ON THE IMPACT OF PRIORITISED RADIO RESOURCE CONTROL ON TCP CONNECTIONS IN A DOWNLINK UMTS CHANNEL

Ramon Ferrús[1], Juan Rendon[2], Anna Sfairopoulou[2], Xavier Revés[1], Ferran Casadevall[1]

[1]Department of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC)
Campus Nord, 08034, Barcelona - Spain
e-mail: ferrus@tsc.upc.es

[2]Technology Department
Pompeu Fabra University
Passeig de Circumvalació 8, 08003, Barcelona - Spain
e-mail: juan.rendon@upf.edu

Abstract. Prioritisation based on traffic class is widely used within radio resource control strategies to provide quality of service (QoS) assurances in 3G wireless systems. Under high load conditions, traffic prioritisation can lead to situations where, momentarily, most of the radio resources are assigned to delay-constrained connections (e.g. conversational users) and, consequently, the available bandwidth for less priority users (e.g. interactive traffic) is significantly reduced. In this paper we analyse the impact of this bandwidth oscillation phenomenon on TCP connections over a downlink UMTS channel. Experimental results have been obtained using legacy TCP connections over a real-time UMTS testbed.

Keywords: TCP, UMTS, prioritisation, RRM.

I. INTRODUCTION

Due to the increasing deployment of wireless networks and the extensive use of the Internet during the last years, the use of Transmission Control Protocol (TCP) applications over wireless networks have gained more and more relevance. The Internet Transmission Control Protocol (TCP) was a protocol initially designed for data transfers over a fixed network and, as a consequence, the performance of TCP in wireless cellular networks constitutes one important research field [1]. The research community has been studying TCP over different cellular networks, such as GSM [2], GPRS [3], IS-2000 [4], etc. and actually some works can also be found for 3G UMTS networks. For example, the impact of different UMTS RLC layer parameters on the TCP protocol has been studied in [5][6], the effect of buffer management mechanisms on TCP for 3G networks is analysed in [7] and the impact on TCP throughput of dedicated and shared channels in the UMTS downlink transmission is studied in [8]. The objective of the paper is the identification and quantification of the effects observed in TCP connections over UMTS networks due to the usage of radio resource management strategies based on traffic class priorization under different system load conditions. To our knowledge, this effect has not yet been described for the case of a transmission over a UMTS network.

Radio resource control 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 the 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 [4]. To study this effect on TCP connections a real time UMTS testbed has been used. The testbed is capable of emulating in real time the behaviour of a UMTS scenario where different advanced RRM algorithms can be implemented and tested [9].

The paper is organized as follows: Section 2 introduces the radio resource assignment mechanisms analysed in our study. 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. PRIORITISATION STRATEGIES FOR TRAFFIC DIFFERENTIATION WITHIN RRM

UMTS networks are designed to fully support multimedia traffic under Quality of Service (QoS) guarantees. In the UMTS air interface, WCDMA technology jointly with efficient RRM functions have been adopted as the key enablers of such complex multimedia radio access network. RRM functions are crucial in WCDMA access networks since capacity is usually limited by the amount of interference in the air interface (soft capacity).

The radio resource assignment algorithm addressed in this analysis is based on the scheduling strategy introduced in [10] for a Downlink Shared Channel (DSCH). The DSCH is a common channel intended to optimise code usage in UMTS downlink by sharing a subtree of OVSF (Orthogonal Variable Spreading Factor) codes among several users [11]. According to the proposed strategy in [10], 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 control is to avoid as much as possible interference peaks in the radio channel which would deteriorate other services with QoS restrictions more sensitive to radio block errors (e.g. radio block errors in conversational services are not usually recovered by retransmissions due to delay restrictions). Thus, within the considered strategy, radio resources are assigned at a frame basis (10ms) following a three-step procedure: capacity requirement, prioritisation and availability check.

A. *Capacity Requirement*

In the capacity requirement phase, the radio resource manager makes an estimation of the amount of resources required for each served user *i*. In WCDMA, resources can be formulated in terms of transmission rate (or equivalently spreading factor SF_i) and transmission power (P_i) in order to meet certain QoS restrictions. Both magnitudes are coupled by means of the following expression:

$$\left(\frac{E_b}{N_0}\right)_i \le \frac{1}{L^{i,k}}\frac{P_i \cdot SF_i}{I+P_N} \qquad \textbf{(1)}$$

where I accounts for the intracell and intercell interference, $L^{i,k}$ is the propagation loss between mobile *i* and base station *k* (BS_k), P_N is the noise power and E_b/N_o is the bit energy over noise ratio that provides the required BLER (Block Error Ratio) target. The SF used in a transmission can be choosen among a set of values (transport formats) associated to each connection. In our study, as further discussed in Section III, interactive services will be allowed to use several SFs and we will use a *rate-oriented strategy* to choose the preferred value of SF (and consequently, using (1), the required power P_i). This strategy selects the appropriate SF to guarantee a certain mean bit rate by means of the "service credit" (SCr) concept. The SCr of a connection accounts for the difference between the obtained and the expected bit rate for this connection.

B. *Prioritisation*

Once the required capacity for each user is known, users are prioritised according to its type of service (first prioritisation level) and, for the same type of service, according to their QoS requirements (second prioritisation level). In our analysis, for the second prioritisation level, users with accumulated credits are served first. So, resource sharing among users will tend to guarantee a mean bit rate for each connection.

C. *Availability check*

After the preordination 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 of the expected load factor and transmission power in each base BS. The expected load factor in BS_k whenever there are *n* transmissions in the system in frame *t* is:

$$\eta_k(n,t)=\sum_{i=1}^{n}\frac{\left((1-\alpha)+f_{DL,i}\right)}{SF_i\cdot\left(E_b/N_o\right)_i^{-1}+(1-\alpha)} \qquad (2)$$

where $f_{DL,i}$ is the measured other-to-own cell interference factor in user *i* and α is the orthogonality factor [11]. Similarly, the expected power in BS_k is given by:

$$P_T^k(n,t)=\frac{P_N}{(1-\eta_k(n,t))}\sum_{i=1}^{n}\frac{L^{i,k}}{SF_i\cdot\left(E_b/N_o\right)_i^{-1}+(1-\alpha)} \qquad (3)$$

In the implemented algorithm, conversational users are always served while, for interactive users, transmissions are carried out only if:

- there are enough codes (Kraft's inequality [12] to check the usage of the OVSF tree devoted to DSCH);
- the estimated load factor falls below a given bound $\eta_k(n,t)\le\phi$;
- the estimated transmitted power falls below a percentage of the maximum power. $P_T^k(n,t)\le\delta\cdot P_T^{MAX}$.

In case any condition is not satisfied, the SF is doubled, or equivalently, the transmission bit rate is reduced by half, and the conditions are checked again. If this is not possible, the request should wait for the next frame.

III. DESCRIPTION OF THE EMULATION PLATFORM

A. *UMTS emulator*

TCP performance results have been obtained using a real time UMTS emulator developed within the framework of the IST ARROWS project [13]. 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 (UE), the UMTS Terrestrial Radio Access Network (UTRAN), and the UMTS Core Network (CN). Figure 1 shows the basic mapping of UMTS components into testbed machines.

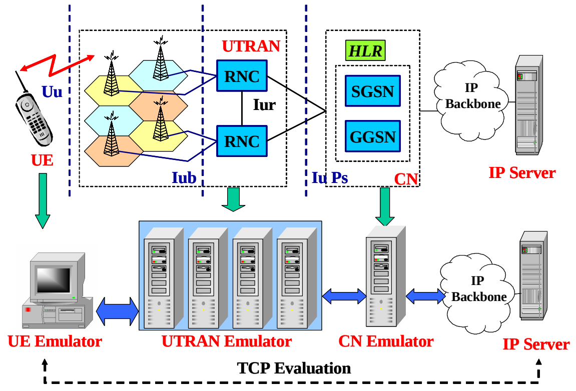

Figure 1. UMTS real-time emulator

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 [14]. Physical layer emulation has been addressed by means of histograms obtained from off-line simulations [13].

B. Settings of the analysed scenario

The UMTS emulator is configured with a 5kmx5km macrocell scenario where cells are omnidirectional and hexagonally distributed with a cell radius of 500m. The system is loaded with a combination of conversational and interactive users. Table 1 provides the Transport Formats Combinations (TFC) used in the correspondent Radio Access Bearers (RABs). Conversational service is offered through dedicated channels (DCH) with two possible Transport Formats (TF): 64kbits/s and no-transmission. For 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). As explained in section II, the selected SF depends on the service credit given to that connection and on the limitations imposed later on by the radio resource manager. In this analysis a 64kbits/s service credit is assigned to each interactive users, including the user under test.

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

Conversational users are assumed to use a videoconferencing service and are modelled by means of constant bit rate sources at 64 kbps with average call duration of 120s. For interactive users, the WEB model provided in [15] has been used. The traffic model parameters have been adjusted to provide a mean generation bit rate of 61.4 kbits/s when downloading a page. Values related to the radio QoS parameters of both services are given in Table 2.

Maximum transmission power of base stations has been fixed to 43 dBm. Neither admission control nor congestion control is considered in the system and handover decisions are taken each 100ms assuring the mobile is always connected to the best cell with a replacement hysteresis of 1dB. Orthogonality factor has been fixed to 0.4 and other-to-own-cell interference factor $f_{DL,i}$ used in the radio resource controller is assumed to be computed by statistical estimation methods among the values experienced by all the users (f_{DL}=0.6 has been considered in our case).

Table 2. Lower layer parameters.

QoS parameters	CONV	WWW.
BLER target	1%	10%
Max ACKs (RLC layer)	0	4
Max SDU Size (bytes)	160	570

TCP behaviour is analysed by executing a browsing application in the UE against a web server running in the IP Server machine. Both the server and the client run Linux version 2.4-16 with the Timestamps TCP option enabled. The impact of the SACK and FACK TCP options on the end-to-end TCP throughput is considered in the study. The TCP options DSACK and Window Scaling are disabled. The maximum receiver window is 64 KB and the Maximum Segment Size (MSS) is 530 bytes. Buffer size in UTRAN RNC is set to 32KB. The *tcptrace* tool was used to analyse the TCP traces captured. The performance of TCP over the UMTS radio interface depends on a large number of parameters and, therefore, the results obtained in this paper are valid for the UMTS settings considered.

IV. RESULTS

A. System level results

Results have been obtained for a scenario loaded with 250 conversational users and either 400 or 600 interactive users. This number of users has been fixed after several experiments since effects on TCP are already observable. The maximum load factor ϕ used in the radio resource scheduler has been varied from 60% to 100% while the percentage of maximum power control has been fixed to δ=100%. Table 3 shows the load factor values and the acceptation of transmissions in the DSCH experienced in the central BS of the considered scenario. From these results, it is clearly shown that, when 400 web users are considered, no limitation in the radio scheduler regarding the load factor (ϕ=100%) leads to a mean load factor of 77% and a ratio of acceptance in the DSCH near to 93%, being in this case the shortage of codes the main reason of DSCH requests rejections.

Table 3. System level results for the selected scenario

	System loaded with 250 Conversational and 400 Interactive Users						System loaded with 250 Conversational and 600 Interactive Users					
	η	Prob η>90%	DSCH Acceptation Requests	DSCH Code Acceptation (%)	DSCH Power Acceptation (%)	DSCH Load Acceptation (%)	η	Prob η>90%	DSCH Acceptation Requests	DSCH Code Acceptation (%)	DSCH Power Acceptation (%)	DSCH Load Acceptation (%)
No WEB Users	0.61	2.72	-	-	-	-	0.61	2.72	-	-	-	-
ϕ=60%	0.74	13.28	32.21	99.11	100.00	33.10	0.77	17.72	19.92	96.87	100.00	23.05
ϕ=70%	0.76	17.20	61.63	97.17	100.00	64.46	0.80	27.04	38.44	88.44	99.99	50.01
ϕ=80%	0.76	18.03	75.51	94.85	99.99	80.67	0.83	34.77	53.50	81.92	99.99	71.60
ϕ=90%	0.78	21.40	87.87	93.87	99.98	94.02	0.84	37.50	64.79	73.78	99.97	91.03
ϕ=100%	0.77	21.40	92.97	94.26	98.77	99.95	0.84	40.19	66.53	69.48	97.18	99.87

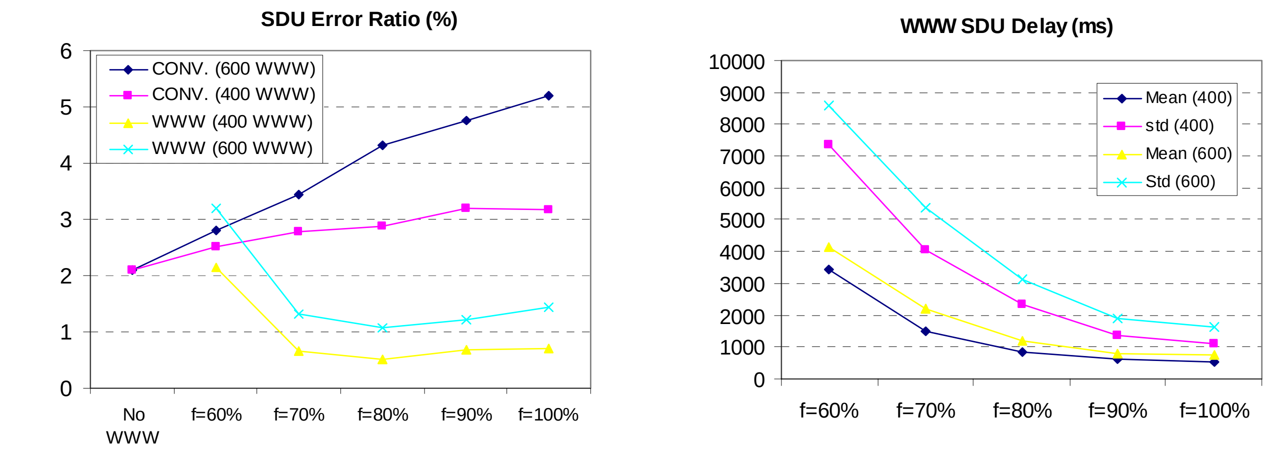

Figure 2. SDU Error Ratio for conversational and interactive traffic (leftmost graph) and interactive delay (mean and standard deviation) normalised for 570 bytes length packets (rightmost graph) versus maximum allowed load factor ϕ.

Otherwise, limiting the load factor bound ϕ the acceptance ratio in the DSCH decreases considerably as well as the load factor of the BS. The same behaviour is observed when 600 Web users are considered. Under such scenario, it is interesting to notice the SDU Error Ratio experienced by both conversational and interactive services. According to leftmost graph in Figure 2, a 2% SDU Error Ratio is achieved for conversational services when no Web users are considered. This value corresponds exactly to have a target BLER=1% and use two radio blocks of 640 bits each one (see Table 1 and Table 2) to carry out a 160 bytes packet. However when interactive users are present in the system, the conversational SDU Error Ratio is degraded but the value of ϕ can be used to control this degradation. Regarding SDU Error ratio of Web traffic, two different effects are observed in the Figure 2 depending on the value of ϕ. For low values of ϕ, the percentage of interactive SDU Error ratio increases as some packets are discarded due to RLC buffer overflow. On the other hand, high ϕ values lead to higher interference level and interactive traffic is degraded as well as we have argued for conversational traffic. Rightmost graph in Figure 2 provides the effect of varying ϕ on the SDU delay for interactive users (the delay is normalised to a 570 packet length). In this case it is observed that, in terms of delay, interactive traffic does well out of increasing ϕ since more acceptance is achieved in the DSCH channel.

So, these results demonstrate how traffic class prioritisation could be used to control SDU Error ratio for conversational services (as it is the main QoS parameter to be guaranteed for this service) at the expense of modifying the delay and the SDU Error Ratio experienced by interactive traffic.

B. Effects over TCP connections

Figure 3 shows the TCP throughput values obtained for the user under test for the different ϕ values. For the transmission with 400 www users the impact of the ϕ value on TCP is remarkable. The TCP throughput increases with the ϕ value. The reason for this behaviour is that the SDU error ratio and the SDU delays decrease with the ϕ (see Figure 2).

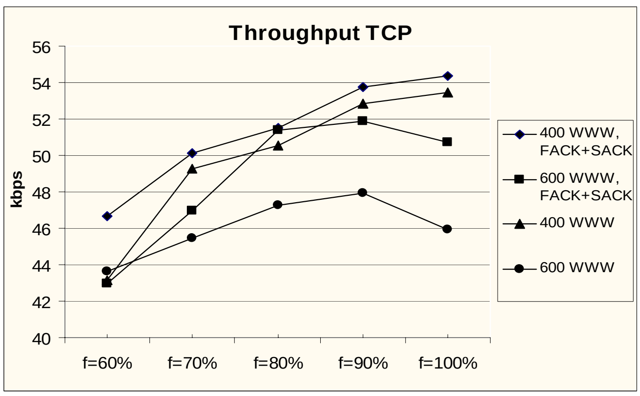

Figure 3. TCP throughput obtained for different ϕ values.

Figure 4 shows 2 examples of transmissions with ϕ values of 90 and 70. For a transmission with a ϕ value of 90 we observe few packet losses, whereas for a transmission with a ϕ value of 70 we observe high delays that sometimes trigger the timers at the sender (see Figure 4).

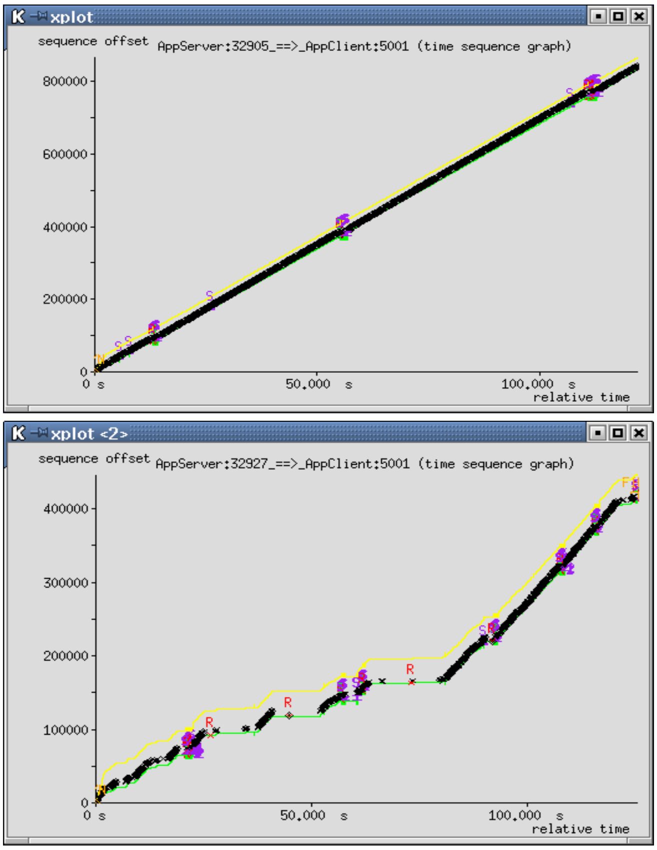

Figure 4. Sequence number evolution of transmissions for WWWusers=400, SACK,FACK enabled and ϕ value=90 (top) and ϕ value=70 (bottom).

For the transmission with 600 users the TCP throughput increases with the ϕ values of 60, 70 and 80, but for the ϕ values of 90 and 100 it decreases in comparison with the ϕ value of 80. The SDU delays decrease with the ϕ value, but the SDU error ratio initially decreases and then, since ϕ=80 (80,90,100), the SDU error ratio starts to increase due to a higher interference level. The increase of the SDU error ratio is not so pronounced for the case of 400 WWW users.

Regarding the impact on the TCP throughput of the SACK and FACK mechanisms, for the case of 400 www users the SACK_FACK mechanism enhances the TCP throughput in overall on the same percentage in all the cases.

When the system has 600 www users the impact of the SACK_FACK options on TCP is clearly observed. When the system has high ϕ values (100,90) the system generates a lot of sporadic losses (only 1 loss), which are well treated by the SACK_FACK mechanism. When the system has low ϕ values (60,70) there are a few burst losses, which are treated by the SACK_FACK mechanism as better as possible, but not with the same effectiveness as with only one loss.

This increase in delay, mainly in the form of delay spikes, can cause unnecessary timer-driven retransmissions at the TCP sender, which leads to a reduction of the TCP throughput.

V. CONCLUSION

This paper has shown the effect on real TCP connections of a radio resource management strategy that prioritizes conversational over interactive traffic and adapts the amount of bandwidth devoted to interactive traffic depending on system load. Under the analysed scenario, results show how a maximum allowed load factor ϕ for interactive traffic could be used to control SDU Error ratio for conversational services (as it is the main QoS parameter to be guaranteed for this service) at the expense of modifying the delay and the SDU Error Ratio experienced.

Regarding the impact of this maximum load factor ϕ on the TCP throughput of a real-time user, we have observed that, when the system load is less than 75-80%, the TCP throughput improve with higher values of ϕ, but when the system is too much loaded (load higher than 80%) we have found that there is an optimum ϕ value that maximizes the TCP throughput.

ACKNOWLEDGMENTS

This work has been realised over a testbed developed under the ARROWS project under the IST framework (IST 2000-25133) and it has been supported in part by the Spanish Research Council under grant TIC 2001-2222.

REFERENCES

[1]. H. Inamura et al., “TCP over second (2.5G) and third (3G) generation wireless networks”, IETF RFC3481, Feb. 2003.

[2]. M. Kojo et al.,”An Efficient Transport Service for Slow Wireless Telephone Links”, IEEE JSAC, September 1997.

[3]. J. Rendon, F. Casadevall, and J. Carrasco, "Wireless TCP Proposals with Proxy Servers in the GPRS network", IEEE PIMRC 2002, Lisboa, Portugal, September 2002.

[4]. M. Yavuz and F. Khafizov, “TCP over wireless links with variable bandwidth”, VTC Fall 2002, Canada, September 2002.

[5]. M. Taferner and E. Bonek, “Wireless Internet Access over GSM and UMTS”, Springer-Verlag, 2002.

[6]. F. Lefevre, G. Vivier, “Optimizing UMTS Link Layer Parameters for a TCP Connection”, VTC Spring 2001, Greece, May 2001.

[7]. M. Sagfors et al., “Queue Management for TCP Traffic over 3G Links”, IEEE WCNC, March 2003.

[8]. B.J. Prabhu et al., “A simulation study of TCP performance over UMTS”, IEEE VTC2003-Fall, USA, October 2003.

[9]. Ramon Ferrús et al., "Real-time emulation of RRM strategies for UMTS bearer services”, IEEE VTC 2002 Fall.

[10]. 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.

[11]. H. Holma and A. Toskala, “WCDMA for UMTS”, John Wiley & Sons, 2000.

[12]. T. Minn, K.Y. Seu, “Dynamic Assignment of Orthogonal Variable Spreading Factor Codes in W-CDMA”, IEEE JSAC, August 2000.

[13]. ARROWS Project WEB site, http://www.arrows-ist.upc.es/

[14]. 3GPP TS 25.301, “Radio Interface Protocol Architecture”

[15]. ETSI TR 101 112 v3.2.0, “Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03)”, ETSI, 1997.