ROHC and Aggregated VoIP over Fixed WiMAX: An Empirical Evaluation

Esa Piri, Jarno Pinola, Frerk Fitzek and Kostas Pentikousis
VTT Technical Research Centre of Finland
Kaitoväylä 1, FI-90571 Oulu, Finland
firstname.lastname@vtt.fi

Abstract

WiMAX has been at the center of attention in wireless communications during the last years. Nonetheless, very few testbed or field trial measurement accounts have been reported in the peer-reviewed literature. We fill this gap by exploring scenarios where fixed WiMAX is employed for VoIP traffic. VoIP packets typically exhibit large header overheads and small total packet sizes. The actual codec payload per packet is very small compared to the total length of headers appended to each voice frame. RObust Header Compression (ROHC) can significantly decrease header size by capitalizing on static or rarely changing header fields. Aggregating multiple voice frames into one packet is another attractive and effective way to increase application goodput and overall bandwidth utilization. We study the effect of ROHC and application layer aggregation on VoIP performance in a fixed WiMAX testbed consisting of one base station and two subscriber stations. We find that ROHC increases the number of simultaneous bidirectional emulated VoIP flows by 6% when compared to plain VoIP. When aggregation and ROHC are employed in unison, they allow for 86% more flows than standard VoIP to be sustained in our testbed.

1 Introduction

The increased use of voice over IP (VoIP) changes the network traffic composition. Typically, a VoIP packet carries a single audio sample, which contains voice fragments lasting 10-60 ms [1]. Thus, most currently used codecs produce voice frames with sizes in the range of 20-100 bytes. In effect, networks have to carry more smaller packets instead of maximum transmission unit (MTU) sized packets. Moreover, the small size of voice data when compared to the size of RTP, UDP, and IP headers in a VoIP packet translate into low application-layer bandwidth utilization. Especially when employing wireless technologies, where data transport over the air is dearer than the processing power of current mobile devices, a large gap between overall channel utilization and application goodput is undesirable. Thus, the communication/computation tradeoff needs to be revisited.

We are particularly interested in the IEEE 802.16 family of wireless standards (see www.ieee802.org/16), which are an attractive option for Internet service providers interested in offering broadband network connections to a larger set of users, including those living in rural areas. A typical deployment scenario for broadband wireless metropolitan area networks (WMANs) considers point-to-multipoint fixed topologies. In this paper, we evaluate VoIP performance in a point-to-multipoint fixed WMAN testbed where multiple bi-directional VoIP flows compete for resources.

A single voice frame may consist of, for example, 20 bytes of data. After adding the RTP/UDP/IPv4 headers (12/8/20 bytes, respectively) the proportion of actual voice data per packet is only 33%. One way to lower header overhead is to employ voice sample aggregation. Alternatively, RObust Header Compression (ROHC) can decrease the size of protocol headers, which comprise fields that change rarely or never during the duration of a VoIP call. ROHC can be used to compress point-to-point traffic and fits well with fixed WiMAX topologies, increasing application-level channel utilization between two base stations (BS) or a BS and a subscriber station (SS). We quantify the benefits of employing voice frame aggregation and ROHC in practice. Our experiments compare the performance of ROHC and non-ROHC VoIP traffic with application layer aggregation.

The rest of this paper is organized as follows. In Section 2, we relate our work to other ROHC studies. Our testbed, used tools, and evaluation methodology are introduced in Section 3. We present our results in Section 4 and conclude this paper in Section 5, outlining items of future work.

2 Related Work

To the best of our knowledge, there are no publicly released studies of ROHC performance over WiMAX, and thus this paper fills this gap with empirical evidence. Jung et al. [2] studied the benefits of ROHC and packet aggre-

978-1-4244-2703-1/08/$25.00 ©2008 IEEE

gation in wireless mesh networks but used simulation, not testbed measurements, to estimate individual and overall gains in single and multiple source/destination scenarios. They concluded that, although the simulated performance improvements with ROHC and aggregation are remarkable, researchers should also take into consideration the processing capacity required from mesh routers. According to their measurements, it takes approximately 10 - 67 μs/packet to compress and 12 - 51 μs/packet to decompress ROHC packets, depending on the processing power of the router. In this study we do not attempt to quantify the additional processing delays introduced by ROHC compressor and decompressor, as this depends on particular implementations and testbed hardware. Instead, we focus on establishing upper bounds on the performance gains by using ROHC and VoIP aggregation.

Rein et al. [3] evaluated the effect of ROHC on the quality of VoIP for a ROHC implementation. The sending and receiving hosts were connected with Ethernet. They used mean opinion score (MOS) to evaluate the quality of VoIP with different uncorrelated bit error probabilities emulated at the link layer, attempting to model the error process in 3G networks. VoIP traffic consisted of three voice samples encoded by GSM codec (33 octets VoIP frame size). Their study indicates that besides the potential gains in bandwidth and delay savings, ROHC could also improve voice quality.

Despite the great interest in the technology, the majority of WiMAX performance studies are conducted using simulation, mostly because of the lack of inexpensive equipment. One of the few measurements performed with real WiMAX equipment studying VoIP is [4] in which Scalabrino et al. performed measurements with a fixed WiMAX testbed located in Turin, Italy. They evaluated with different emulated VoIP codecs in the presence of best-effort background traffic, the performance of the WiMAX network for VoIP traffic in its saturation region. Unfortunately ROHC or aggregation are not studied in [4]. As the performance comparison between WiMAX and Wi-Fi has also attracted a following we have evaluated their performance with VoIP and IPTV in [5]. The results indicate that Wi-Fi clearly underperforms fixed WiMAX for VoIP. ROHC for Wi-Fi is studied in [6].

3 Methodology

Our testbed (Fig. 1) consists of one Airspan MicroMAX-SoC base station (BS), two Airspan subscriber stations, EasyST (SS1) and ProST (SS2), and three PCs (GNU/Linux Ubuntu 7.04, kernel ver. 2.6.20-16). Two of the PCs are symmetrically connected to SS1 and SS2, respectively, acting as traffic sources and sinks. The third PC runs PTPd [7] which is an easy to use and light weight open source implementation of the IEEE 1588 Precision Time Protocol (PTP) [8]. PTP can synchonize COTS PC clocks with an accuracy

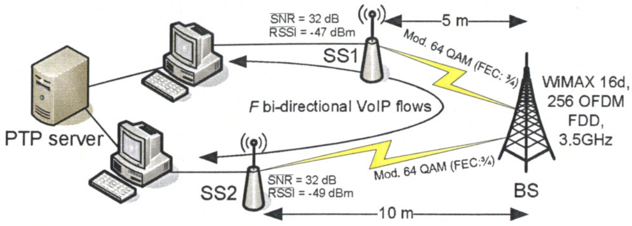

Figure 1. Schematic of the WiMAX testbed

of up to 10 μs, and in ideal cases even more accurately. In our topology, all hosts were synchronized with an accuracy of tens of μs which is close to GPS-based synchronization accuracy. PTP is based on synchronization messages sent over a network. In our case, messages are sent over separate, dedicated Ethernet interfaces on all hosts. No synchronization traffic was transported over the WiMAX link.

We experimented with direct line-of-sight (LOS) conditions for both SS1 and SS2 in a laboratory environment. The key parameters of our testbed during the measurements are shown in Fig. 1. Although laboratory conditions are static by nature, the SS1 downlink FEC modulation (64 QAM) switched between 3/4 and 2/3, albeit rarely. We monitored SS1 for a longer period of time and the alternation between the two FEC modes was random and very short in duration; its effect on our measurements was marginal.

Both PCs connect to their respective SS using 100BaseT Ethernet, which forces us to use VoIP codecs that generate voice frames with sizes more than 20 bytes. Shorter frames are padded by the Ethernet interface to the minimum Ethernet frame size of 64 bytes, including 4 bytes of Frame Check Sequence and 14 bytes of Ethernet header.

3.1 RObust Header Compression

ROHC [9] compresses the headers of RTP, UDP, and IP packets capitalizing on their redundancy. Several fields in these protocol headers change rarely or never during the duration of a flow, such as, source and destination addresses. By removing these fields from most of the messages, exploiting dependencies and the predictability of other fields, the total header size can be reduced significantly depending on the profile and mode used. In order to decompress the ROHC headers the receiving side must keep track of the context of every sender. During decompression, each sender's context includes relevant information from previously received headers and other data describing each packet flow. Static header information is basically sent only during the initialization phase.

A header compression profile is defined for every kind of packet flows over different link types. In our study, we used the RTP profile which defines the methods for compressing RTP/UDP/IP headers. RFC 3095 [9] specifies three different modes for compressors and decompressors: Unidirectional (U), Optimization (O), and Reliable (R). In the O and R modes, communication between compressor and decompressor is bidirectional employing a feedback channel. In the U mode communication is unidirectional, taking an optimistic approach. Due to the lack of feedback channel in the U mode, the compressor sends periodically a packet with an uncompressed header in order to refresh the context.

In our empirical evaluation, we use the ROHC U mode with only the Initialization and Refresh (IR) and Second Order (SO) states. We ignore the First Order (FO) state, in which header information is at least partially compressed.We generate similar packet sending patterns for the IR and SO states to the one Wang and Seah use in [10]. In the IR state, the static part of the context is initialized in a slow-start fashion, where the amount of compressed header packets sent between the IR packets increases by one in a linear manner starting from zero. In the IR state, the sequence indices of uncompressed IR packets can be calculated using Eq. (1), where i is the index of the calculated IR packet and k is the amount of IR packets sent so far. All other indices are for compressed header (CH) packets. We remain in the IR state until every N^{th} sent message is an IR message.

$$i_k = \begin{cases} 0 & \forall k = 0 \\ i_{k-1} + k & \forall 1 \leq k \leq N \end{cases} \quad (1)$$

The total amount of messages sent in the IR state is given by Eq. (2):

$$n_{tot} = 1 + \sum_{k=1}^{N+1} k = 1 + \frac{N * (N+1)}{2} \quad (2)$$

In our study, $N = 30$, which means a total of 465 messages are sent in the IR state. The IR state has a relatively long period, but decreasing N to, say, 10 has only a marginal impact on overall efficiency.

After the IR state, the packet sending pattern moves to the SO state where compression is optimal. In other words, every sent packet is compressed besides periodically sent uncompressed context repetition IR packets. In our case the length of repetition period is N packets. Both IR and SO states are depicted in Fig. 2. The ROHC context size used in IR packets is calculated according to the rules of [9] and is 40 bytes. The context size varies a lot depending on the protocol settings. For the CH packet we used the UO-0 packet format which compresses the RTP/UDP/IP headers including UDP checksum to four bytes only.

3.2 Traffic Generation

When selecting the VoIP codec to be used in the scenarios experimented with in this study, we decided to emulate a more sophisticated and up to date codec than those

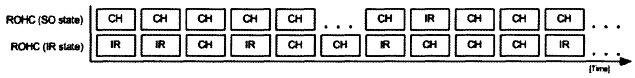

Figure 2. Packet sending patterns for SO and IR ROHC states

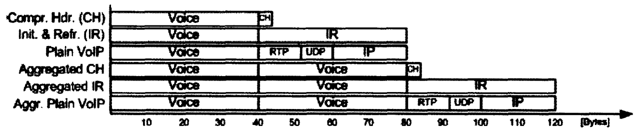

Figure 3. Comparison of packet sizes

used in [2,4,10]. Based on its capabilities as a recent wideband codec with sophisticated layered structure, we chose the ITU-T G.729.1 [11] audio codec. Based on the ITU-T G.729 [12] narrowband codec, G.729.1 serves as an extension to it and provides scalable wideband speech and audio compression capabilities on top of the narrowband coding. This interoperability also makes the G.729.1 codec suitable for seamless transition from narrowband to wideband VoIP. The layered structure of the encoding in G.729.1 gives the codec twelve possible bitrates ranging from 8 kb/s to 32 kb/s. The core layer carrying 20 ms of voice is sent at 8 kb/s and it is fully bitstream compatible with G.729. In fact, only the core layer is needed to successfully decode the received VoIP packets to understandable speech. Extra layers are providing the wideband extension on top of the narrowband layer and improve the received audio quality.

We do not evaluate the adaptive features of G.729.1 in this paper. Instead, we chose one of the operation modes provided by the codec and kept it fixed during each measurement run. The chosen operation mode for the codec was the four-layer coding, which led to codec payload of 40 bytes at a bitrate of 16 kb/s. After adding the RTP header and accounting for the UDP and IP headers, the total network packet size is 80 bytes. When aggregating two voice frames into one packet the total packet size is 120 bytes. The packet sizes used in the measurements are illustrated in Fig. 3. Evaluating the adaptivity features of G.729.1 is left for future study.

Considering the features needed from a traffic generator, we opted to use JTG [13] to inject synthetic VoIP traffic into our testbed. JTG is a simple, versatile and efficient open source traffic generator. JTG can generate both constant bit rate (CBR) and variable bit rate (VBR) traffic. Measurements without ROHC were CBR whereas ROHC measurements were VBR because of the periodically sent IR packets. The accuracy of JTG depends on the GNU/Linux kernel which is not natively a real-time kernel. The packet transmission interval when using the C select() call has, on average, 10 ms accuracy which is sufficient for our measurements. JTG appends UDP/IPv4 (28 bytes) headers to every sent packet and RTP header size (12 bytes) needs to be added to the payload. ROHC compression is taken into account by changing payload size in JTG accordingly.

4 Results

For every measurement run, a different amount of emulated G.729.1 encoded VoIP flows was sent over the WiMAX link and the amount of dropped packets and the one-way delay (OWD) of the received packets was recorded into JTG log files together with the bit rate of the received packet flows. Measurements with particular VoIP flow amount was repeated three times and the median was taken to the presented results. Each of the measurement runs lasted 100 seconds.

Before we conducted our performance measurements with emulated VoIP traffic, we quantified the maximum capacity of the WiMAX link. These baseline measurements were conducted in direct LOS conditions with Ethernet MTU sized (1500 bytes) UDP packets separately to both uplink (UL) and downlink (DL) directions. The maximum achieved goodput with negligible packet loss (<0.1%) was 9.4 Mb/s for the DL and 5.5 Mb/s for the UL.

We experimented with emulated bidirectional VoIP traffic, that is, for each VoIP call two packet flows were generated destined to the opposite direction, symmetrically. Unfortunately, in our equipment only best effort service is fully and correctly implemented. Since our measurements involve bidirectional traffic and no QoS guarantees or traffic differentiation is provided by our WiMAX equipment, the bottleneck of the testbed's performance is the UL, which has lower capacity than the DL. Thus, the goodput results are compared to the maximum performance of the UL direction. The VoIP quality metrics used in this section are chosen after assessing different quality measurement methods and models from [1] for similar codecs. The threshold values for voice frame loss rate (5%) and end-to-end (e2e) OWD (200 ms) should be such, that the resulting received quality of the speech is at an adequate level. However, it is important to bear in mind that the actual voice quality of a real-world VoIP call depends not only on the overall loss rate and the e2e OWD, but also on the particular loss distribution and loss recovery mechanisms involved. Our ongoing work involves further measurements which take interpacket delay and loss into account and deliver a voice quality evaluation using standardized mean opinion scores.

In the following subsections, the performance of the tested equipment under varying loads of emulated VoIP traffic, with and without ROHC and aggregation, is presented. For each case, we pay closer attention to the region around the saturation point of the WiMAX link. This way, the important information is not blurred by trivial extra measurement points. Individual measurements at SS1 and SS2 are presented with two different marks (× and +) and their average is depicted as a dotted line for the measurement runs done without ROHC and with a solid line for those done with ROHC.

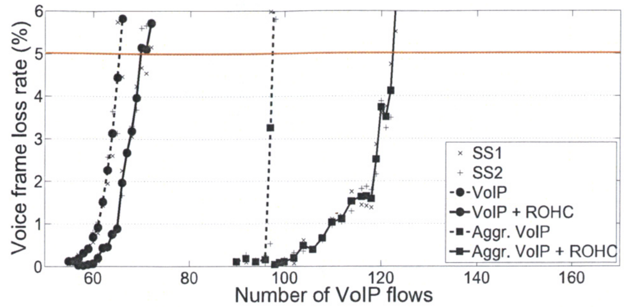

Figure 4. Voice frame loss

4.1 VoIP with and without ROHC

An issue with VoIP when operating over wireless links is the large amount of small-sized packets. The large header overhead diminishes goodput and utilization and, thus, the overall performance of the wireless link. Fig. 4 shows that with plain VoIP packets, each containing one VoIP codec frame, the average voice frame loss rate of the received voice flows crosses the 5% threshold when the number of simultaneous bidirectional VoIP flows, F, exceeds 65. The resulting goodput for $F = 65$ is slightly over 1.35 Mb/s, which is only 24.5% of the baseline UL capacity.

When using ROHC, the overhead is decreased and our WiMAX equipment can handle $F = 69$ flows with voice frame loss rate less than 5%. With ROHC employed the testbed can carry 6.2% more VoIP calls. However, Fig. 5 illustrates that when ROHC is used, the e2e OWD grows more moderately as F, the number of simultaneous VoIP conversations, increases. The average OWD remains under 200 ms for the plain VoIP packets only when $F < 66$. When ROHC is used, OWD stays under 80 ms even when $F = 72$. At this point, the voice frame loss rate is almost 6%.

4.2 VoIP Aggregation and ROHC

Another way to decrease overhead for small packets is to aggregate a number of voice frames into one (larger) packet and send it to the network with one set of headers [14]. For VoIP applications aggregation can be done to the encoded voice frames before they are packetized and handed down to the transport layer (RTP). However, when an aggregated packet is lost during transmission, all the consecutive voice samples included in it are lost at once. This can lead to

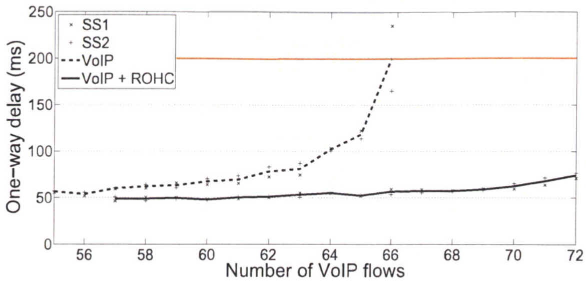

Figure 5. VoIP e2e OWD with & without ROHC

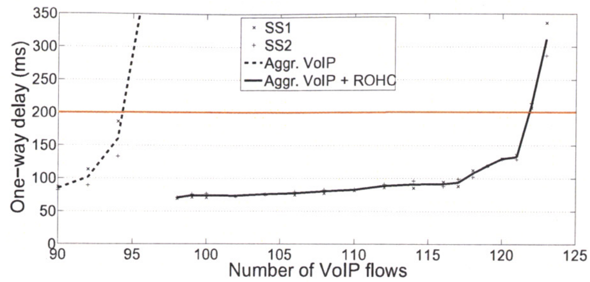

Figure 6. Aggregated VoIP e2e OWD with & without ROHC

severe quality deterioration when layered VoIP codecs are used. Moreover, the effect of additional delays encountered by the voice samples waiting in the sending buffer for the subsequent voice samples to be aggregated in the same packet, must be taken into account. For this study, we emulated the application layer VoIP frame aggregation by putting two codec payloads into one packet. This is a tradeoff where the positive effects of the decreased network overhead are clearly measurable while the delay induced by the aggregation scheme is at a reasonable level.

The frame loss rates with packets containing two voice frames each, can be seen in Fig. 4. The performance gains are remarkable. The WiMAX link can now sustain $F = 97$ bidirectional VoIP flows with average voice frame loss rate under 5%. On the other hand, the usage of aggregation of two G.729.1 encoded voice frames into one packet leads to an extra e2e delay of 20 ms for the first voice frame, while it is queued for the second frame to be sampled. The delays depicted in Fig. 6 are measured for this first voice frame and thus are illustrating the maximum average e2e OWD experienced by the speech samples in one VoIP packet. The figure shows, that OWD exceeds 200 ms already when $F > 94$. In this case, the e2e OWD delay is the limiting factor, not loss (which allows for $F = 97$ emulated VoIP calls), as the quality of the conversation would be unacceptably poor due to the large delays. In short, when accounting for both voice frame loss and OWD, aggregation allows our testbed to carry 44.6% ($F = 94$) more emulated calls. The voice frame loss for aggregated VoIP increases less gracefully than for plain VoIP. It jumps from almost zero to over 3%, when the number of flows is increased from 96 to 97 and then another 7%, when $F = 98$. This kind of behavior is usually encountered in association with buffer drop tail events. With aggregated VoIP, two voice frames are lost for each packet dropped resulting in higher frame loss rates.

Fig. 4 shows evidence that the use of ROHC alongside sample aggregation boosts performance substantially. With aggregated VoIP and ROHC, the system can sustain $F = 122$ emulated VoIP calls with voice frame loss rate less than 5%. Nonetheless, when OWD is also considered

Table 1. Number of simultaneous emulated G.729.1 VoIP calls (F). Voice sample loss less than 5%; e2e OWD less than 200 ms. The relative gain is shown in parentheses.

Voice frames per packet	Standard Headers		Employing ROHC	
1	65		69	(+6.2%)
2	94	(+44.6%)	121	(+86.1%)

the aggregated VoIP packets using ROHC record delays exceeding 200 ms when $F > 121$. Thus, when ROHC is employed for aggregated VoIP our testbed can handle 30.1% more emulated simultaneous calls than when only aggregation is used. Overall, aggregation and ROHC allow our testbed to sustain 86.1% more emulated simultaneous VoIP calls with e2e OWD less than 200 ms and loss less than 5%. This is summarized in Table 1. In effect, the UL utilization nearly doubles.

When the performance gains of ROHC and aggregation are compared, we note a striking difference between the cases. ROHC allows for only 6.2% more emulated calls when aggregation is not employed. But when aggregation is used, ROHC allows for 30.1% more calls. This is an intriguing result. More analysis and validation via simulation is needed but at this stage we conjecture the interarrival process of aggregated VoIP packets can be handled better by the tested WiMAX equipment than the process governing plain VoIP. The equipment buffer management algorithm is unknown. Verification of this conjecture is part of our ongoing work.

Finally, based on the measurements reported by Grondalen et al. [15], the results presented in this paper might be applicable, to a certain extent, to outdoor environments and for considerably longer BS-SS distances (even up to 5 km), even though our measurements were done in laboratory conditions. Of course, this requires verification with real outdoor measurements in the future.

5 Conclusion and Future Work

We empirically evaluated the impact of employing ROHC for VoIP on a fixed WiMAX testbed consisting of two SSs and one BS. Our results indicate that ROHC can improve performance in terms of packet loss rate and one-way delay allowing for more simultaneous emulated VoIP calls. When considering the packet loss rate for standard VoIP, we managed to get an increase of four VoIP calls with ROHC to the total amount of simultaneous uncompressed VoIP flows. When aggregation is used, the difference is even bigger. ROHC solely increases the number of simultaneous flows by 6.2% compared to plain VoIP. If voice sample aggregation and ROHC are employed, 86.1% more emulated calls can be sustained with less than 5% loss and less than 200 ms e2e OWD. Notably, the delay increase when aggregating two voice frames is, in most cases, not detrimental. The biggest impact of ROHC seems to be on OWD. With standard headers the delay correlates greatly with the drop rate function whereas with ROHC OWD stays within tolerable levels even when drop rate threshold is exceeded.

We injected synthetic VoIP traffic into our fixed WiMAX testbed, emulating the ROHC compressor and ignoring its processing overhead. The first item on our future work agenda is to repeat our measurements using real ROHC implementations. Of course, we are interested in further analysis and validation of our results in other testbeds and via simulation and, in particular, experimenting with different queue management mechanisms. Finally we are looking forward to evaluating ROHC with IEEE 802.16e equipment in the near future.

Acknowledgement

This work was carried out in the framework of the IST FP6 Integrated Project WEIRD (IST-034622), which is partially funded by the Commission of the European Union. The views expressed in this paper are solely those of the authors and do not necessarily represent the views of VTT, WEIRD, or the Commission of the European Union.

References

[1] A. Raake. *Speech Quality of VoIP : Assessment and Prediction.* John Wiley & Sons Ltd., 2006.

[2] S. Hong S. Jung and P. Park. Effect of RObust Header Compression (ROHC) and Packet Aggregation on Multi-hop Wireless Mesh Networks. In *Proc. IEEE CIT'06*, Seoul, Korea, September 2006.

[3] S. Rein, F.H.P. Fitzek, and M. Reisslein. Voice Quality Evaluation for Wireless Transmission with ROHC. In *Proc. IMSA*, pages 461–466, Honolulu, USA, August 2003.

[4] N. Scalabrino, F. De Pellegrini, R. Riggio, A. Maestrini, C. Costa, and I. Chlamtac. Measuring the quality of VoIP traffic on a WiMAX testbed. In *Proc. TRIDENTCOM*, pages 1–10, May 2007.

[5] K. Pentikousis, J. Pinola, E. Piri, and F. Fitzek. A measurement study of Speex VoIP and H.264/AVC video over IEEE 802.16d and IEEE 802.11g. In *Proc. MediaWiN*, July 2008.

[6] R. Pries, A. Mäder, and D. Staehle. Performance of header compression for VoIP in wireless LANs. Technical report, University of Würzburg, Institute of Computer Science, April 2007.

[7] K. Correll, N. Barendt, and M. Branicky. Design Considerations for Software Only Implementations of the IEEE 1588 Precision Time Protocol. Proc. Conference on IEEE 1588, October 2005.

[8] International Electrotechnical Commission. *Precision clock synchronization protocol for networked measurement and control systems.* Reference numbers: IEC 61588:2004(E), IEEE 1588-2002(E), September 2004.

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

[10] H. Wang and K.G. Seah. An analytical model for the ROHC RTP profile. In *Proc. IEEE WCNC'04, vol. 1.*, pages 126 – 131, March 2004.

[11] *G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729.* ITU-T Recommendation G.729.1, 2006.

[12] *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP).* ITU-T Recommendation G.729, 1996.

[13] J. Manner. *Jugi's Traffic Generator (JTG).* Available: http://hoslab.cs.helsinki.fi/savane/projects/jtg.

[14] K. Pentikousis, E. Piri, J. Pinola, F. Fitzek, T. Nissilä, and I. Harjula. Empirical evaluation of VoIP aggregation over a fixed WiMAX testbed. In *Proc. TRIDENTCOM*, Innsbruck, Austria, March 2008.

[15] O. Grondalen, P. Gronsund, T. Breivik, and P. Engelstad. Fixed WiMAX field trial measurements and analyses. In *Proc. 16th IST Mobile and Wireless Communications Summit*, pages 1–5, July 2007.