

# Buffer Management for Multimedia QoS Control over HSDPA Downlink

Suleiman Y. Yerima, Khalid Al-Begain  
Mobile Computing, Communications and Networking RG  
Faculty of Advanced Technology, University of Glamorgan  
Pontypridd (Cardiff) CF37 1DL, Wales, UK  
E-mail: {syerima,kbegain}@glam.ac.uk

## Abstract

*HSDPA specifications include support for a flexible framework for QoS management. In this paper, it is shown how buffer management could be incorporated into HSDPA QoS framework for 'multimedia' traffic QoS control in the MAC-hs of the Node-B. A Time-Space-Priority (TSP) scheme is proposed as viable buffer management scheme to this effect. Comparative simulation study with other schemes is presented, demonstrating the effectiveness of the TSP buffer management scheme for 'multimedia' service QoS control in HSDPA Node-B data buffers.*

## 1. Introduction

Improved support for packet-switched data in UMTS mobile networks introduced in recent 3GPP releases presents opportunities for service providers to introduce novel broadband services. High-Speed Downlink Packet Access (HSDPA) is the 3GPP release 5 specifications that enhances UMTS networks to provide higher data rates and increased capacity for greater support of broadband services like multimedia conferencing, VoIP, audio/video streaming, mobile multimedia gaming, high-speed internet access, etc. The ability to support high data rates enables application developers to create content rich 'multimedia' applications, typically consisting of a number of classes of media or data- with different QoS requirements- being concurrently downloaded to a single user [1].

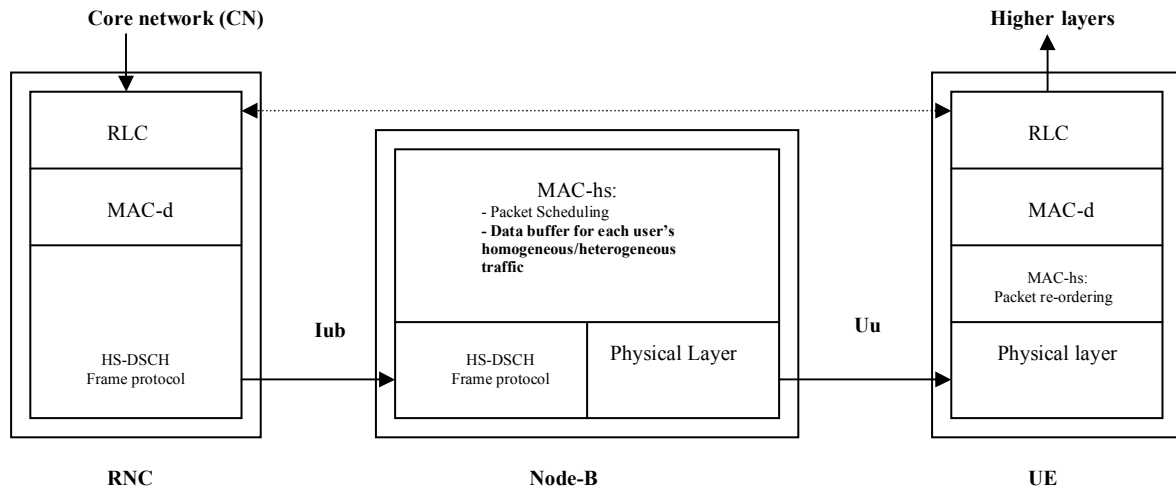
HSDPA significantly reduces downlink transmission latency, enabling theoretical data rates of up to 14.4 Mbps in addition to a three-fold capacity increase in WCDMA UMTS networks [2, 3]. A shared downlink channel is utilized, which adapts transmission capacity to changing radio propagation conditions (fast link adaptation). Fast link adaptation

employs adaptive modulation and coding (AMC) whereby different modulation and coding schemes are selected for transmission of traffic to the User Equipments (UE) within the serving HSDPA cell. AMC scheme selection is based on the experienced radio channel quality of the UE. Other features of HSDPA include HARQ for error control and channel-dependent Fast Scheduling. Minimum allocation time or TTI (Transmission Time Interval) on the shared channel is 2ms which improves the tracking of fast channel variations.

Not only have these new features been added to the Physical and MAC layers of in the HSDPA specifications, but they have also been implemented in the base station (Node-B). Furthermore, unlike in the UMTS Release 99 architecture, the packet scheduling is moved from the centralized RNC to the base station and embedded in a new MAC entity called MAC-hs.

Recent studies, in [4] for example, emphasize the flexible framework for efficient QoS differentiation featured in the Release 5 specifications. Thus, the inclusion of packet scheduling capability (i.e. MAC-hs) in the Node-B presents opportunity to utilize buffer management schemes for QoS management. Hence, in this paper, a combined Time-Space Priority (TSP) buffer management strategy is proposed for multimedia traffic QoS control over HSDPA downlink in Node-B buffers. In particular, the potential of TSP for efficient QoS control of heterogeneous multimedia traffic- comprising flows with diverse QoS requirements- is demonstrated through performance comparisons with the non-priority First-come-first-serve (FCFS), Space Priority (SP), and Time Priority (TP) buffer management schemes.

Figure 1 shows aspects of the HSDPA user-plane protocol stack relevant to buffer management based QoS control. Due to the functional split between RNC and Node-Bs, the buffer management parameters could be computed and set by the RNC as an additional Resource Management (RM) function in the control plane as shown in Figure 1.



**Figure 1. HSDPA user plane protocol stack depicting aspects relevant to buffer management based QoS control**

The scope of the paper is limited to heterogeneous traffic buffered in the Node-B and destined for a single user in a HSDPA cell. However, the concept can be extended to inter-user MAC-hs QoS scheduling of multimedia traffic over the downlink shared channel. Ongoing work is currently being undertaken to investigate this.

The rest of the paper is organized as follows. Section 2 discusses the buffer management schemes investigated in the paper. Section 3 describes the simulation model, while in section 4 performance results are presented to demonstrate the potential of the TSP scheme for multimedia QoS control and optimization in the Node-B. Finally, the concluding remarks are given in section 5.

## 2. Buffer Management in HSDPA Node-B

Several buffer management schemes [5] are discussed in the literature but not all of the existing schemes can be effectively applied to MAC-hs QoS control in HSDPA Node-B. The Node-B provides the radio interface (Uu interface) that connects the UE to the network for reception of the heterogeneous multimedia packet streams. In the uplink direction, measurements made by the Node-B indicating the channel conditions of each UE within its coverage area, are passed on to the RNC. The RNC can then use these to configure a parameter set for each UE's data buffer in the Node-B MAC-hs for QoS control and scheduling.

QoS control and scheduling are not part of UMTS R99 Node-B functionality. Instead, as noted in [6], the

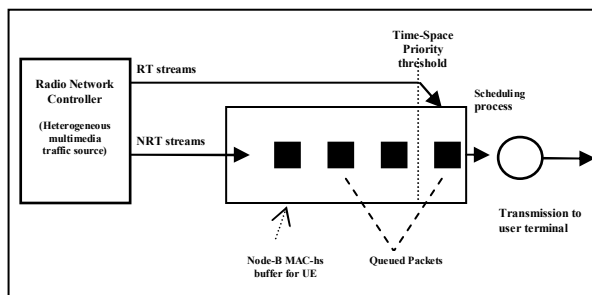
RNC in UMTS R99 utilizes QoS parameters such as traffic class (TC), traffic handling priority (THP) and allocation retention priority (ARP) etc for QoS control and QoS-aware packet scheduling. This information however, is unavailable to HSDPA Node-B MAC-hs. Thus a new QoS interface has been defined for HSDPA between the RNC and Node-B [4]. HSDPA-specific QoS parameters that are set by the RNC and sent across this new QoS interface to the MAC-hs include: guaranteed bit rate (GBR), service priority indicator (SPI) and discard timer (DT). Note that the RNC sets these parameters based on TC, THP and ARP.

The GBR indicates the average bit rate that will guide the MAC-hs scheduling, while DT specifies the maximum allowable time to buffer a packet in the Node-B before discarding. SPI is expressed as an integer value between 0-15, with a high number indicating higher priority. SPI, therefore, is a crucial parameter that could be used (by the RNC) for QoS differentiation when implementing a QoS-aware or QoS-based buffer management scheme in the Node-B's MAC-hs.

Consequently, QoS-based buffer management schemes in Node-B should be designed to utilize these parameters while also taking into account the RNC-Node-B functionality split. In addition, the scheme must be able to fulfil the individual QoS requirements of the diverse flows within the heterogeneous traffic. Our proposed solution, the TSP, is a QoS-based scheme that meets both requirements. It can be applied to efficient QoS control of heterogeneous multimedia with real-time and non-real-time flows. For example, a service with real-time (RT) conversational class traffic

and non-real-time (NRT) background class traffic such as a VoIP/Video session taking place simultaneously with file transfers (FTP).

The basic idea of Time-Space Priority concept [7]-[9] is that RT flows are given service priority because of their stringent delay requirements; while NRT flows have buffer space priority to minimize loss. This concept is illustrated in figure 2. When applied to a UE data buffer in the Node-B, RT packets (having higher SPI) will be queued in front of the NRT packets (with lower SPI) to receive priority scheduling for transmission on the shared channel. NRT packets will only be transmitted when no RT packets are present in the buffer. This way, the RT QoS delay and jitter requirements would not be compromised. However, in order to fulfil the QoS of the loss sensitive NRT flow, the number of RT packets that is admitted into the buffer queue is restricted so that more buffer space is allocated to the NRT flow. The maximum buffer capacity  $N$  and the TSP threshold  $Th$  are the parameters computed in the RNC RM based on the QoS requirements of the diverse flows, current load, and channel conditions. Other possible buffer management parameters include flow control thresholds where flow control mechanism is implemented.



**Figure 2. TSP scheme in Node-B buffer allocated to a UE receiving multimedia traffic**

Comparative simulation performance evaluation proves the Time-Space priority concept an efficient solution as will be illustrated by results of the experiments conducted. Other schemes to which the TSP concept is compared are explained next.

## 2.1. Non-prioritized FCFS

This is a non-prioritized scheme with no QoS differentiation i.e. does not take into account the QoS classes but buffers the packets in the MAC-hs in their order of arrival. Transmission is scheduled on a first-come-first serve basis. Although sometimes the scheme may be fair in terms buffer space allocation, this scheme cannot be used where strict QoS requirements for real-time flows need to be met. Any arriving packet

is rejected or blocked if a full buffer is encountered on arrival.

## 2.2. Time Priority buffer management

With the Time Priority (TP) scheme, when heterogeneous streams arrive in the data buffer, packets of the higher priority class are queued at the front to be scheduled for transmission. TP thus favours the priority class packets while being biased against the non-priority packets in scheduling for service. An arriving priority packet is guaranteed a place in buffer as long as the total number of priority packets  $R$  in the buffer is less than the total buffer capacity,  $N$ . Thus, if  $R < N$  and a full buffer is encountered by an arriving priority packet, a *push-out* mechanism is used to discard a non-priority packet to accommodate the newly arrived priority packet.

Unless a threshold is set in the buffer to limit the number of admitted arriving priority packets, the scheme could be unfair to the non-priority packets especially under high load conditions, leading to excessive loss of non-priority packets.

## 2.3. Space Priority buffer management

With the Space Priority (SP) scheme, when the mixed traffic arrives in the data buffer, packets of the higher priority class are admitted such that a larger portion of the buffer is allocated to them. A pure SP scheme does not prioritize service allocation, but instead, aims to limit the admission of non-priority packets so that more of the priority packets can be retained in the buffer. Depending on the traffic intensity, this scheme could be unfair to the non-priority packets in terms of buffer space allocation.

Packets are queued in order of arrival, but the non-priority packets are restricted to a maximum of  $P$  where  $P < N$ , the maximum buffer capacity. Any arriving priority packet will be admitted provided the buffer is not full. On the other hand, an arriving non-priority packet will be blocked if  $P$  non-priority packets are already queued in the buffer whether it is full or not.

## 3. Simulation Model

Our simulation model is implemented in C programming language. It is an event driven simulator designed for comparative analysis of the buffer management schemes and also validation of previous analytical queuing models developed by us. The simulation is limited to events and mechanisms occurring within a MAC-hs data buffer allocated to a

single UE, since this will suffice for comparative analysis of the schemes considered in this study. Hence, the physical layer air interface, MAC-hs HARQ error control, other higher layers are not explicitly simulated.

The model employs a traffic generator that generates different random arrival processes to model traffic flows into the buffer. The buffer is represented as a finite queue whose queuing behaviour is defined by and modified by altering the queue admission criteria and the service scheduling policy. Furthermore, the design allows for computation of various performance measures of interest such as mean queue length, average throughput, mean queuing delay, utilization, and packet loss probability etc. These measures are readily obtained by varying required system and traffic parameters, such as offered traffic load, buffer capacity, control thresholds, transmission (service) capacity etc. defined in the simulation framework.

The performance of the buffer management schemes are compared under the same traffic load and system conditions. The traffic source model is assumed to be mixed multimedia flow comprising VoIP stream from a speech conversation and FTP download stream of a file transfer occurring during the VoIP conversation. The VoIP traffic model employed in the simulation is implemented as described in [10], while for FTP we used the ETSI WWW model presented in [11] with one packet call for the duration of the multimedia session. Thus the multimedia stream is assumed to arrive at the Node-B comprising VoIP packets arriving every 20ms during on periods and FTP packets arriving at (geometrically distributed) time intervals corresponding to the download bit rate allocated for the session. Furthermore, we assume a receiving UE belonging to the category 8 terminal capability class. Thus the UE can support 10 parallel HS-DSCH codes, 14,600 channel bits per 2ms TTI resulting in achievable maximum data rate of 7.2 Mbps.

The buffer queuing behaviour differs from one scheme to another. Two main properties that distinguish one buffer management scheme to another are the admission criteria (including buffer space allocation) and the scheduling strategy. These were implemented for each scheme according to the descriptions given in section 2.

## 4. Results and Discussions

In this section we present and discuss the results of experiments conducted using the simulation model. A multimedia session consisting of conversational and

background traffic is simulated. Thus the scenario models a category 8 UE assumed to be having a VoIP conversation with the source whilst simultaneously downloading a file via FTP from the source. We investigate the loss and delay performance of the real-time (VoIP) and non real-time (FTP) flows by varying the data arrival rate of the FTP flows in the mixed multimedia traffic scenario. Performance measures are taken at FTP data rates of 8, 16, 32, 56 and 64 kbps respectively. The buffer parameters used in the simulation are given in Table 1.

**Table 1. MAC-hs Buffer parameters**

Scheme	Buffer parameters
FCFS	Total buffer capacity=20
SP	Total buffer capacity=20 Space priority threshold=4
TP	Total buffer capacity=20
TSP	Total buffer capacity=20 Time-Space priority threshold=4

### 4.1 VoIP performance in the mixed traffic

The impact of the NRT traffic load variation (represented as various FTP download rates) on VoIP performance is shown in figures 3 and 4. Results are presented for FCFS, TP, SP and TSP buffer management schemes for comparative analysis. The buffer parameters used in the simulation are as given in Table 1.

As depicted in figure 3, increasing FTP download rates is seen to have greater impact on the FCFS, TSP and SP than on TP, as far as VoIP packet loss is concerned. Varying FTP rates between 8 and 64 kbps resulted in zero packet loss for the TP scheme and also showed SP to have the worst VoIP loss performance. The excellent VoIP loss performance of TP is as a result of its displacement policy which drops buffered FTP packets in favour of arriving VoIP packets when the buffer is full. Whereas in the SP scheme FTP packets were accorded space priority over VoIP packets which accounts for the poorer VoIP loss performance compared to other schemes. For this particular scenario, the TSP buffer partition threshold of 4 assumed in the experiment resulted in a better FCFS VoIP loss performance compared to TSP. Since TP is a special case of TSP with threshold  $Th=N$ , the buffer capacity, the VoIP loss performance for TSP will approach that of TP when the TSP threshold is increased. This implies that TSP VoIP loss could be

better than that of FCFS if the TSP threshold is made larger.

Figure 4 illustrates the impact of FTP arrival rates on VoIP delay. FCFS is seen to have the poorest VoIP delay performance. TSP and TP, on the other hand, have better delay performance than FCFS and SP, as expected, due to the priority access to transmission resources implemented for the former category. TSP has the best VoIP delay performance of all the schemes. This clearly demonstrates the potential that TSP has for meeting real-time QoS delay requirements compared to the other schemes.

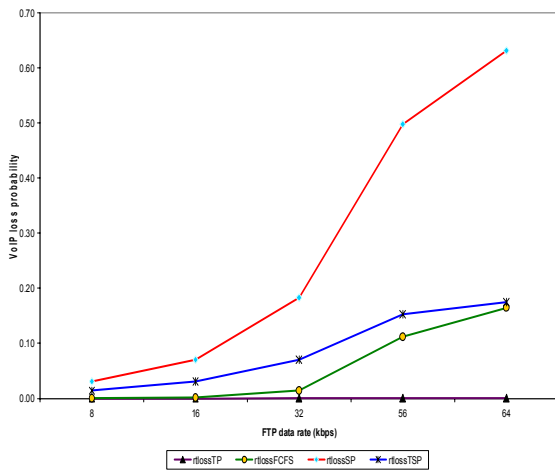


Figure 3. VoIP loss Vs FTP download rate

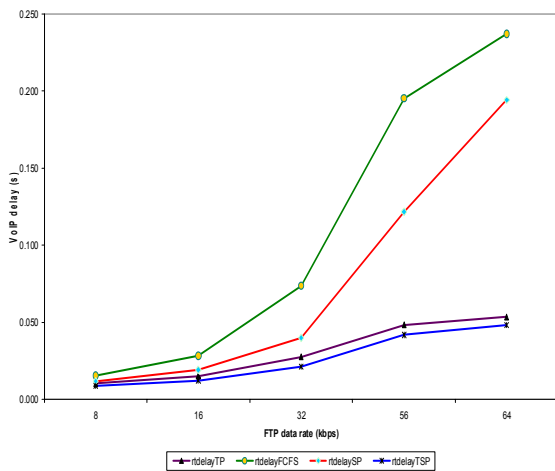


Figure 4. VoIP delay Vs FTP download rate

#### 4.2 FTP performance in the mixed traffic

For the NRT packets in the ‘multimedia’ stream, the most crucial QoS parameter is the loss or blocking

probability which must be minimized. Looking at figure 5, SP gives excellent loss performance for the FTP traffic over the FTP download rate range investigated. Also, up to 32 kbps, TSP depicts almost identical FTP loss performance. This is because of better retention of the non priority FTP packets in the buffer when SP and TSP schemes are used. FCFS and TP give worse FTP loss performance as the FTP data rate increases. Displacement of FTP packets in favour of arriving VoIP packets, accounts for this observation in the case of TP. Whilst in the case of FCFS it is because buffer space is not guaranteed for arriving FTP packets unlike in SP and TSP where the thresholds guarantee a certain amount of buffer space depending on the number of FTP packets already in the buffer.

The implication of figure 5 results is that implementing TP or FCFS as a buffer management scheme for mixed traffic will generally result in greater loss of NRT lower priority packets. In order to minimize the loss/blocking of lower priority packets, the TP scheme needs to incorporate a threshold to restrict the number of priority packets buffered. This effectively turns a TP scheme into a TSP scheme.

As FTP belongs to the background TC with lower SPI, loss probability of the packets is a more relevant performance index than delay in the context of QoS control of MAC-hs buffered multimedia traffic. However, FTP delay performance is investigated to gain better insight and for experimental verification purposes. The results are illustrated in figure 6. As expected, SP and FCFS show better delay performance than the other two. The reason for this is the absence of service prioritization thus allowing non-real-time FTP packets to enjoy fair access to transmission resources. TSP and TP both implement service priority for VoIP packets so that more FTP packets have to wait before being transmitted, hence their relatively poorer FTP delay performance.

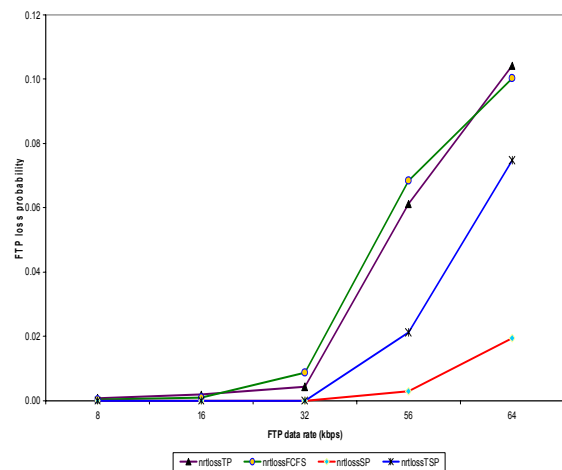


Figure 5. FTP loss Vs FTP download rate

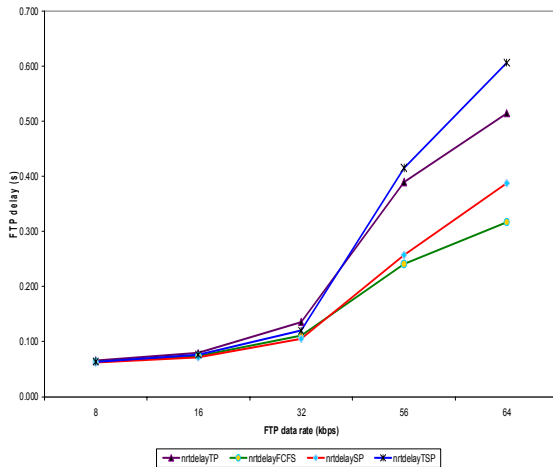


Figure 6. FTP delay Vs FTP download rate

## 5. Concluding Remarks

This paper has examined some revisions to UMTS R99 introduced in HSDPA Release 5 to support a flexible QoS control framework. The role and importance of buffer management in QoS control has been emphasized while illustrating how efficient buffer management schemes can be designed to fit into the framework and effectively utilize existing standardized architecture, QoS interface and parameters.

A Time-Space priority scheme is proposed as an example buffer management scheme that could be used for QoS control of heterogeneous multimedia service in the Node-B MAC-hs data buffer. Results of a comparative simulation study with FCFS, TP, and SP schemes given in the paper strongly indicate this potential. This is evident from the service priority attribute which allows it manage the real-time delay and loss performance to meet QoS constraints. The results show that TSP gives the best real-time delay performance amongst the schemes compared while also not excessively jeopardizing the loss performance.

Further work will entail the extension of the TSP to incorporate flow control thresholds in order to improve non-real-time traffic loss performance.

## References

- [1] A. Galoup, O. Holland and A. H. Aghvami, "Concept and Optimization of an effective Packet Scheduling Algorithm for Multimedia Traffic over HSDPA" *Proc. IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications*, Berlin, Germany, 2005.
- [2] S. Parkvall, E. Englund, M. Lundevall, and J. Torsner, "Evolving 3G Mobile Systems: Broadband and Broadcast Services in WCDMA," *IEEE Comms. Mag.*, Vol. 44, No.2, pp 68-74, Feb. 2006.
- [3] T.E. Kolding et al. "High Speed Downlink Packet Access: WCDMA Evolution," *IEEE Vehic. Tech. Soc. News*, Feb. 2006.
- [4] K. I. Pedersen, P. E. Mogensen, T.E. Kolding, "Overview of QoS Options for HSDPA" *IEEE Communications Magazine*, pp 100-105, July 2006.
- [5] Y. Bai, M. R. Ito, "Application-aware Buffer Management Schemes: new metrics and techniques" *IEEE Transactions on Broadcasting*, Vol 51, No. 1, March 2005.
- [6] S. Dixit, Y. Guo, and Z. Antoniou, "Resource Management and Quality of Service in Third-Generation Wireless Networks," *IEEE Commun. Mag.*, vol. 39, pp. 125-133, Feb. 2001.
- [7] K. Al-Begain, A. Dudin, and V. Mushko, "Novel Queuing Model for Multimedia over Downlink in 3.5G Wireless Networks," In *Proc. 12<sup>th</sup> Int. Conf. on Analytical and Stochastic Techniques and Applications*, Riga, Latvia, pp. 111-117, June 2005.
- [8] Al-Begain, K.; Dudin, A.; and Kizimirsky, A. "QoS Optimisation of Multimedia Services in 3.5G Wireless Networks", *Proc. 9th International Conference on Modelling and Simulation*, Al-Begain, Orsoni, A-Dabass (Eds.) Oriel College, Oxford, UK, pp. 27-32. April 2006.
- [9] S. Y. Yerima, K. Al-Begain, "Analysis Of  $M_2/M_2/1/R$ , N Queuing Model for Multimedia Over 3.5G Wireless Network Downlink" *Proc. European Modelling Symposium*, London, Sept. 2006.
- [10] W. Bang, K. I. Pedersen, T.E. Kolding, P.E. Mogensen, "Performance of VoIP on HSDPA", *IEEE Proc. VTC*, Stockholm, June 2005.
- [11] E. Anderlind, and J. Zander, "A traffic Model for Non-real time Data Users in a Wireless Radio Network", *IEEE Communication. Letters*, 1(2) 1997.