A Dynamic Buffer Management Scheme for End-to-End QoS Enhancement of Multi-flow Services in HSDPA

Suleiman Y. Yerima and Khalid Al-Begain
Integrated Communications Research Centre
Faculty of Advanced Technology, University of Glamorgan
Pontypridd (Cardiff) CF37 1DL, Wales, UK
E-mail: syerima@glam.ac.uk

Abstract

End-user multi-flow services support is a crucial aspect of current and next generation mobile networks. This paper presents a dynamic buffer management strategy for HSDPA end-user multi-flow traffic with aggregated real-time and non-real-time flows. The scheme incorporates dynamic priority switching between the flows for transmission on the HSDPA radio channel. The end-to-end performance of the proposed strategy is investigated with an end-user multi-flow session of simultaneous VoIP and TCP-based downlink traffic using detailed HSDPA system-level simulations. Compared to an equivalent static buffer management scheme, the results show that end-to-end throughput performance gains in the non-real-time flow and better HSDPA channel utilization is attainable without compromising the real-time VoIP flow QoS constraints.

1. Introduction

High Speed Downlink Packet Access (HSDPA), has been standardized by the Third Generation Partnership Project (3GPP) to improve packet-switched services support on WCDMA UMTS mobile networks. HSDPA increases the available downlink peak data rates per cell from 2 Mbps to 14.4 Mbps enabling the deployment of new mobile applications and services.

HSDPA utilizes a downlink shared channel to transmit data to the User Equipments (UE) in the cell. It provides increased capacity per cell and better end-user experience, with shorter connection and response times. HSDPA consists of three interacting domains; Core Network (CN), UMTS Terrestrial Radio Access Network (UTRAN) and the UE i.e. the receiver. The Core Network is responsible for switching, transit and routing of user traffic. UTRAN provides the air interface access for the UEs and handles all radio related functionalities. UTRAN consists of a Radio Network Controller (RNC) and base station or Node-B. In HSDPA, additional functionalities have been introduced in the Node-B which include fast link adaptation based on adaptive modulation and coding (AMC), hybrid automatic repeat request (HARQ), and a shorter minimum transmission time interval (TTI) of 2ms. AMC utilizes different modulation and coding schemes which are selected for transmission of traffic to the UE based on the experienced radio channel quality of the UE reported to the Node-B via a channel quality indicator (CQI). Furthermore, the packet scheduler is moved from the RNC to the Node-B. See [1]-[4] for further details on HSDPA architecture, protocols and performance evaluation.

Packet scheduling and HARQ retransmissions necessitate buffering in the Node-B. This provides opportunity to employ buffer management strategies to enhance performance of downlink packet switched services. Certain HSDPA sessions could be characterized by multiple flows with diverse QoS requirements being concurrently downloaded to a single user. Such multi-flow services would benefit from advanced buffer management strategies applied in the Node-B to manage the flows' QoS requirements which in turn impact on their end-to-end performance.

Hence, in this paper we propose and evaluate the performance of a dynamic buffer management strategy for such multi-flow services in HSDPA. The scheme extends the active Time-Space-Priority (TSP) buffer management proposed and analyzed in [5], by incorporating dynamic time (transmission) priority switching between the flows (based on RT flow delay budgets). TSP is a novel queuing concept in which flows of the same session downloaded to a UE are classed into real-time (RT) and non-real-time (NRT) with the former accorded time priority while the latter is given space priority [6].

The new scheme, termed Dynamic Time-Space-Priority (D-TSP) is investigated with an end-user mul-

978-0-7695-3333-9 /08 $25.00 © 2008 IEEE
DOI 10.1109/NGMAST.2008.43

ti-flow session of simultaneous real-time VoIP and non-real-time TCP-based file download traffic using detailed HSDPA system-level simulation. Compared to active TSP, it is shown that better end-to-end NRT throughput and improved HSDPA channel utilization is attained with D-TSP.

In the next section, we describe HSDPA buffer management with active TSP and D-TSP respectively. Section 3 presents HSDPA simulation methodology while experimental results are discussed in section 4. Finally, concluding remarks are given in section 5.

2. HSDPA Buffer Management

In HSPDA Node B, separate data buffers are provided for each user in the Medium Access Control (MAC-hs) entity. In each user's MAC-hs buffer, MAC Protocol Data Units (PDUs) are received from the RNC over the Iub interface [7]. In order to achieve efficient resource utilization whilst also improving end-user experience, advanced Node-B buffer management strategies are necessary in HSDPA, more so for multiple flow sessions. In this section we describe buffer management strategies for handling multi-flow sessions; TSP, and a new dynamic version D-TSP whose comparative performance analysis are given in section 4.

2.1 Time-Space Priority Scheme

Time-Space priority (TSP) is a hybrid priority queuing mechanism that combines time priority with space priority into a multi-flow buffer management scheme using threshold(s) to control the QoS of diverse flows within a multi-flow session. With TSP, diverse flows destined to a single user are classed into RT and NRT flows and the RT packets (such as video or voice packets) are queued ahead of NRT packets (such as email or FTP packets) for priority transmission on the shared channel (i.e. Time Priority).

Furthermore, as illustrated in Figure 1, given a maximum buffer space allocation of N packets for a multi-flow user, NRT packets get space priority via restriction of queued RT packets with a threshold R, which allows the loss tolerance of the RT flow to be exploited in favour of loss-sensitive NRT packets. Thus the maximum number of admitted RT packets into the buffer at any given time, is R. Whereas space priority allows up to a maximum of N NRT packets in the queue thereby minimizing loss.

Due to loss sensitivity, arrival of NRT packets at the RNC necessitates the use of RLC Acknowledged Mode (AM) for onward transmission over the Iub interface to the Node-B. RLC AM packets require acknowledgement from the peer RLC entity in the UE. This feedback is sent by the RLC UE entity via a STATUS message, which is triggered by a POLL message from the RNC RLC entity [8]. Thus the loss of AM NRT PDUs due to Node-B buffer overflow will trigger retransmissions leading to waste of Iub resources, Node-B buffer space as well as air interface transmission resources. In addition, RLC round trip time increases, resulting in overall end-to-end delay of NRT packets, and hence degradation in end-to-end throughput for TCP-based applications.

In addition to R, an active-queue-managed TSP solution includes additional thresholds L and H as shown in Figure 1, where the feedback via the *NBAP Iub* signalling issues grants to the RLC entity to control the arrival rate of the NRT PDUs and hence minimize losses due to Node-B buffer overflow. Using the grant allocation, NRT PDU arrival rate λ is reduced by a factor δ when average queue length, Ave_Q, is between L and H and set to zero when Ave_Q exceeds H. The overall active TSP solution enables reduced RNC and Node-B buffer requirements by keeping the queue lengths small and minimizing NRT round trip time ultimately improving NRT end-to-end throughput.

A potential problem with the active TSP solution, is that stalling of NRT packets in the Node-B buffer could occur at high RT arrival rates, increased shared channel load, or deteriorating radio channel conditions. This is because transmission of RT packets is always prioritized in order to meet delay constraints. Stalling of NRT packets could reverse the gain of active Iub flow control by causing large queue build up in the RNC resulting in increased round trip time and hence NRT throughput degradation. A possible solution to this problem is to incorporate priority switching in a Dynamic TSP scheme as described next.

2.2 Dynamic Time-Space Priority Scheme

The dynamic TSP (D-TSP) scheme extends the aforementioned active TSP strategy by incorporating dynamic switching of transmission (time) priority between RT and NRT flows. For a given transmission opportunity assigned by the Packet Scheduler, when there is no danger of Head-of-Line (HOL) queuing delay of RT packets exceeding a given delay budget, transmission priority is switched to the NRT flow. If RT HOL delay is greater than or equal to the delay budget or no NRT packets are present in the queue, transmission priority remains with the RT flow. The delay budget can be expressed in terms of the number of queued RT packets via a parameter k where:

Delay budget = RT packet inter-arrival time $\times$ k

Thus, k = 2 and an RT packet inter-arrival time of 20 ms is equivalent to a delay budget of 40 ms.

Let *MAX_delay* represent the maximum allowable queuing delay to enable end-to-end QoS delay guarantee for RT flow. A Discard Timer (DT) is set on arrival of RT packets to the MAC-hs buffer. DT is configured to time-out after a period of *MAX_delay,* triggering the dropping of HOL RT packet(s) queued for up to *MAX_delay* seconds. DT is cancelled on transmission of RT packet(s). We can therefore express the time priority switching strategy as follows.

IF RT packets < k *AND RT HOL delay*< *MAX_delay AND NRT packets* > *0*
Time Priority = NRT flow
Generate Transport Block from NRT PDUs
ELSE
Time Priority = RT flow
Generate Transport Block from RT PDUs

It must be stressed however, that D-TSP solution necessitates the use of a play-out buffer in the UE which should be configured as suggested in [9]. This will eliminate the effects of jitter which may result from VoIP PDU bundling during transmission.

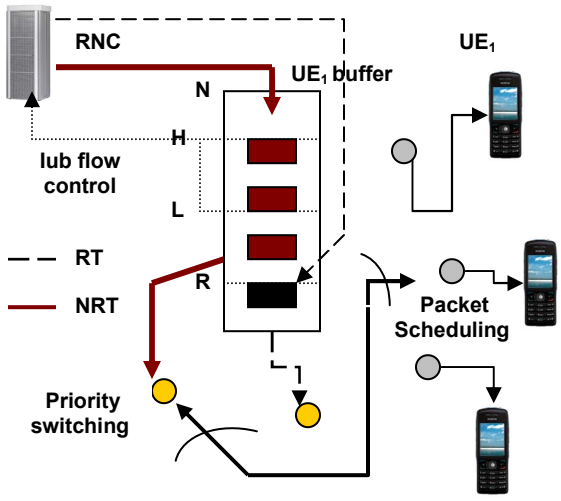

Figure 1. Active TSP with dynamic priority switching (D-TSP) in UE$_1$ Node B MAC buffer

3. HSDPA Simulation Model

In order to evaluate the performance of the proposed D-TSP scheme, a system level HSPDA simulation model was developed using OPNET. The simulation model is shown in Figure 2. The multi-flow source node includes a full implementation of VoIP ON/OFF source with the same parameters employed in [9], and a

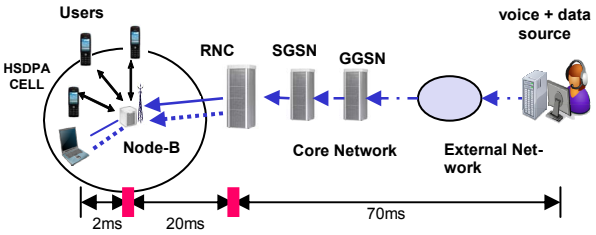

Figure 2. HSDPA Simulation model

customizable NRT source with TCP Reno implementation.

Other aspects of HSDPA modeled in detail include: RNC, with packet segmentation, RLC MAC queues, RLC AM and UM modes including ARQ for AM mode. RNC – Node-B Iub signaling is also modeled. In the Node-B, MAC-hs queues (applying TSP and D-TSP), HARQ processes, AMC schemes, and Packet Scheduling on the HSDPA air interface are modeled. In the receiver, we included SINR calculation and CQI reporting, HARQ processes, RLC modes with ARQ for AM, packet reassembly queues, peer TCP entity, and an application layer.

In the experiments, a test user equipment (UE 1) was assumed to be receiving multi-flow traffic of simultaneous RT (VoIP) and NRT (FTP) during a 180s simulated voice conversation and file download session. VoIP packets were being received while file download was taking place using FTP over TCP. The overall set up models a single HSDPA cell with Round Robin packet scheduling to m users. A summary of the HSDPA parameters used are given in Table 1.

Buffer scheme configuration include : R=10 PDU, L=100 PDU, H=150 PDU, N = 200 PDU.

We assume that maximum allowable one way VoIP delay is 250 ms [10], [11]. For the VoIP flow, RNC queuing delay is negligible since active TSP grants ensures prompt transmission to Node-B. Hence, from Figure 2, estimated maximum queuing delay budget is given by: 250 – (External +CN delays) – Iub delay = 160 ms. We therefore set *MAX_delay* for discard timer to 160ms. The D-TSP parameter k, was varied from 2, 4 , 6 and 8 corresponding to delay budget settings of 40, 80, 120 and 160ms respectively, since VoIP PDU inter-arrival time is approximately 20 ms during ON periods. Performance metrics observed include:

- *End-to-end NRT throughput*: the end-to-end TCP throughput at the test UE 1 during file download in the multi-flow session.
- *RT PDU Discard Probability*: defined as the number of late HOL RT PDUs discarded from the (D-TSP or TSP) MAC-hs queue as a result of DT timeout.

Table 1. Simulation parameters

HSDPA Simulation Parameters	
HS-DSCH TTI	2ms
Path loss Model	148 + 40 log (R) dB
Transmit powers	Total Node-B power=15W, HS-DSCH power= 50%
Shadow fading	Log-normal: $\sigma = 8$ dB
AMC schemes	QPSK ¼, QPSK ½, QPSK ¾, 16QAM ¼, 16 QAM ½
Number of HS-DSCH codes	5
CQI letency	3 TTIs (6ms)
HARQ processes	4
HARQ feedback latency	5ms
Packet Scheduling	Round Robin
MAC PDU size	320 bits
Iub (RNC-Node-B) delay	20ms
External + CN delays	70ms
TCP (Reno)	MSS =536 bytes, RWIND = 64

- *Percentage air interface utilization*: calculated from Transport Block Size transmitted divided by maximum Transport Block Size allowable by the selected AMC scheme, measured at every transmission opportunity.

4. Results and Discussions

Figures 3 to 8 depict the end-to-end NRT flow throughput of an end user (UE 1) terminal running a multi-flow session of simultaneous VoIP and TCP-based file download on HSDPA channel. The average throughput over a session period of 180s are plotted in the graphs for various VoIP delay budget settings of the *dynamic TSP* and are compared to that of the TSP buffer management. Each of the Figures depict results obtained with different number of users sharing the HSDPA channel in a single cell. In all scenarios, UE 1 is assumed to be stationary and located 0.2 km from the base station while other users (where applicable) are placed at random positions within the cell.

Figure 3 gives the NRT throughput of UE 1 terminal when it occupies the HSDPA channel alone. Consequently, it is being allocated all the available channel codes in every TTI. We observe that increasing the dynamic TSP parameter ***k***, with a 160ms Discard Timer setting does not yield a significant increase in average throughput compared to the TSP scheme. This can be explained by the fact that (depending on radio conditions) scheduling transmission every TTI for the UE 1 curtails the accumulation of RT PDUs in the buffer reducing the possibility of loss of transmission

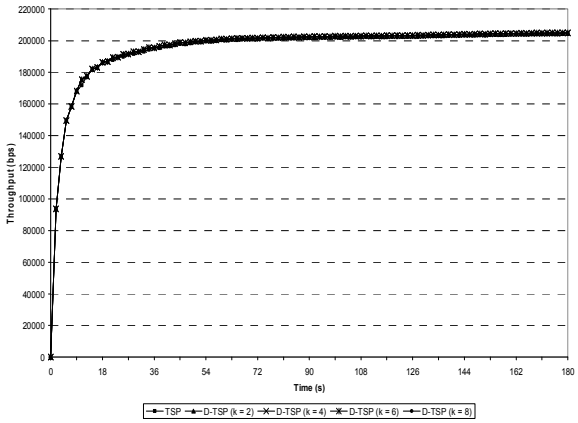

Figure 3. UE 1 NRT throughput for various VoIP delay budget settings when utilizing HSDPA channel alone

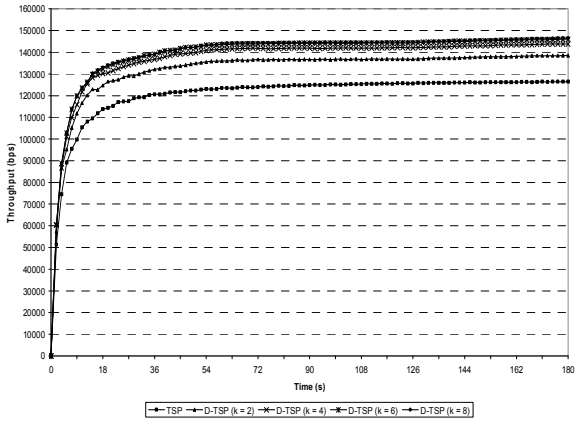

Figure 4. UE 1 NRT throughput for various VoIP delay budget settings with 5 users sharing HSDPA channel in cell

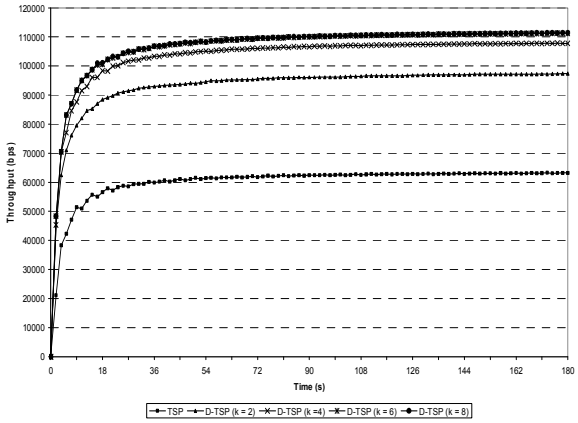

Figure 5. UE 1 NRT throughput for various VoIP delay budget settings with 10 users sharing HSDPA channel in cell

opportunity for the NRT PDUs in TSP. As a result, application of D-TSP buffer management with even the most relaxed delay budget setting can only yield marginal improvement in NRT throughput.

In contrast, noticeable performance gain is observed with the D-TSP as more users occupy the HSDPA channel. In Figure 4, the throughput of UE 1 is plotted for a scenario with a total of 5 users connected to the HSDPA channel with Round Robin scheduling employed by the packet scheduler. The TSP scheme achieves a steady state peak average throughput of about 125 kbps, whereas the D-TSP scheme with k = 8 gives a peak throughput of 145 kbps.

The experiment is repeated for scenarios with the same simulation settings but with 10, 20, 30 and 50 users on the HSDPA channel and the results are depicted in Figures 5 -8 respectively. From Figure 4, average UE1 NRT throughput with TSP is around 60 kbps and increases to about 110 kbps with D-TSP (k = 8) in the 10-user scenario. Figure 6 shows increase in UE 1 NRT throughput from about 42 kbps with TSP, to 71 kbps with D-TSP and k = 8, in the 20-user scenario. For the scenario with 30 users, Figure 7 shows increase in UE 1 NRT throughput from 32 kbps with TSP, to 50 kbps with D-TSP and k = 8. Lastly, Figure 9 shows increase in UE 1 NRT throughput from 18 kbps with TSP, to nearly 32 kbps with D-TSP and k = 8 in the scenario with 50 users.

4.1. VoIP flow QoS in the multi-flow session

Since a Discard Timer is used to discard Head-of-Line VoIP PDUs with delay exceeding the *MAX_delay setting* of 160ms, PDUs violating the delay deadline bound will not be received at the UE 1. Thus, as a measure of the UE 1 VoIP QoS in the multi-flow session, we consider the VoIP PDU discard probabilities for the aforementioned scenarios for both TSP and D-TSP with the various k settings. The results are illustrated in Figure 9. Generally, more VoIP PDUs are discarded from the UE 1 MAC-hs queue as more users are scheduled on the HSDPA channel and also with higher k settings which correspond to more relaxed delay budget. Assuming a maximum of discard ratio of 2% is acceptable for VoIP QoS, Figure 9 shows that VoIP QoS is satisfied in all cases of D-TSP and TSP for the 1-user, 5-user, 10-user scenarios. (Note that DT mechanism is also applied in TSP). Whereas for the 20-user scenario the maximum setting of k for the D-TSP is 6. For the 30-user scenario maximum acceptable k = 4 while in the 50-user scenario maximum acceptable setting for k is 2.

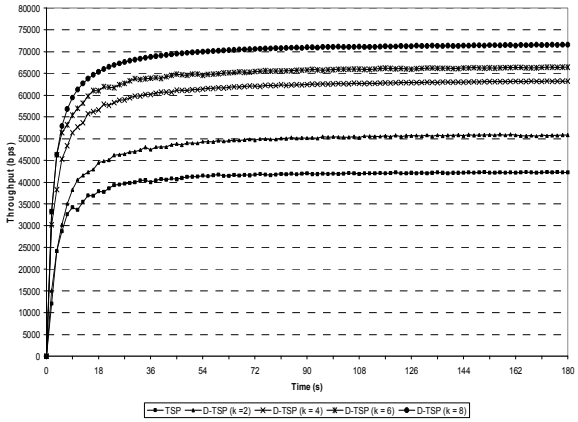

Figure 6. UE 1 NRT throughput for various VoIP delay budget settings with 20 users sharing HSDPA channel in cell

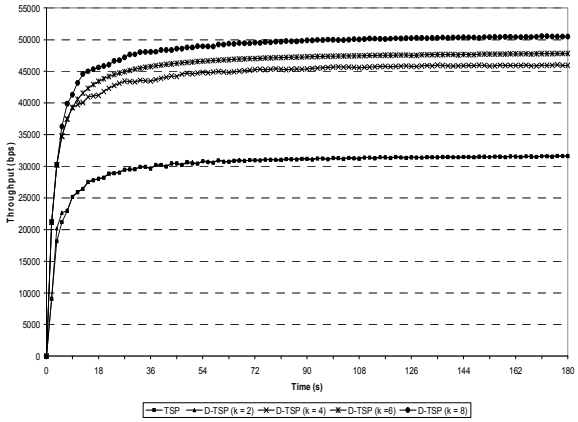

Figure 7. UE 1 NRT throughput for various VoIP delay budget settings with 30 users sharing HSDPA channel in cell

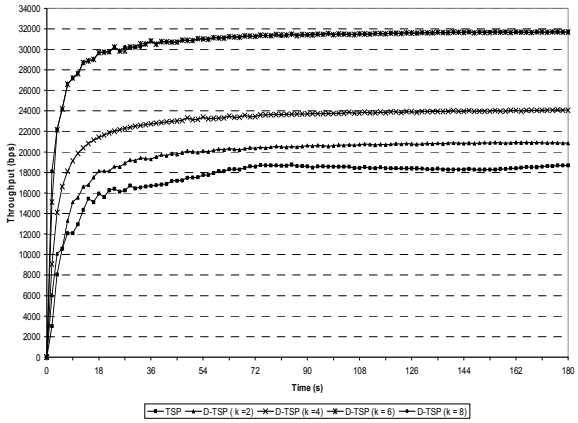

Figure 8. UE 1 NRT throughput for various VoIP delay budget settings with 50 users sharing HSDPA channel in cell

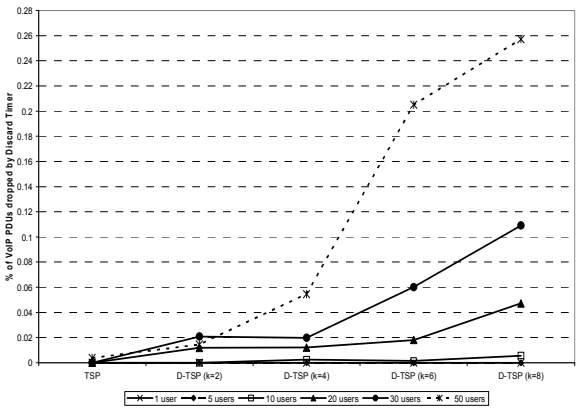

Figure 9. % VoIP PDU Discarded for UE 1

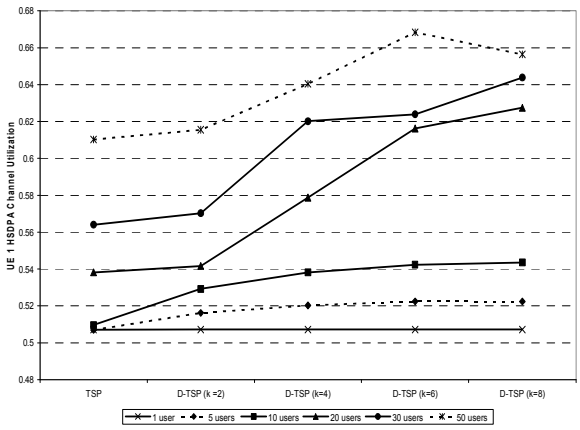

Figure 10. UE 1 HSDPA channel utilization for the delay budget settings.

4.2 HSDPA channel utilization

In addition to the throughput performance gain, D-TSP also improves the air interface utilization on the HSDPA channel compared to TSP. As seen from Figure 10, the higher the number of users being scheduled on the air interface with the Round Robin scheme, the better the air interface utilization. For instance, in the 20-user scenario, total channel utilization is 54 % for both RT and NRT flows in the multi-flow session of the UE 1 when using TSP. With D-TSP and k = 6, on the other hand, utilization of almost 62% is achieved. This is due to VoIP PDU bundling in the Transport Block during transmission.

5. Concluding Remarks

This paper proposed a dynamic buffer management scheme, D-TSP, for end-user QoS management of multi-flow sessions with concurrent RT and NRT flows over HSDPA downlink. D-TSP incorporates dynamic time priority switching to active TSP, a time-space priority queue management with Iub flow control mechanisms. The priority switching is controlled via a parameter k related to the RT flow delay budget, while a discard timer drops RT packets likely to violate the end-to-end maximum QoS delay constraint.

Comparative performance study between D-TSP and TSP is undertaken via extensive system-level HSDPA simulations. End-to-end TCP-based NRT throughput is observed in a test receiver, including cases where multiple users share the HSDPA channel with RR scheduling. The experiments reveal that throughput gain is achieved (with higher HSDPA channel load) with D-TSP compared to TSP, and, depending on the setting of k, VoIP packet discard can be kept within QoS bounds. Finally, D-TSP not only increases UTRAN resource utilization by averting potential stalling of the NRT flow, but also improves HSDPA channel utilization. Further work will include analysis with other possible multi-flow traffic and packet scheduling algorithms.

References

[1] H. Holma and A. Toskala, Eds., *HSDPA/HSUPA for UMTS*. John Wiley & Sons, 2006.

[2] T.E. Kolding et al. "High Speed Downlink Packet Access: WCDMA Evolution," *IEEE Vehic. Tech. Soc. News*, Vol 50, No. 1, pp. 4-10, Feb. 2006.

[3] R. Love et al, " HSDPA Perfromance", IEEE Proc. VTC 2001 Fall, September 2001.

[4] H. Van den Berg, R. Litjens, J. Laverman, " HSDPA Flow Level Performance: The Impact of Key System and Traffic Aspects", *ACM MSWim' 04 Proc*., pp. 283-292, Oct. 2004, Venezia, Italy.

[5] S. Y. Yerima and K. Al-Begain, "An Enhanced Buffer Management Scheme for Multimedia Traffic in HSDPA" in Proc. *NGMAST 2007*, Cardiff, Sept. 2007.

[6] K. Al-Begain, A. Dudin, and V. Mushko, *Novel Queuing Model for Multimedia over Downlink in 3.5G Wireless Networks*, Journal of Communications Software and Systems, Vol 2, No 2, June 2006.

[7] 3GPP TS 25.435 "UTRAN Iub Interface User Plane Protocols for Common Transport Channel data streams " V ersion 5.7.0, March 2004.

[8] 3GPP TS 25.322 " Radio Link Protocol (RLC) protocol specification, V 5.13.0, December 2005.

[9] W. Bang, K. I. Pedersen, T.E. Kolding, P.E. Mogensen, "Performance of VoIP on HSDPA", *IEEE Proc. VTC*, Stockholm, June 2005.

[10] ITU Recommendation G.114, "One-way Transmission Time", 2003.

[11] R. Cuny, A. Lakaniemi, " VoIP in 3G Networks: An end-to-end quality of service analysis", IEEE Proc. VTC spring, vol. 2, pp. 930-934, April 2003.