Impact of 4G Wireless Link Configurations on VoIP Network Performance

Stefan Alfredsson, Anna Brunstrom
Department of Computer Science
Karlstad University
SE-651 88 Karlstad, Sweden.
Email: {Stefan.Alfredsson, Anna.Brunstrom}@kau.se

Mikael Sternad
Signals and Systems
Uppsala University
SE-751 20 Uppsala, Sweden.
Email: Mikael.Sternad@signal.uu.se

***Abstract*—The performance of applications in wireless networks is partly dependent upon the link configuration. Link characteristics varies with frame retransmission persistency, link frame retransmission delay, adaptive modulation strategies, coding, and more. The link configuration and channel conditions can lead to packet loss, delay and delay variations, which impact different applications in different ways. A bulk transfer application may tolerate delays to a large extent, while packet loss is undesirable. On the other hand, real-time interactive applications are sensitive to delay and delay variations, but may tolerate packet loss to a certain extent. This paper contributes a study of the effect of link frame retransmission persistency and delay on packet loss and latency for real-time interactive applications. The results indicate that a reliable retransmission mechanism with fast link retransmissions in the range of 2-8 ms is sufficient to provide an upper delay bound of 50 ms over the wireless link, which is well within the delay budget of voice over IP applications.**

I. Introduction

The demand for higher capacity and wider coverage of wireless network access is increasing. As the third generation mobile systems are becoming commercialized, research focus has shifted towards 4G systems. The main transmission technology in 4G proposals such as evolutions of the 3GPP long term evolution (LTE) [1] and IEEE 802.16 [2] is orthogonal frequency division multiplexing (OFDM), which uses multiple carrier frequencies dedicated to a single data source.

One early 4G system proposal based on OFDM was developed within the "Wireless IP" project at Uppsala University, in cooperation with Chalmers University of Technology and Karlstad University [3]. The main focus is to cover wide areas to service vehicular users, in excess of speeds of 100 km/h with a 30-fold bandwidth increase compared to UMTS/3G. To realize this goal, adaptive OFDM is used in combination with channel prediction. Transmissions can then be scheduled to maximize the total satisfaction of the users, depending on their current channel quality. This is combined with increased cross layer interaction, link level automatic repeat-request (ARQ), and other mechanisms.

The Wireless IP system is conceived as a testbed for wireless all-IP Internet traffic. It should be designed to provide a good service to the network and transport layers and it is thus important to consider the system level implications of lower layer design decisions. In order to allow performance measurements on the interaction between the physical/link layer design and upper layers the Wireless IP emulator (WIPEMU) was developed. WIPEMU is able to emulate a range of wireless link parameters, and enables studies of the resulting effects on network traffic.

Applications can have different requirements on the network traffic properties. For example, a bulk transfer application has low requirements on packet delay and delay variations, but is sensitive to data loss. The main interest is to achieve as high throughput as possible. A real-time interactive application has low requirements on data throughput, as long as a minimum throughput can be upheld. Compared to a bulk transfer application, it is more sensitive to delay and delay variations but less sensitive to packet loss. The terms "elastic" and "inelastic" are often used to describe such application classes, where the elasticity refers to the ability to adapt to varying network conditions.

From a research standpoint, inelastic applications are interesting to study because they have stricter requirements on the network, and in some aspects pose a greater challenge than elastic applications that adapt to the provided resources. In this paper, a voice over IP (VoIP) application is used to represent the class of inelastic, real-time interactive applications that is sensitive to delay but may tolerate packet loss. The resulting packet loss and delay is studied in relation to the link layer frame retransmission delay and the maximum number of link layer retransmissions. The results indicate that a short link frame retransmission delay enables link reliability with a resulting packet delay that is acceptable for real-time interactive class applications.

The remainder of the paper is structured as follows. A description of the studied application class and its properties is given in section II. Then a description of the wireless link configuration settings follows in section III. The experiment setup and execution is presented in section IV, followed by the results in section V. Finally, section VI contains the conclusions.

II. Application network requirements

A real-time interactive flow, such as a VoIP application, has constraints on the total delay, and implicitly on the delay variations. The delay impacts the user's perception of interactiveness. If the delay of for example a phone conversation exceeds a couple of hundred milliseconds, this causes annoyance with the user [4]. The delay we consider consists of two parts; the delay in the fixed network, and the delay over the wireless link. The former is hard to improve as it depends on the distance between the communicating partners, congestion levels in the network, and other factors. The latter, the wireless delay, is the topic of interest here.

When the wireless link provides reliability, it also increases the delay variation. This must be compensated with an increased playout buffer, which in turn increases the total delay before playback.

Therefore, voice applications traditionally do not use retransmissions, but employ FEC coding to achieve a reasonable bit or packet error rate. An example of this is the "conversational" QoS class in UMTS/3G [5]. This is not optimal from a capacity standpoint, and it requires a more complex configuration and tuning of ARQ/FEC for different traffic classes.

A tradeoff between providing full reliability and single transmission with FEC is to provide partial reliability. This can be achieved by performing a limited amount of retransmissions. The delay variation will decrease at the expense of decreased reliability in the form of packet loss. As discussed later, packet loss can to some extent be tolerated by the application/user.

Normally IP/UDP/RTP is used for real-time interactive flows. UDP does not provide retransmissions on the transport layer, which means that the application can decide whether to do retransmissions or not. This is normally not done for real-time interactive applications, as the playback buffering needed to capture the retransmission could cause the user tolerable delay to be exceeded.

While end-to-end retransmissions may take too long, link-level retransmissions can be fast enough to provide reliability while still keeping below delay limits. In the existing 3G systems, the system design has caused high retransmission delays which have had a negative impact on the resulting packet transmission delay. 4G systems with short delay loops therefore give new possiblities, and is interesting to study in combination with link ARQ and the effect on packet transmission characteristics. 3GPP LTE [1] has a target of 8 ms HARQ RTT, while the WINNER system [6] has a target of 2 ms HARQ RTT.

The performance of VoIP applications can be measured in a number of ways, for example with regard to delay, jitter and packet loss. There are also perceptual models and metrics that attempt to capture the experience of the user, as a result of the different delays and losses. Examples of these include Mean Opinion Score (MOS) [7], Perceptual Evaluation of Speech Quality (PESQ) [8], and the E-Model [9]. In this paper the packet loss and delay are examined without correlating to speech quality, other than recognizing recommendations of keeping a low delay limiting the amount of packet loss. In [4] (fig. 2), delays up to 150 ms are considered "very satisfactory", with a steep slope at 175 ms after which the quality is considered "satisfactory", "some users dissatisfied", "many users dissatisfied" and "exceptional limiting case", on a declining scale up to 500 ms. The explanation for this is that in a normal conversation, there is about 200 ms between speaking turns. Additional delay can be interpreted as hesitation, cause over-talking, or break-in problems. Packet loss on the other hand more directly impact the speech quality. A loss of a packet can cause pops, clicks or noise depending on the packet loss concealment method employed by the application. As shown in [4] (table 2), a standard G.711 (pulse code modulation or PCM) voice codec with packet loss concealment can tolerate about 3% packet loss while still providing "satisfactory" user experience.

III. Wireless link emulation and configuration

The wireless link used for the experiments is based upon the "Wireless IP" 4G downlink system proposal [10], [11].

The downlink uses OFDMA coded transmissions divided in time and frequency: 1500x25 frames per second, over a 5 MHz channel in the 1900 MHz band. Each time-frequency slot, or *bin*, consists of 108 symbols. The symbols are modulated with 1 to 8 bits/symbol corresponding to uncoded BPSK to 256-QAM modulation.

The modulation is adaptive on a per-bin time scale. A predicted channel strength is used to determine the modulation level from a table. This table is optimized for maximum throughput when an unlimited number of uncoded link frames are allowed [12]. The amount of data within a bin can therefore vary from 108 to 864 bits due to the adaptive modulation.

For each bin a channel prediction with a normalized mean square error (NMSE) of 0.1 is used to decide which modulation level to use. As the frame is transmitted, the true channel is used to estimate the symbol, bit and frame error rate. If the frame is received erroneously, it is retransmitted until a maximum limit is reached. The retransmission delay depends upon a number of factors such as the feedback delay of the link loss notification mechanism, channel scheduling decisions, and retransmission queuing and priority. In this paper we investigate the effect of different retransmission delays, independently of their origin.

Link retransmissions cause reordering between frames which may also cause packet reordering. Unordered packets are not efficient for TCP, which will issue duplicate acknowledgements for reordered packets. These may be interpreted by the sender as a sign of packet loss and lead to a decreased sending rate. For UDP, packet reordering can cause packet discard at the receiver, for example if the packet playout time has already passed. Therefore, packets are ordered at the wireless receiver before further processing.

With regards to queueing, link queueing is in this case negligible since the transmissions operate below the link capacity[1].

A compilation of the system parameters is shown in table I.

Fixed network	
Fixed network delay	40 ms RTT, 20 ms one-way
Network queue size	unlimited
Wireless Downlink	
Frame transmission time	0.667 ms
Channel model	12-tap Jakes typical urban fading model @ 75 km/h, 16 dB SNR + AR(1) shadow fading with variance 4 dB and pole at 0.74
Modulation	Adaptive BPSK,4-256 QAM with switching adjusted for a prediction error of NMSE 0.1 [12].
Frame size	108 symbols
Coding	Uncoded M-QAM
Scheduling	best frame
Link ARQ	0 to 30 retransmissions
Link retr. delay	2-16 ms
Ordering	Unordered link retransmissions, ordered output queue
Wireless uplink	
Channel model	imposed bandwidth limit and delay
Packet loss	0% (lowest modulation level assumed)
Capacity	20 kbit/s
Delay	2 ms
Transport layer parameters	
Protocols	UDP (Linux 2.6.8)
Transferred data	10000 packets á 172 bytes, 50 pkts/s

TABLE I
EXPERIMENT PARAMETERS

IV. Experiment Setup

The experiment setup consists of an emulated scenario with a sender in a fixed network transmitting to a mobile receiver. This is emulated with three computers; one sender, one gateway running the WIPEMU [13] emulator, and one receiver. Traffic that flows through WIPEMU is subjected to an emulation of a number of link characteristics such as queueing, channel prediction, scheduling, adaptive modulation, ordering, link frame retransmission and delay.

The real-time interactive traffic is represented by a voice over IP application. With 8 kHz PCM sampling, 20 ms packetization delay [14] and RTP header, this gives a packet stream of 50 packets/s with 172 byte payload. The traffic is generated and captured with the `mgen` [15] tool, and analyzed with `trpr` [16].

V. Results

The main concerns for real-time interactive traffic is packet loss, delay and delay variation. There is a tradeoff between these parameters. By providing a reliable link without packet loss, delay is increased since the link will retransmit frames to achieve reliability. Likewise, decreased reliability can lead to a lower delay.

[1]If this was not the case, i.e. transmission at near or above link capacity, it would lead to increased queueing, delay and packet overflow loss in relation to the capacity overload. This is not a stable operating point for the targeted class of applications, and therefore not considered.

Traditionally such traffic has been sent without retransmission, and instead used FEC coding to handle transmission errors. This means that different link configurations are needed for different traffic classes. For the experiments in this paper, a configuration optimized to provide maximum throughput for a bulk transfer application was used, to see if a common configuration for multiple service classes would be possible.

Figure 1 shows the packet loss percentage for varying link reliability. When only one frame transmission is done (zero retransmissions) there is a resulting packet loss rate at over 45%. This is clearly unacceptably high. After a maximum of three retransmissions, loss percentage is down to about 2%. At a maximum of 10 retransmissions the loss percentage is about 0.02%. As mentioned earlier, an acceptable packet loss rate is below 3%, which is thus reached within three link retransmissions. Figure 1 also shows that the packet loss rate is independent of the link retransmission delay.

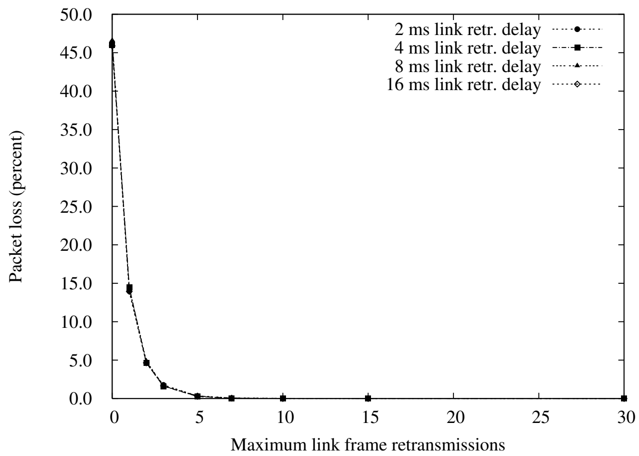

Fig. 1. Packet loss as a function of maximum link retransmissions and retransmission delay

The question is how these retransmissions impact the resulting packet delay on the application layer. Figure 2 shows the percentage of packets that arrive below a certain delay threshold, for varying link retransmission delay, when a maximum of three link frame retransmissions are used. Since the delay in the fixed network is 20 ms, no packets arrive before this time. After about 100 ms (80 ms over the wireless link), all packets have arrived even for the scenario with the highest tested link retransmission delay of 16 ms. The figure also shows that shorter link retransmission delays lead to reduced end-to-end packet delays.

This was the case for an unreliable channel, which had about 2% packet loss due to the limit of maximum three retransmissions. It is therefore interesting to study how increased reliability affects the delay outcome. Figures 3, 4 and 5 shows the results for a maximum of 5, 10 and 30[2] link retransmissions, respectively. Because of the increased retransmissions, packets are slightly more delayed than in the

[2]Representing an “infinite” number of retransmissions.

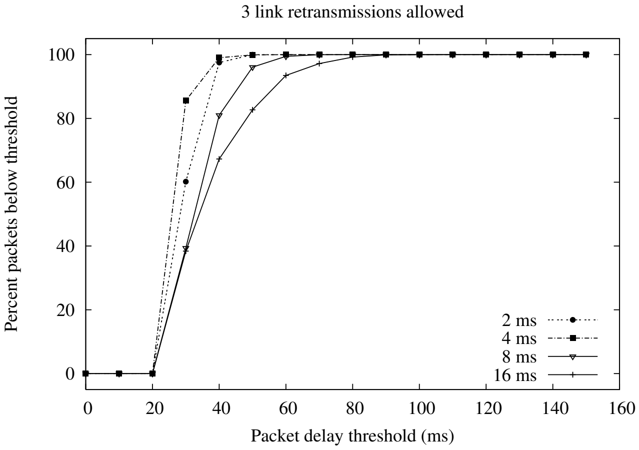

Fig. 2. Percentage of packets below certain delay thresholds for varying link retransmission delays and a maximum of 3 link layer retransmissions.

previous figure. Still at 100 ms including the 20 ms delay offset almost all packets have arrived except for the case with the highest link delay.

If a maximum retransmission delay of 8 ms is used, all packets are delivered within 50 ms (excluding the 20 ms delay offset), compared to the 150 ms delay bound for a "very satisfactory" voip user experience discussed earlier.

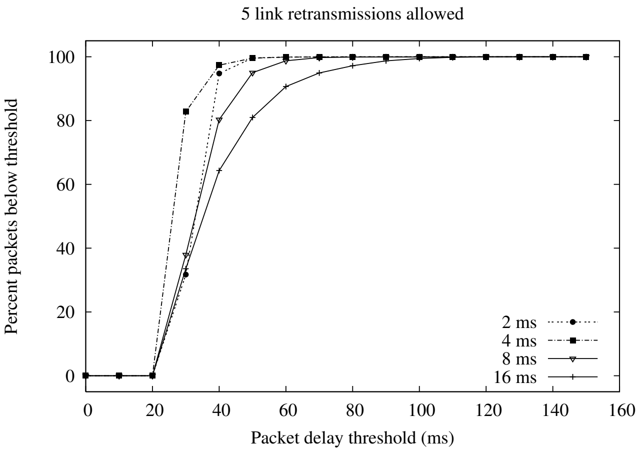

Fig. 3. Percentage of packets below certain delay thresholds for varying link retransmission delays and a maximum of 5 link layer retransmissions.

Figures 6, 7 and 8 shows a reorganization of the data with 2, 8 and 16 ms delays. There is a clear tendency of increasing packet delay as the link retransmission delay increases. The conclusion from these figures is that the link transmission delay seems to be of more importance than the link reliability. The reason for this is that the link reliability is expressed as the upper retransmission limit, and the actual number of retransmissions are mostly below this limit, especially for the higher limits. This can also be seen from the amount of packet loss in Figure 1, where there is little difference between using a maximum of 10, 15 or 30 retransmissions. A closer investigation reveals that the maximum number of

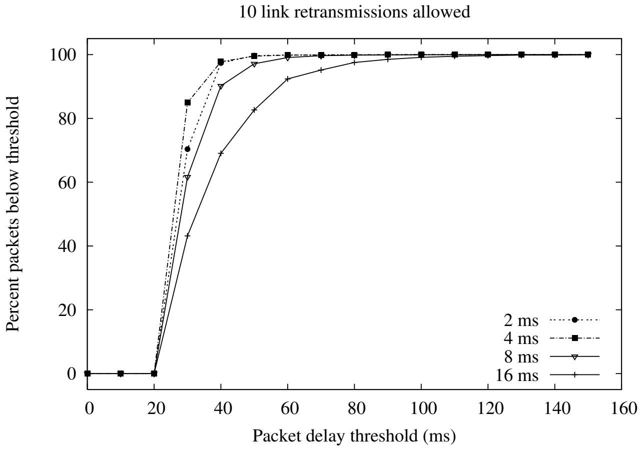

Fig. 4. Percentage of packets below certain delay thresholds for varying link retransmission delays and a maximum of 10 link layer retransmissions.

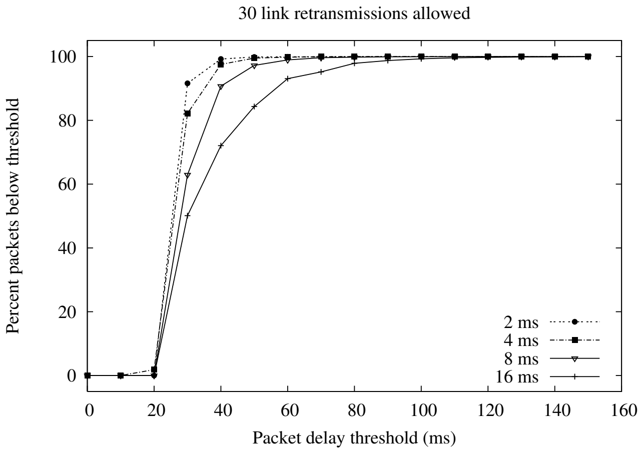

Fig. 5. Percentage of packets below certain delay thresholds for varying link retransmission delays and a maximum of 30 link layer retransmissions.

retransmissions reached was 12 for the used combination of channel prediction and adaptive modulation switch levels.

VI. Conclusions

As mobile networks are shifting from voice to packet data centric, it is interesting to study how voice data transmission will perform in such networks. This paper investigated the VoIP packet loss and delay performance in a 4G evaluation system, with regards to link level reliability and frame retransmission delay.

The results indicate that short delay loops allows full link reliability while keeping an acceptable delay bound. A link retransmission delay of 2 ms resulted in a packet delay of 20 ms over the wireless link, and 8 ms link retransmission delay resulted in a packet delay of 50 ms. The highest link retransmission delay tested, 16 ms, resulted in an 80 ms upper delay bound which is a good margin to the "very satisfactory" user criteria of 150 ms end-to-end delay for voice over IP.

Further, the link configuration used settings optimized to provide maximum throughput at the expense of link retransmissions. The results indicate that the short delay loop gives new opportunities and could allow a single ARQ configuration that optimizes network capacity to be used for all services, elastic as well as inelastic.

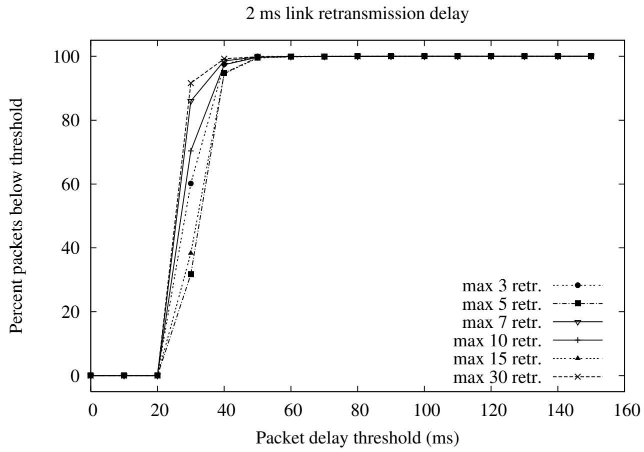

Fig. 6. Percentage of packets below certain delay thresholds for varying maximum link retransmissions and a retransmission delay of 2 ms.

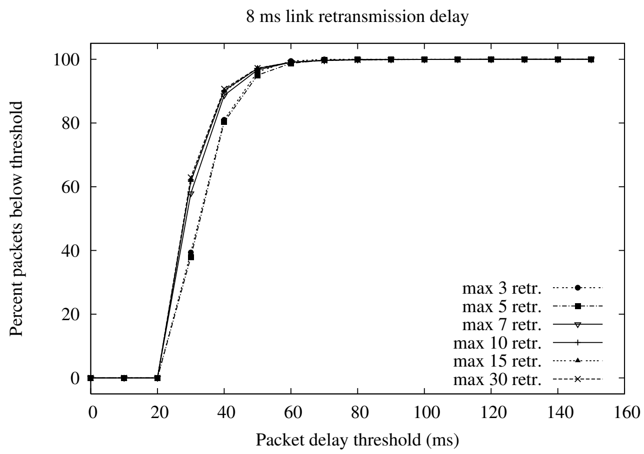

Fig. 7. Percentage of packets below certain delay thresholds for varying maximum link retransmissions and a retransmission delay of 8 ms.

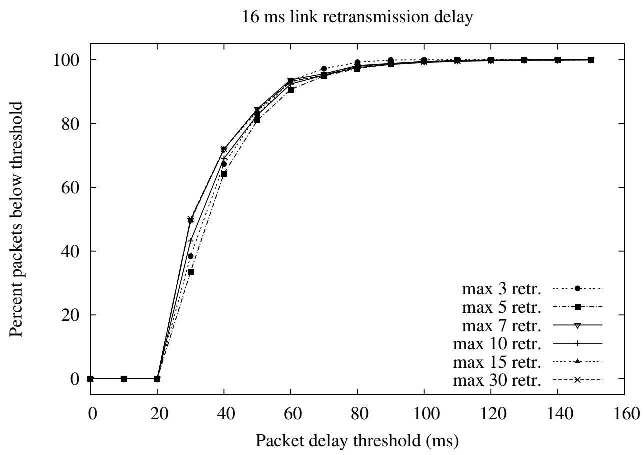

Fig. 8. Percentage of packets below certain delay thresholds for varying maximum link retransmissions and a retransmission delay of 16 ms.

REFERENCES

[1] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution - HSPA and LTE for mobile broadband*. Academic Press, July 2007.

[2] IEEE Standard for Local and Metropolitan Area Networks, "Part 16: Air interface for fixed broadband wireless access systems," *IEEE*, Feb 2005.

[3] M. Sternad, T. Ottoson, A. Ahlén, and A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks," in *Proceedings of IEEE Vehicular Technology Conference VTC2003-Fall*, Orlando, FLA, Oct 2003. Available: http://www.signal.uu.se/Publications/pdf/c0309.pdf

[4] Telecommunications Industry Association, "TIA/TSB-116-A: Voice quality recommendations for IP telephony," march 2006. Available: http://www.tiaonline.org/standards/technology/voip/documents/TSB116-Afinalforglobal.pdf

[5] 3GPP TS 23.107, "Quality of Service (QoS) concept and architecture (release 4)," *3GPP Technical Specification Group Services and System Aspects*, Jan 2003.

[6] Auer, Döttling, Doppler, Häfen, Johansson, Kalliojärvi, Lestable, Luo, Nyström, Olsson, Osserian, Sternad, Svensson, and Wijting, "WINNER II system concept description," *Deliverable D6.13.14 available at http://www.ist-winner.org/WINNER2-Deliverables/D6.13.14%20v1.1.pdf*, Jan 2008.

[7] ITU-T, "Recommendation P.800: Methods for objective and subjective assessment of quality," *ITU-T*, 1996.

[8] ITU-T, "Recommendation P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," *ITU-T*, 2001.

[9] ITU-T, "Recommendation G.107: The E-Model, a computational model for use in transmission planning," *ITU-T*, 2000.

[10] A. Svensson, A. Ahlén, A. Brunstrom, T. Ottosson, and M. Sternad, "An OFDM based system proposal for 4G downlinks," in *Proceedings of Multi-Carrier Spread Spectrum Workshop*, Oberpfaffenhofen, Germany, 2003. Available: http://www.signal.uu.se/Publications/pdf/c0309.pdf

[11] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunstrom, "Towards systems beyond 3g based on adaptive ofdma transmission," in *Proceedings of the IEEE*, vol. 95, no. 12, 2007, pp. 2432–2455.

[12] M. Sternad and S. Falahati, "Maximizing throughput with adaptive M-QAM based on imperfect channel predictions," in *Proceedings of IEEE PIMRC*, Barcelona, Sep 2004.

[13] S. Alfredsson, A. Brunstrom, and M. Sternad, "Emulation and validation of a 4G system proposal," in *Proceedings of Radiovetenskap och Kommunikation*, Linkoping, Sweden, June 2005.

[14] Audio-Video Transport Working Group and H. Schulzrinne, "RFC 1890: RTP profile for audio and video conferences with minimal control," Jan. 1996.

[15] U.S. Naval Research Laboratory, "Multi-generator MGEN, version 4," 2008. [Online]. Available: http://cs.itd.nrl.navy.mil/work/mgen/

[16] U.S. Naval Research Laboratory, "trpr: Trace plot real time," 2008. [Online]. Available: http://cs.itd.nrl.navy.mil/work/proteantools/