Fast ARQ in High Speed Downlink Packet Access for WCDMA Systems [*]

Mainak Chatterjee, Giridhar D. Mandyam[†] and Sajal K. Das

Center for Research in Wireless Mobility and Networking (CReWMaN)
Department of Computer Science and Engineering
University of Texas at Arlington
Arlington, TX-76019
E-mail:{chat,das}@cse.uta.edu

[†] Nokia Research Center
6000 Connection Drive
Irving, TX-75039
E-mail: giridhar.mandyam@nokia.com

ABSTRACT

In this paper, we study the reliability as offered by the medium access control (MAC) and the radio link control (RLC) for the high speed downlink packet access (HSDPA) in a Wideband-CDMA (WCDMA) system. The Automatic Repeat reQuest (ARQ) mechanism in the RLC usually has considerable amount of delay associated with it which might not be able to sustain real-time communication with strict delay requirements. However, if retransmissions are done at a lower layer, such as the MAC, the response time is quicker which enhances the performance of the system. We perform simulation experiments with synthetically generated HTTP traffic and show how the delay and throughput vary as observed from the RLC, for finite number of retransmissions at the MAC layer. We observe that there is a substantial gain in the performance with the incorporation of fast *retransmission at the MAC layer, which incurs a delay of only 2 ms (equivalent to 3 slots in a WCDMA frame).*

1 INTRODUCTION

Transport control protocol (TCP) is still the major suite for the Internet Protocol (IP) and provides reliable end-to-end transmission [11] in the wireline domain. With the proliferation of the World Wide Web (WWW) in our daily life, a number of services also need to be supported in the wireless domain. Typical examples of such services include speech, audio, video streaming, file and web downloading. Although, several advancement in wireless communications made this possible, most of these wireless data technology still depend on the IP-based network to leverage the already existing and most dominant IP-based infrastructure.

To support WWW traffic to the mobile devices, it is important that a suitable protocol or standard be chosen to cater to the growing demands of data services over wireless channels. One such proposition is Wideband CDMA (WCDMA), which is specified by the 3rd Generation Partnership Project (3GPP) as the rapidly emerging global 3G radio access technology. Current WCDMA specifications support data rates up to 2 Mbps in indoor/small-cell-outdoor and up to 384 Kbps with wide-area coverage, which is in full agreement with the IMT-2000 requirements. Both high-rate packet data and high-rate circuit-switched data are currently being supported. The frame structure for the WCDMA downlink is shown in Figure 1. Each 10 ms frame is divided into 15 slots, each of which is 2/3 ms, containing both data and control information. Data is carried over the dedicated physical data channel (DPDCH) while control information is carried over dedicated physical control channel (DPCCH). The control channel comprises the transport power control (TPC), the transport format configuration indicator (TFCI) and the pilot channel.

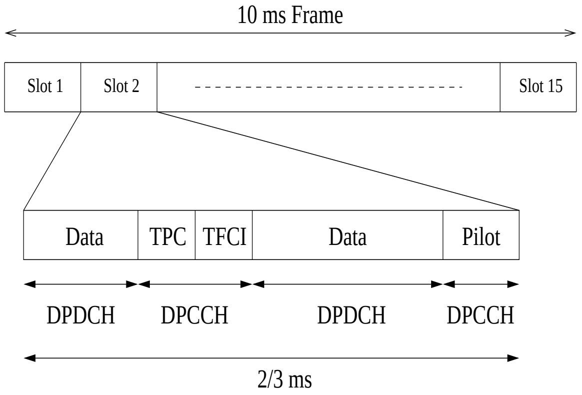

Figure 1: Frame structure for WCDMA

A big challenge in 3G systems is how to handle a wide variety of multimedia services with different quality of service (QoS) requirements. It will be beneficial to improve the WCDMA air interface beyond the third gen-

---

[*] This work is partially supported in part by Texas Telecommunication Engineering Consortium (TxTEC) and Nokia Research Center, Irving, Texas.

eration (3G) requirements because of the increasing demands on packet data services. One of the major propositions to satisfy these demands in WCDMA systems is the evolution of *high-speed downlink packet access* (HSDPA). The main goal of HSDPA is to allow instantaneous bit rates upto 10 Mbps for best-effort packet data services with certain bounds on delay and capacity. The HSDPA channel is a resource that is shared among several users in the mobile communication system. By using a fast scheduler located at the base station, the HSDPA channel can be assigned to the user with the currently best channel, i.e., the user that can transmit with the highest data rate. The idea behind this is that all users shall only use the HSDPA channel when their own downlink is good, and let other users utilize the channel when it is anyway bad.

Some of the basic principles used in HSDPA are fast link adaptation, fast scheduling and fast retransmissions of erroneously received packets. In packet data services, when the receiver detects a packet in error, it requests for a retransmission. However, the receiver does not discard the received soft information associated with the incorrectly-received packet. Rather, it buffers the data and coherently combines the buffered data with the received soft information of the retransmitted bad packet [10]. This type of packet combining mechanism provides increased reliability in CDMA systems. The error recovery mechanism is initiated by the *radio link control* (RLC) which triggers retransmissions to salvage the damaged packets. The delay associated with retransmissions at the RLC might not be small enough to sustain a real-time application. However, if retransmissions at a lower layer (for example MAC) is supported, then the delay would be appreciably smaller.

Our main motivation behind our work is to study the reliability offered jointly by the RLC and the MAC layers of WCDMA systems in supporting HTTP traffic. We model the wireless channel as ITU-A channels and consider soft packet combining at the receiver which effectively lowers the frame error rate (FER). We take into account the possibility of misinterpretations of acknowledgments (though a small fraction) at the MAC layer due to fast decoding. Simulation experiments are also conducted to obtain the delay and throughput as experienced by the MAC and the RLC, for finite number of retransmissions. It is observed that there is tremendous gain if the MAC has at least one retransmission, however, the gain does not increase substantially if the number of retransmissions is more than two. The recovery rate of missing frames at the RLC is prominent at higher FER and also when the misinterpretations of acknowledgements are high.

The rest of the paper is organized as follows. Section 2 discusses the architecture and protocol stack for WCDMA systems. Section 3 deals with the retransmissions at the RLC and MAC layers. The simulation model and experimental results are presented in Section 4. Section 5 shows how the results are affected when some of the ACKs are misinterpreted. Conclusions are drawn in the last section.

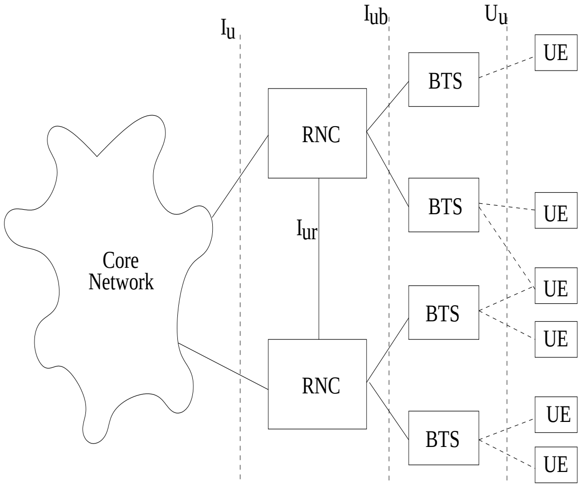

Figure 2: UTRAN Architecture

2 WCDMA Architectural Support

The basic architecture proposed for UTRAN (Universal Terrestrial Radio Access Network) [7] is shown in Figure 2. A number of RNCs (Radio Network Controllers) are connected to the core network. The RNCs are connected among themselves via the I_{ur} interface. Each RNC supports multiple base stations or base transmitter systems (BTS). The user equipment (UE) communicates directly with the BTS through the U_u interface.

The radio interface protocol stack for WCDMA, shown in Figure 3, primarily consists of three layers. Layer 1 is the physical layer (PHY). Layer 2 has four sublayers: the medium access control (MAC), the radio link control (RLC), the broadcast/multicast control and the packet data convergence protocol. Layer 3 contains the radio resource control (RRC) which is mainly responsible for the radio resource allocation to the UEs. It also does all the control plane and user plane signalling between the UTRAN and the UEs, and tries to deliver the negotiated quality of service (QoS) to the UEs. In order for the RRC to do so, it sends control signals to all the sublayers in Layer 2 through the service access points (SAPs) indicated by ovals. Any application can also provide data directly into the RLC layer through the appropriate SAP.

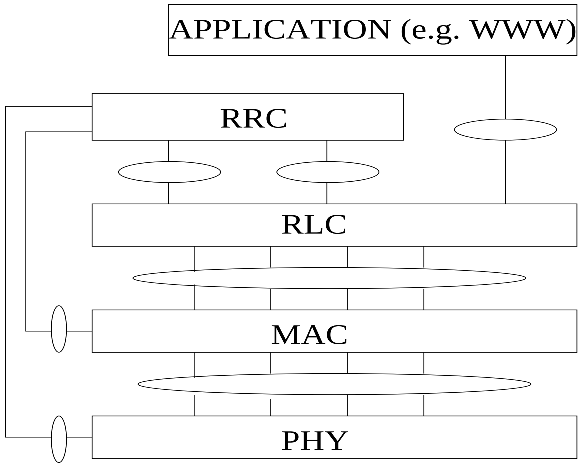

Figure 3: Radio Interface Protocol Stack

The RLC layer provides transparent, unacknowledged or acknowledged mode of data transfer to the upper layers. The main functions of RLC are ciphering of data and *Automatic Repeat reQuest* (ARQ). The MAC layer maps the logical channels of the RLC to the transport channels at the physical layer. The MAC upon receiving information from the RRC about the quality of service to be supported, does the resource allocation and the priority handling of different data flows. Since, this paper mainly deals with the retransmissions at the MAC and the RLC, we discuss them in the next section.

3 WHY RETRANSMISSIONS?

TCP is the most widely used protocol suite for IP which provides reliable end-to-end transmission [11]. The design of TCP has been done in such a way that it performs well in wireline networks where the channel error rates are extremely low and whatever congestion occurs is due only to loss of packets. However, when TCP is used in the wireless domain which is characterized by high bit error rate, the performance of TCP severely degrades. Any packet loss at the wireless link is interpreted as congestion by TCP; and it responds to it by reducing the transmission window size, initiating the congestion control mechanism and resetting the retransmission time [6]. The congestion control mechanism built for wireline networks causes an unnecessary reduction in the TCP throughput. Several schemes have been proposed to alleviate the effects of non-congestion related losses over wireless links [1, 3, 8]. *Retransmissions* of damaged packets is one such scheme which recovers erroneous frames before the TCP timer expires and leaves the TCP throughput unaffected.

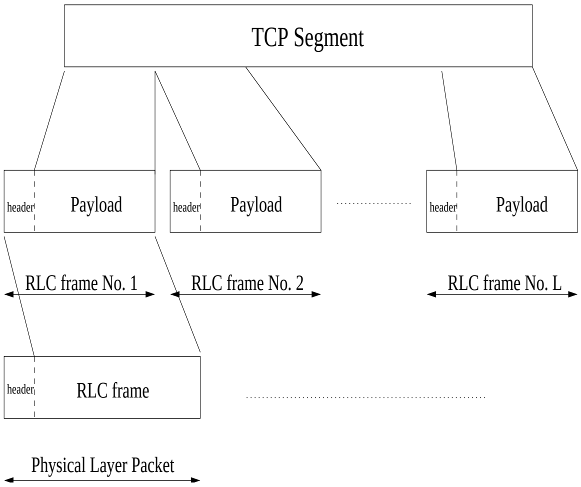

Figure 4: Fragmentation of TCP segments

3.1 Retransmissions at RLC Layer

The objective of the radio link control is to shield the effect of the loss over wireless links from the TCP layer [2, 5]. The RLC segments an upper layer packet (an IP packet or a TCP segment) into several RLC frames before transmitting over the wireless channel. The fragmentation as shown in Figure 4 is done to increase the granularity of the transmission, i.e, in case of any error, an RLC frame which is of a smaller size is affected rather than the whole TCP segment. In case of an RLC frame loss during transmission, the RLC uses an ARQ error recovery mechanism to retrieve the lost RLC frame. The process for recovery of erroneous frames is initiated by the receiver by requesting retransmission of the missing or damaged frames. The recovery of erroneous frames should be done before the TCP timer expires for the TCP throughput to remain unaffected.

3.2 Retransmissions at MAC Layer

Radio link control protocols are usually sufficient to shield the physical layer impairment from the TCP, but might fail to do so if the application has very strict delay requirements. The delay associated with retransmissions at the RLC might not be small enough to sustain a real-time application. Similar problems will arise if we deal with *interactive* real-time traffic. In such scenarios, the TCP timer might time-out before the RLC recovers a missing frame. To avoid the delay associated with retransmissions at the RLC, a lower layer fast MAC retransmissions can be used. The faster retransmissions can provide a better round trip time for real-time applications. Since the number of transmissions allowed at the MAC layer is finite, it does not completely eliminate the possibility of having missing or damaged frames. If the fast ARQ fails to deliver a frame correctly even after retransmitting the maximum allowed number of times, then the responsibility is passed on to the RLC to retrieve the frame. Thus we attain two layers of retransmission reliability as shown in Figure 5. Let us discuss the principal mechanism which leads to delay reduction by MAC layer retransmission.

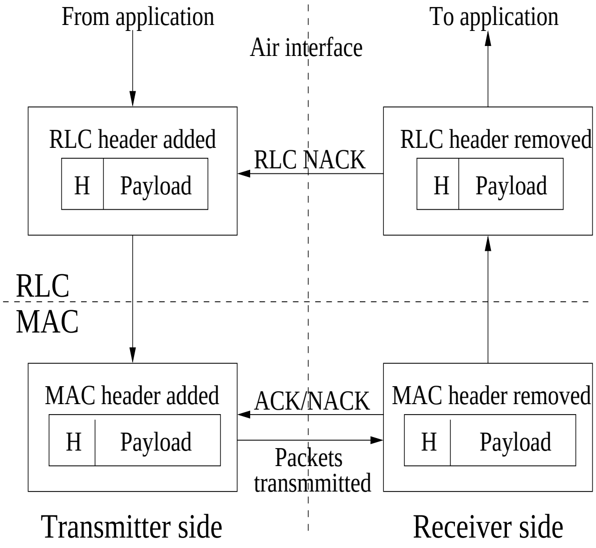

Figure 5: Two layer reliability

A larger delay at the RLC is due to the fact that the RLC detects a bad RLC frame when it detects a “hole" (i.e. a missing number or a sequence of numbers). This could take several frames if, for instance, the mobile is in a deep fade for a long time. Only after detecting a hole can an RLC NACK be sent by the receiver. The RLC NACK is often routed to a centralized part of the network (accounting for the 80 - 100 ms round-trip time). The reason RLC oftentimes is terminated in a central part of the network is due primarily to the possibility of handovers. To make RLC terminate in the base station would result in the entire RLC instance being moved from one base station to another when the mobile executes a fast cell selection. This would mean all the RLC frames which the base station has buffered in anticipation of NACKs, which can possibly come at any time, would be wasted. This implies a large throughput requirement on base station-to-base station communication during handovers.

The HSDPA fast ARQ mechanism delivers an ACK/NACK synchronously with the ARQ phases received. Therefore, if a NACK has to be sent, the mobile does not wait for a hole but sends the NACK within a fixed time offset from receipt of the frame in error. The base station receiver decodes the NACK and sends the retransmission also in HSDPA; it does not route the NACK to some central entity such as a BSC. This is possible since for HSDPA, the mobile station receives frames from only one base station at a time. When the mobile station moves from one base station to another, it simply restarts the fast ARQ mechanism with the new base station.

4 SIMULATION MODEL

We perform simulation experiments to obtain the average delay and throughput. In our simulation model, we assume that the RLC payload is of constant size and each RLC frame maps to one physical layer frame. We choose HTTP as the application to be supported because supporting real-time WWW traffic on the mobile terminals is a major challenge. Let us briefly discuss the assumptions on soft-information combining and the model for HTTP traffic.

4.1 Packet Combining

In order to ensure that the receiver does not try to combine packets from one ARQ phase with another, it is assumed that outband signaling is sent on a forward control channel concurrently with each frame the receiver receives on the forward shared channel called HS-DSCH, or the high-speed downlink shared channel. It can be noted that the FER experienced by a packet during its first transmission is not the same as that during the second transmission. It is less in the latter case because there is at least a 3-db gain in the bit energy-to-noise ratio (E_b/N_o) due to packet combining [10].

4.2 Effective FER

Whenever a packet is not correctly decoded at the receiver, the packet is not discarded but stored in the buffer and is used to re-combine packets resulting in a gain in the E_b/N_o value. If the packet is not correctly decoded even after the packet combining, it is retransmitted for the second time, if allowed. With the second retransmitted packet the gain in E_b/N_o would be 6-db. That is, with every retransmission there is a gain of 3-db. If we consider indoor A channel with $\frac{3}{4}$ rate QPSK modulation, then according to simulations the effective FER for successive retransmission (RET) would be as shown in Table 1. These values were obtained through simulation of the said channel. For example, if the FER is 0.1 for the initial transmission (Tx), then on successive retransmissions the effective FER would be decreasing as 0.045, 0.021, 0.011 and so on. Although the decrease in the effective FER could not be generalized, it works well with ITU pedestrian A and ITU vehicular A channels.

Table 1: Effective FER

Initial Tx	First RET	Second RET	Third RET
0.05	0.023	0.012	0.005
0.10	0.045	0.021	0.011
0.15	0.072	0.032	0.017
0.20	0.094	0.042	0.0205
0.25	0.128	0.057	0.027
0.30	0.140	0.071	0.031

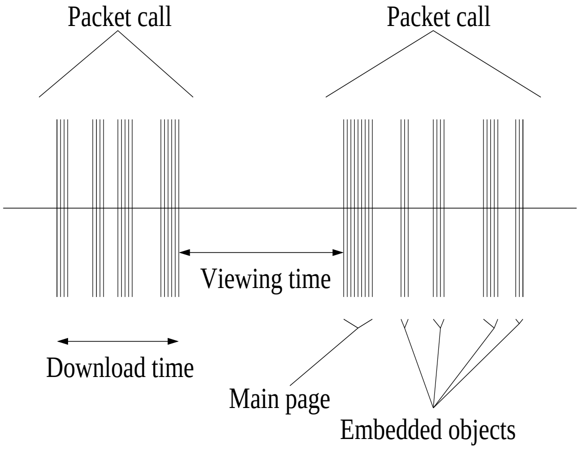

Figure 6: Web page traffic

4.3 HTTP Model

Instead of trying to investigate the nature of HTTP traffic, we synthetically generate HTTP traffic by using the results obtained in [4]. The basic model of HTTP is shown in Figure 6. A packet call represents the download of a web page requested by a user. It usually has a main page followed by some embedded objects. A new request (packet call) is immediately generated after the expiration of the viewing period. The model is similar to an ON/OFF source where the ON state represents the activity of a page request and the OFF state represents a silent period after all objects in that page are retrieved. The download time of a page follows Weibull distribution, the mean of which depends on the underlying bandwidth of the wireless channel. We have considered a data rate of 76.8 Kbps (9600 bytes/sec). Other statistics and parameters used to generate the HTTP traffic are shown in Table 2.

Table 2: Statistics for HTTP Traffic

Component	Distribution	Mean
Main page size	Lognormal	10710 bytes
Embedded object size	Lognormal	7758 bytes
No. of embedded objs	Pareto	5.55
Viewing time	Weibull	40 ms

Table 3: Simulation Parameters

Number of ARQ phases (N)	6
Max. number of retrans at MAC (M)	1, 2 and 3
FER (p)	0 - 30%
Frame duration (T)	10 ms
RLC payload	400 bytes

4.4 Fast ARQ

Each object (main page and embedded objects) is fragmented into multiple equal-sized RLC packets. The transmitter transmits one RLC packet in each 2 ms (3 slots) physical layer frame and waits for the ACK. If the ACK does not arrive in 3 slot timings, the frame is retransmitted immediately. The ACK timer for that packet is again reinitialized. If we consider the downlink, then it is not necessary that the mobile station will deliver the ACK/NACK precisely at the slot boundaries. The actual physical layer boundary is more precise. This is due to the fact that normally when using coherent receivers in the reverse link, the base-station suffers from some processing delay. The number of retransmission trials allowed is varied between 1 and 3. If a packet is not successfully received or combined at the receiver even after the maximum number of MAC retransmissions, then the RLC retransmission is triggered. We assume that the MAC frame is of equal size as that of an RLC frame. The information on ARQ instance is sent through outband signaling and therefore does not affect the throughput. Also, the WCDMA MAC header is normally less than an octet, thus adding minimal overhead. The other parameters for the simulation are given in Table 3.

4.5 Simulation Results

Figure 7 shows the delay when the maximum number (M) of allowed retransmissions at the MAC is varied. The delay is tremendously high when only the RLC retransmissions (i.e., M = 0) is considered. With just one retransmission (M = 1) at the MAC, there is a tremendous improvement in the delay. However, there is not much difference between M = 2 and M = 3. This is due to the fact that after two transmissions, most of the packets are recovered and the third retransmission is hardly required. Figure 8 shows the throughput of the system. By *throughput* we mean the combined throughput due to MAC and RLC. It can be seen that the throughput decreases as the FER increases. This is obvious because a larger FER damages more packets during transmission and thus the number of retransmitted packets becomes large. In Figure 9 we observe the efficiency of the RLC recovering the missing packets. The y-axis indicates the fraction of the packets recovered by RLC. If the fast ARQ mechanism at the MAC is turned off, which effectively means M = 0,

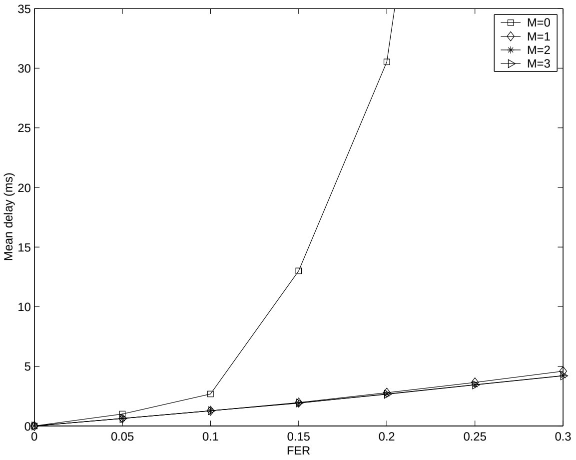

Figure 7: Delay

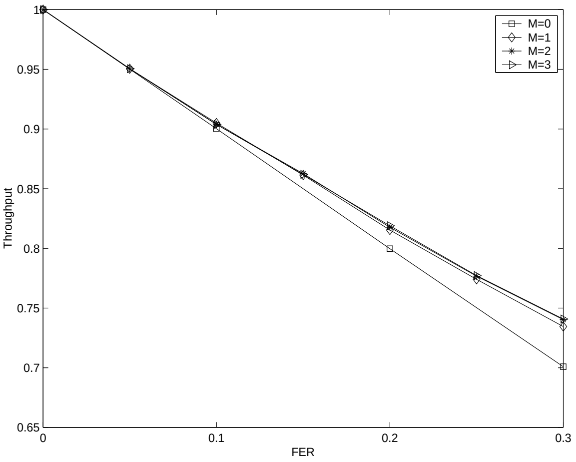

Figure 8: Throughput

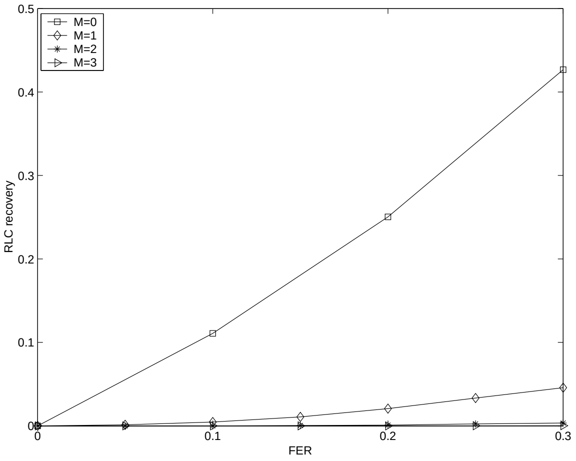

Figure 9: RLC Recovery

all the missing packets are recovered by RLC. With M = 1, there is a considerable drop in the RLC recovery because the MAC does most of the recovery with just one retransmission. With M = 2 or 3, the RLC recovery is almost zero due to the fact that virtually all the packets are recovered with three retransmissions.

5 MISINTERPRETATION OF ACK/NACK

There is a small probability that some of the ACKs and NACKs from the fast ARQ will be misinterpreted. The decoder at the transmitter might interpret a NACK as an ACK, and hence would not transmit the packet assuming correct reception. The RLC at the receiver will detect the missing packet and trigger its own retransmission mechanism to recover the packet. On the other hand, if an ACK is decoded as a NACK, a retransmission will be triggered by the MAC if the number of retransmissions has not reached the maximum limit. If the number of allowed retransmissions at the MAC is exhausted, then the RLC will recover the packet. No matter which layer does the recovery, it results in duplicate retransmission of a previously (correctly) received packet. It can be noted that, any kind of wrong interpretation at the MAC can only be detected by the RLC and thus the reliability of the RLC cannot be ignored. This is where the real benefit and importance of RLC is manifested. The percentage of error recovery by the RLC depends on the percentage of the misinterpretations, which we call the “falseACK".

5.1 Results with $falseACK$

We conduct our simulations incorporating a certain percentage of misinterpretations of ACKs. The misinterpretations percentage is varied from $f = 1\%$ to $f = 5\%$. Figure 10 shows the mean delay for M = 2. It can be noted that the delay is more compared to Figure 7 where f = 0. Figure 11 shows the degradation in the system performance in terms of throughput for M = 2. The percentage of RLC recovery for M = 2 is shown in Figure 12. With M = 1, the RLC recovers more packets than with M = 2. This happens because with two retransmissions at the MAC layer, almost all the packets are recovered. The RLC mostly recovers the ones which were misinterpreted. It can be expected that the RLC recovery will be more for even greater FER values. But we did not consider more than 30% FER because at that high error rates, there will be other severe consequences with maintaining the link.

6 CONCLUSIONS

As the demand for real-time wireless data communication increases, more efficient and fast protocols are being designed. High speed downlink packet access (HSDPA) is one of the major propositions for WCDMA systems which can allow bit-rates upto 10 Mbps. It is also important to recover the lost or damaged packets through retransmissions at lower layers than TCP. In this paper, we showed how the combination of the RLC and MAC retransmissions can enhance the performance of WCDMA systems. The fast retransmissions at the MAC sublayer quickly recovers most of the damaged packets before the RLC can trigger its retransmission scheme. Due to soft

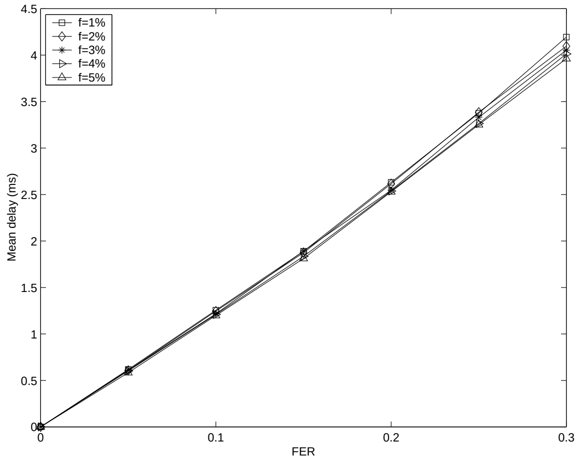

Figure 10: Delay

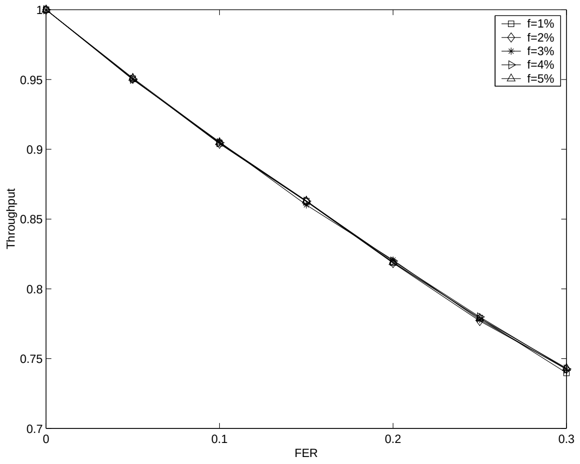

Figure 11: Throughput

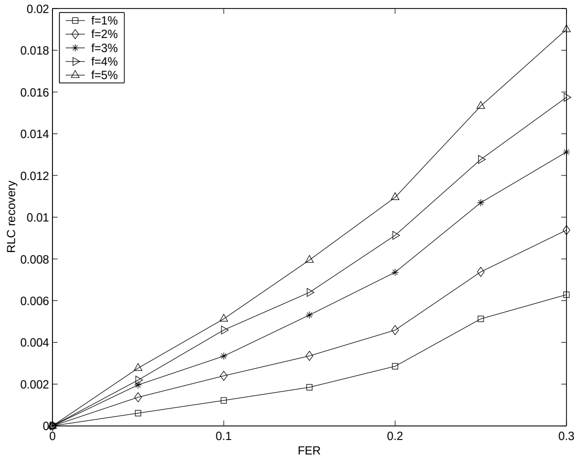

Figure 12: RLC Recovery

packet combining, the effective FER experienced by a packet undergoing retransmission is lowered for every successive retransmissions. From our simulation results, we observe that the fast retransmission mechanism at the MAC layer enhances the overall system performance. However, there is not much improvement if the maximum number of retransmissions at the MAC is greater than two. The recovery by the RLC becomes crucial if the FER is high and also where the ACKs are misinterpreted at the MAC sublayer.

REFERENCES

[1] A. Bakre and B.R. Badrinath, “I-TCP: Indirect TCP for Mobile Hosts", *Proc. 15th Internationals Conference on Distributed Computing Systems*, May 1995, pp. 136-143.

[2] G. Bao, “Performance evaluation of TCP/RLC protocol stack over CDMA wireless links", *ACM Wireless Networks Journal*, Vol. 2, 1996, pp. 229-237.

[3] R. Caceres and L. Iftode, “Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments", *IEEE Journal on Selected Areas in Communications*, Jun 1995, Vol. 13, pp. 850-857.

[4] H.K. Choi and J.O. Limb, “A Behavioral Model of Web Traffic", *International Conference of Network Protocols* (ICNP) 1999, pp. 327-334.

[5] A. Chockalingam and G. Bao, “Performance of TCP/RLC protocol stack on correlated fading DS-CDMA wireless links", *IEEE Transactions on Vehicular Technology*, Volume 49, Issue 1, Jan 2000, pp. 28-33.

[6] D.E. Comer, *Internetworking with TCP/IP*, Vol. 1, Prentice-Hall, 1991.

[7] H. Kaaranen, A. Ahtiainen, L. Laitinen, S.K. Naghian, V. Niemi, “UMTS Networks: Architecture, Mobility and Services”, Wiley & Sons Publishing, 2001.

[8] TIA/EIA/IS-707-A.2, “Data Service Options for Spread Spectrum Systems: Radio Link Protocol", March 1999.

[9] S. Parkvall, E. Dahlman, P. Frenger, P. Beming and M. Persson, “The evolution of WCDMA Towards Higher Speed Downlink Packet Data Access", in Proceeding of IEEE Vehicular Technology Conference (VTC) 2001.

[10] S. Souissi and S.B. Wicker, “A Diversity Combining DS/CDMA System with Convolutional Encoding and Viterbi Decoding", *IEEE Transactions on Vehicular Technology*, Vol. 44, No. 2, May 1995, pp. 304-312.

[11] W.R. Stevens, “TCP/IP Illustrated", Addison-Wesley, 1994.