increase the capacity of the PDCCH, it is important to limit the sizes of the control signaling messages in the standard and/or make sure an efficient scheduling framework is being agreed on in 3GPP. When selecting the scheduling framework, it is also important to not only consider low overhead but also the possibility to make channel dependent scheduling decisions and apply link adaptation.

Besides receiver diversity, MIMO will not enhance the downlink capacity further for VoIP. The reasons are found both in the small fraction of resource blocks transmitted at low SINR, the minimum resource of two resource blocks per user and TTI, and the increased overhead.

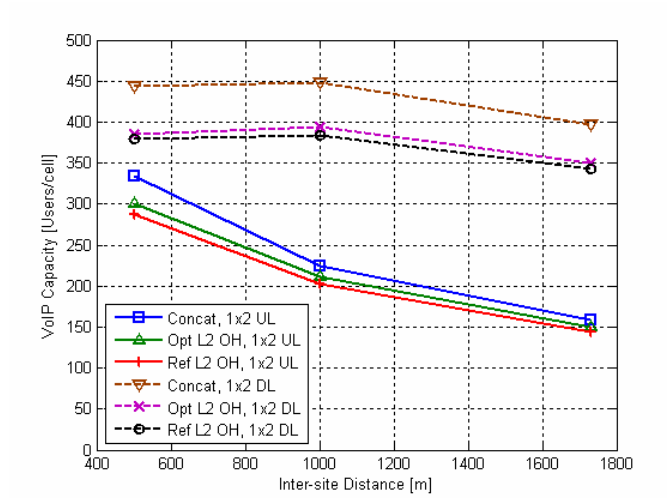

Figure 5. The capacity when considering different layer 2 protocol overhead reductions, with receiver diversity and dynamic scheduling.

One important capacity enhancement is to concatenate two voice frames in each RLC PDU. Capacity gains of up to 28% and 70% were shown for the uplink and the downlink, respectively, when considering receiver diversity. The gain can be derived both from the decreased protocol overhead and less transmitted signaling messages. Reducing the RLC/MAC protocol overhead further only gave a minor contribution to the capacity improvements, here up to about 5%.

ACKNOWLEDGMENT

A special thanks to Dr Anders Furuskär at Ericsson Research for evaluation methodology support and for providing base simulation results.

REFERENCES

[1] 3GPP TS 26.114, "Multimedia Telephony; Media handling and interaction", 2007.

[2] 3GPP TR 23.401, "GPRS enhancements for E-UTRAN access", 2007.

[3] J. Rosenberg, et al., "SIP: Session Inititation Protocol", RFC 3261, June 2002.

[4] M. Handley, V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.

[5] 3GPP TS 36.323, "Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) protocol specification", 2007.

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

[7] 3GPP TS 36.322, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Protocol (RLC) protocol specification", 2007.

[8] 3GPP TS 36.321, "Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Protocol (MAC) protocol specification", 2007.

[9] 3GPP TS 36.211, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation", 2007.

[10] E. Dahlman, et al., "The 3G Long-Term Evolution – Radio Interface Concepts and Performance Evaluation", In Proceedings of VTC 2006 Spring, May 2006.