HSDPA Performance in Live Networks

Marko Jurvansuu, Jarmo Prokkola, Mikko Hanski and Pekka Perälä
Converging Networks Laboratory
VTT Technical Research Centre of Finland
FIN-90540 Oulu, Finland
Email: Marko.Jurvansuu@vtt.fi

***Abstract*—The first HSDPA (High-Speed Downlink Packet Access) networks have been recently deployed for operational use. We evaluate and compare live HSDPA operational network performance from the end-user perspective, especially related to Voice over IP and web browsing applications. We compare goodput performance with user datagram protocol (UDP) and transmission control protocol (TCP) in both Wideband Code Division Multiple Access (WCDMA) and HSDPA. We also address one-way delay measurements to show how up- and downlink asymmetry affects delay and jitter. We found that HSDPA and WCDMA goodput correlate well with the advertised maximum performance values. However, with small payload sizes HSDPA does not bring a substantial improvement over TCP. With UDP the situation is clearly better except with payload sizes commonly used by VoIP. We verify that HSDPA delay is substantially smaller than with WCDMA, as well as the occurrence of delay spikes caused by link level retransmissions. We found that despite 2 ms time division in HSDPA, the main jitter level is at 10 ms. The results can be exploited by operators and equipment manufacturers for service optimisation.**

I. INTRODUCTION

Widespread deployment of WCDMA technology has opened up new possibilities for operators to offer their customers IP-based voice and video services. This is mainly due to increased bandwidth and higher user capacity compared to 2G systems. The next phase in 3G evolution is the introduction of high-speed downlink packet access [1,2] in 3rd Generation Partnership Project (3GPP) release 5 (hereafter referred to as HSDPA). HSDPA is basically an upgrade to the UMTS release 4 (hereafter referred to as WCDMA) radio interface, providing bandwidth up to 14 Mbps with very low jitter and latencies below 100 ms. It is thus suitable for real-time services such as Voice over IP (VoIP) and video applications. In addition, the high system capacity enables the use of HSDPA technology as a wireless replacement for a fixed digital subscriber line (xDSL) broadband. As an example, one of the Finnish cellular operators offers a monthly fee, flat-rate pricing-based wireless broadband connection using HSDPA.

The first HSDPA networks are now being brought into operational use. It is not, however, clear how significant the difference in performance is between WCDMA and HSDPA for end-user applications. Most of the studies related to HSDPA technology are based on simulations and analytical evaluation, and some on measurements performed in test or laboratory networks. The performance evaluation of live operational HSDPA networks from the end-user perspective seems to be very rare in the literature. This is a significant gap, since new technologies such as HSDPA are usually marketed by their theoretical maximum performance values, which are not necessarily fulfilled in live networks.

In this article, we present the measurement results of public live operational networks, both HSDPA and WCDMA. Aspects of the user experience are emphasized, as the measurements are done with hypertext transfer protocol HTTP/TCP and VoIP applications. For the same reason we use *goodput* as our main performance metric and use ordinary terminals, i.e. laptops equipped with wireless network cards.

We present maximum goodput performance with TCP and UDP in WCDMA and HSDPA live operational networks. We discuss the reasons behind the correlation between goodput and payload size. We also discuss how this affects user applications. We compare measured delay and jitter values using *one-way measurements*. This measurement type is especially important in wireless networks, where up- and downlink performance is asymmetrical. With one-way measurements we demonstrate how the Round Trip Time (RTT) is composed of up- and downlink portions and compare our results with the RTT values measured in [3]. We discuss how one-way link performance affects TCP and UDP. We also show examples how link level re-transmissions create bursts in packet delivery and how Iub-interface flow control affects jitter. Related to this, more detailed study is given in [4] about how data traffic type can affect the output of delay and jitter measurements.

The rest of the paper is organised as follows. In Section II, an overview of HSDPA technology and in Section III description of the measurement set-up is given. In Section IV we present the measurement results and discuss their meaning for end-user services. The conclusions are given in Section V.

II. HSDPA TECHNOLOGY OVERVIEW

One of the key components of HSDPA is a new shared downlink channel, HS-DSCH. The resources in this channel are divided into 2 ms transmission time intervals (TTI). A short TTI reduces the overall delay and enables rapid changes in transmission parameters based on radio link quality. The overall HSDPA delay was also decreased considerably by moving some of the radio network controller (RNC) functionalities like scheduling and link layer acknowledgements to base station (nodeB).

1-4244-0353-7/07/$25.00 ©2007 IEEE

Fast link adaptation continuously adjusts the modulation (quadrature phase-shift keying QPSK or quadrature amplitude modulation 16-QAM) and coding scheme in every TTI separately. This is based on link quality measurements for each terminal. One or more users can be allocated to the same TTI, sharing the code space. The spreading factor is fixed at 16 and the maximum number of parallel codes is 15. Depending on the terminal capabilities, the number of supported codes may be smaller, e.g. some of the first phase terminals can use 5 codes simultaneously.

Along with link adaptation, *packet scheduling* is performed by nodeB. The packet scheduler decides for each HS-DSCH TTI to which user(s) packets are transmitted and at which data rate. The way in which the resources are shared between the users depends on the scheduler algorithm. The basic round-robin packet scheduling shares the resources between users equally. However, a large gain in overall cell capacity can be obtained by exploiting link quality information, for instance using proportional fair scheduling. One of the key features is also *hybrid* Automatic Repeat-reQuest (ARQ), which enables a terminal to request rapid retransmission of erroneous transport blocks directly from nodeB instead of RNC. However, RNC retransmission can be used if HARQ fails. HARQ can also use soft combining of retransmitted bits, being able to exploit parts of erroneous frames.

III. Measurement Set-up and Equipment

Measurements were carried out in a commercial cellular operator network in Finland. Both WCDMA and HSDPA technologies were evaluated within the same operator. We used two different measurement set-ups, one specially designed for TCP and one for UDP measurements. The TCP set-up was geared towards revealing user application-level effects, whereas the UDP set-up was focused on getting detailed information on network performance with millisecond resolution.

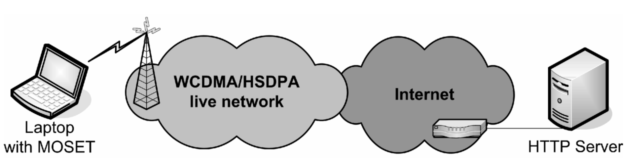

Figure 1. TCP measurement set-up with MOSET tool.

TCP goodput measurements were performed with the MOSET software developed at VTT Technical Research Centre of Finland [5]. In the measurements illustrated in Fig 1, we used HTTP GET messages to start a file download from a web server to a local laptop. The file sizes ranged from 32 B to 2 MB. Goodput (G) was calculated in a straightforward fashion as G(b/s) = S(b) / RTT(s), where S is the payload size and RTT is the application level round-trip-time. RTT was measured as a time difference between sending an HTTP GET request from the terminal to the web server at the moment when the downloaded file was fully received. In total, 1.8 GB of data was transferred with WCDMA (0.9 GB) and HSDPA (0.9 GB) during the measurements.

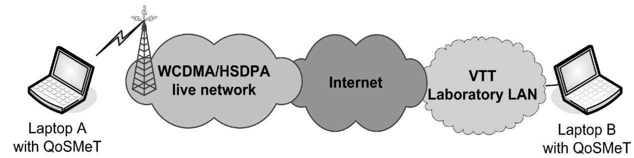

Figure 2. UDP measurement set-up with QoSMeT tool.

UDP goodput measurements were performed with the QoSMeT tool [4, 6], which has also been developed at VTT. The set-up is shown in Fig. 2. QoSMeT monitors one-way packet throughput, delay and jitter by capturing data link-level packets and comparing receiver and sender side Global Positioning System GPS-synchronised packet timestamps. A Mobile IP (MIP) [7] solution was used in order to get our measurement traffic through the operator's firewall. Goodput was calculated from the 1 second average throughput and by omitting the extra headers caused by MIP UDP-in-IP encapsulation. The total amount of headers was: IP (20 B), UDP (8 B), MIP (4 B), IP (20 B), and UDP (8 B), followed by a variable payload size. UDP traffic was generated with the D-ITG traffic generator [8] and with application-level payload sizes from 32 B to 65 kB. The maximum goodput with each packet size was found by offering a 1.1-1.3 Mbps load to the network with the traffic generator. A total of 300 MB and 3.8 GB of data were transmitted with WCDMA and HSDPA respectively.

We also performed delay and jitter measurements with a VoIP application and QoSMeT (Fig. 2). VoIP traffic was generated with SJphone [9], using a GSM voice codec. Without setup, SJphone produces bidirectional constant-bit-rate traffic of 50 packets/s regardless of voice activity. VoIP traffic consisted of voice payload (33 B), MIP UDP-in-IP encapsulation and real-time transport protocol RTP headers (72 B).

Both measurement setups' hardware consisted of laptops running Windows XP with an Option Globetrotter WCDMA or HSDPA data card providing the air interface. The HSDPA card was from category 12, capable of using QPSK modulation and 5 simultaneous codes. The maximum achievable L2 data rate for the data card is 1.6 Mbps [1]. Therefore it did not present a bottleneck in our measurements, since the operator has limited the per-user throughput to 1 Mbps at the network level.

We decided to examine the maximum performance of the network and to see how well it correlates with advertised performance values. Thus the results presented herein are an average of 10 percent of the best measured goodput values within each measurement point (payload size). We also found out that there was a relatively large variation (10-20 %) in the results depending on which time of day the measurements were performed, see Table I for delay variation.

Other users in a network cell can cause unpredictable effects on the measurements. In test measurements, the bandwidth was 1 Mbps with one HSDPA connection, whereas with two simultaneous connections bandwidth was 500 kbps for each connection. This implies that the resources in a cell were shared in round-robin-fashion. Therefore, those measurements where the data rate dropped suddenly were ruled out from the analysis, as they were understood to be another HSDPA user entering the cell.

Due to link adaptation, the transport block size depends on the channel link quality and favors high block sizes in good link conditions [3]. According our data card software, the signal strength was at maximum level and thus large block sizes were presumably used.

IV. RESULTS

A. Goodput performance with TCP and UDP

The measured TCP and UDP goodput is shown in Fig 3. As anticipated, the HSDPA outperforms WCDMA with both UDP and TCP protocols. It can also be noticed that the maximum goodput values are quite close to the advertised throughput of 384 kbs for the WCDMA downlink and 1 Mbps for HSDPA (operator-limited). Naturally, goodput is lower than the actual throughput, since only the effective payload transmission is considered. This difference is illustrated in Fig. 3 in the case of UDP transmission over HSDPA.

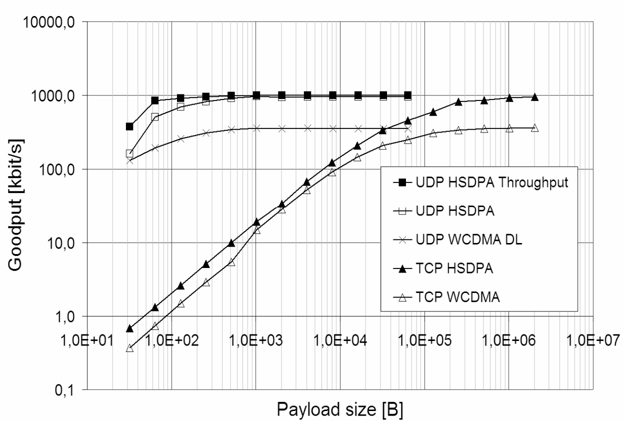

Figure 3. HSDPA and WCDMA goodput for UDP and TCP traffic in a live network. HSDPA UDP throughput is shown for comparison.

With UDP the maximum measured goodput was 955 kbps for HSDPA and 366 kbps for WCDMA. The maximum value was reached with payload sizes above 1 kB and it is constant over the measurement range up to 65 KB. Below 1 kB, UDP goodput is clearly lower, affecting most VoIP applications that use payloads of these sizes. However, UDP goodput is substantially better than with TCP over the measurement range. One of the main reasons for lower UDP goodput below 1 kB is the large proportion of header overhead versus actual payload size. This is especially visible in our measurements due to Mobile IP UDP-in-IP encapsulation. Our results imply that MIP is not optimal with VoIP applications and small payloads. The effects can be even more profound with IPv6 that has larger header sizes than IPv4. However, there are MIP related techniques that can decrease header overhead.

With TCP, payload sizes of hundreds of kilobytes are needed to utilise the wireless link effectively, as can be seen from Fig. 3. For WCDMA, this was verified by our earlier measurements [10]. Compared with WCDMA, HSDPA brings improvement to TCP performance with all payload sizes, but the effect is not as profound as one might expect.

The main reasons for such behaviour are the TCP connection establishment hand-shake and following slow start procedures that consume a large portion of the actual connection time. After the hand-shake, TCP probes available bandwidth by increasing the congestion window size, which increases the send rate if there is more bandwidth available. The longer the RTT, the longer it takes to increase the rate to match the link throughput. Unfortunately, with small payload sizes, the whole content may have been transmitted without reaching the link maximum throughput. In addition, link level retransmissions can falsely be interpreted as congestion by TCP, thus initiating the congestion control algorithms. There are also other factors contributing to TCP performance as discussed in [11].

The strong correlation between the payload size and the application-level goodput naturally has implications at the user level as well. The typical web page contains a number of separate downloadable items ranging from a few kilobytes to hundreds of kilobytes. These payload sizes are in the region where TCP goodput is not optimal. If each of these required an individual TCP-session, it would result in poor user service quality, especially with small payload sizes. However, modern web browsers can overcome this by using HTTP/1.1 persistent connections where the page content is downloaded in one or several simultaneous TCP connections. The content is integrated into larger chunks of payload that are more favourable for link utilisation than small payloads.

It seems that an even bigger improvement is needed for the RTT in order to overcome TCP inefficiency in wireless links. It is good to keep in mind that UDP performance correlates with one-way characteristics, whereas TCP performance is dependent on both up- and down-link performance. For UDP, HSDPA is thus a clear improvement to downlink performance. With TCP, the effect is not as pronounced as with UDP, because WCDMA uplink performance does not match the HSDPA downlink. Fortunately, the upcoming High Speed Uplink Packet Access (HSUPA) technology will improve the situation.

B. Delay

The delay characteristics are tightly linked to network load and with the operator pre-defined or dynamic network settings. The general simplifying assumption is that if the network is crowded, it usually shows as a longer average delay and increased jitter. In the end-to-end path used here, the delay is composed of sender, radio access, radio network, Internet and receiver-related delays. The most dynamic and largest delay source originates from the radio access network, while the other sources can be roughly estimated as being constant. Fig. 4 shows an example of a typical one-way delay measurement performed at 9 a.m. At later hours of the day, the performance was usually somewhat worse. The figure consists of two separate 400-second measurements (WCDMA UL/DL and HSDPA) integrated into one plot. The measurements were done by transmitting VoIP traffic between a terminal connected to the internet via VTT's laboratory LAN and a terminal with a WCDMA/HSDPA connection. The amount of Internet delay is ~20 ms, which was estimated by using the *traceroute* tool. The following figures use receiver time scale, i.e. the time when packets arrived to receiver.

The figure shows that, as expected, the lowest one-way delay is with HSDPA (46 ms), followed by the WCDMA uplink (74 ms) and WCDMA downlink (102 ms). By subtracting Internet delay from both uplink (WCDMA) and downlink delays, we come up to 80 ms for RTT for HSDPA. This result correlate well with 76 ms RTT found in literature [3].

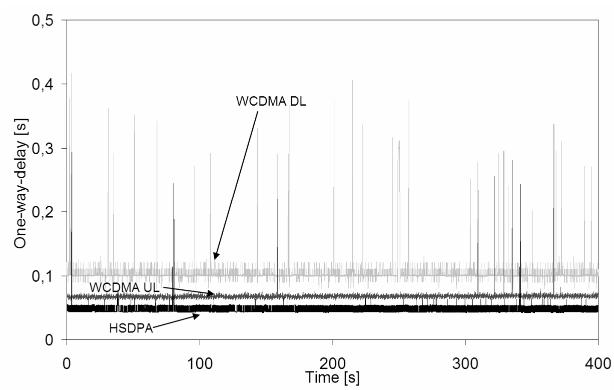

Figure 4. HSDPA and WCDMA UL/DL one-way delay.

A prominent feature of Fig 4 is the large number of delay spikes in the WCDMA downlink, with few spikes in the WCDMA uplink and HSDPA. This behavior also occurred with other measurements. In the spikes related to the WCDMA downlink, the delay varies from 200 – 430 ms, which is rather high considering the requirements of packet-based real-time services. Fig. 5 shows a more detailed structure of delay spikes. Taking a closer look, each delay spike is actually a burst of packets that are received after a non-activity time interval when no packets have been detected by our measurement software. Such behavior can be explained by the Automatic Repeat reQuest (ARQ) in WCDMA [12]. When a packet data unit (PDU) from nodeB is not successfully decoded at the receiver, an RNC retransmission procedure is initiated. While the lost radio link-level data packet is being retransmitted, several other packets can be received and placed in a receiver buffer. Once the lost data PDU is received, the buffer is emptied to the upper layers in one burst as ordered delivery is used. This can be verified from Fig 4 for WCDMA downlink delay spikes. For example, after a 300 ms period of inactivity, 15 packets are received in a burst, which corresponds to the 20 ms send rate used by VoIP.

With HSDPA the retransmissions should not introduce as long delays as with WCDMA. This is mainly because retransmissions are controlled by hybrid ARQ located at nodeB instead of having ARQ at RNC. Also, the short TTI offers a substantial improvement since the retransmission cycle can be finished much faster than with longer TTI. Results show that indeed, number of delay spikes as well as their length is reduced in HSDPA. In Fig 5, we see only two 200 – 300 ms delays with HSDPA, from which the first one has similar shape as with WCDMA delay spikes. We comprehend this delay spike as a consequence of rare situation when HARQ fails and retransmission has to be performed by RNC. The other HSDPA delay spikes have the shape where delay is first increasing packet by packet and then decreased to base level, but not in one burst. This behavior could be related to HARQ and/or RNC retransmissions.

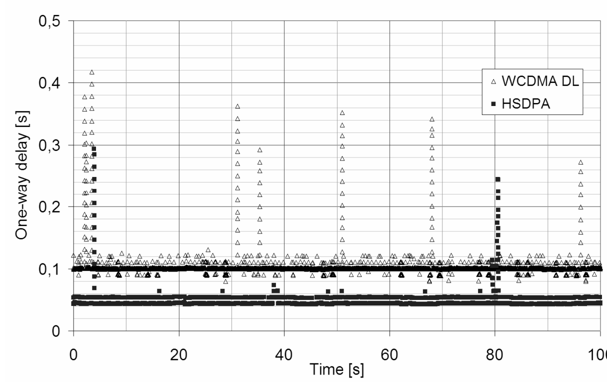

Figure 5. One-way delay with WCDMA downlink and HSDPA.

Our measurement results are summarised in Table 1. 20 measurements for WCDMA and 20 for HSDPA were performed between 6th of June and July 3rd. Each measurement period was 10 minutes and was done at different times of a day between 9 a.m. and 4 p.m. First average delays were calculated from each 10 minute measurement and then minimum, average and maximum values among these were extracted to Table 1.

TABLE I. MEASURED GOODPUT AND DELAY

	Max Goodput [kbps]	***Min delay*** [ms]	***Average delay*** [ms]	***Max delay*** [ms]
WCDMA DL	366	98	111	131
WCDMA UL	n/a	64	76	95
HSDPA	955	44	50	61

In the light of our results in Table 1 and experiences learned during the tests, some remarks can be made. Delay and throughput affects one or two-way traffic types differently. With UDP, non real-time video applications can compensate low one-way throughput and delay with buffering techniques. For VoIP the one-way delay in cellular network (usually below 130 ms) provides adequate VoIP quality. As matter of fact, there is evidence suggesting that delays of up to 600 ms can be tolerated by experienced VoIP users without a significant degradation in the conversational quality [14]. For real time applications such as push to talk (PoC) or gaming, the delay has big impact to user experience. For two-way traffic such as TCP, the delay is a crucial factor since downlink throughput can be limited by the uplink delay in handshake procedure or in acknowledgements, especially with small payload sizes.

During the measurements we found cellular network related behaviour that affects user experience. Mobile terminals can be in different activity modes, for instance, idle mode can be used to reserve battery. Also, there are throughput or packet size thresholds to assign the user with shared, dedicated channel or to dedicate user with either WCDMA or HSDPA. There are also timers that expire after inactivity period and change the connection from HSDPA to WCDMA.

C. Jitter

HSDPA jitter is extracted to Figure 6. The HSDPA delay does not vary considerably; the jitter is mainly concentrated at 0 and 10 ms, except for RNC retransmission-originated delay spikes. Intuitively, one would expect to have jitter levels corresponding either 2 ms TTI or HARQ related minimum retransmission interval of 12 ms [1]. One explanation for 10 ms jitter could be related to *Iub* flow control [1, 13]. With HSDPA, there are two data buffers, one with nodeB and one with RNC. Depending on flow control mechanism, the data is transferred between buffers in certain time interval. If the vendor specific flow control parameterisation is the main cause for jitter behaviour, our results would imply that Iub flow control interval was 10 ms. This jitter level was also found with WCDMA uplink and downlink measurements, where 10 ms TTI has to be taken into account as well as Iub flow control. With good signal conditions and with low interference from other users, there may be no need for retransmissions. It is also possible that our measurement set-up with 20 ms packet interval was too coarse to show detailed HARQ behaviour.

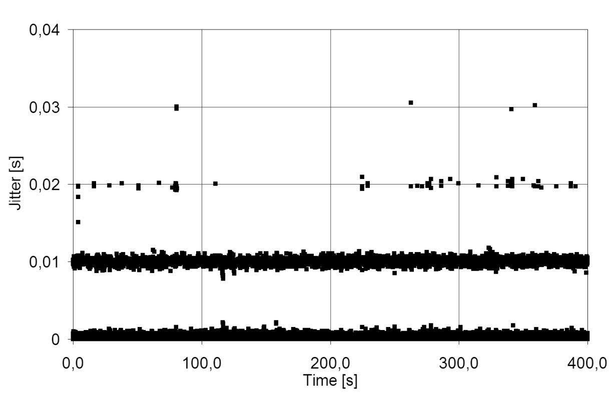

Figure 6. HSDPA jitter.

V. SUMMARY AND CONCLUSION

We evaluated the user-perceived Quality of Service in a live operational network utilising WCDMA and HSDPA technologies. We found out that, as expected, HSDPA offers a clearly higher throughput and lower delay compared to WCDMA. The measured maximum goodput values correlate well with advertised performance values in HSDPA and WCDMA. There is a strong correlation between payload size and goodput. This is especially important with TCP, where the maximum capacity of the link is reached only with relatively large payload sizes even with HSDPA. The upcoming integration of high-speed uplink HSUPA to operator networks is likely to improve TCP performance in this region. With UDP the present situation is clearly better and already payload sizes above 1 kB reach maximum goodput. Below 1kB, UDP performance is not optimal, which may affect VoIP applications especially when used together with MIP. We also found that link-level retransmissions in WCDMA create delay spikes of order of hundreds milliseconds. With HSDPA and HARQ retransmissions the delay spikes are very rare or below our measurement set-up resolution. This implies that the user-perceived service level is more stable with HSDPA than with WCDMA.

Further work is needed to analyse the long-term variation in network performance, the details of HARQ related jitter as well as the effect of upcoming HSUPA technology for TCP performance.

ACKNOWLEDGMENT

We are grateful to Dr. Harri Holma from Nokia and Dr. Kostas Pentikousis, Dr. Martin Varela from VTT for commenting this work. This work was supported by EUREKA ITEA Easy Wireless project, partially funded by Finnish Funding Agency for Technology and Innovation (Tekes).

REFERENCES

[1] H. Holma, A. Toskala (Editors), *HSDPA/HSUPA for UMTS*, John Wiley and Sons, first edition, 2006.

[2] S. Parkvall, E. Englund, M. Lundevall and J. Torsner, *Evolving 3G Mobile Systems: Broadband and Broadcast Services in WCDMA*, IEEE Comm. Mag., vol. 44, Feb. 2006, pp. 68-74.

[3] H. Holma and J. Reunanen, "3GPP Release 5 HSDPA Measurements", in proc. PIMRC'06, Helsinki, Finland, September 11-14, 2006.

[4] J. Prokkola, M. Hanski, M. Jurvansuu and M. Immonen, *Measuring WCDMA and HSDPA Delay Characteristics with QoSMeT*, in proc. of ICC'07, Glasgow, UK, 24-28 June 2007.

[5] M. Palola, M. Jurvansuu, and J. Korva, *Breaking down the mobile service response time*, in Proc. IEEE ICON 2004, Vol. 1, Nov. 2004, pp. 31 – 34.

[6] M. Hanski, *A Tool for Monitoring End-to-end Quality of Service in IP Networks*, Master's thesis, University of Oulu, Department of Electrical and Information Engineering, Oulu, Finland. 2005.

[7] Secgo Mobile IP software, http://www.secgo.com/

[8] Distributed Internet Traffic Generator, http://www.grid.unina.it/software/ITG/

[9] SJphone, http://www.sjlabs.com/

[10] K. Pentikousis, M. Palola, M. Jurvansuu and P. Perälä, *Active Goodput Measurements from a Public 3G/UMTS Network*, IEEE Comm. Lett, Vol. 9, issue 9, Sept. 2005 pp. 802-804.

[11] J. Alcaraz, F. Cerdan and J. Garcia-Haro, *Optimizing TCP and RLC interaction in the UMTS Radio Access Network*, IEEE Network, March/April 2006, Vol. 20, iss. 2, pp. 56- 64.

[12] J. M. Cano-Garcia, E. Gonzales-Parada and E. Casiliri, *Experimental Analysis and Characterization of Packet Delay in UMTS Networks*, in Proc. of NEW2AN 2006, St. Petersburg, Russia, May/June 2006.

[13] M. C. Necker and A. Weber, *Parameter Selection for HSDPA Flow Control*, in Proc. of the 2nd International Symposium on Wireless Communication Systems, Siena, 5-7 Sept. 2005, pp. 233 – 237.

[14] Martín Varela *"Pseudo-Subjective Quality Assessment of Multimedia Streams and its Applications in Control"*, Ph.D. thesis, November 2005. University of Rennes 1, France.