Impact of UMTS load control mechanisms on TCP

Juan Rendon, Ramon Ferrus, Anna Sfairopoulou, and Ferran Casadevall

***Abstract*— Radio Resource Management (RRM) algorithms are responsible for efficient utilisation of the air interface resources in UMTS networks. Network load control is one important task of the RRM functionality aimed at ensuring system stability around the targeted network load defined by the radio network planning. A common load control action is to reduce the bandwidth assigned to TCP traffic in overload situations. This paper studies, with the help of a UMTS emulator, the impact of a UMTS load control mechanism on the performance of TCP-based traffic. Results show that, even in a highly loaded UMTS system, TCP traffic is effectively managed without incurring negative effects associated with the bandwidth oscillation produced by the load control mechanism.**

***Index Terms*— TCP, UMTS, load control, RRM**

I. Introduction

Universal Mobile Telecommunications System (UMTS) is a third generation wireless cellular system that supports the transmission of voice, data and multimedia traffic. In the UMTS system the Transmission Control Protocol (TCP) is used by several applications such as e-mail and World Wide Web. A key piece within UMTS is the Radio Resource Management (RRM) functionality, which is in charge of the allocation of resources by means of scheduling algorithms, besides other key functions such as admission, congestion, power and handover control [1]. RRM strategies used in the UMTS air interface, which are usually based on service class prioritisation and packet data scheduling, can have a great impact on TCP performance. Under high load conditions, the available bandwidth for interactive traffic could be dramatically reduced when most radio resources are assigned to conversational traffic to satisfy its more stringent QoS requirements. This radio resource control strategy may produce a kind of bandwidth oscillation effect on TCP since an amount of TCP interactive traffic would not be able to obtain radio resources temporarily in periods of high load. This bandwidth oscillation could lead to significant throughput degradation due to factors such as spurious TCP retransmissions [2].

This work was supported in part by the Spanish government in the CICYT projects TIC2004-00518 and TIC2003-09279.

J. Rendon and A. Sfairopoulou are with the Technology Department, Pompeu Fabra University, Passeig de Cirvumvalacio 8, 08003 Barcelona, Spain (e-mail: juan.rendon@upf.edu).

R. Ferrus and F. Casadevall are with the Signal Theory and Communications Department, Polytechnic University of Catalonia, c. Jordi Girona 1-3, 08034 Barcelona, Spain (e-mail: ferrus@tsc.upc.edu).

In this sense, RRM functions can significantly affect TCP performance. One of the RRM functionalities that can have an important impact on TCP behaviour is the load control mechanism, since the available bandwidth for TCP traffic could be reduced momentarily when most radio resources are assigned to real-time traffic.

Regarding previous work on the research field of TCP over UMTS, different authors have studied the interaction of TCP with a few UMTS parameters. For example, the impact of queue management mechanisms on TCP for 3G networks is studied in [3][4], the effect of different UMTS Radio Link Control (RLC) layer parametes on the TCP protocol is analysed in [5][6][7] and the impact of scheduling mechanisms on TCP is presented in [8][9].

However, to the best of the author's knowledge, there has not been any experimental research that studies the impact of the UMTS load control mechanism on TCP traffic. In this paper we study, with the help of a UMTS emulator, the impact of a load control proposal on TCP connections. We presented some preliminary results in [10], but for this article we have implemented the UMTS radio interface protocols much more in detail.

The paper is organized as follows: Section 2 introduces the load control strategy used in our study and presents some characteristics of the radio interface in UMTS. Next, in section 3, the architecture of the UMTS emulator used for evaluating the TCP performance and the scenario set-up are described. Section 4 presents the obtained results, and finally, conclusions are addressed in Section 5.

II. Load Control Strategy and UTRAN Interface in UMTS

The radio resource assignment algorithm addressed in this analysis is based on the scheduling strategy introduced in [11] for a Downlink Shared Channel (DSCH). The DSCH is a common channel intended to optimise code usage in UMTS downlink by sharing a sub-tree of Orthogonal Variable Spreading Factor (OVSF) codes among several users. According to the proposed strategy, the number of allowed transmissions in the DSCH, apart from being limited by the number of allowed codes, must be limited by controlling the total power consumption and the load factor increase produced by this channel. The goal of this load control strategy is to avoid, as much as possible, interference peaks in the radio channel that would deteriorate other services with QoS constraints more sensitive to radio block errors (e.g. radio block delays in conversational services should be kept as low as possible). Thus, within the considered strategy, radio resources are allocated to every 10ms-frame following a three-step procedure which is described below (see Figure 1).

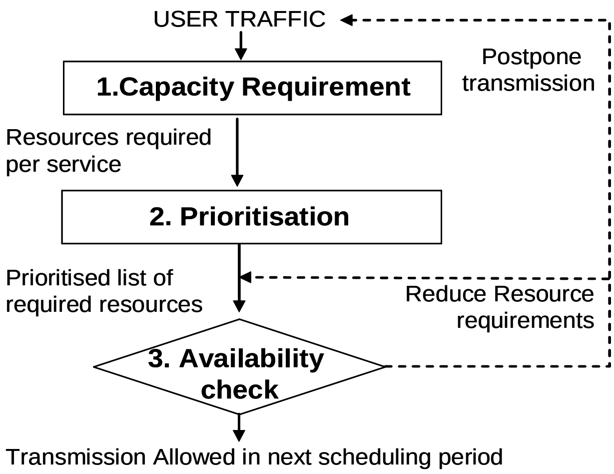

Fig. 1. Resource allocation strategy

1. Capacity Requirement. The radio resource manager makes an estimation of the amount of resources required for each served user. A guaranteed rate criterion is used at this stage to decide the amount of resources required for a TCP connection.

2. Prioritisation. Once the required capacity for each user is known, users are prioritised according to the type of service (first prioritisation level) and, for the same type of service, according to their Quality of Service (QoS) requirements (second prioritisation level). In our analysis, for the second prioritisation level, users are ordered so that connections having rates below their guaranteed targets are served first. So, resource sharing among users will tend to guarantee a mean bit rate for each connection.

3. Availability checks and load Control. After the prioritisation of the users to be served, the algorithm checks whether or not this selection is possible depending on the available resources and modifies it accordingly. Available resources are estimated in the radio resource controller by computing the expected load factor and transmission power in each Base Station (BS). In the implemented algorithm, voice users (i.e. conversational traffic class) are always served, while for TCP-based users (i.e. interactive traffic class), transmissions are carried out only if the estimated network load factor (η) falls below a given threshold (η_{TH}). Figure 2 shows the available capacity for serving TCP traffic in terms of network load factor as well as a situation where TCP traffic is not served since the network load η goes above the network load threshold η_{TH}.

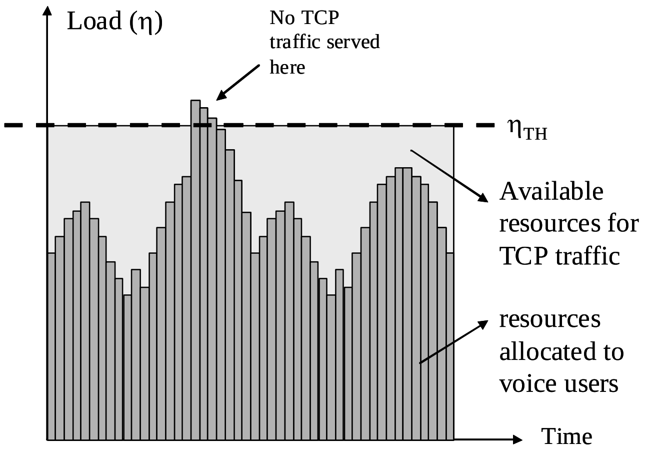

Fig. 2. Network load factor η and network load threshold η_{TH}.

RRM strategies are supported by the protocol stack of the UMTS Terrestrial Radio Access Network (UTRAN). The radio interface of the UTRAN (see Figure 3) is layered into three protocol layers: the physical layer (L1), the data link layer (L2), and the network layer (L3). Additionally, L2 is split into two sublayers, RLC and medium access control (MAC). On the other hand, the RLC and L3 protocols are partitioned in two planes, user and control. In the control plane, L3 is partitioned into sublayers where only the lowest sublayer, denoted radio resource control (RRC), terminates in the UTRAN. Connections between RRC and MAC as well as RRC and L1 provide local interlayer control services, and allow the RRC to control the configuration of the lower layers. In the MAC layer, logical channels are mapped to transport channels. A transport channel defines the way in which traffic from logical channels is processed and sent to the physical layer. The smallest entity of traffic that can be transmitted through a transport channel is a Transport Block (TB). Once in a certain period of time, called a Transmission Time Interval (TTI), a given number of TBs will be delivered to the physical layer in order to introduce some coding characteristics, interleaving, and rate matching to the radio frame. The set of specific attributes are referred as the Transport Format (TF) of the considered transport channel. Note that the number of TBs transmitted in a TTI indicates that different bit rates are associated with different TFs. As the User Equipment (UE) may have more than one transport channel simultaneously, the TF Combination (TFC) refers to the selected combination of TFs. The list of allowed TFCs to be used is referred to as the Transport Format Combination Set (TFCS).

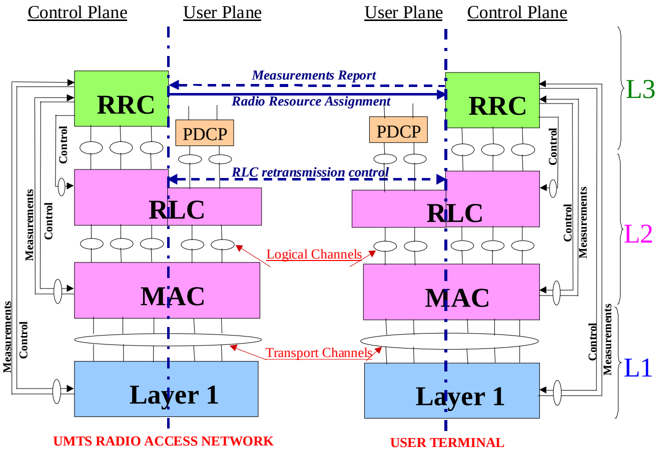

Fig. 3. Radio Interface Protocol architecture

III. UMTS EMULATION PLATFORM

TCP performance results have been obtained using a real time UMTS emulator developed within the framework of the IST ARROWS project [12]. The UMTS emulator is a real-time operation HW/SW platform that includes multimedia terminals, UMTS elements and IP connectivity. Among the main features of the platform we can remark the possibility of validating RRM strategies under complex scenarios and the possibility of testing the impact of these strategies over the end-to-end behaviour of legacy IP-based multimedia applications with Quality of Service (QoS) requirements. The external organisation of the UMTS testbed is composed by the User Equipment, the UTRAN, and the UMTS Core Network (CN). Figure 4 shows the basic mapping of UMTS components into testbed machines.

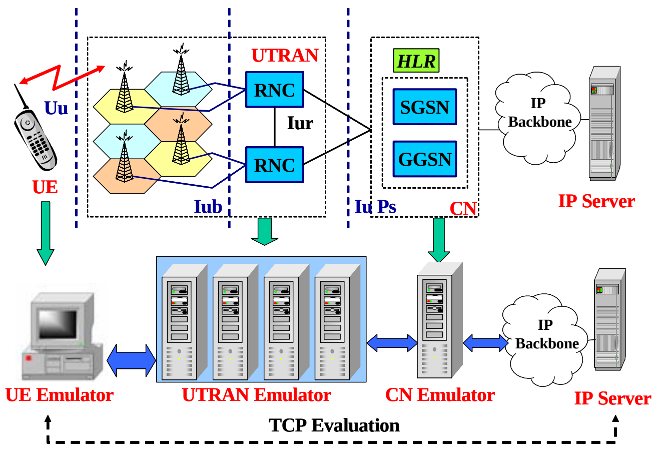

Fig. 4. UMTS real-time emulator

Different applications such as WEB browsing are executed in the UE and IP Server for a reference user under test. The rest of users are emulated by means of traffic models inside the UTRAN Emulator. A complete protocol stack has been developed for the reference user in accordance to 3GPP specifications [13].

The radio layer is modelled by means of histograms of the Block Error Rate (BLER) versus E_b/N_o obtained from off-line simulations [12][14]. The link-level simulator includes both the propagation environment and the transmission and reception processes (i.e. channel encoding and decoding, RAKE receiver, diversity, fast power control, etc.). Simulator outputs characterise the relevant transport channels considered in the current study in terms of transport BLER as a function of the global interference existing in the radio channel. Thus, different histograms have been obtained for the different environments (outdoor 120 km/h, outdoor 50 km/h or pedestrian 3 km/h), for different channels (DSCH and Dedicated Channels (DCHs)), for each Transport Format, and for different values of E_b/N_o (where N_o includes all the sources of noise).

The UMTS emulator is configured with a 5kmx5km macrocell scenario where 14 omnidirectional cells are hexagonally distributed with a cell radius of 500m. The system is loaded with a combination of conversational and web browsing users uniformly distributed in the service area.

Table 1 provides the TFC used in the correspondent Radio Access Bearers (RABs). Conversational service is offered through DCHs with two possible Transport Formats: 64kbits/s and no-transmission. For www interactive users, transmissions are carried out through a DSCH channel and a total of 7 TFs are considered ranging from data rates from 256kbits/s down to the no-transmission case. One fourth of the OSVF code tree has been assigned to the DSCH channel in each BS (This means that, according to allowed TFs in table 1, the maximum number of simultaneous users in DSCH could take any combination between 16 users at 16kbits/s and 2 users at 256 kbits/s). Web sessions consist of an exponential distributed random number of successive TCP persistent connections.

Table 1. Transport formats for the considered RABs.

Service		WWW	VIDEOPHONE
Channel type		DSCH	DCH
Transport Block (TB) sizes		336 bits (320 payload)	640 bits
TFC	TF0, bits	0×336	0x640
	TF1, bits	1×336 (16 Kb/s, SF=128)	2x640 (64 kb/s, SF=32)
	TF2, bits	2×336 (32 Kb/s, SF=64)	-
	TF3, bits	4×336 (64 Kb/s, SF=32)	-
	TF4, bits	8×336 (128 Kb/s, SF=16)	-
	TF5, bits	12×336 (192 Kb/s, SF=8)	-
	TF6, bits	16×336 (256 Kb/s, SF=8)	-
TTI, ms		20	20

The base station maximum transmission power is set up so that capacity is mainly limited by code usage and network load. The BLER target is 1% for voice users and 10% for TCP traffic. For this latter traffic, the maximum number of ARQ retransmissions for 320-bit RLC blocks is set to 4.

Uplink transmissions are carried out over DCHs. In this link, DCH channels are configured with 4 Transport Format Combinations allowing 0,8,16,32 and 64 kbps instantaneous data rates. This data rates are measured in a TTI of 20ms. So, mobile terminals may select any of these values according to a distributed algorithm based on a rate-oriented approach. The guaranteed bit rate in this channel is set up to 32 kbits/s for each user. So, as far as uplink transmissions do not exceed such value, user are allowed to make use of the assigned DCH resources. The RLC layer works in Acknowledged Mode (AM) with a maximum of 4 retransmissions per transport block. These settings make the uplink channel not to introduce any disturbances in the obtained results (exclusively focused on downlink transmissions) other than the latency incurred in uplink DCH transmissions, which is equal to the 20ms-TTI value in most cases.

The objective of this study is to evaluate the impact of the load control on downlink TCP transmissions in a highly loaded UMTS system. In this sense, two scenarios were considered: the first one, with 400 www TCP users and 250 voice users, and the second one, with 600 www TCP users and 250 voice users.

For the user under test, live TCP traffic is generated by using the *iperf* tool [14]. This tool is designed for TCP network testing and allows to establish persistent TCP connections and transfer data at the maximum allowed network rate during a certain amount of time. Both the server and the client machines running *iperf* operate under Linux version 2.4-16 with the Selective Acknowledgment (SACK) and Timestamps TCP options enabled for better TCP performance [16]. The maximum receiver window is 64 KB and the Maximum Segment Size (MSS) is 530 bytes. Buffer size in the UTRAN Radio Network Controller (RNC) is set to 64 KB. TCP results are derived by evaluating 50 consecutive TCP connections for the user under test in each analysed configuration.

IV. RESULTS

The results shown in Figures 5, 6 and 7 correspond to results obtained from downlink transmissions. Figure 5 displays the packet loss ratio experienced in UMTS by both conversational and interactive services versus the selected network load threshold value η_{TH}. According to Figure 5, as expected, the lower the load threshold is, the better the behaviour of voice packets error ratio is since less data traffic is served in the DSCH channels. So, network operators should use a proper value for the network load threshold in order to satisfy QoS requirements for conversational traffic. Notice that a η_{TH}=100% means that no load restrictions are applied on DSCH channels. On the other hand, for interactive traffic, if η_{TH} is below 70% the performance of TCP traffic significantly decreases due to an excessive packet loss. This packet loss is due to buffer overflows since the available bandwidth for DSCH is drastically reduced. This situation also leads to reduced TCP throughput values as shown in Figure 6. However, if the network load threshold ranges between 70-90% the TCP packet loss is kept at a minimum value (1%) and TCP throughput almost reaches its maximum allowed value (achieved when no load restrictions are applied). Therefore, using network load threshold values above 90% has no advantage at all since TCP traffic throughput does not improve and voice quality is even further degraded, especially in very high traffic load conditions (600 www users). Moreover, as shown in Figure 7, the delay values difference above 90% is not significant. These delays refer to an IP packet, which is the Service Data Unit (SDU) delivered to the RLC layer. This effect also indicates that there is an optimum margin for the network load threshold value that enables to offer a certain QoS level for voice users while handling TCP traffic as if no load limitations were applied.

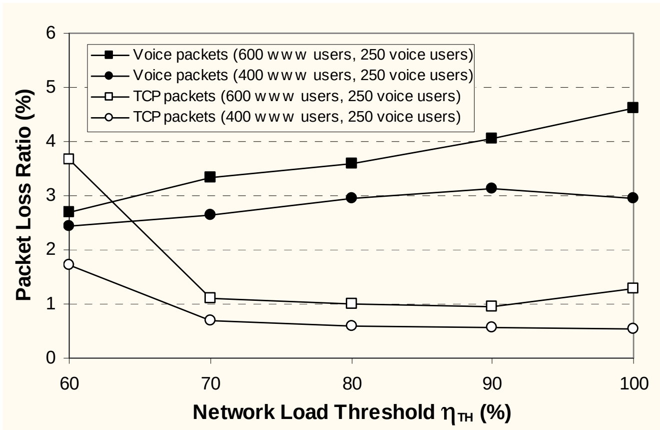

Fig. 5. Packet Loss Ratio

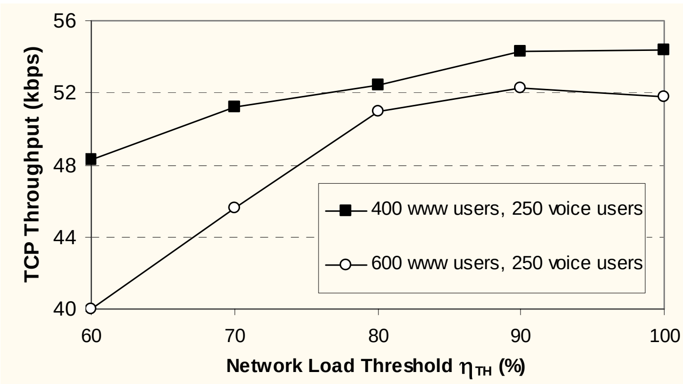

Fig. 6. TCP downlink throughput

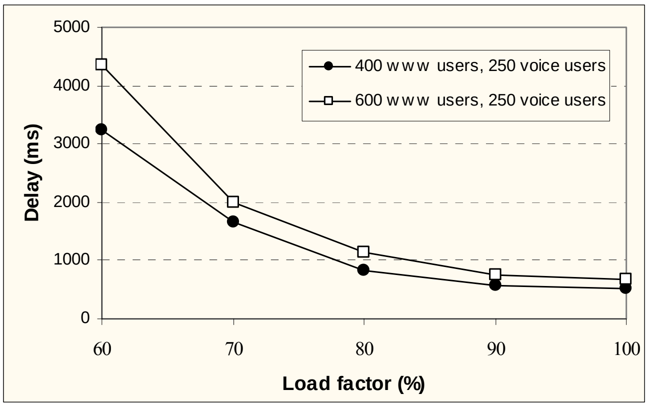

Fig. 7. TCP delay values

By analyzing the TCP traces we did not find phenomena such as peak delays nor burst errors that could affect the TCP performance negatively. In most cases, the TCP sender could enter into the congestion avoidance phase immediately and there were few timer-driven retransmissions activated by peak delays.

V. CONCLUSIONS

The effect of the UMTS load control mechanism on TCP has been evaluated. Results show that in a realistic and highly loaded UMTS network where voice users are prioritized over TCP users, TCP performance could be maintained if proper network load threshold values are selected. Under the considered scenario a good trade-off between packet loss ratio for voice users and TCP throughput is obtained when the network load threshold value is around 80%.

REFERENCES

[1]. H. Holma and A. Toskala, “WCDMA for UMTS”, John Wiley & Sons, 2000.
[2]. M. Yavuz and F. Khafizov, “TCP over wireless links with variable bandwidth”, VTC Fall 2002, Canada, September 2002.
[3]. M. Sagfors et al., “Queue Management for TCP Traffic over 3G Links”, IEEE WCNC, March 2003.
[4]. R. Bhattacharya et al., "Intelligent queue management protocol for TCP traffic over 3G links", IEEE ICPWC 2005.
[5]. J. Alcaraz et al., "Optimizing TCP and RLC interaction in the UMTS radio access network", IEEE Network, March/April 2006.
[6]. M. Taferner and E. Bonek, “Wireless Internet Access over GSM and UMTS”, Springer-Verlag, 2002.
[7]. F. Lefevre, G. Vivier, “Optimizing UMTS Link Layer Parameters for a TCP Connection”, VTC Spring 2001, Greece, May 2001.
[8]. X. Wang et al., "A cross-layer scheme to improve TCP performance in UMTS with packet scheduling", IEEE VTC 2005 Fall.
[9]. M. Malkowski and S. Heier, "Interaction between UMTS MAC scheduling and TCP flow control mechanisms, ICCT 2003.
[10]. R. Ferrus et al., "On the impact of prioritised radio resource control on TCP connections in a downlink UMTS channel", IEEE PIMRC 2004.
[11]. O. Sallent et al, “An emulator framework for a New Radio Resource Management for QoS Guaranteed Services in W-CDMA Systems”, IEEE JSAC, October 2001.
[12]. ARROWS Project WEB site, http://www.arrows-ist.upc.es/
[13]. 3GPP TS 25.301, “Radio Interface Protocol Architecture”.
[14]. J. Olmos and S. Ruiz, “UTRA-FDD Link Level Simulator for the Arrows Project”, IST’01, Conference Proceedings, pp. 782-787.
[15]. Iperf tool web page, http://dast.nlanr.net/Projects/Iperf/]
[16]. H. Inamura et al., “TCP over second (2.5G) and third (3G) generation wireless networks”, IETF RFC 3481, February 2003.