Project WISQY: A Measurement-based End-to-End Application-Level Performance Comparison of 2.5G and 3G Networks

Peter Reichl
ftw. - Telecommunications Research Center Vienna
Donaucitystr. 1,
A-1220 Vienna, Austria
reichl@ftw.at

Martina Umlauft
ftw. – Telecommunications Research Center Vienna
Donaucitystr. 1,
A-1220 Vienna, Austria
umlauft@ftw.at

Joachim Fabini
Technical University Vienna
IBK (E388)
Favoritenstr. 9/388
A-1040 Vienna, Austria
joachim.fabini@tuwien.ac.at

Reinhard Lauster
Kapsch CarrierCom AG
Am Europlatz 5
A-1120 Vienna, Austria
lauster@kapsch.net

Günther Pospischil
Mobilkom Austria AG & Co KG
Obere Donaustrasse 29
A-1020 Vienna, Austria
g.pospischil@mobilkom.at

Abstract

The imminent transition from 2.5G to 3G networks is generally considered to be a keystone for the future economic success of mobile telecommunications. This paper presents a practical comparison of GRPS and UMTS network performance, based on measurements in the live network of Mobilkom Austria and comparing the results to dedicated lab trials. Starting from raw TCP/IP download and upload transfer rates and ICMP Ping round-trip-times, we end up with page response times and file transfer rates for simulated HTTP/1.0 scenarios. The presented results are especially interesting for a realistic calibration of forthcoming 3G and beyond simulation tools.

1. Introduction

Following the tremendous success of 2G/2.5G mobile networks like GSM or GPRS, for the next couple of years the prosperous economic future of the telecommunications industry will depend heavily on the transition towards 3G and beyond networks, e.g. UMTS with its enormous expected performance gain compared to GPRS. Whereas from a conceptual perspective, standardization has already made remarkable progress towards this goal (especially within 3GPP), in most countries the practical implementation of 3G networks is well behind schedule. There are only few exceptions, one of them being most notably Austria, where Mobilkom Austria has become the first European network operator to launch a national UMTS network in September 2002, followed by the customer launch for this 3G network in April 2003. This has provided us with the unique opportunity to gain realistic measurement data from a live 3G network. In this paper we report on the main results of this evaluation which has been performed within the project "WISQY" (Wireless Inter-System Quality-of-Service) [1], an application-oriented research activity undertaken at the Telecommunications Research Center Vienna, Austria, as part of the Austrian Kplus Competence Center program.

The UMTS/GPRS measurements in Mobilkom's live network (generally known as "A1") have additionally been repeated in a dedicated laboratory environment provided by Kapsch CarrierCom (KCC), an Austrian system synnovator of communication technology solutions for fixed, mobile, and data network operators. The KCC test system is a simple but complete and fully functional GSM/GPRS and UMTS network (conformant to 3GPP R99 specifications), which can be used for end-to-end tests with real GSM/GPRS or UMTS equipment. Note that, because of KCC's long-standing system integration relationship with Mobilkom Austria, there is an almost perfect match between the A1 network and the KCC lab in terms of hard- and software.

Beyond providing a realistic comparison of GPRS and UMTS performance, our quantitative results will be especially useful for the future configuration of realistic simulation tools for 3G networks.

Applicability of such tools is highly dependent on a careful parametric calibration derived from live network tests involving the transfer of huge amounts of data.

The remainder of the paper is structured as follows: Section 2 introduces testbed architecture, measurement scenarios, and evaluation tools. Section 3 presents the UMTS/GPRS live and lab performance evaluation results for TCP/IP, ICMP and HTTP/1.0. Section 4 concludes the paper with summarizing remarks.

2. Testbed setup and tools

Figure 1 sketches the setup for our live UMTS/GPRS network measurements. As mentioned before, we tested Mobilkom's live UMTS network with standard APN (DCH: DL 384 kbps, UL 64 kbps). We used a Nokia 6650 UMTS phone, which has been connected to the mobile client (HP Omnibook laptop with Windows XP SP1) via USB cable, thus preventing bandwidth bottlenecks and other potential distortions due to IrDA or bluetooth connections. For the GPRS measurements, we used GPRS mobile phones like the Siemens S55 or TelMe T919 which support GPRS class 10 (total of 5 timeslots, dynamically 4DL + 1UL or 3DL + 2UL).

All application servers have been running on a SUSE Linux-based server, which has been connected to the TUNet IPV4 network (for live measurements) and to the KCC Gi LAN (for lab measurements). RNC, SGSN, GGSN, VLR, and HLR were located in the A1 network environment (for live measurements) and in the KCC UMTS test network environment (for lab measurements). In the KCC lab, all tests have been performed as exclusive single-user tests, giving no other mobile phones the possibility to allocate resources within the mobile network or create interfering traffic. In order to guarantee undisturbed radio conditions, the mobiles have been put into an RF shield box. Finally, also the Gi LAN has been kept free of any external traffic.

In order to derive representative results, all live measurements have been evenly distributed over a period of more than a week. The measurements were performed using a stationary mobile (i.o.w. no mobility.)

We have used IPerf 1.7.0 [2] to measure raw TCP/IP network throughput and the standard ping utility to measure ICMP packet round-trip times (RTT) with varying payload size. Local and remote systems were synchronized via SSH client/server communications. Automated testing was implemented using the Windows task scheduler and Perl scripts.

To generate realistic HTTP traffic, we have used the WebSim traffic generator developed at FTW [3]. WebSim consists of a client and a server which together simulate HTTP/1.0 traffic according to the SURGE model [4]. While the WebSim server simulates an HTTP server, the WebSim client emulates user behavior by requesting files according to the HTTP/1.0 standard. Pages typically consist of an HTML document with several embedded objects (eg. pictures embedded with the <IMG> tag.) With HTTP/1.0 browsers open a new TCP connection for each embedded object, usually multiple simultaneous TCP connections in parallel. WebSim also simulates user think times between page requests and parsing times of the browser. For further information on the measurement setup, we refer to [5].

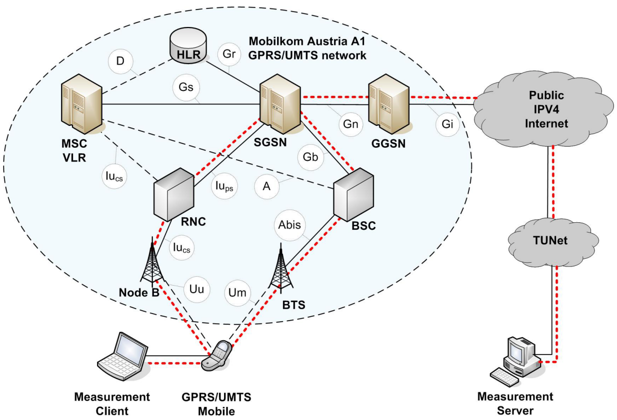


Figure 1: Testbed setup for A1 live UMTS/GPRS network

3. Measurement results

As pointed out above, our performance analysis reflects the current status of GPRS and UMTS networks from an end-user perspective, both in live networks and in test laboratory environments. Relying on a strict end-user view, we regard the mobile network as a black box with no user-configurable parameters and restrict optimization to the selection and configuration of mobile devices and to the fine-tuning of operating system parameters.

Note that following the technical specifications, UMTS networks should outperform GPRS networks by at least one order of magnitude both in terms of throughput and round-trip-time. On the other hand, UMTS networks have just started to become operational, whereas GPRS networks are already highly optimized; this might reduce our results concerning the benefit end-users can expect from a high-performance infrastructure like UMTS compared to GPRS as of yet.

3.1. TCP/IP raw network performance

Our first comparison of GPRS and UMTS investigates TCP/IP throughput of 30 second bursts with respect to the TCP window size. From Figure 2 we can derive how TCP/IP throughput in GPRS networks depends on the number of GPRS timeslots allocated for upload, for download, and the GPRS coding scheme. As a secondary conclusion we see that with satisfying conditions for the GSM provisioning channel, the user equipment etc. has only minor impact on TCP/IP throughput. Note that with GPRS, the theoretical maximum number of timeslots available to the user for data transfer is limited by the mobile device's GPRS class and by the maximum number of GPRS timeslots that the operator grants to one mobile device in uplink and in downlink direction. Depending on the current network load an operator can dynamically decide to allocate fewer timeslots to the user. In our case, the A1 GPRS network fully supports mobile phones up to GPRS class 10, thus enabling the allocation of a total of 5 timeslots, either as 4 download + 1 upload or as 3 download + 2 upload. Depending on the predominant data transfer direction, the allocation changes dynamically. From Figure 2 we may conclude that the A1 network uses the (4+1) scheme for TCP/IP download and switches to (3+2) for the upload. Combined with GPRS coding scheme CS 3/4, the A1 network transfers an average of up to 60 kbit/s.

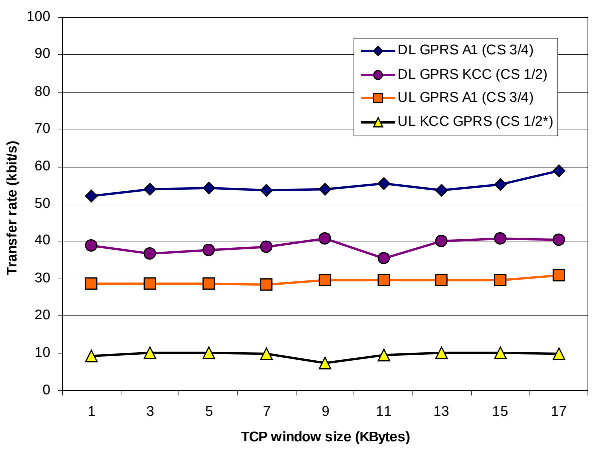


Figure 2: IPerf TCP/IP: download and upload transfer rates for GPRS

These results have been compared to measurements in the KCC GPRS test network, which has been configured to support GPRS class 8 with CS 1/2, i.e. a maximum of 4 timeslots for download and only 1 timeslot for upload. The resulting upload and download throughput for GPRS networks with CS 1/2 and CS 3/4 is presented in Figure 2. Note the remarkable performance gain of GPRS CS 3/4 compared to CS 1/2 both in the upload and download direction. The operator restriction simulated in the upload direction (where the GPRS network grants only one upload timeslot to the mobile) shows how conservative operator settings can impact GPRS performance. In Figure 2 the TCP/IP upload throughput for KCC amounts to half the value the mobile device is theoretically capable of at CS 1/2.

Figure 3 demonstrates that UMTS outperforms GPRS both in TCP/IP upload and download direction (note that Figure 3 differs from Figure 2 in terms of the y-axis scale by a factor of five). We see that UMTS upload throughput peaks at 55 kbit/s, equivalent to about twice the throughput of a GPRS mobile with CS 3/4 and full operator support (two timeslots in upload direction).

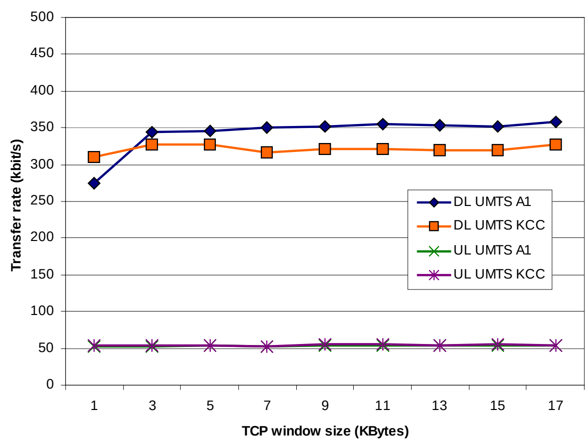


Figure 3: IPerf TCP/IP: download and upload transfer rates for UMTS

In contrast, our UMTS download measurements show a peak of 357 kbit/s throughput, i.e. close to the theoretical throughput limit for UMTS network micro-cells. Finally note that the KCC UMTS Lab network configuration has not been optimized for maximum throughput, resulting in slightly lower download results. The UMTS download throughput amounts to six times the maximum GPRS download performance.

3.2. ICMP round-trip time

The focus of our next experiment was on average ICMP packet round-trip times (RTTs) for varying packet payloads (the payload being identical for sent and returned packets.) In Figure 4 we show the average, minimum, and maximum ICMP RTT for GPRS and UMTS networks. The complete ICMP test has been performed 65 times for any payload size between 100 and 1450 bytes, distributed over a one-week interval, and each resulting point averages a total of 20 ping packets. Thus, the results are a good approximation of expected RTTs for UDP packets over GPRS and UMTS networks.

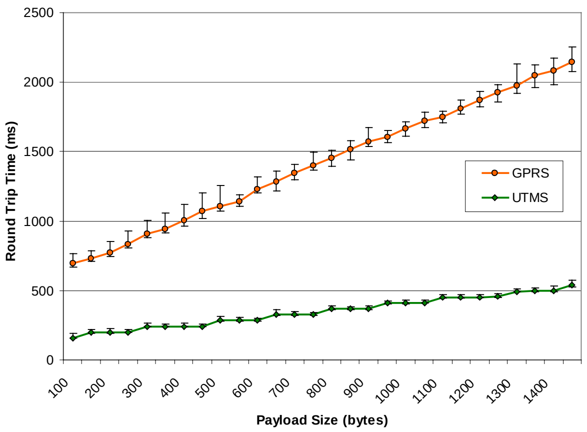


Figure 4: ICMP ping RTT for UMTS and GPRS

Compared to GPRS, the round-trip performance that we measured for UMTS networks can be considered excellent, as UMTS ranges from less than 150 ms for 32 byte and 160 ms for 100 byte packets to 537 ms for 1450 byte packets while GPRS RTT peaks at 2142 ms for 1450 byte packets.

As a practical conclusion, typical SIP/IMS messages carrying a payload of around 800 bytes will incur a live UMTS network RTT of approximately 360 ms with a maximum jitter of approximately 22 ms (-5ms + 17ms). Note finally that the measurement results for UMTS are much more deterministic than for GPRS, leading to only very small differences between the minimum and maximum RTT for UMTS as shown in Figure 4.

3.3. Web response times and transfer rates

Based on the previous results on raw TCP/IP performance, we chose HTTP/1.0 for our application-level performance evaluation. Figure 5 shows response times (download latency) for complete HTML pages including the transfer of embedded objects for GPRS (top) and UMTS (bottom).

As can be seen in the figure, GPRS, with a minimum of 1.3 sec is still above the important perception barrier of 1 sec for "fast" responses. In UMTS page response times as low as 300 ms (which is approximately the human reaction time) can be achieved and a significant number of responses is below the important 1 sec mark.

Of course, there is still room for improvement, esp. for pages containing several small embedded files. With HTTP/1.0 which opens a separate TCP connection for each file such pages suffer a significant decrease in performance.

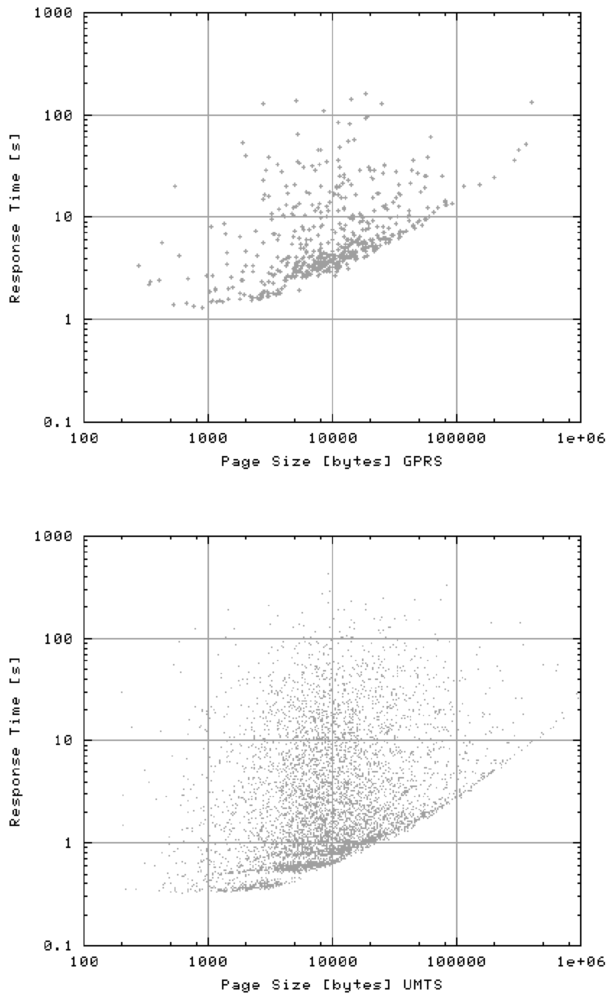
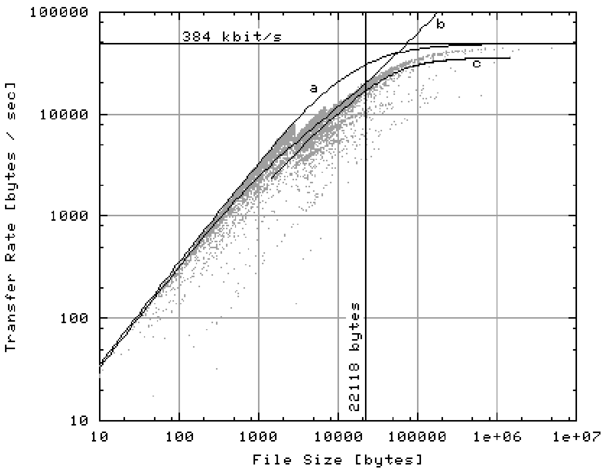


Figure 5: Page response times for HTTP/1.0: GPRS (above) vs. UMTS (below)

Figure 6 shows the perceived transfer rates for UMTS as measured for single files (top) and complete pages made up of these files (bottom).

To judge the quality of the network we first derive an upper bound on the achievable performance as follows: ideally, the download of a file of length x would take $t_{DL}^{ideal} = \frac{x}{B}$ with nominal bandwidth B = *384 kbps*. Since the TCP 3-way handshake and start of the file transfer add an initial delay of at least τ = *2*RTT*, the perceived transfer rate is bounded with $\tilde{R} = \frac{x}{\tau + x/B}$. Using the RTT of 150 ms from our ICMP measurements, this bound is depicted as curve (a) in Figure 6.



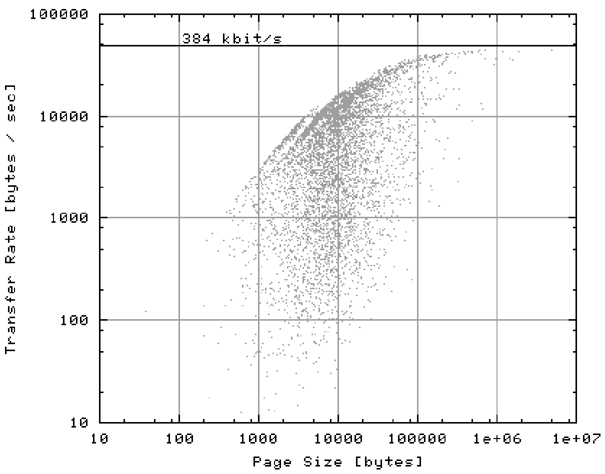


Figure 6: Perceived file (above) and page (below) transfer rates for HTTP/1.0 in the UMTS live network

To account for TCP Slow Start we compare our measurements to the optimal model derived in [6], Equation (1) (indicated by curve (b) in Figure 6). Note that this model is valid up to a file size of $x = \frac{B * RTT * r - w}{r - 1} = 22118\, bytes$ for our bandwidth-delay product of *384 kbps*150 ms* and the typical values of r = *1.5* and w = *1* for standard TCP, which is perfectly matched by our results.

To benchmark our results for filesizes larger than 22118 bytes, we compare our data to the model in [7] which is an extension of [8]. Estimating p with the well-known $1/\sqrt{p}$ law [9] yields $p = \left(\frac{MSS * C}{B * RTT} \right)^2 = 0.0392$. Using this we have plotted the model of [7] as curve (c) in Figure 6. As can be seen, the models work well for a UMTS network under low load using RLC AM (acknowledged mode) and a stationary mobile.

4. Conclusions and future work

In this paper we have reported on a couple of measurement results allowing detailed insight into the end-to-end application-level performance of live 2.5G and 3G networks. After providing quantitative results on TCP/IP throughput and ICMP Ping RTTs, we focus on HTTP response times and transfer rates, comparing our measurements with related modelling work.

As shown, the UMTS live network yields a TCP performance according to the estimation models for fixed networks [6, 7, 8, 9] – given low load, a fixed mobile and RLC AM. Therefore, to further improve user QoS perception, efforts should focus on RTT reduction and improvements in application level protocols (e.g. using HTTP/1.1 instead of HTTP/1.0.) No measurements for the high-load, moving mobile case were performed but it can be expected that spurious RTOs due to RTT jitter caused by excessive retransmissions on the RLC layer or handover will have significant impact on TCP performance. Future work in the course of ftw project N0 [10] will address these issues.

For further work performed within the WISQY project (e.g. IPv6-based mobility management) we refer to [1]. Current and future work deals with applying our tools and measurement methods for IP applications in the context of the future 3GPP IMS (IP Multimedia Subsystem).

5. Acknowledgements

This work has been funded by the Austrian Kplus competence center program.

6. References

[1] A3-WISQY: Wireless InterSystem QoS. URL: www.ftw.at/projects/A3_engl.html.

[2] NLANR/DAST Iperf Version 1.7.0. URL: http://dast.nlanr.net/Projects/Iperf/ , March 2003.

[3] E. Hasenleithner, H. Weisgrab, T. Ziegler: *A0 Deliverable 5-2, Final Report Testbed Phase 1*. FTW Technical Report A0 D.5-2, Vienna, Austria, Feb. 2002.

[4] P. Barford, M. E. Crovella: *Generating Representative Workloads for Network and Server Performance Evaluation*. ACM Sigmetrics '98, pp. 151-160, Madison (WI), 1998.

[5] P. Krüger, N. Jordan, M. Teufel: *WISQY Test Scenario Definition.* Techn. Report FTW-TR-2004-012, Feb 2004.

[6] N. Cardwell et al.: *Modeling the Performance of Short TCP Connections*, Technical Report, Computer Science Department, University of Washington, Nov. 1998.

[7] N. Cardwell, S. Savage, T. Anderson: *Modeling TCP Latency*. INFOCOM 2000, Tel Aviv, Israel, March 2000.

[8] J. Padhye, V. Firoiu, D. Towsley, J. Kurose: *Modeling TCP Throughput: A Simple Model and its Empirical Validation.* ACM Comp. Comm. Rev., pp 303-314, 1998.

[9] M. Mathis, J. Semke, J. Mahdavi, T. Ott: *The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm*, ACM Comp. Comm. Rev., ACM SIGCOMM 27(3), July 1997.

[10] N0: Mechanisms and Performance of Packet Networks, www.ftw.at/ftw/research/projects/ProjekteFolder/N0