TD(05)002

2nd MC Meeting, Colmar, France, Feb. 3-4 2005

Performance evaluation of internet access over a precommercial UMTS network

M. Catalan, C. Gomez, D. Viamonte, J. Paradells, A. Calveras, F. Barcelo
Dept. of Telematic Engineering, Technical University of Catalonia (UPC),
Jordi Girona 1-3, C-3, 08034 Barcelona, Spain
{mcatalan, carlesgo, viamonte, teljpa, acalveras, barcelo}@entel.upc.es

Abstract

3G wireless networks such as UMTS are currently providing wide area Internet access wth bit rates close to those from ADSL. However, the Internet protocols exhibit a suboptimal performance in such scenarios. In this paper we first characterize a live UMTS link from the IP level. We identify critical features such as high latency, delay spikes, high packet loss probability and connectivity gaps, especially when the user moves between cells. Next, we test a set of TCP configurations, demonstrating that TCP performance can be optimized over UMTS by the usage of adequate end-to-end mechanisms and parameter values. The last part is devoted to the evaluation of services such as web access, streaming or remote terminal on top of UMTS bearers. The most relevant conclusion is the improvement in terms of throughput and latency compared to GPRS. This improvement allows using services not correctly carried out with previous cellular solutions.

Keywords

Experimentation with real networks, TCP, IP, UMTS, Internet services, web browsing, streaming, optimization

Working Group 1

I. Introduction

Significant advances in wireless technologies are building the path towards an ubiquitous Internet. Nowadays, mobile users can access the Internet through Third Generation (3G) networks, such as Universal Mobile Telecommunications System (UMTS), providing data bearers designed to address the needs of a mobile Internet user in the wide area.

The Internet as a fixed network has experimented a big success that relies on the two elements that constitute the heart of the Internet protocol stack: the Internet Protocol (IP) [1] and the Transmission Control Protocol (TCP) [2]. IP is a network layer protocol that contains addressing and control information, enabling packets to be routed. IP has two primary responsibilities: providing connectionless, best-effort delivery of datagrams through a network and performing fragmentation and reassembly of datagrams to support data links with different Maximum Transmission Unit (MTU) sizes. At the transport layer, TCP provides reliability to data delivery between two end entities over an IP network. The TCP behaviour is a key factor for application performance and user experience, since Internet applications are mainly based on the usage of TCP.

In this paper, we focus on analyzing and optimizing the performance of IP and TCP over a live UMTS network by testing a set of different protocol configurations. Organizations such as IETF and WAP Forum have already produced TCP parameter configuration proposals for its usage over 2.5G and 3G networks [5],[10]. We have only been able to compare our results with those provided by the xMotion project [11] and a recent paper by Chakravorty et al. [19], since all the rest of related information available is mainly based on simulation studies. The last part of the paper is devoted to the evaluation of services such as web access, streaming or remote terminal on top of UMTS bearers. These tests allow us to derive subjective conclusions, closer to the user perception. The most relevant conclusion is the improvement in terms of throughput and latency compared to GPRS and the benefit of soft handover in front of the GPRS cell reselection that introduces long connectivity gaps.

The rest of the article is organized as follows. In the next section we survey and discuss the main IP and TCP parameters and mechanisms which may affect data performance over 3G links. Section III describes the real scenarios where our trials are held. Section IV is devoted to an IP level characterization of the tested live UMTS network. In Section V we test a set of TCP parameter configurations, taking advantage of the knowledge of the IP level UMTS link behaviour. Section VI covers the tests of web browsing, streaming and remote terminal with UMTS packet data bearers. Finally, we summarize our conclusions in Section VII.

II. Relevant IP and TCP parameters and mechanisms

This section describes the parameters and mechanisms that may have some impact on the performance of IP and TCP protocols over UMTS. We focus on the ones that can be configured under virtually any operating system.

A. IP parameters and mechanisms

1) Maximum Transmission Unit (MTU). In an error-prone link, a small MTU value increases the chance of successful transmission, since the frame damage probability is reduced. Furthermore, in low speed links large MTU values can affect other flows in which human perceptible delays (100-200

milliseconds) are relevant, such as interactive or real time applications. On the other hand, if link layer protocols assure low Bit Error Rate (BER) values by the usage of Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC) mechanisms, larger MTU sizes can be appropriate to balance overhead versus data information [3], [5].

2) Fragmentation. Fragmentation at the IP level is not recommended because it increases processing time. Furthermore, the loss of a single fragment implies the retransmission of the entire original datagram. Reference [5] recommends the usage of Path MTU (PMTU) Discovery mechanism [4] in order to avoid fragmentation over 3G links.

B. TCP parameters and mechanisms

1) Maximum Segment Size (MSS). The MSS value must fit into the IP MTU in order to avoid fragmentation at the TCP layer. Therefore, the choice of an appropriate MSS depends on the MTU used at the IP layer. However, the MSS may interact with some TCP mechanisms. For example, since the congestion window is counted in units of segments, large MSS values allow TCP congestion window to increase faster.

2) Maximum transmission window. This parameter is defined by the Reception WINdow (RWIN) advertised to the sender in each TCP segment. The maximum transmission window must be adequate to the Bandwidth Delay Product (BDP) of a path to fully utilize its capacity. However, authors in [15],[16], [17] suggest using larger windows over 2.5G and 3G networks in order to assure maximum link utilisation even if the sender enters the Fast Recovery phase [6]. BDP is also related to the MSS, since the number of segments in the transmission window affects TCP congestion mechanisms; at least a window size of four segments is needed for applying Fast Retransmit and Fast Recovery mechanisms.

3) Selective Acknowledgements. This option is useful when multiple losses in a single window occur, allowing to recover the lost segments in a single Round Trip Time (RTT). Published results [12] aligned with IETF and WAP Forum recommendations [5],[10] state that SACK option helps improving performance over GPRS networks.

4) Timestamps Option. In 3G networks, this option may be useful since a more accurate Retransmission TimeOut (RTO) should help avoid spurious retransmissions and the activation of congestion avoidance mechanisms. Nevertheless, the effect of the twelve byte overhead on TCP header should be analysed.

5) Initial window. When starting a TCP connection there is a initial windows that by default has de size of two segments. It is possible to increase the size [7] to speed up the slow start phase of a TCP connection and it is convenient when having lots of shorts TCP connections. Usually this is a parameter difficult to modify.

III. Test scenarios

Next, we describe the scenarios and equipment involved in our live UMTS network tests.

A. Scenarios

Our tests are held in two scenarios, namely: Static Scenario and Mobile Scenario. Results obtained in static scenarios constitute an upper bound on the expectable IP performance. Mobile scenarios may perform worse since critical procedures such as handovers must be carried out if the user moves from one cell to another one. Furthermore, the radio signal may experience problems due to multipath, fading and other mobility related phenomena in a varying environment, leading to poor Signal to Interference (S/I) ratio values.

1) Static Scenario. It consists of an indoor static client accessing a local server. The client is a notebook connected to the Internet via a precommercial Spanish public UMTS network with the Background Quality of Service (QoS) class and using a dedicated channel. The Received Signal Strength Indication (RSSI) varies between – 105 dBm and – 81 dBm.

2) Mobile Scenario. The endpoints and UMTS access are the same as the ones used in the Static Scenario. We perform trials in an urban environment such as the one shown in Fig. 2, where the location of two node Bs, as well as the path followed by the mobile client, is shown in the two subscenarios defined next.

2.1) Pedestrian Mobile Scenario. Trials are held with a mobile client walking outdoors, moving from spot 1 to spot 2 and back again as shown in Fig. 2. The RSSI measured along the quoted path falls in the range from -97 dBm to -65 dBm.

2.2.) Vehicular Mobile Scenario. In this case, the client moves by car at an average speed of 30 km/h, following the dashed path indicated in Fig. 2. RSSI values inside the car are in average 4 dBm smaller than those observed in the previous scenario.

It is relevant to note that macrodiversity was not supported between the two node Bs involved in the Mobile Scenario at the moment of the trials.

B. Equipment

In our tests, the notebook is connected to the Internet via a Beta version of a Merlin T510 UMTS PCMCIA card, which supports 64 kbps in the uplink and 384 kbps in the dowlink. The notebook runs Microsoft Windows XP Professional v5.1.2600 – Service Pack (SP) 1. The local server acting as the fixed endpoint runs Microsoft Windows 2000 Server v.5.00.2195 - SP4. TCP/IP settings not explicitly mentioned are the default ones in all cases.

IV. IP level characterization and optimization over UMTS

We devote this section to an analysis of the perceived behavior of UMTS from the IP layer and the influence of parameters and mechanisms of this protocol. It must be noted that effects of fragmentation are not presented on this paper since it has not occurred in any of the test cases. We describe our measurement methodology in part A and provide our results over UMTS in part B.

A. Definition of IP measurements

The target of our trials is the following set of performance metrics: end-to-end delay, jitter, throughput and round trip loss probability of IP traffic.

1) Static Scenario measurements. Two main kinds of experiments are performed in this scenario as described next.

1.1) Delay measurements. They are based on an RTT analysis using the ping tool in both uplink and downlink directions. Note that RTT measured at this level includes transmission time components in both uplink and downlink directions. Several experiments are held, using the following set of MTU values: 128, 256, 512, 1024 and 1500 bytes. In each experiment, 100 ping packets are sent, at a rate of one packet per second. The waiting time that determines when a packet is considered to be lost is set to 10 seconds.

1.2) Throughput measurements. Trials are held to measure the obtained IP level throughput at the mobile side in both cases, uplink and downlink, under overflow conditions, that is, sending data at a higher rate than the capacity of the UMTS link. For this purpose a User Datagram Protocol (UDP) [9] flow of packets is sent at a constant rate for the set of MTU values already mentioned.

1.3) Round trip loss measurements. Loss percentage for UMTS is derived from unsuccessful ping trials.

2) Mobile Scenario measurements. Our goal is to analyse to what degree delay jitter and losses increase with respect to those obtained in static scenarios. In particular, cell reselection and handover mechanisms are critical situations that require special attention since connectivity gaps and severe delay fluctuations may appear in such circumstances.

Measurements are derived from experiments sending a UDP stream of packets to the mobile side at a constant data rate avoiding wireless link capacity overload in the Mobile UMTS scenario. The sending time between consecutive packets is set to 100 ms.

We measure gap times as the difference of times between the instant in which the first packet is received after a burst of lost packets and the instant in which the last packet is received before that burst. We extract delay jitter measurements from the Packet Interarrival Delay (PID) at the mobile side.

B. IP level results

1) Static Scenario

1.1) Delay measurements. Uplink and downlink ping trials results obtained in the Static Scenario are shown in Fig. 1. We can state that RTT grows linearly with the packet size and, in some cases, the RTT standard deviation is higher than 50% of the mean value, probably due to temporary poor radio conditions. A significant conclusion is that interactive applications are expected to perform better over UMTS than over 2.5G technologies such as GPRS [12], since UMTS RTT values are close to achievable values in the fixed Internet. For example, the mean RTT with 128 byte packets is 205 ms. Therefore, UMTS may support services such as Voice over IP (VoIP) applications, since typical one-way delay requirements for VoIP services are below 150 ms. Note that our experiments are performed using the Background QoS class. UMTS defines a Conversational class which may be more suitable for voice communications.

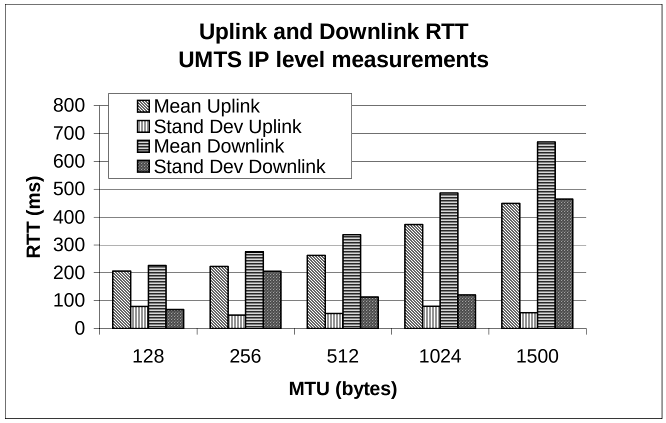

Figure 1. RTT measurements of uplink and downlink ping trials over UMTS

1.2) Throughput measurements. As Table I illustrates, the maximum IP level throughput value for the uplink is around 56 kbps, whereas for the downlink, the top throughput is around 372 kbps. These results agree with theoretically expected values taking into account nominal capacities, radio interface protocol overhead and other inefficiency factors. In any case, the measured IP bitrate figures open the door to the possibility of deploying certain services which require a higher amount of bandwidth than that available with GPRS, with speeds under 50 kbps [12], [15]. One example is Packet-switched multimedia Streaming Services (PSS).

Another conclusion is that overhead due to packet headers degrades throughput performance for small MTU values in both uplink and downlink measurements, as shown in Table I.

MTU (bytes)	Uplink throughput (kbps)	Downlink throughput (kbps)
128	45.1	230.5
256	51.4	318.5
512	55.6	356.8
1024	56.6	355.3
1500	55.8	370.9

Table I. Uplink and downlink IP level throughput. Static scenario.

1.3) Round trip loss measurements. A significant result is that 1% to 5% losses occur, even in a static situation. The reason may be the inability of UMTS link layer mechanisms to deliver data units in temporary poor signal quality conditions.

2) Mobile Scenario.

We first report network coverage and handover issues, describing next some related effects such as connectivity gaps and delays.

2.1) Network coverage and handover zones. Fig. 2 shows the path followed when the user moves at a pedestrian speed. The zones where handovers and relevant problems have occurred are labelled as “A” and “B”.

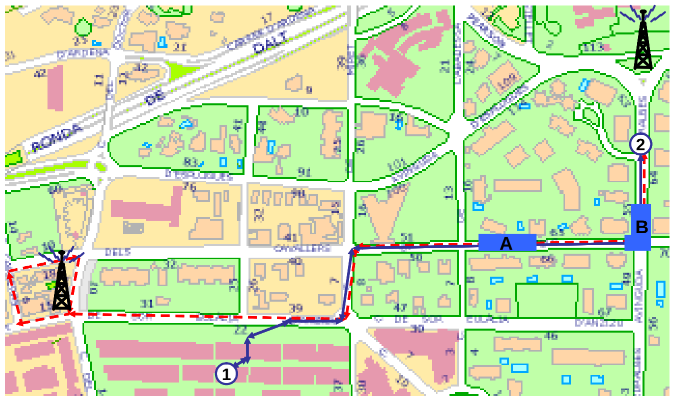

Figure 2. Handover zones. Pedestrian and vehicular mobile UMTS scenarios

When the user moves from spot 1 to spot 2, the handover occurs in the B zone, while it happens in the A zone when the user returns to spot 1. Thus, a hysteresis effect occurs. A remarkable conclusion is that no coverage holes exist along the path. Furthermore, the soft handover procedure is not performed, since macrodiversity is not supported. Therefore, handovers lead to connectivity gaps of several seconds. Such phenomena will occur in commercial UMTS networks when the user moves between cells using different carriers or controlled by different Radio Network Controllers (RNCs) without support for macrodiversity. Otherwise, soft handover will take place.

2.2) Gap and delay periods. We observe two kind of phenomena related with the Radio Link Control (RLC) retransmission mechanism. The first one is gap periods, where one or more packets are lost. The reason may be buffer overflow at some network element, since data cannot be delivered during radio channel unavailability, together with a discarding mechanism. Measured IP connectivity gaps have significant durations, which are actually longer than those measured in live GPRS networks [12]. The second undesired effect is delay periods, where no packets are lost, but the time between two consecutively received packets is up to one order of magnitude greater than the expected one (in this case 100ms). The reason may be the inability of RLC to correctly deliver data at the first transmission, requiring several retransmissions until a successful one occurs.

For example, gaps of up to 20 seconds have occurred, as well as delay periods of up to 8 seconds in our trials. Fig. 3 illustrates the reception of data at the mobile side in the Pedestrian Mobile Scenario.

Every vertical line represents a received packet. The period labelled as "Delay" lasts 3.95 seconds, and the one labelled as "Gap" has a 12.89 second duration. The behaviour observed in the Vehicular Mobile Scenario is similar to this one, but in this case, measured gap values are even greater reaching a maximum of 55 seconds.

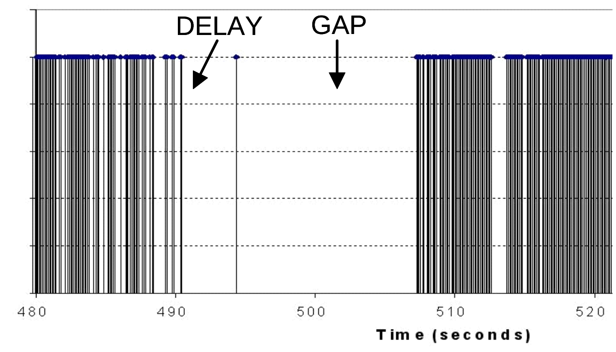

Figure 3. Detailed view. Pedestrian mobile scenario.

Additionally, in some specific trials, bursts of packets are received in small amounts of time when the handover procedure is finished and the communication is recovered, leading to a higher perceived bandwidth than the actual capacity provided by the UMTS network (for example, peak rates of around 1,20 Mbps). The traffic pattern gets back to a smoother rate after the connectivity gap and this burst of packet arrivals. These bitrate peaks are due to buffering at the driver on the receiver side.

2.3) Delay jitter. We analyze the PID of each trial and compare it to the ideal one. Losses are not taken into account, in order to measure only the delay between consecutive packets.

Figure 4 shows an example of the evolution of the PID of a trial performed in the Pedestrian Mobile Scenario. As it can be seen, PID slightly fluctuates around the expected value in most of the trial, when the user is far from the border of the cell. However, when the user approaches the handover zone, PID increases, and suddenly reaches a high peak value (8 seconds in this trial) just before performing a handover. Therefore, delay jitter grows dramatically in the proximity of handover zones. Therefore, it is more likely that spurious TCP retransmissions will occur as a consequence of these delay spikes in UMTS.

The majority of PID values are close to the interval between sent packets (in our trials, 100 ms), though a significant percentage of packets fall in the area below 100 ms. This feature corresponds to situations where the receiver buffer at the end point rapidly delivers packets stored during coverage outages.

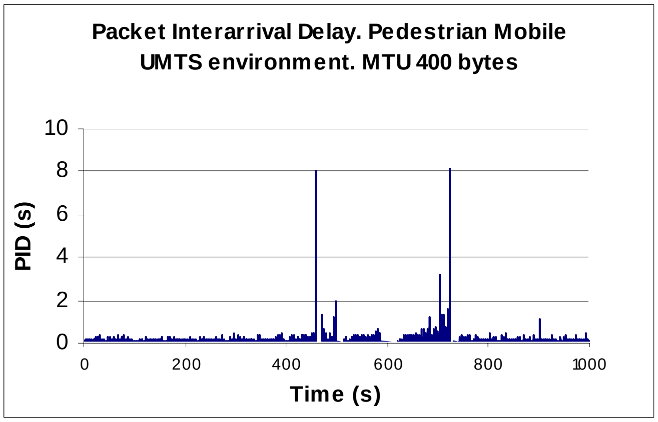

Figure 4. Packet interarrival delay. Pedestrian mobile scenario

V. TCP analysis and parameter optimization over UMTS

A. Definition of TCP measurements

We focus on performance of long lived TCP flows. One reason is the fact that web browsing is the main service over current GPRS networks in terms of data volume [18]. With the widespread support for HyperText Transfer Protocol (HTTP) version 1.1, several objects can be downloaded through the same TCP connection, thus leading to large TCP flows in many cases.

TCP throughput is the main performance metric in our TCP tests, which consist of File Transfer Protocol (FTP) [13] single transfers using a 995 kB file. We perform three trials with each set of TCP parameters as defined next.

1) Static Scenario measurements. We analyze the impact of MSS and RWIN configuration in the Static Scenario. We only test one small, one medium and one large MTU value since results for intermediate values can be extrapolated from the knowledge of the IP level UMTS characterization. From our IP measurements, we derive the optimal RWIN to be used with each MSS value from computing the BDP as the product of the RTT and the obtained IP bandwidth. Note that, as previously mentioned, this value may not yield the highest bitrate. We will denote this RWIN value by RWIN-BDP. RTT used for BDP computation is the same as the one measured with the ping tool, extracting one transmission time component, since we assume TCP ACKs transmission time is negligible. Table II shows the calculated optimal RWIN values for different segment sizes.

MSS (bytes)	Uplink RWIN-BDP (bytes)	Downlink RWIN-BDP (bytes)
88	792	4048
472	1888	10856
1460	2920	21900

Table II. RWIN-BDP values for uplink and downlink over UMTS.

2) Mobile scenario. We measure performance degradation while the user is on the move, with RWIN-BDP for the same MTU values shown in Table II. We test as well the timestamps option in a mobile scenario. As seen in Section IV, data flows may suffer significant jitter in such an environment. Hence, timestamps might help the TCP sender obtain a better knowledge of the RTT values and avoid spurious retransmissions. We perform trials only in the Pedestrian Mobile Scenario, since similar results have been obtained in the Vehicular Mobile Scenario at the IP level.

B. Results

1) Static Scenario.

1.1) MSS and RWIN. In this scenario we perform MSS and RWIN optimization trials. Fig. 5 and Fig. 6 show the achieved TCP throughput values for both, uplink and downlink, transmissions and for each MSS tested. In both cases, optimal MTU size appears to be 1500 bytes. In almost all cases, FTP transfers with the highest maximum transmission window perform better.

In uplink direction, high and medium MTU values use the full available capacity of the link since a small number of losses have occurred in the trials. Using an MTU of 128 bytes TCP, performance is affected by header overhead.

For downlink, TCP performs under the maximum achievable throughput in almost all cases due to retransmissions which lead to frequent reductions of the transmission window. Due to this fact, for MTU 1500 bytes the highest performance is achieved with optimal RWIN.

1.2) SACK option. Table III shows the measured mean throughput and losses from transfers enabling and disabling SACK [8]. This option does not seem to have a significant impact on performance, probably because the situations in terms of loss patterns where SACK should improve TCP behaviour have not occurred in the trials (notice the low mean loss rate). However, it does not either degrade performance. Therefore, we recommend its usage in TCP flows over UMTS since it does not decrease the obtained throughput and may help increase it when multiple losses take place in the same transmission window.

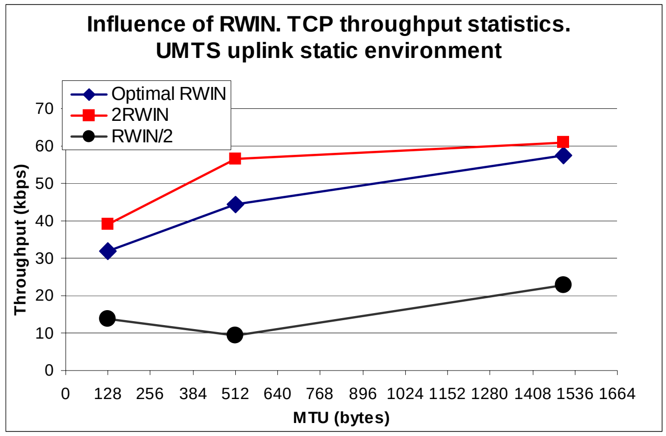

Figure 5. MSS and RWIN influence on TCP throughput over UMTS. Uplink measurements in the Static Scenario

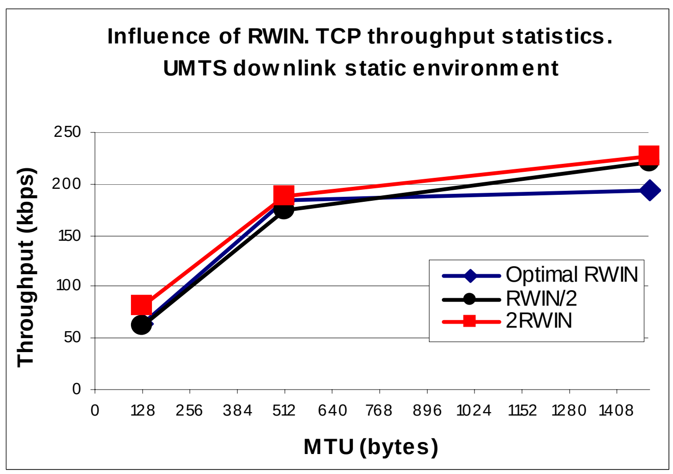

Figure 6. MSS and RWIN influence on TCP throughput over UMTS. Downlink measurements in the Static Scenario

	Uplink		Downlink	
SACK	Mean Throughput (kbps)	Lost packets (%)	Mean Throughput (kbps)	Lost packets (%)
Yes	61,9	0	266,8	0,01
No	61,9	0,01	236,7	0

Table III. SACK usage main statistics. Static scenario.

2) Mobile Scenario

2.1) MSS and RWIN. In our trials TCP performance in mobile scenario is dramatically affected by packet losses, specially in downlink direction where throughput values sink to less than 50% of the throughput measured in the static scenario with the same TCP configuration. For example, for MTU 1500 bytes and RWIN-BDP, we obtain an average throughput of 98,13 kbps and a rate of 11,4% lost packets. Analysing traces of such trials, we observe that most retransmissions are triggered by duplicate ACKs. Another fact contributing to lower throughput is the delay increase for packets successfully delivered after several link layer retransmissions near cell reselection zones. Table IV shows throughput values and lost packet rates obtained in our trials.

	Uplink		Downlink	
MSS (bytes)	Mean Throughput (kbps)	Lost packets (%)	Mean Throughput (kbps)	Lost packets (%)
88	23,3	0,6	25,1	3,4
472	32,4	0,9	64,2	6,6
1460	43,3	0,5	98,1	11,4

Table IV. MSS and RWIN. Main results. Mobile scenario.

2.2) Timestamps option. We have observed that, unfortunately, when timestamps option is enabled and data has to be retransmitted, these data, that originally were sent in one single packet, are splitted at the sender and retransmitted in two packets of smaller size. Therefore, the implementation of timestamps option in the operating system used in our trials, does not allow obtaining representative results on its usage when a high number of losses takes place. Table V provides results in a set of uplink trials where the mean number of retransmissions is 3% and 0,8%, enabling and disabling timestamps, respectively. No clear conclusion can be derived from the usage of timestamps due to the problem mentioned.

Timestamp option	Mean Throughput (kbps)	Max. Throughput (kbps)	Mean Retransmissions
Enabled	59,8	60,2	3 %
Disabled	61,2	61,6	0,8 %

Table V. Timestamps usage main statistics over UMTS. Uplink trials.

VI. Internet services over UMTS.

Apart from the characterisation of the UMTS bearer to support IP and upper protocols such UDP and TCP it is interesting to perceive the effect of such bearers on user application and services. Not all the degradations measured so far have the same impact on the user; it will depend on how they affect the different services. In order to evaluate this effect several key services (web access, streaming and remote terminal) have been tested. These services are selected in terms of popularity (future usage on a mobile device) and in terms of requirements. Web access is demanding in terms of delay and losses, streaming in terms of throughput and remote terminals demands low delay and good throughput for interactivity

A. Web browsing

Web browsing is a good example of popular TCP based application. Table VI shows the main features of a set of three web pages selected under the criteria of popularity and content types, reproduced in a local server for avoiding server load effects on the measurements. The three pages offer a mixture of number of objects and type of objects suitable to derive conclusions from the tests.

The HTTP server runs IIS 5.0 over a Windows 2000 Server platform.

Web page	Numb. of objects	HTML (%)	GIF (%)	JPEG (%)	CSS (%)	Java script (%)	Total size (kB)
Yahoo!	15	77.7	13.8	8.4	0	0	102.4
Star Wars	70	24.7	46.2	26.1	0.7	2.0	430.2
CNN	70	28.4	39.5	17.6	14.4	0	178.1

Table VI. Web pages main features

We use Microsoft Internet Explorer as configured by default in the operating system tested in our trials, that is, using HyperText Transfer Protocol (HTTP) 1.1 [14] and up to 2 parallel TCP connections. We use Microsoft Network Monitor for traffic capture and analysis. We obtain average download times from 5 trials in each case, testing additionally the server-side standard compression option. Table VII shows the main results from such trials.

	No compression		Compression	
Web page	Download time (s)	Throughput (kbps)	Download time (s)	Throughput (kbps)
Yahoo!	13.9	60.3	9.2	91.2
Star Wars	27.6	127.7	26.3	134.0
CNN	20.0	72.9	19.7	74.0

Table VII. Web page download times and throughput

A first conclusion is that, although two simultaneous TCP connections are open during downloads, UMTS link is severely under-utilized. The default Get-Reply pattern leads to significant transmission stalls.

On the other hand, compression provides significant benefit only in the Yahoo! page download. The reason is that 78% of this page is textual content that is compressed at the server. Note that GIF and JPEG content are not further compressible with standard HTTP server compression mechanisms. Moreover, most objects can be sent in a single segment, leading to minor improvement once compression is performed.

We have studied the usage of pipelining to minimise the effect of the Get-Replay pattern. As opposed to the benefit seen in GPRS [23], in UMTS pipelining does not provide a significant improvement in the download times (see figure 7). There in no clear conclusion about the usage of pipelining, further work should be done to identify the reason of this unexpected underperformance.

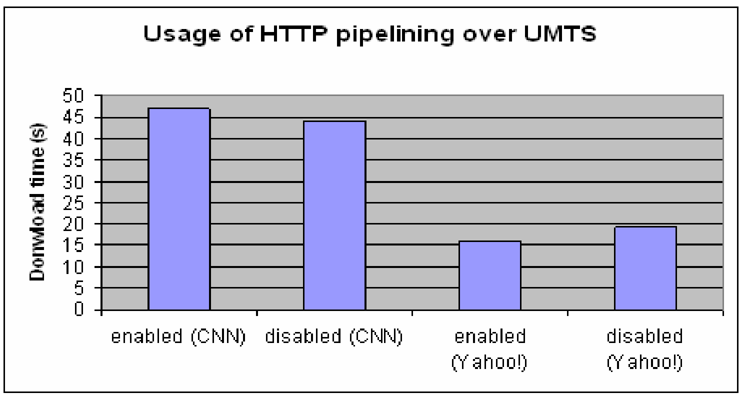

Figure 7. Download web page time over UMTS with and without pipelining activated.

Other alternative to minimise the effect of Get-Replay on the download time is to use several parallel TCP connection for the same web page download. Meanwhile an object is requested the link can be used by the rest of connections. Figure 8 shows for the CNN web page (that has a significant number of objects) gets reduction close to the 43% on the download time when moving from 2 to 10 TCP connections. The usage of several connections has some drawbacks since some browsers only support a limited number of connections and web servers try to limit the number on connections per user for capacity reasons.

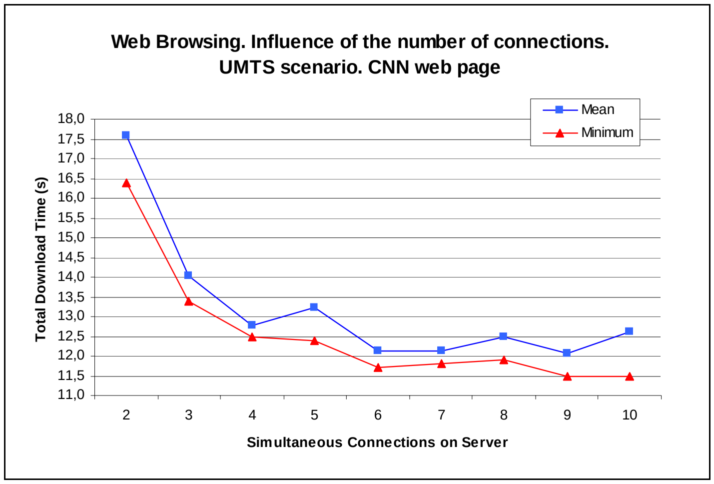

Figure 8. Download web page time over UMTS when using different number of TCP connection simultaneously.

B. Graphical remote access

One of the examples of prohibitive applications over GPRS that become feasible over UMTS are graphical remote access applications. We have tested the Microsoft Terminal Server/Client software over a UMTS connection, comparing it with the usage of a 10 Mbps Ethernet connection. The main result is that user perception on the achieved interactivity is close to that of the fixed Ethernet. For example, a remote file opening action is performed in 1 second over UMTS, except in sporadic situations where TCP retransmissions increase this time.

C. Streaming

Streaming applications are expected to perform better over UMTS since bandwidth conditions are less restrictive than in GPRS environments. However, the effect of packet loses, bandwidth or delay fluctuations in the proximity of handover zones may affect the appearance of the displayed information in the mobile client.

We have tested Microsoft Windows Media Services v.9 as the most popular streaming solution. The media server runs on Windows 2003 Server OS. The client application is Windows Media Player v.9. This platform includes a set of options, named Fast Streaming that allows the server to profit from favourable link conditions sending media information at the maximal available rate, independently of the rate at which media was coded. This information is cached in the mobile client device and in presence of poor link conditions the client may display the cached information while the system recovers from the impairments.

Streaming trials consist of real time transmissions of a video codified at 148kbps in the mobile scenario described in figure 2 with macrodiversity support. The selected bit rate is suitable for streaming a video on a small device screen such smartphone or a PDA. The media server runs

Windows 2003 Server OS. The client application is Windows Media Player v.9. Real Time Protocol (RTP) is tested using TCP and UDP as the transport protocols.

Fast Streaming mechanism chooses TCP as the default transport protocol for media delivery. This policy differs from the one used by other streaming platforms like QuickTime and from the standard 3GPP PSS [20] which defines UDP as the transport protocol for media content delivery. The utilization of TCP could be justified if a large amount of packets get lost during the transmission. Usually, UDP will not retransmit lost data. However we have observed that Windows Media platform implements mechanisms for notification (by means of the Real Time Control Protocol, RTCP) and recovery of erroneous RTP packets. The platform uses separate connections for the delivery and retransmission of packets, both with the same Synchronization SouRCe (SSRC) identifier. The rate of retransmitted data ranges between 1% and 2% in our trials. Results show that not all lost RTP packets are recovered during transmission. However, the user does not perceive degradation on the received media. Even when moving in the area close to the border between cells softhandover performs noticeably and results in a continuous flow of bits.

Moreover, we have stated that at the beginning of the transmission the source sends three RTP packets. This process leads to an initial estimation of the available bandwidth [22]. The same connection is later used for the retransmission of RTP packets. However, we have observed that the SSRC values of these three initial packets differ from the SSRCs of the rest of the transmission. The variability of this field is relevant if we consider the use header compression protocols such as ROHC [21] to improve performance of streaming services. ROHC assumes that the field SSRC will remain constant for all RTP packets sent during transmission as it identifies the source of the communication. A change of this field in a packet would suppose the renegotiation of the full context of the compression protocol. This process would reduce the expected performance.

As a main conclusion, results show that streaming services are feasible over UMTS if media is coded at rates below maximal available throughput and macrodiversity is supported.

VII. Conclusions

3G technologies, such as UMTS, provide wide area Internet wireless access. The success of the mobile Internet relies in part in the ability of the Internet protocols to cope with wireless links features and in an adequate protocol configuration.

In this paper, we have presented three sets of trials. The first one is an IP level UMTS characterization. It must be noted that a precommercial UMTS network and a beta UMTS card have been used. We have obtained throughput, losses, RTT and jitter measurements in both static and mobile scenarios. RTTs are close to fixed Internet values and up to 371 kbps are available at the IP level according to our measurements. Losses occur even in static scenarios, with round trip loss values of the order of 10-2. IP level measurements reveal that medium and high MTUs yield the maximum throughput values. Mobility, specially near cell reselection and handover zones, leads to delay spikes and connectivity gaps with durations of several seconds if macrodiversity is not supported. Such features may lead to undesired effects on TCP, such as spurious and serial retransmissions, with an exponential growth of the RTO value. Our results so far show that soft handover should be supported and UMTS coverage improvement and fine network tuning is needed.

The second part of our work is devoted to testing several TCP parameter configurations, analyzing the impact of MSS, RWIN, SACK and timestamps over UMTS. TCP performance is highly dependant on the parameter configuration used. The highest throughput is obtained using large MSS values and a maximum transmission window larger than the BDP of the path. We recommend the usage of SACK for UMTS, since its overhead has a negligible effect on performance and, on the other hand, the option may help improve performance when multiple losses occur within the same transmission window. However, we cannot provide firm conclusions from our tests on the timestamps option.

A significant problem found in UMTS is the gaps due to cell reselection if macrodiversity is not supported and outages that result in losses and delay spikes. A way to cope with this problem may be to use explicit link failure mechanisms or a proxy between the wireless and wired parts.

The third part, related to the performance of selected services on top o UMTS it is noticeable the reduction of the RTT and the increase of available data rate compared to GPRS. Also the availability of soft handover assures a continuous flow of bits instead of the connection gaps of GPRS cell reselections. Adding up all these aspects result in a significant user perception improvement. Applications such terminal servers are now feasible with UMTS and streaming is able to guarantee a continuous video even in front of the user mobility. A very significant result is the poor performance when accessing the web with UMTS. The perceived throughput is close to 60 kbps, quite far from the nominal capacity of the bearer (384 kbps dedicated channel for packet service). The usage of HTTP on top of UMTS give chances to further improvement to the Internet protocol stack when used with UMTS

Acknowledgment

This work was supported in part by FEDER, the Spanish Government through project TIC2003-01748 and Telefónica Móviles through Plan de Promoción Tecnológica del UMTS.

References

[1] J. Postel, “Internet Protocol”, RFC 791, September 1981.

[2] J. Postel, “Transmission Control Protocol”, RFC 793, September 1981.

[3] S. Dawkins, G. Montenegro, M. Kojo, V. Magret, “End-to-end Performance Implications of Slow Links”, RFC3150, July 2001.

[4] J. Mogul, S. Deering, “Path MTU discovery”, RFC 1191, November 1990.

[5] H. Inamura, Ed., G. Montenegro, Ed., R. Ludwig, A. Gurtov, F. Khafizov, “TCP over Second (2.5G) and Third (3G) Generation Wireless Networks”, RFC3481, February 2003.

[6] W.R. Stevens, “TCP Slow start, congestion avoidance, fast retransmission and fast recovery algorithms”, RFC 2001, January 1997

[7] M. Allman, S. Floyd, C. Partridge, “Increasing TCP's Initial Window”, RFC3390, October 2002.

[8] M. Mathis, J. Mahdavi, S. Floyd, A. Romanow, “TCP Selective Acknowledgement Options”, RFC2018, October 1996.

[9] J. Postel, “User Datagram Protocol”, RFC 768, August 1980

[10] “Wireless Profiled TCP. Version 31-March-2001. Wireless Application Protocol. WAP-225-TCP-20010331-a”

[11] A. Timm-Giel “UMTS/GPRS Measurements and Evaluations”, FG521 Workshop, Hamburg, February 2004

[12] A. Calveras, J. Paradells, C. Gomez, M. Catalan, J.C. Valles, “Optimizing TCP/IP parameters over GPRS and WLAN networks”, IEEE Vehicular Technology Conference, VTC 2003-Fall, October 2003

[13] J. Postel, J. Reynolds, “File Transfer Protocol”, RFC 959, October 1985

[14] R. Fielding et al., “Hypertext Transfer Protocol – HTTP 1.1”, RFC 2616, January 1997

[15] P. Benko, G. Malicsko, A. Veres, “A Large-scale, Passive Analysis of End-to-End TCP Performance over GPRS”, IEEE INFOCOM’2004, March 2004.

[16] R. Chakravorty, J. Cartwright, I. Pratt, “Practical Experience with TCP over GPRS”, IEEE GLOBECOM’2002, November 2002.

[17] M. Sagfors, R. Ludwig, M. Meyer and J. Peisa, “Queue management for TCP traffic over 3G links”, IEEE WCNC 2003, March 2003.

[18] R. Valden, T. Varga, B. Wouters, B. Sanders, “Wireless Service Usage and Traffic Characteristics in GPRS networks”, 18th International Teletraffic Congress, ITC-18, September 2003.

[19] R. Chakravorty, S. Banerjee, P. Rodriguez, J. Chesterfield, I. Pratt, “Performance Optimizations for Wireless Wide-area Networks: Comparative Study and Experimental Evaluation”, ACM MOBICOM’04, September, 2004.

[20] 3GPP TS 26.234 V6.0.0 (2004-06) Technical Specification 3rd Generation Partnership Project; “Technical Specification Group Services and System Aspects; Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (Release 6)”

[21] “RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed”, C. Bormann, et al., RFC3095, July 2001, (Status: Standard Track.)

[22] “Measurements of the Congestion Responsiveness of Windows Streaming Media”, James Nichols, Mark Claypool, Robert Kinicki, and Mingzhe Li. International Workshop on Network and Operating System Support for Digital Audio and Video. Pag. 94-99. 2004.

[23] ”Recommendations for Internet Access from Mobile Terminals. Initial Results from RIU253”, J. Paradells, A. Calveras, X. Hesselbach, C. Gomez, J. C. Vallés, M. Catalán, M. Estévez, P. Armas. IST Mobile & Wireless Communications Summit 2003. Aveiro, Portugal.