

Empirical Evaluation of VoIP Aggregation over a Fixed WiMAX Testbed

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

ABSTRACT

The WiMAX Reference Network Architecture can be used in point-to-point and point-to-multipoint network topologies, and is suitable for providing last-mile, building-to-building, and residential broadband connectivity. Another major application, and the main focus of this study, is the use of fixed WiMAX as backhaul for voice and data services. We evaluate voice over IP (VoIP) performance over a fixed WiMAX testbed and quantify the benefits from employing application- and network-level aggregation. We examine such aggregation schemes using our fixed WiMAX testbed and report the results for both uplink and downlink. If we use objective mean opinion scores (MOS) as the main gauge of overall performance, application-layer aggregation appears to be the best scheme, allowing our fixed WiMAX testbed to sustain nearly three times more flows in the downlink and over two times more flows in the uplink than when no aggregation is used, at comparable MOS values.

Categories and Subject Descriptors

C.2.5 [Computer-Communication Networks]: Local and Wide-Area Networks; C.4 [Performance of Systems]: Measurement techniques

General Terms

Experimentation, Measurement, Performance

Keywords

Network Measurements, WiMAX, IEEE 802.16, WiMAX testbeds, VoIP, VoIP Aggregation, G.723.1

1. INTRODUCTION

The IEEE 802.16 family of wireless local and metropolitan area network standards (see [11,12] and www.ieee802.org/16) and its potential to change the field of telecommunications

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Tridentcom 2008 March 18–20, 2008, Innsbruck, Austria
Copyright 2008 ACM 978-1-60558-009-8 ...\$5.00.

operations and business models has attracted significant attention during the last years. Along with the WiMAX Forum extensions (see [2] and www.wimaxforum.org), this wireless LAN/MAN technology emerges as a promising proposal for building next-generation wireless networks. The WiMAX Forum has been developing a Reference Network Architecture [25,26] which can be used in point-to-point and point-to-multipoint network topologies, and is suitable for providing last-mile, building-to-building, and residential broadband connectivity [9]. Outdoor deployment is of central interest, but solely indoor use is also foreseen. Another major application, and the main focus of this study, is the use of fixed WiMAX as backhaul for voice and data services.

Key vendors and major telecommunication operators have joined the WiMAX Forum, in an effort to establish concrete standards, and facilitate and certify equipment compatibility. In the December 2007 IEEE Spectrum Insider Webinar, Christian John of Intel anticipated that in only two years' time (2010), WiMAX operators will cover areas inhabited by 650 million people. By 2012, WiMAX may reach 1.3 billion people. Arguably, all of them could become WiMAX users. According to Maravedis, a market analysis firm, by the end of September 2007 there were less than 1.4 million WiMAX subscribers worldwide [18]. Interestingly, 55% of all customer premises equipment (CPE) are based on proprietary solutions (such as Motorola Canopy), 29% are IEEE 802.16-2004 [11] compatible, and less than 12% are IEEE 802.16-2005 [12] compliant; 3.8% of all CPEs were not disclosed [18]. Maravedis notes an almost three-fold rate of subscriber growth between January 2006 and September 2007, with Europe leading the way with some 38% of worldwide deployments. Clearly, some vendors at least, expect an explosive growth in WiMAX adoption.

Nevertheless, despite the buzz, there is very little data on what can actually be done with WiMAX in practice today. There are, of course, several vendor reports and demonstrations where capacities exceeding 100 Mb/s are highlighted. Yet the CPEs and base stations (BSs) that are currently available deliver an order of magnitude less application-level throughput (also called goodput). In fact, as we will see in Section 2, there is only a handful of empirical studies of with actual WiMAX equipment in the peer-reviewed literature. To this extend, this paper, which reports measurements from a fixed WiMAX testbed in Oulu, Finland, adds to our knowledge about real-world WiMAX performance using off-the-shelf hardware.

In particular, we evaluate voice over IP (VoIP) performance over fixed WiMAX using synthetic traffic genera-

tion. We quantify both uplink and downlink performance for G.723.1 [13] emulated VoIP traffic in terms of cumulative goodput, packet loss, and mean opinion scores, based on the ITU-T E Model [14,24]. Moreover, we empirically investigate the benefits from employing application- and network-level VoIP aggregation when using fixed WiMAX as a backhaul. Aggregation appears as a promising approach for increasing the overall network efficiency and resource utilization. Hoene et al. [10], for example, find that it is better to change the packet rate rather than the coding rate when a VoIP flow encounters limited capacity links. There have been several proposals about aggregating VoIP flows for different wireless technologies ranging from WLAN to cellular networks. We put these aggregation schemes to the test in our fixed WiMAX testbed and compare their performance with non-aggregated VoIP. We find that aggregation can more than double the VoIP backhaul capacity of our fixed WiMAX link in terms of flows.

This paper is organized as follows. Section 2 compares and contrasts our work with previously published results. Then, Section 3 presents our experimental facility, provides some background on WiMAX and discusses current developments in this area. Section 4 explains the methodology adopted in our study. Sections 5 and 6 present and discuss our results and outline future research items. Finally, Section 7 concludes this paper.

2. RELATED WORK

Pioneering and closely related work to ours has been recently published by Scalabrino et al. [21,22]. They report results from a fixed WiMAX testbed in Turin, Italy. The testbed comprises three CPEs and one BS (Alvarion BreezeMAX equipment) and is deployed outdoors. The CPE-BS distance is between 8.5 and 13.7 km. Scalabrino et al. focus on the QoS aspects of their WiMAX testbed and evaluate VoIP performance over WiMAX in particular when service differentiation is employed. They generate synthetic VoIP traffic and report throughput, loss and R-scores, but observe that in their setting the downlink is the bottleneck. Moreover, since they focus on QoS differentiation, they examine how many VoIP flows can be sustained if the traffic class committed information rate (CIR) for VoIP is set to only 300 kbits/s. We argue that evaluating VoIP with a best effort WiMAX profile is more realistic as most VoIP applications typically do not (attempt to) take advantage of differentiated services, and their flows are effectively appearing as best-effort sessions. In addition, due to the limited CIR, their results involve an order of magnitude fewer flows than in our evaluation. Finally, the benefits of VoIP aggregation are not studied in [21,22].

Grondalen et al. [8] report on fixed WiMAX field trial results from 15 different locations in the metropolitan Oslo, Norway region. They measure RSSI and SNR values and compare them with the maximum attainable FTP/TCP and UDP throughput with both line- and non-line-of-sight conditions. They report that their WiMAX deployment with the same modulation and FEC as ours can deliver a throughput of up to 9.6 Mbits/s when a single (greedy) flow is using the downlink. They do not study VoIP performance, let alone different aggregation schemes. However, their results indicate that this throughput level is possible to attain at a distance up to 5 km from the BS so, to some degree, our results (see Section 5) should be applicable to outdoor environ-

Table 1: Testbed Configuration

Base station	Airspan MicroMAX-SoC
Subscriber station	Airspan ProST
MAC scheduling	Best Effort
PHY	WiMAX 16d, 256 OFDM FDD
Modulation (DL & UL)	64QAM (FEC: 3/4)
Frequency band	3.5 GHz
Channel bandwidth	3.5 MHz
BS and SS Tx power	1.0 dBm
Link condition	Direct line-of-sight
BS-SS distance	10 m

ments. This, of course, needs to be verified in future work. Grondalen et al. do not mention which vendor equipment they employ, but do note that the manufacturer’s manual lists the maximum nominal bitrate as 12.71 Mbits/s.

Neves et al. [19] discuss the merits of using WiMAX for emergency services, such as environmental monitoring and fire prevention. They describe their testbed, deployed in the mountainous area near Coimbra, Portugal. After planning the site using simulation, they perform measurements for their video surveillance application and report throughput of 1.1 Mbits/s and 1.3 Mbits/s, for uplink and downlink, respectively. For this link, the BS-SS distance was 22.78 km, with direct line-of-sight (LOS), and a channel of 3.5 MHz. They also measured throughput for another link with BS-SS distance was 18.5 km LOS, and same channel bandwidth. They report 2 Mbits/s and 2.4 Mbits/s for uplink and downlink, respectively. Neves et al. [19] do not study VoIP performance in their testbed.

Ng et al. [20] discuss their “multiplex-multicast” VoIP aggregation scheme and evaluate it over wireless LANs but do not report actual results from a fixed WiMAX testbed. In addition, they only consider a small number of flows and do not consider application-level aggregation. Nonetheless this is very interesting work, which gave us several ideas about how to continue our work in the future.

There are several papers that examine WiMAX- and VoIP-related topics using simulation and modeling or analytical tools. Due to space restrictions we shall confine ourselves to the most recent results and simply sample a few on VoIP aggregation: Sengupta et al. [5,24], demonstrate the benefits of VoIP aggregation for WiMAX using simulation; Komolafe and Gardner [16] explore aggregation from a teletraffic approach focusing on 3G and mobility scenarios; and Kim et al. [15] study different VoIP aggregation algorithms

3. TESTBED AND BACKGROUND

Fig. 1 illustrates the experimental facility used in our empirical performance evaluation study. The testbed includes an Airspan MicroMAX-SoC fixed WiMAX base station (BS) and an Airspan ProST subscriber station (SS). The testbed employs Frequency Division Duplexing (FDD) in the 3.5 GHz band. The transmitting power of the BS during the experiments was set to 1 dBm, and the BS-SS distance was 10 m. The BS uses a 60° directional antenna, see Table 1 for more details.

As shown in Fig. 1, we connect GNU/Linux (kernel ver. 2.6.20-16, Ubuntu 7.04) PCs with Broadcom NetXtreme

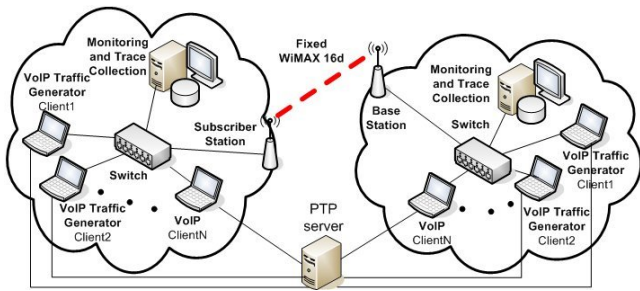


Figure 1: Schematic of our WiMAX testbed

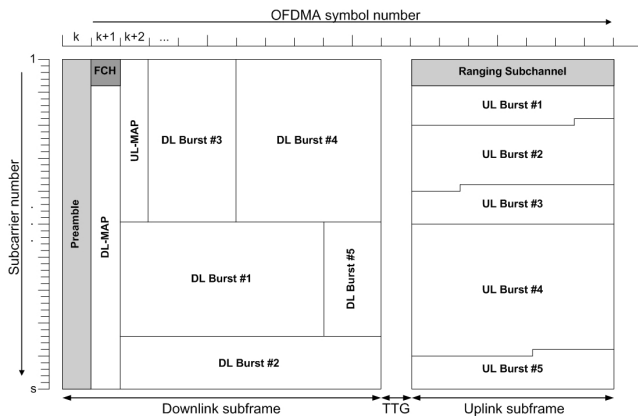


Figure 2: An example of an OFDMA frame

BCM5754 1 Gb/s Ethernet PCI cards symmetrically on the BS and SS side. Effectively we define a classic dumbbell topology with a fixed WiMAX link as the bottleneck. The tests were performed in laboratory, direct line-of-sight (LOS) conditions. The conditions are assumed to be relatively static, even though there may be some light variations in a wireless link even in laboratory conditions. This lab-based setup is aimed at determining the upper bound on what can be achieved in practice with commodity, off-the-shelf fixed WiMAX equipment.

In the WiMAX systems based on Orthogonal Frequency Division Multiple Access (OFDMA) as specified in the IEEE 802.16 standards [11, 12], the generation of both the downlink (DL) and uplink (UL) frames is controlled by the BS MAC. As can be seen in Fig. 2, the transmission of the OFDMA downlink frame begins with the *DL_MAP* and *UL_MAP* right after the preamble and frame control header (FCH). The *DL_MAP* is generated by the BS MAC, and it contains the information of the downlink frame structure and the specifications for constructing the downlink bursts. These can contain either control information or payload data intended for different users. The uplink frame (see Fig. 2) does not bear any additional sections besides the UL bursts and the ranging subchannel intended for closed-loop time, frequency and power adjustment, as well as bandwidth requests. Thus, the allocation of the uplink data is controlled by the BS MAC and communicated to the SS via the *UL_MAP* message.

Naturally, this kind of frame construction mechanism generates some challenges for the WiMAX physical layer (PHY). In order to be able to transmit in a given slot, the SS must

be perfectly synchronized with the other SS:s, or otherwise the orthogonality of the UL frame will be destroyed and the reception of the UL frame is not possible at the BS. Therefore, the BS will provide feedback to the SSs, telling them how to adjust their synchronization.

In order to allow different, more advanced techniques to be used in the downlink transmission, the DL frame can be divided into zones. The first zone, always present in the transmission and illustrated in Fig. 2, uses only single antenna transmission and techniques that can be received by all the SSs. The other, optional zones (not shown in Fig. 2) can be used for space-time coded signaling, adaptive antenna systems, adaptive modulation and coding, and advanced frequency allocation strategies.

Before commencing and during the measurements we employed PTPd [7] which is a light-weight, open source implementation of IEEE 1588 Precision Time Protocol [1] to synchronize the COTS PCs of our testbed. High precision time synchronization of the end hosts in the magnitude of tens of microseconds was achieved during the measurements. This compares with an average round trip time (RTT) which exceeds 40 ms in the WiMAX link. Note that we used two network cards for each PC, so that the clock synchronization traffic and the traffic under measurement did not interfere (as shown in Fig. 1).

We opted to use JTG [17], an open source, simple, flexible and configurable network traffic generator to inject synthetic VoIP traffic in our testbed topology. JTG, unlike other more sophisticated traffic generators, such as, for example, D-ITG [4], does not come with pre-build traffic models. However, it can be easily integrated in custom-made scripts since it can be invoked in a command-line fashion. JTG allows us to generate different traffic patterns on demand. Packet generation is configured by setting transmission rates and packet sizes, by providing trace files for replaying traffic, for example, or defining arbitrary traffic patterns. In this paper, we generate synthetic VoIP traffic based on the output of G.723.1 codec [13] as explained in the following section.

4. METHODOLOGY

Before proceeding with the measurements reported in Section 5, we conducted baseline experiments to determine the maximum throughput that can be attained in our lab testbed. We saturated the fixed WiMAX link and measured the maximum application-level throughput, also called goodput, on the testbed downlink and uplink, i.e., from the BS to SS and vice-versa, respectively. We experimented with various maximum transmission units (MTU), and obtained the best results for application payloads of 1472 bytes (MTU = 1500 bytes, the recommended MTU size for IEEE 802.16 standard-compliant equipment [11]). For the uplink, the average maximum measured goodput was 5.515 Mb/s; for the downlink 9.395 Mb/s. We also measured goodput using TCP with a maximum segment size of 1460 bytes (MTU = 1500 bytes after adding the TCP and IPv4 headers), and recorded 5.417 Mb/s and 9.213 Mb/s for the uplink and downlink, respectively. These measurements serve as a reference as we evaluate different VoIP aggregation schemes in the remainder of the paper.

4.1 Traffic Generation

For this empirical study, we generate synthetic traffic load based on the ITU-T G.723.1 codec [13]. This is a low-rate



Figure 3: G.723.1 sample application-layer aggregation

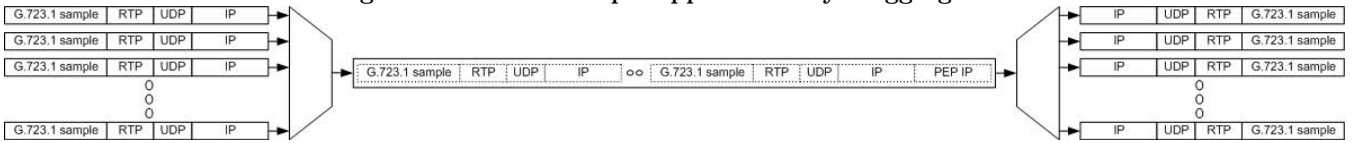


Figure 4: VoIP flow aggregation at the network layer

codec, for real time encoding and decoding, and with good voice quality. A single G.723.1 frame can be 24 octets, which translates into an application-only transmission rate of 6.3 kb/s, or 20 octets (5.3 kb/s), or, finally, 4 octets only. The latter type of payloads corresponds to Silence Insertion Descriptor (SID) frames, used to specify comfort noise parameters, whereas the former two are used to transmit actual voice data. We do not consider SID frames in this study. The higher bitrate variant offers better voice quality compared to the lower bitrate one, and is our chosen variant for these tests. Note, however, that it is possible to switch between the two rates on the fly in state of the art VoIP applications. Studying the dynamic rate adaptation features of the codec is beyond the scope of this paper.

Typically, a single G.723.1 sample is encapsulated in an RTP [23] packet and sent over UDP/IP. When the RTP, UDP, and IPv4 standard headers, consisting of 12, 8, and 20 bytes, respectively, are appended to the 24 octets codec payload, the total actual packet size is 64 bytes. This raises the overall VoIP transmission rate to 17 kb/s. In our setup, a single VoIP flow generated by JTG injects a 64-byte packet (carrying a single audio sample) every 30 ms. This constant bit rate (CBR) traffic constitutes what we call “non-aggregated VoIP flow” in this paper. The G.723.1 small frame size and low bandwidth requirement provides a very good benchmark for investigating the gains of VoIP aggregation.

We inject multiple parallel traffic streams using scripts we developed so that testbed hosts can act as VoIP senders and receivers. A single VoIP flow is managed by a single JTG instance running at its own port, so that traces for each stream can be post-processed in detail. Although it is possible to use one JTG instance for receiving more than one stream, it is not possible to separate the logs in a straightforward manner and be able to accurately apportion loss and delay to each flow. Note that the JTG transmission accuracy depends on the accuracy of the system clock, in our case managed by the GNU/Linux kernel, which is not a (natively) real-time kernel. In our experiments we used `select()` calls for sending packets. This yields a granularity of 10 ms, which is sufficient for the case studies we examine in this paper.

Fig. 3 illustrates application-layer VoIP aggregation. Here, a single RTP packet carries more than one G.723.1 samples, which is explicitly allowed by the RTP specification. In our tests, we experimented with two levels of application-level aggregation carrying two or three samples in a single RTP packet. For “L1 aggregation”, an 88-byte packet is generated every 60 ms. This means that every other sample is buffered at the application for 30 ms. When the next sam-

ple is available, it is packetized using RTP along with the preceding sample, and is sent over UDP. In the case where three G.723.1 samples are aggregated (“L2 aggregation”), a 112-byte packet is generated every 90 ms. Although we experimented with higher aggregation rates, we do not see particular value in considering them for realistic applications as the buffering delays for some of the audio samples become excessive.

We also experimented with network-based VoIP flow aggregation using a VoIP performance enhancing proxy (PEP), inserted at the two ends of the fixed WiMAX link. This PEP can put together two or more complete VoIP-carrying packets from *multiple* flows, as shown in Fig. 4, possibly in coordination with a router. Although one can argue that this can be achieved at a lower layer (see, for example, [24]), in our testbed, with no access to the hardware, the only possible way to reliably and realistically measure aggregation below the application layer is via a software-based solution in the network. In the case where aggregation can be performed at a lower layer, the additional PEP-introduced IP header is not needed because the packet is transmitted over a single hop only, from the BS to the SS, and significant performance benefits may be possible. Unfortunately this is beyond the capabilities of our lab testbed and, thus, the scope of this paper.

4.2 Performance Metrics

For the purposes of this study we consider three main metrics: (a) cumulative goodput, (b) packet and audio sample loss, and (c) the objective mean opinion score (MOS). MOS expresses the quality of the VoIP transmission in a single number in the 1-5 range (1: worst quality; 5: best quality).

We calculate MOS values based on the well-known transmission rating factor R , or “R-Score” [24], which is derived from packet loss and delay measurements as follows [14]:

$$R = R_0 - I_s - I_d - I_e + A \quad (1)$$

where R_0 represents the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise. I_s captures the effect of impairments to the voice signal, and I_d factors in impairments due to delays. I_e takes into account the effects caused by using low-bitrate codecs. The expectation factor A compensates the impairment in the case that the user has access to alternatives. In conventional (wired) environments A is equal to 0. Besides A , the rest of the factors can be subdivided if needed. In our case there is no need to so, as we assume default values for speech transmission. Therefore, we simplify Eq. (1) as follows

$$R = 94.2 - I_d(d) - I_e(c, l) \quad (2)$$

I_e is a function of the used codec (c), and the loss rate (l)

$$I_e(c, l) = \gamma_1 + \gamma_2 \cdot \log(1 + \gamma_3 \cdot l)$$

where $\gamma_1, \gamma_2, \gamma_3$ are specified from the codec. For G.723.1, we have $\gamma_1 = 15$, $\gamma_2 = 90$, and $\gamma_3 = 0.05$ [22].

I_d depends on the delays (d) [14]:

$$I_d(d) = 0.024d + 0.11(d - 177.3)H(d - 177.3)$$

$H(x)$ is the step function ($H(x) = 1$ if, $x \geq 0$ and 0 otherwise.) [6].

The ITU E-Model also specifies the non-linear mapping to the (listening) MOS:

$$MOS = 1 + 0.035R + 7 \cdot 10^{-6} R(R - 60)(100 - R) \quad (3)$$

5. RESULTS

As mentioned earlier, the main goal of this study is to separately quantify the maximum downlink and uplink performance of the tested AirSpan WiMAX equipment for G.723.1 encoded VoIP data traffic in a laboratory environment. We followed a “stress test” approach. That is, we started with a small traffic load yielding no losses and excellent MOS, and increased gradually the offered load measuring the gradual degradation. We continued well beyond the maximum capacity of the tested link in order to see how the system behaves when its maximum capacity is exceeded. As a result from the measurements, the amount of G.723.1 encoded VoIP flows the link can support without seriously decreasing the received voice quality, was identified. Each of the test runs lasted 100 seconds.

5.1 Cumulative Goodput

The small size of single-sample G.723.1 VoIP packets has a substantial effect on the cumulative goodput and the number of flows that can be supported on the downlink and uplink of our WiMAX testbed. Fig. 5 shows that the maximum downlink goodput for non-aggregated VoIP traffic is no more than 3.3 Mbits/s. In uplink, the goodput saturation point is reached already at 1.5 Mbits/s, as can be seen in Fig. 6. This very poor performance is mainly attributed to the overhead introduced by the RTP, UDP and IP headers, when compared to the size of the actual VoIP codec payload. Indeed, every 64 byte VoIP packet carries only 24 bytes of application payload (total overhead of 167%).

When application-layer aggregation is used, the negative effects of, on the one hand, excessive header overhead, and on the other, the more frequent packet generation, can be mitigated. Figures 5 and 6 show that application-layer VoIP aggregation outperforms the simple audio sample encapsulation in both downlink and uplink. When employing L1 aggregation (two codec samples in an IP packet) the header overhead drops to 83%. As a result, 650 flows can be injected into the downlink, instead of 400 in the non-aggregated runs. In addition, L1 aggregation increases the amount of bandwidth used for user data transmission by more than 60%.

When L2 application-layer aggregation is used (3 codec samples/IP packet), we need 800 flows to saturate the fixed

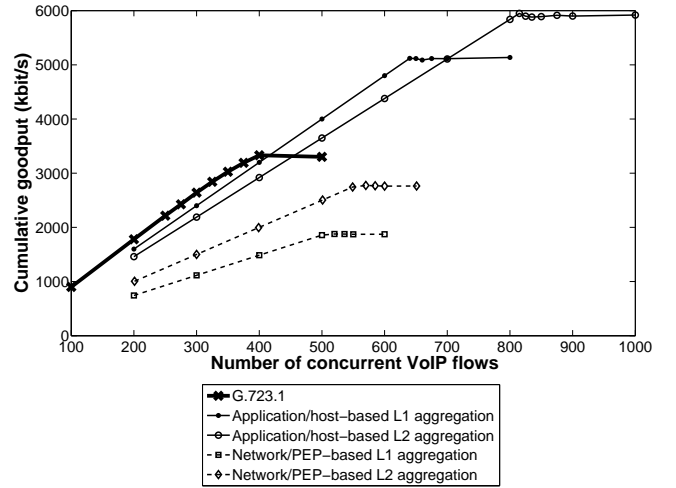


Figure 5: Cumulative downlink goodput

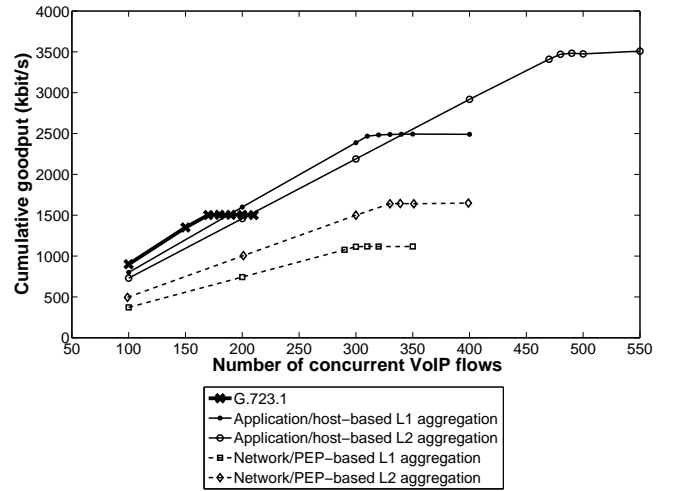


Figure 6: Cumulative uplink goodput

WiMAX downlink. This means that L2 aggregation effectively doubles the number of VoIP flows that can be served in the downlink. This is a direct result of limiting the header overhead to 55%. The L2 cumulative goodput is nearly 6 Mbits/s.

In the uplink, the gains in terms of VoIP flows that can be served at the cumulative goodput saturation point are even greater. With L1, more than 300 flows can be sustained instead of only 150 in the non-aggregated VoIP scenario. With L2, the number of VoIP flows effectively more than triples. In terms of cumulative goodput, L2 can deliver almost 3.5 Mbits/s.

We note that the simple network-layer aggregation schemes we are testing cannot attain any gains in terms of cumulative goodput. At L1 network-layer aggregation, the overhead for forwarding two VoIP packets in a single PEP-encapsulated IP packet is 208%. At L2 aggregation, the overhead is 194%. This amount of overhead by far exceeds the overhead of a single G.723.1 VoIP packet. Nevertheless, as we will see in the following subsections, network-layer PEP-based aggregation can deliver higher VoIP capacity than no aggregation

at all. The maximum cumulative goodput for downlink is 1.8 Mb/s and 2.7 Mb/s for L1 and L2 network-layer aggregation, respectively. The corresponding uplink cumulative goodput are 1.1 Mb/s and 1.6 Mb/s for L1 and L2, respectively.

Although the goodput performance gains attained with application-layer aggregation are remarkable, the total goodput over the fixed WiMAX link with several hundred of VoIP flows is substantially lower than our baseline measurements with MTU-sized packets. Even with L2 application-layer aggregation, the cumulative downlink goodput is roughly only 2/3 of the single-flow baseline downlink goodput of nearly 9.4 Mb/s, measured with 1500-byte packets. This is a significant reduction. The case in the uplink is similar (5.5 Mb/s vs. 3.6 Mb/s).

We also conducted a limited number of experiments with even higher levels of aggregation, packing more codec samples or complete VoIP packets, for application-layer and network-layer aggregation, respectively. We recorded further gains in terms of cumulative goodput results which are closer to the maximum goodput of the wireless link. However, when more samples are bundled into one packet, we experience diminishing gains in terms of cumulative goodput. Moreover, the additional delay introduced while voice samples are buffered at the VoIP client can significantly depreciate the received speech quality in real-world applications, as each sample has a very tight deadline to make. Thus, the advantage of higher level aggregation is effectively “cancelled out” by the unacceptably long packet delays.

5.2 Packet and Sample Loss

Figures 7 and 8 illustrate the measured packet loss rates as we increase the number of concurrent flows in the fixed WiMAX downlink and uplink, respectively. When no aggregation is employed, our testbed downlink can sustain only approximately 200 concurrent VoIP flows with negligible loss. In the uplink about 150 flows can be sustained with negligible loss.

If we introduce more flows, the uplink emerges as a clear bottleneck. With 375 flows, the downlink loss rate raises to 4% and with 400 flows we exceed the threshold of 5% drops. In general, drop rates less than 5% are tolerable for G.723.1 VoIP clients. Drops in excess of 10% are unacceptable. In the uplink, injecting 175 flows is enough for crossing the 5% threshold. The drops in the uplink are distributed very unevenly between the concurrent flows. When some flows experience near-zero packet loss, others may end up losing up to 90% of their packets. This very interesting behavior appears, especially, with small packets, such as non-aggregated VoIP packets. Further study of our testbed measurement results and comparison with results obtained using simulation, which especially focuses on the uplink scheduler and queue management behavior, is necessary.

When we consider application-layer aggregation both L1 and L2 outperform the straightforward single-sample transmission. With L1, no packets are dropped until we start introducing 650 concurrent flows in the downlink and nearly 300 flows in the uplink. With L2, drops start to occur only after we inject 820 flows into the downlink and more than 400 flows into the uplink. The 5% packet loss threshold is exceeded with L1 aggregation with 680 flows in the downlink and nearly 330 flows in the uplink. With L2 aggregation we need to inject more than 840 and 500 flows in the down-

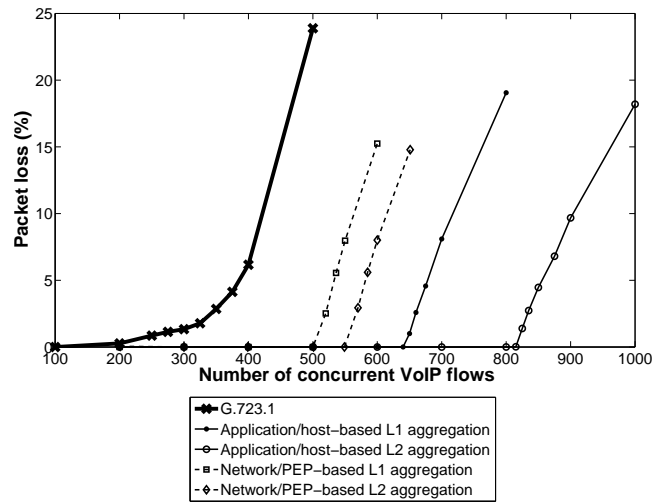


Figure 7: Average downlink packet loss

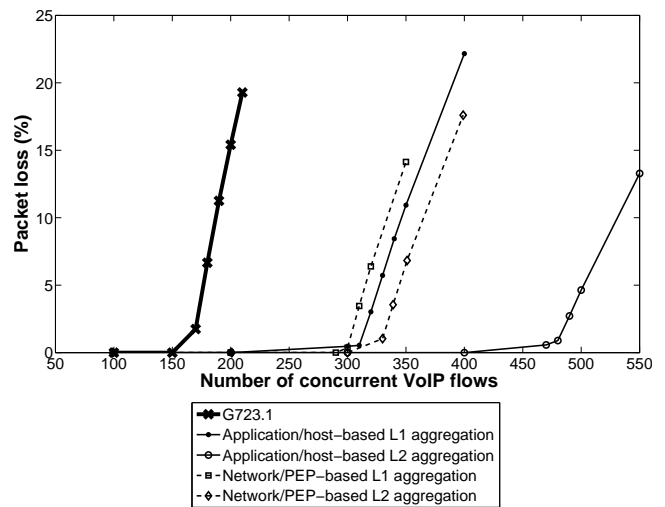


Figure 8: Average uplink packet loss

link and uplink, respectively, in order to start observing an average of 5% packet loss.

Network-layer aggregation does not match the performance of application-layer aggregation, but nonetheless nearly doubles the capacity of the WiMAX link in terms of flows with negligible loss. L1 network-aggregation allows the downlink to sustain 500 flows with zero loss. L2 network-layer aggregation modestly outperforms L1. In the uplink, however, L2 network-layer aggregation outperforms L1 application-layer aggregation.

However, since we are aggregating samples, it is fairer to compare the actual sample loss rates of the different schemes. That is, when considering L1 aggregation, for example, a packet loss rate of $x\%$ corresponds to an effective sample loss rate of $2x\%$. For L3, the effective sample loss rate is $3x\%$. The effective loss rate for downlink is shown in Fig. 9 whereas corresponding uplink values can be seen in Fig. 10.

For L1 network-layer aggregation, the effective sample loss rate in the downlink exceeds 5% when we inject 530 flows.

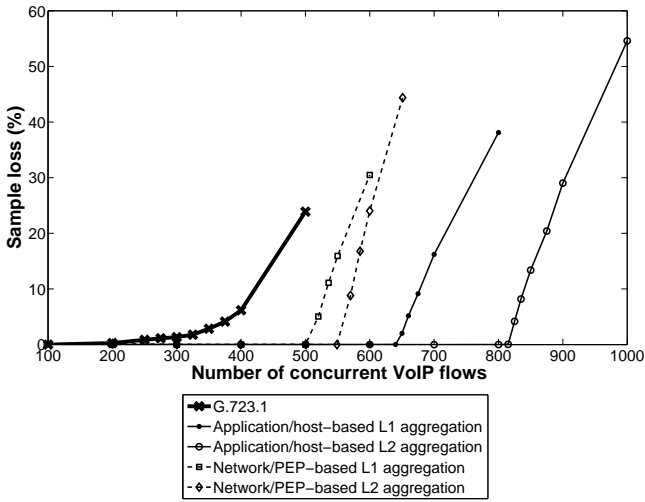


Figure 9: Average downlink sample loss

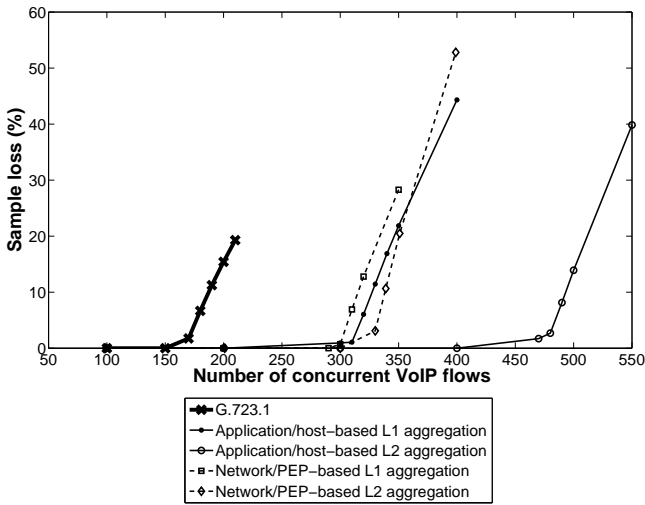


Figure 10: Average uplink sample loss

For L2, over 570 flows lead to unacceptable sample loss rates ($>10\%$). Nevertheless, L1 can more than double the number of VoIP flows in the WiMAX downlink, while L2 records a gain of 275% more flows with zero-loss. Interestingly, although L2 outperforms L1, when we compare the gains attained by the former are in the order of 10% only, despite injecting 50% fewer packets in the network. In addition, L2 suffers from a significantly higher effective sample loss rate than L1 does. Recalling Fig. 5, we note that at 500 flows L1 attains a cumulative goodput of 1.7 Mb/s saturating the link. For L2, with 550 flows the link is also saturated and the cumulative application goodput exceeds 2.6 Mb/s. The difference between network layer aggregations in the uplink is not significant. When staying in the acceptable loss area ($<10\%$), L2 outperforms L1 only by 20 flows and when considering the delay increase for the first packet waiting for bundling can conclude that the benefits of using L2 instead L1 are minor also in the uplink.

The clear winner with respect to sustaining more flows without any loss is application-layer aggregation in both

the uplink and the downlink. In the downlink with L1 application-layer aggregation more than 640 flows can be sustained without any drops; with L2, more than 800 flows can be sustained. These represent two- and three-fold improvements over the non-aggregated results. More importantly, perhaps, this is achieved while saturating the testbed WiMAX link, and using more wireless resources for actual user data (recall Fig. 5 and 6).

If the WiMAX link operator is willing to accommodate more flows at a slightly degraded service level, L1 aggregation can sustain nearly 680 flows with less than 10% sample loss rate, while L2 aggregation raises the same bar to higher than 840 flows. In the uplink, L2 application-layer aggregation increases the tolerable flow amount by nearly 50% when compared with the corresponding L1 results: L2 can allow for up to 490 flows to be sustained. With zero loss rate, L2 application-layer aggregation allows the WiMAX link to sustain 270% more flows than when no VoIP aggregation is employed. If the one-way end-to-end delay, over and above the buffering delay introduced by the aggregation scheme is in the order of 60 to 90 ms, application-layer aggregation is the best solution for increasing VoIP capacity according to our measurements.

Note that the tested WiMAX equipment does not have any built-in intelligent algorithms to assure QoS to individual data flows. When the link is operating above its maximum capacity, dropped packets and additional delays caused by congestion in the test network, affected all active VoIP flows sharing the link, as would be the case in any best effort network. In our testbed, no restrictions are imposed by the BS to new data flows attempting to use the WiMAX link when it is operating beyond its capacity.

5.3 Mean Opinion Score

Based on our packet loss and one-way delay measurements, we calculate the objective mean opinion scores for the received speech quality using Eq. (3). As Chatterjee and Sengupta [5] explain, the packet loss rate experienced by a VoIP flow is a more critical factor when calculating the R-score that delay deterioration and, thus, has significant impact on the MOS values calculated at the receiving end. Even slight increases in loss, decrease the R-score more rapidly than would similar (in proportion) increases in delay. Packet loss in our tests occurs when the bottleneck link, i.e. the fixed WiMAX link, is overwhelmed with traffic.

Fig. 11 presents the dramatic increases of downlink one-way delay in all cases we explored in this paper. Congestion on the best-effort WiMAX downlink will translate in rapid declines in MOS, especially when aggregation is employed in the downlink. In the uplink, the one-way-delay stayed in all cases under 100 ms, even if the packet loss rate exceeds 20% (not shown). Thus, the effect of delay in uplink MOS is not remarkable. These considerably smaller delays are most probably because of smaller buffer sizes in the WiMAX SS equipment.

Improvements in the measured packet loss rates, when aggregation is used, lead directly to better MOS values. In Fig. 12, MOS values for different number of concurrent VoIP flows in the downlink are shown. MOS for VoIP packet flows without aggregation declines to unacceptable levels after injecting 375 concurrent flows. For network-layer aggregation, MOS remains at a very high level with more than 500 flows. L1 network-layer aggregation seems to be degrading in a

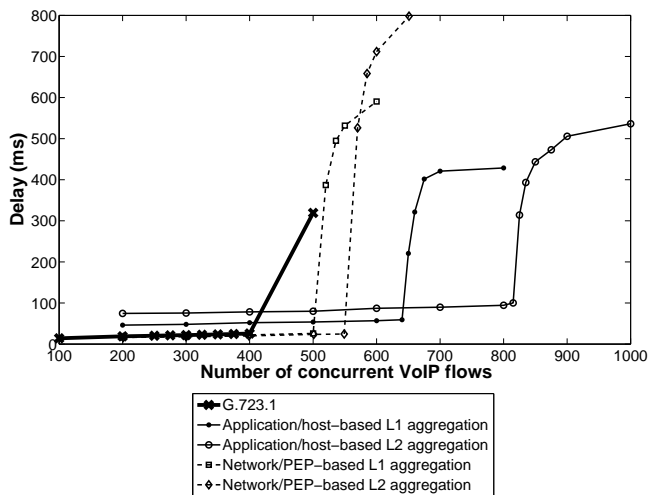


Figure 11: Downlink one-way delay

slightly more orderly fashion in terms of MOS when compared to L2. With application-layer aggregation, our testbed WiMAX link can sustain more than 650 flows with high MOS values. For L2 application-layer aggregation, MOS remains above 3.5 even when more than 800 flows are injected.

As can be seen from Fig. 13, MOS values in the uplink do not have as sharp descents as in the downlink because of the smaller one-way delays. Without aggregation, we can only inject 175 concurrent flows in order to keep MOS at tolerable levels; after that threshold, MOS declines sharply. L1 application-layer aggregation and L2 and L3 network-layer aggregations sustain MOS values above unacceptable level until 320-330 flows are injected. The far and away best results can be achieved by using L2 application-layer aggregation which lets our testbed sustain up to 480 high quality VoIP flows in the uplink.

If we use MOS as the only gauge of overall performance, application-layer aggregation seems to be the best scheme, allowing our fixed WiMAX testbed to sustain nearly three times more flows in the downlink and over two times more flows in the uplink than when no aggregation is used, at comparable level of MOS values. When comparing network- to application-layer aggregation, we note that the latter outperforms the former by 25-47% in the downlink. In the uplink, L2 application-layer aggregation outperforms the corresponding network-layer aggregation by 30%.

6. DISCUSSION AND FUTURE WORK

Our testbed results indicate that “applications know better” and that with proper tuning they can outperform any simple PEP or other middlebox, if configured appropriately. This may not always be the case, however. There is already a host of G.723.1-compliant handsets in the market, with pre-configured profiles, and they may not be upgradable. Our results show that a simple PEP can increase the effective VoIP capacity (when measured in numbers of lossless flows) of the fixed WiMAX testbed in our lab. Nonetheless, the employed network-layer aggregation scheme is neither too sophisticated nor we claim it to be the most efficient. Yet, it is simple to develop and easy to deploy and can double the effective VoIP capacity without any changes to exist-

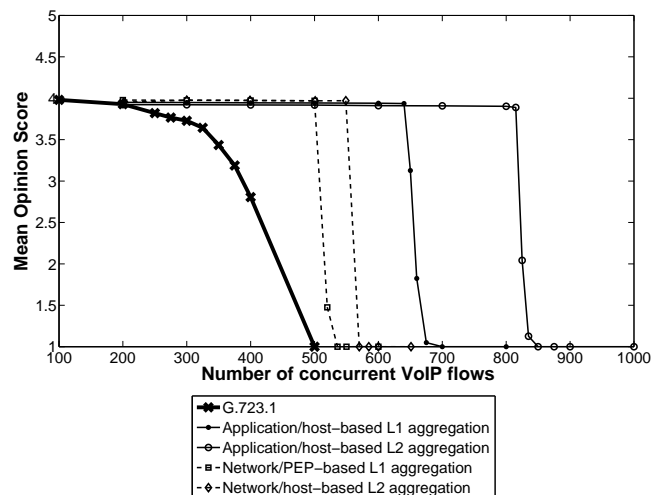


Figure 12: Downlink MOS

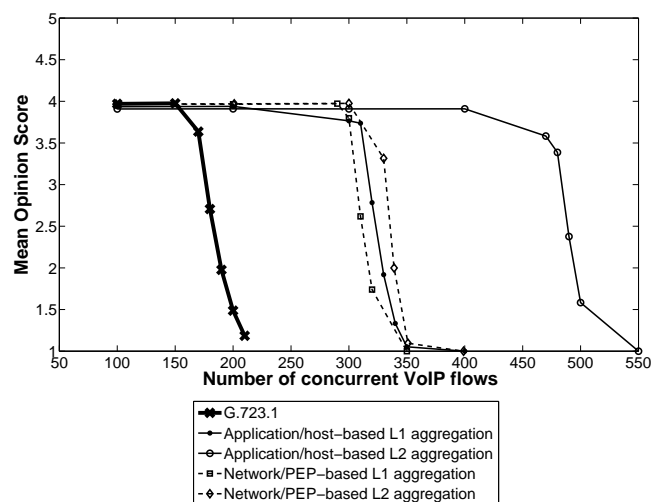


Figure 13: Uplink MOS

ing applications or handsets. Robust header compression (ROHC) can increase performance even more. MAC-layer aggregation may perform better, but it requires cross-layer design. Our future work aims at evaluating other more sophisticated VoIP aggregation schemes in our testbed.

Moreover, we would like to evaluate the performance of codecs other than G.723.1 in our testbed. Studying the performance of other multimedia services is also in our agenda, as is evaluating VoIP adaptation schemes in the presence of congestion and loss on the WiMAX segment. Finally, we are already investigating the performance of bidirectional VoIP flows. In this paper, we empirically investigated VoIP performance while considering the WiMAX downlink and uplink in isolation. Although this is an important first step to address the lack of information about WiMAX performance in practice, it will be even more interesting to examine how many “VoIP calls” can our testbed sustain.

The testbed employed in this measurement campaign is based on fixed WiMAX equipment following a subset of the physical layer specifications presented in the IEEE 802.16-2004 standard for fixed metropolitan area networks [11].

The tests were performed in a lab environment which yields maximum system throughput and reliability. In the future, these tests could be carried out in a more demanding radio environment to verify the performance under different conditions. Moreover, it will be important to repeat these and other measurements using IEEE 802.16-2005 [12] equipment and consider mobility scenarios.

We would be very interested in repeating the measurements with some of the more advanced error correcting codes [12] as well, such as turbo codes or low-density parity check codes in more demanding conditions. Another feature specified for WiMAX systems that is not employed in the current testbed is the use of adaptive modulation and coding (AMC), which allows the base station to adjust its modulation and coding parameters on the fly according to the measured channel condition. Perhaps the most intriguing feature missing from the current setup is the use of multiple-input multiple-output (MIMO) techniques. The current testbed includes only a single, although directional BS antenna and single receiving directional SS antenna. MIMO techniques (space-time coding, spatial multiplexing, and beamforming) employ several transmitting and receiving antennas at both ends of the transmission link improving system throughput, transmission reliability and transmission range [3]. It will be interesting to see tests similar to the ones carried out in this paper in a system equipped with, for example, adaptive beamforming and AMC, and study how it behaves in different radio environments.

7. CONCLUSION

Despite the significant interest in WiMAX technology and developments, WiMAX equipment is yet to become readily available at affordable prices. As such, most studies of WiMAX are performed using simulation and modeling. Few are the studies publicly reporting results from testbeds or field trials. In this paper we considered the performance of aggregated and non-aggregated VoIP over a fixed WiMAX testbed. We measured the performance of different transmissions schemes in terms of cumulative goodput, packet and sample loss rates, and calculated the objective mean opinion scores using the R-Score specified by ITU.

We found that VoIP flows carrying single sample payloads generated by the G.723.1 codec are clearly underperforming in both uplink and downlink. Indeed the header overhead introduced by RTP, UDP and IP lead to significant waste in wireless bandwidth resources. The strategy of sending a single sample encapsulated in an RTP/UDP/IP packet is, to say the least, suboptimal. We found that application-layer VoIP aggregation can more than triple the number of lossless VoIP flows in the downlink, without any network or hardware support. The results are of the same order for the uplink as well. A simple network-layer VoIP aggregation algorithm can also increase the effective capacity of conventional (non-aggregated) VoIP flows. Network-layer aggregation can more than double the number of lossless flows that can be sustained in our WiMAX testbed, and could be easily adopted in cases where application-layer aggregation cannot be implemented.

Acknowledgement

This work was conducted at the VTT Converging Networks Laboratory within the framework of the IST Sixth Frame-

work Programme Integrated Project WEIRD (IST-034622), which is partially funded by the Commission of the European Union. The authors also received support from the IST FP6 Integrated Project ANEMONE (IST-035072). Study sponsors had no role in study design, data collection and analysis, interpretation, or writing the report.

The views expressed in this paper are solely those of the authors and do not necessarily represent the views of VTT, the WEIRD project, or the Commission of the European Union. We thank the anonymous reviewers for their comments and suggestions; they greatly improved the quality of this paper.

8. REFERENCES

- [1] *Precision clock synchronization protocol for networked measurement and control systems*. International standard IEC 61588, 2004.
- [2] J. G. Andrews, A. Ghosh, and R. Muhamed. *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall, 2007.
- [3] H. Bolcskei. MIMO-OFDM wireless systems: basics, perspectives, and challenges. *IEEE Wireless Commun.*, pages 31–37, August 2006.
- [4] A. Botta, A. Dainotti, and A. Pescapè. Multi-protocol and multi-platform traffic generation and measurement. In *INFOCOM 2007 DEMO Session*, May 2007.
- [5] M. Chatterjee and S. Sengupta. *WiMAX Applications*, chapter VoIP over WiMAX. CRC Press, 2008.
- [6] R. G. Cole and J. H. Rosenbluth. Voice over IP performance monitoring. In *Proc. ACM SIGCOMM*, volume 31, pages 9–24, April 2001.
- [7] K. Correll, N. Barendt, and M. Branicky. Design Considerations for Software Only Implementations of the IEEE 1588 Precision Time Protocol. Conference on IEEE 1588, 2005.
- [8] O. Grondalen, P. Gronlund, 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.
- [9] N. Gupta and G. Kaur. *WiMAX Applications*, chapter WiMAX Technology for Broadband Wireless Communication. CRC Press, 2008.
- [10] C. Hoene, H. Karl, and A. Wolisz. A perceptual quality model intended for adaptive VoIP applications. *International Journal of Communication Systems*, 19(3):299–316, 2006.
- [11] IEEE 802.16 Working Group, editor. *IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems*. IEEE Std. 802.16-2004, October 2004.
- [12] IEEE 802.16 Working Group, editor. *IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems. Amendment 2: Physical and Medium Access Control Layer for Combined Fixed and Mobile Operation in Licensed Bands*. IEEE Std. 802.16e-2005, December 2005.
- [13] ITU-T. *Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbs*. ITU-T Recommendation G.723.1, 1996.

- [14] ITU-T. *The E-Model, a computational model for use in transmission planning*. ITU-T Recommendation G.107, December 1998.
- [15] K. Kim, R. I. S. Ganguly, and H. Sangjin. On packet aggregation mechanisms for improving VoIP quality in mesh networks. In *Proc. VTC 2006-Spring*, volume 2, pages 891–895, May 2006.
- [16] O. Komolafe and R. Gardner. Aggregation of VoIP streams in a 3G mobile network: A teletraffic perspective. In *Proc. European Personal Mobile Communications Conference*, pages 545–549, April 2003.
- [17] J. Manner. *Jugi's Traffic Generator (JTG)*. Available: <http://hoslab.cs.helsinki.fi/savane/projects/jtg>.
- [18] Maravedis. WiMAXCounts Quarterly Report, November 2007.
- [19] P. Neves, P. Simoes, A. Gomes, L. Mario, S. Sargento, F. Fontes, E. Monteiro, and T. Bohnert. WiMAX for emergency services: An empirical evaluation. In *Proc. International Conference Next Generation Mobile Applications, Services and Technologies*, pages 340–345, September 2007.
- [20] P. C. Ng, S. C. Liew, and C. Lin. Voice over wireless LAN via IEEE 802.16 wireless MAN and IEEE 802.11 wireless distribution system. In *Proc. IEEE WirelessCom*, July 2005.
- [21] N. Scalabrino, F. D. Pelegrini, I. Chlamtac, A. Ghittino, and S. Pera. Performance evaluation of a WiMAX testbed under VoIP traffic. In *Proc. WINTECH '06*, pages 97–98, September 2006.
- [22] N. Scalabrino, F. D. 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.
- [23] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. IETF Request for Comments: 3550, July 2003.
- [24] S. Sengupta, M. Chatterjee, S. Ganguly, and R. Izmailov. Improving R-Score of VoIP streams over WiMAX. In *Proc. IEEE ICC 2006*, volume 2, pages 866–871, June 2006.
- [25] WiMAX Forum. *WiMAX End-to-End Network Systems Architecture Stage 2: Architecture Tenets, Reference Model and Reference Points*. Release 1.1.0. June 2007.
- [26] WiMAX Forum. *WiMAX End-to-End Network Systems Architecture Stage 3: Detailed Protocols and Procedures*. Release 1.1.0. June 2007.