Performance Analysis of General Packet Radio Service Protocol Stack

A THESIS

SUBMITTED FOR THE DEGREE OF

Master of Science (Engineering)

IN THE FACULTY OF ENGINEERING

by

K Premkumar

Department of Electrical Communication Engineering

INDIAN INSTITUTE OF SCIENCE

BANGALORE - 560012

August, 2001

My beloved Parents

Abstract

Second generation cellular systems, including Global System for Mobile communications (GSM), provide primarily voice communications to mobile users. Recently, there has been a growing interest in packet data communications over wireless to provide fast and efficient Internet access to mobile users. Mobile radio channels exhibit poor bit error rate performance due to channel imperfections including time-varying distance and shadow losses, and multipath fading. Commonly used higher layer protocols (e.g., transport layer protocols like TCP) have been designed to perform well in wireline networks where the channel error rates are very low. Thus, the lower layers in wireless protocol stacks must be designed to address such higher layer performance concerns. In particular, efficient and robust data link layer (LL) and media access control (MAC) layer protocols are crucial to system performance in wireless data communications. This thesis addresses the performance analysis of various layers in a wireless protocol stack and suggests techniques to enhance performance.

We consider the protocol stack proposed in the General Packet Radio Service (GPRS) system. GPRS is a wireless system that provides packet data communications to mobile users using GSM cellular infrastructure. The radio link control/medium access control (RLC/MAC) layers in the GPRS protocol stack essentially are responsible for the way in which the GSM/GPRS radio resources (frequency-time slot pairs) are shared by various mobile users. The uplink (mobile-to-base station link) channel resources are shared based on a *request-reservation* mechanism. Performance of the GPRS RLC/MAC layers under various radio channel and traffic load conditions influence the overall GPRS network performance. Several studies have investigated the performance of the GPRS RLC/MAC layers, but mainly through simulations.

A new contribution in this thesis is the modeling and *analytical* evaluation of the performance of the RLC/MAC protocols in GPRS, considering the uplink *request-reservation mechanism*. Using theory of Markov chains, we derive expressions for the uplink throughput and delay performance of the GPRS MAC protocol. Our analysis quantifies the throughput-delay performance as a function of system traffic load, number of random access channels (to carry resource requests from mobiles), and number of traffic channels (to carry data traffic) on the uplink. The results can be of use in choosing the optimal mix of number of random access channels and traffic channels for a given traffic load. Analytical performance results are found to closely match with the simulation results.

Another key contribution in this thesis is the proposal and the performance analysis of *slot level retransmission* at the RLC layer to enhance performance. A packet that is scheduled to be transmitted over the GPRS air interface is formatted into one or more logical link control (LLC) frames, which are then segmented into several RLC blocks. The radio resource allocation in GPRS is performed in units of RLC blocks. Each RLC block occupies four physical layer time slots. A selective repeat (SR) automatic repeat request (ARQ) mechanism is provided at the RLC layer to recover RLC block errors using a block check sequence (BCS) and block level retransmission. The drawback with block level retransmission is that even if error occurs in only one slot, the entire block of four slots needs to be retransmitted, which degrades performance. In this thesis, through analysis and simulations, we show that slot level retransmission significantly improves the RLC layer performance, particularly when the channel error rates are high. In addition, slot level retransmission enables resource allocation at the individual time-slot level to achieve a finer resource allocation granularity and greater flexibility.

Unrecovered block errors at the RLC layer are handled by another ARQ mechanism at the LLC layer which performs retransmission of ‘erroneous LLC frames’. A question in this regard is how many retransmission attempts (in the event of a block/frame errors) can be allowed at the RLC layer and the LLC layer. Our performance results show that it is more beneficial to keep a larger number of retransmission attempts at the RLC layer than at the LLC layer.

We further study the performance of transport layer protocols over GPRS. Transmission control protocol (TCP) is an end-to-end transport layer protocol that provides reliable, in-sequence

delivery of packets. We evaluate the performance of TCP on the uplink in GPRS through simulations. Both block level as well as slot level retransmission at RLC are considered. An ON-OFF traffic model, such as the web and e-mail traffic, is considered. The OFF period distribution is modeled to follow a Pareto distribution. It is shown that TCP performance improves with slot level retransmission as compared to block level retransmission. Effect of various TCP parameters like fast retransmit threshold on the throughput performance are evaluated.

Acknowledgments

I wish to thank my advisor *Dr. A. Chockalingam* for initiating me into this project and for being a source of constant inspiration. I am grateful to all my instructors for taking great interest in my progress. In particular, my sincere thanks go to *Profs. Anurag Kumar, Kumar N. Sivarajan, Utpal Mukherji, B. Sundar Rajan,* and *A. Makur*. My special thanks go to Profs *Vinod Sharma* and *Narahari* for their encouragement in my general test. I thank *Prof. A. Selvarajan* and *Prof. G. V. Anand* and all other faculty in the department for providing a stimulating atmosphere for research.

I would like to thank *Department of Science and Technology, New Delhi*, for sponsoring this research project through scheme Ref:DSTO/EEC/ACM/461.

Thanks to the *Department of Human Resources, Government of India* for providing financial assistance throughout my stay in IISc.

I enjoy the company of Soni, Shankar, Prabhu, Kalyan, Ramesh, and Suren in ***W**ireless **R**esearch **L**ab*, for sharing our ideas, jokes and what not ...

My stay in the beautiful IISc campus was a delightful and pleasurable period, thanks to my friends, C. Venki, Ananth, Venki Baabs, Rishi, Mukundh, Chandra, Arun, Satheesh, Moorthy, Vishwanath, R. Venki, Naveen, Maaran, Vijay, Vasu, R. Hariharan, K. Hariharan, Sanal, Guha, Sharan, ... the list is endless. I can never forget the discussions made with CV, Rishi, Venki Baabs,... for the feasibility of some thing which does not seem to be clear even now.

I would like to thank people at Tea Kiosk and A-mess who did their job to my utmost satisfaction.

Musa, Mohan, Ganesh, Aayilyan, Balaji, Raghu, Rama, Badri of *Infosys* are instrumental in making me choosing my path.

Bhaskar, Chinthu, Ramshankar, Balamurali, Ramkumar, Krishnakumar, Manikkam, Lokesh, Mukundhan, Narasimha sundaresan ... the list of close friends in ACCET is endless, but without them life as a whole would not have been half as great or half as much of fun. I would like to mention Vijaya, Radhika, Anitha, Thavamani for being a source of confidence. Ramya deserves special mention for throwing light on (non linear) time varying systems, which I could not understand even now.

My heartfelt thanks go to my parents, *Karumbunathan* and *Rajeswari*, and my brother, *Elangovan*, for their constant love, support, and encouragement, and to my sisters' kids, *Poojha* and *Bharu*, for bringing so much joy into my life.

As a small token of my gratitude, I dedicate this work to my parents.

I would like to recollect the famous lines of *Robert Frost*

And miles to go before I sleep,
And miles to go before I sleep.

Contents

List of Figures

Chapter 1

Introduction

The last decade has witnessed widespread and successful commercial use of wireless mobile communications. Wireless mobile communications has been an extremely rich field for research, due to the many difficulties that the wireless environment presents, and due to the ever increasing users' demand for more, newer, and better services. The past few years have seen tremendous advances in the research, development, and design of mobile radio systems, and many more advances are expected in the near future. The key technical features that make the wireless environment so challenging are *channel unreliability*, *limited bandwidth*, and *support for mobility* [1].

1.1 Wireless Channels

Due to a number of physical factors, signals propagating through wireless mobile channels are subject to severe impairments. The effect of propagation on the transmitted signal can be adequately described by three phenomena, namely, *distance loss*, *shadowing*, and *multipath fading* [2].

Distance loss accounts for the signal attenuation due to the physical distance, d, between the transmitter and the receiver. In free space, the received power is proportional to d^{-2}, whereas in ground wave propagation there is an additional effect due to the combination of the direct (free-space) propagated wave and its replica as reflected by the earth surface. In this case, the distance loss behavior is better described by the function $d^{-\nu}$, where ν depends on a number of properties of the environment, and usually takes values between 2 and 5. The shadow loss, on the other hand, accounts for obstructions along the propagation path (e.g., buildings or trees), which may cause significant signal attenuation. Shadow loss is often modeled as having a Gaussian distribution when expressed in dB. This model (also called log-normal shadowing), has been found to match field measurements very well. Both distance loss and shadow loss do not vary much over distances on the order of many wavelengths of the carrier. The increased distance loss and shadow loss in a mobile radio environment calls for careful link design with adequate mobile transmit power and receiver sensitivity.

A more serious channel impairment is the loss due to multipath fading. Due to various objects and reflectors in the propagation environment, the transmitted signal is scattered in many directions so that multiple versions of it, coming from different angles and with different path delays and attenuations, reach the receiver. The different delays encountered by the multiple versions of the signal result as phase differences which add either constructively or destructively at the receiver, leading to deep fades (signal losses) when the phases add destructively. These fades can be as high as 60 dB of loss in mobile radio environments. Since the mobile terminal can move, the received signal envelope will vary with respect to time (even if the terminal does not move the environment around them may change over time resulting in a time-varying signal envelope). The received signal variation due to multipath fading varies rapidly depending on the mobile speed, and typically follows a Rayleigh distribution [2]. Typical techniques which are often used to counteract the effect of multipath fading include equalization, diversity and power control.

All the above physical layer impairments make wireless mobile communication very different from communications over more stable and predictable channels such as cables and fibres. *Specifically, these physical layer impairments contribute to high bit error rates on the wireless channel, which, in turn, affect the design and performance of wireless protocol stacks.*

1.2 Limited Spectrum and Cellular Concept

Another important difference between wireline and wireless systems is the amount of bandwidth available. Today's wireless mobile systems mostly operate at UHF frequencies (typical bands are

900 MHz and 1.8 GHz). Additional frequency bands are being considered for future systems in the 1–6 GHz range. Spectrum at 17 GHz has been designated for wireless LANs in Europe, and there are proposals for systems at frequencies up to 60 GHz. The amount of bandwidth available in these ranges is not large. This is because the radio spectrum is an inherently public resource, and is already crowded due to the presence of other services in these bands (e.g., broadcast TV, military communications, and point-to-point radio links). Therefore, wireless systems are limited by interference, which dictates the amount of bandwidth available and imposes restrictions on how it can be used[1]. The consequences of limited spectrum are the constraints on the system capacity (the number of users the system can support for a given bandwidth) as well as the mobile user's capabilities (the bandwidth which may be instantaneously available to a user). Hence, bandwidth efficient techniques to increase system capacity and user capabilities are important in wireless mobile systems.

Cellular concept is a widely adopted and successful means to increase capacity in wireless mobile systems. Invented by the Bell Laboratories in the 1970s, cellular systems are based on the same concept used for many years in broadcasting, where the same channel frequency can be used in non-interfering geographic areas (called as cells). The available system bandwidth is split into a number of channels, and these channels are assigned to cells in such a way that cells using the same channels are separated geographically by a sufficiently large distance so that certain co-channel interference constraints are met. This allows a capacity increase as the cell size is decreased. How the system bandwidth is sub-divided into channels depends on the multiple access technology of choice. Frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA) are widely used multiple access methods in cellular systems [3].

A key issue in cellular systems, however, is the need to support user mobility. Two basic problems need to be solved to guarantee effective mobility management, namely, user location identification and call handoff. The location problem arises because the system needs to know

[1] Although wireline networks too face interference issues (e.g., crosstalk), the effects are less significant in limiting the capacity compared to wireless systems.

where a user is at any time, in order to be able to correctly forward incoming calls. In cellular systems, the network is divided into service areas, each composed of a number of cells managed by a central unit, called Mobile Switching Center (MSC). Each MSC has two databases, called Home Location Register (HLR) and Visitor Location Register (VLR). The HLR contains information regarding users who are registered as subscribers in the area, whereas the VLR contains information regarding users who are registered as subscribers somewhere else but happen to be roaming in the area. Every time a user registers as a roamer in a new area, signaling takes place between the MSCs involved, and the databases are appropriately updated. The information contained in those registers allows the network to locate the user, i.e., to identify the set of cells where the user can be found at any given time. Whenever the network needs to locate a user, it looks up the HLR at the user's home MSC. If the user is not in that area, the HLR contains information on where the user can be found. If an incoming call for that user arrives, it may be redirected to the MSC of the area in which the user can be found. The MSC will then instruct the central transmitters, called Base Stations (BS), serving all cells in that area to send a paging message for the user, who will reply to the BS of the appropriate cell. This response completes the location procedure, and the call setup can be completed at the designated BS.

The other main issue is how to handle ongoing calls when, because of mobility and/or changing channel conditions, the user needs to change its point of attachment to the network. This may occur when a user moves away from the BS at which the call was established and moves towards another BS. At some point, the attenuation experienced on the path towards the first BS may become too severe, to the extent that the transmission quality is no longer adequate. A system with no support for mobility would drop the call at that point. In cellular systems, however, this situation would trigger a handoff procedure, whereby the mobile terminal and/or the serving BS look for an alternative BS which can guarantee adequate transmission quality. If no such alternative can be found, the call is dropped; otherwise a handoff occurs, i.e., the connection to the old BS is released and a connection with the new BS is established. In general, handoff procedures involve periodic measurements of the channel quality (typically, the signal strength or the signal-to-interference ratio) on the currently used mobile-to-BS link and on the paths to a number of other candidate BSs. Whenever the need arises, these measurements make it possible to determine which BS should be chosen as the new BS to which call must be handed over. The decision about when a handoff is performed and which alternate BS is selected can be made either by the network alone, or with the cooperation of the mobile terminal (known as Mobile Assisted Handoff – MAHO).

1.3 Cellular Systems Evolution

The main components of a typical cellular system is shown in Fig. 1.1. They include

1. mobile terminal, which contains a radio transceiver and a processor;
2. base station subsystem (BSS), which manages communications to and from the users belonging to its own cells, and is composed of a radio part known as the base station (BS) and a base station controller (BSC);
3. mobile switching center (MSC), which controls a number of BS (typically a few tens), provides additional functionalities such as call management and diagnostics and fault recovery, and is the interface to and from the public switched telephone network (PSTN);
4. databases, such as the already mentioned HLR and VLR, plus the equipment identity register (EIR) and authentication center (AuC), used for security and authentication purposes.

In addition to traffic channels to carry voice/data information, the following control channels are present in cellular systems:

1. *broadcast channels* (BS-to-mobile), where general control information is transmitted to all mobiles;
2. *paging channels* (BS-to-mobile), used to notify a mobile user of an incoming call;
3. *random access channels* (mobile-to-base), used for sending mobile-initiated call/resource requests.

These elements provide essential communication services and network management functionalities, which enable proper system operation.

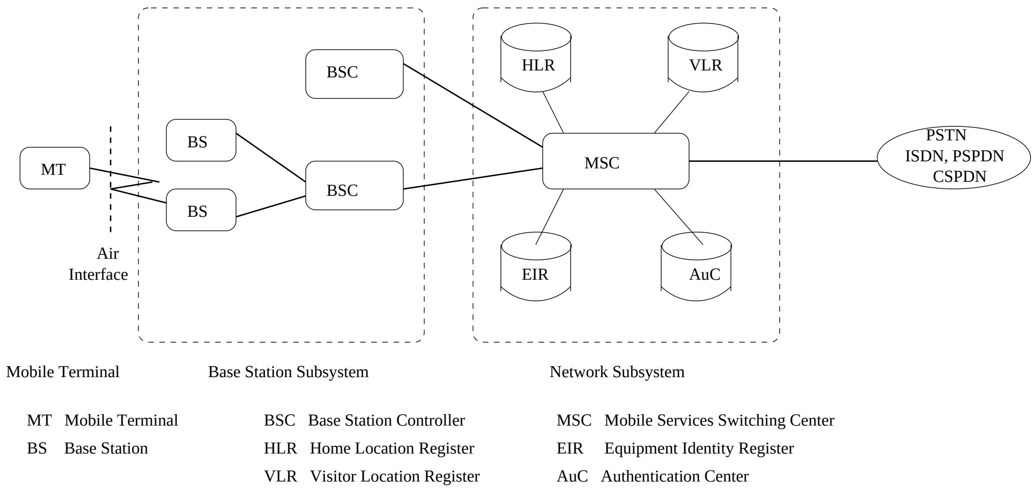


Figure 1.1: Cellular system components

The early cellular (first generation–1G) systems were based on analog FM, primarily for providing voice communications to mobile users. Examples of such systems include the Advanced Mobile Phone Service (AMPS) in the United States and the Total Access Communications System (TACS) in Europe. Since the modulation format is analog, FDMA is the multiple access method used.

The second generation (2G) of cellular systems were based on digital modulation formats, which offered considerable flexibility in handling the information signals, allowing the use of error control coding, source coding and compression, and multiple access methods like TDMA and CDMA. Most importantly, the digital transmission format is more suitable for data communications, which is expected to be a major player in future wireless mobile systems. There are basically two different types of second generation systems, depending on whether they are TDMA-based or CDMA-based. The TDMA-based systems include Global System for Mobile communications (GSM) in Europe, IS-54/136 (the 'Digital AMPS') in the United States, and PDC in Japan [3]. The IS-95 system, widely deployed in the United States and Asia, is a CDMA-based second generation cellular system [4].

The first and second generation cellular systems, including AMPS, GSM, IS-95, are circuit-switched systems, designed primarily for voice communications. In addition to voice services, circuit-switched data communications at low rates (maximum 14.4 Kbps) are supported by 1G and

2G systems (e.g., Cellular Digital Packet Data (CDPD) on AMPS, Digital Fax and Asynchronous Data transfer on GSM, IS-95). However, owing to the accelerated growth of the Internet and Internet based applications, there has been a growing need to provide faster and efficient Internet access to mobile users, on existing and future wireless mobile systems.

It is noted that data communications on 1G and 2G systems are essentially provided on *circuit-switched* air links. That is, an over-the-air traffic channel is brought up and held through the entire duration of the data call. No changes are needed to the physical layer to provide data services in this case. However, suitable protocol stacks need to be provided at the mobile station, base station and the mobile switching center in order to enable data services. At call origination, the service option parameter can specify a certain data service instead of voice service.

It is well known that the computer generated traffic in packet data networks is typically bursty in nature, for which circuit-switched systems are not efficient. It is also known that for bursty data traffic, packet switched communications can result in a much better utilization of traffic channels. This is because traffic channel resources can be allocated only during active periods in communications and can be released during idle periods, enabling statistical multiplexing of channel resources among multiple mobile users. With the increasing demand on mobile data services at high speeds, the evolution of 2G systems to next generation systems has been focused on supporting 'packet switched operation over-the-air' in order to make more efficient use of the radio resources. For example, General Packet Radio Services (GPRS)[2] and Enhanced Data rates for Global Evolution (EDGE) are systems that upgrade and use GSM infrastructure to provide high-speed (270 Kbps in GPRS and 384 Kbps in EDGE), packet-mode data communications. Third generation (3G) systems are envisaged to to offer greatly enhanced performance and service characteristics including multimedia services (voice, data, high quality image, video, etc) at high data rates (up to 2 Mbps) with different quality of service requirements.

[2] For a brief overview of the GPRS system, refer Chapter 2.

1.4 Protocol Stacks for Wireless

Since wireless channels are characterized by high error rates as mentioned earlier, efficient protocol stacks need to be designed and put in place at the mobile terminal side as well as the cellular infrastructure side, in order to provide reliable packet data communications over-the-air. Particularly, sophisticated error control mechanisms need to be employed to improve the reliability of data transfer over wireless links. Forward error correction (FEC) and automatic repeat request (ARQ) are commonly used error control techniques. In FEC, redundant bits are added to information bits to detect as well as correct channel induced errors. In ARQ, on the other hand, error control is achieved not through error correction but through retransmission of erroneous data packets. While FEC is applied at the physical layer, ARQ can be applied at different layers of the protocol stack. Errors uncorrectable by FEC at the physical layer can be handled by ARQ at higher layers.

Commonly used higher layer protocols (e.g., transport layer protocols like TCP) have been designed to perform well in wireline networks where the channel error rates are very low. TCP is the transport protocol used for popular Internet applications like file transfer (ftp), web browsing (http), and e-mail (SMTP). The window adaptation strategy in TCP is devised to recover from packet losses due to network congestion [5]. The transmission window size is reduced when excessive packet losses occur. Since TCP can not differentiate between packet losses due to congestion and channel imperfections (fading, etc), the window size gets reduced quite often due to channel errors when TCP is used on wireless links. This severely degrades the TCP throughput performance. In order to alleviate this problem, ARQ techniques at the lower layers (e.g., link layer) in wireless becomes essential in order to make the data seen by the TCP less erroneous.

1.4.1 Choice of ARQ Technique

There are several ARQ techniques, including Stop-and-Wait (SW) ARQ, Go-Back-N (GBN) ARQ and Selective Repeat (SR) ARQ.

Stop-and-Wait ARQ: In this scheme, after sending a data frame, the transmitter waits for a positive or negative acknowledgment (ACK/NACK) before it sends the next frame. This strategy, though simple to implement, is not widely used in modern data networks because of its highly inefficient use of the communication links. The channel is left idle most of the time while the transmitter waits for an ACK/NACK.

Go-back-N ARQ: In this scheme, there is a transmission window of size N. The transmitter can send up to N frames without waiting for an ACK from the receiver. Suppose that at time t, the receiver expects frame with sequence number M from the transmitter, meaning that frames with sequence numbers from 0 to $M-1$ are successfully received and that the ACK for $(M-1)^{th}$ frame is sent to the transmitter. The transmitter on receiving the ACK for the $(M-1)^{th}$ frame updates its window to $[M, M+N-1]$. The transmitter can then send frames with sequence numbers in the above range. A timeout timer is started when $(M+N-1)$th frame is transmitted. If a NACK comes before exhausting the window, or if the timeout timer expires, the transmitter starts retransmitting the packets from the lower edge of the window. GBN ARQ performs much better than SW ARQ. However, in GBN, since retransmission starts from the packet at lower edge of the window, retransmission of already correctly sent packets may also occur.

If p is the probability of frame error, then the achievable throughput in GBN, η_{GBN}, is bounded by [6].

$$\eta_{GBN} \leq \tfrac{1-p}{1+p\beta}, \tag{1.1}$$

where β is the expected number of frames in one round-trip delay interval.

Selective Repeat ARQ: This strategy is similar to the Go-back-N ARQ except that only erroneous frames are retransmitted. Hence, the feedback from the receiver must include the sequence numbers of the erroneous frames. SR ARQ typically performs better than GBN ARQ as SR ARQ does not do retransmission of correctly sent frames. The throughput achieved by SR ARQ is bounded by [6]

$$\eta_{SR} \leq 1-p. \tag{1.2}$$

The choice of ARQ depends on various parameters like frame error rate of the channel, round-trip delay between the transmitter and the receiver, available buffers, etc. For example, from Eqns. (1.1) and (1.2) it can be observed that GBN can perform as good as SR in terms of throughput when the frame error rate (p) and the round trip delay (β) are very small. However, SR can outperform GBN when p and β are large. *Since the frame error rate p on wireless links is large, SR is the desired ARQ technique for use in wireless protocol stacks.* Similarly, for large round-trip delay (large β) systems too, like satellite links, SR ARQ is preferred.

1.5 Problem Statement

In this thesis, we are interested in the performance analysis of various layers in a wireless protocol stack. Specifically, we consider the protocol stack proposed in the General Packet Radio Service (GPRS) system which provides packet data communications to mobile users using GSM cellular infrastructure [7]. We are also interested in protocol enhancements that can improve performance in a wireless mobile environment.

One key technical issue that is crucial in designing packet mode wireless systems is the efficiency with which the available radio resources are shared by multiple mobile users. This aspect is largely related to the media access control (MAC) protocols employed [6]. In the GPRS protocol stack, the radio link control/medium access control (RLC/MAC) layers in the GPRS protocol stack essentially are responsible for the way in which the GSM/GPRS radio resources (frequency-time slot pairs) are shared by various mobile users [8],[9]. The uplink (mobile-to-base station link) channel resources are shared based on a *request-reservation* mechanism. Performance of the GPRS RLC/MAC layers under various radio channel and traffic load conditions influence the overall GPRS network performance. Several studies have investigated the performance of the GPRS RLC/MAC layers, but mainly through simulations [19]-[26]. *In this thesis, we propose to carry out the modeling and analytical evaluation of the performance of the RLC/MAC protocols in GPRS, considering the uplink request-reservation mechanism.*

In the GPRS stack, a selective repeat (SR) ARQ mechanism is provided at the RLC layer (above the MAC layer) to recover RLC block errors using a block check sequence (BCS) and block level retransmission. Each RLC block occupies four time slots. The drawback with block level retransmission is that even if error occurs in only one slot, the entire block of four slots needs to be retransmitted, which degrades performance. *We propose to analyze the performance of slot level retransmission at the RLC layer and compare it with that of block level retransmission.*

The GPRS stack provides a logical link layer (LLC) layer above the RLC layer. The LLC layer too provides another SR ARQ mechanism at the LL frame level (each LL frame consists of several RLC blocks). A question in this regard is how many retransmission attempts (in the event of a block/frame errors) can be allowed at the RLC layer and the LLC layer. *We propose to investigate the optimal choice of parameters (e.g., number of retransmission attempts) for the ARQ schemes at the RLC and LLC layers*.

Internet protocol (IP) layer is provided above the LLC layer. The GPRS users thus are essentially provided with IP layer (and below), above which transport layer protocols and applications can be used. In this context, the interaction between the ARQs at the lower layers (RLC and LLC) and higher layers (e.g., TCP) is of interest. *We propose to study the performance of TCP on the uplink in GPRS.*

1.6 Literature Survey

The survey paper by Bettstetter *et al,* [18], provides a good introduction to GPRS architecture, protocols and system operation. Tutorial treatment of GPRS concepts and architecture are presented in in [19],[7],[39].

In [19], the GPRS protocol performance is evaluated through simulations using WINLAB GPRS simulator. This study assumes the railway traffic model (characterized by packet sizes that conform to a truncated exponential distribution with a mean value of 170 bytes), which is one of the different traffic load models defined in the ETSI evaluation guidelines. This study shows that multislot operation and capture reduces the blocking rate and improves both throughput and delay performance. In [7],[25], simulation results on the throughput and blocking rate performance of the GPRS MAC protocol, for all the three ETSI traffic models, namely, railway model, FUNET model and MOBITEX model, are compared. The FUNET traffic model is characterized by packet sizes conforming to a truncated Cauchy distribution with a mean value of 800 bytes. The MOBITEX traffic model is characterized by a uniform packet size distribution with a mean value of 30 bytes.

On RLC/MAC: In [28], the interaction between the ARQ schemes at the RLC/MAC and

LLC is analyzed by modeling the channel by a two state Markov model. However, the analysis does not take into account the request-reservation mechanism at the MAC layer. In [26], a hybrid FEC/ARQ mechanism using Reed-Solomon code to reduce the number of control blocks is used for the RLC acknowledgement mechanism. It is shown that packet transmission delay is reduced by using this scheme. In [38], the feasibility of supporting integrated services, such as packet voice, web browsing, and best-effort data using enhanced features to GPRS is discussed. Additional capabilities like $a)$ fast uplink access during an ongoing session, $b)$ fast resource assignment for both uplink and downlink, and $c)$ the ability to differentiate services at the BSS, are proposed. This paper also highlights the merits in going for slot level retransmission at the RLC, but does not provide any performance results[3].

On Internet/web performance: In [39], the performance of web browsing through GPRS facilities is evaluated. In particular, it presents an insight to the capacity of GPRS and the number of web users that can be supported with certain allocated resources. Simulations are used to evaluate the performance. In [32], the delay performance of web browsing on GPRS is studied. Different delay components associated with data transfer are identified. The average delay versus channel utilization and packet generation rate performance is evaluated, assuming a Poisson packet arrival process. It is shown that using multi-slot operation offers higher throughput and better delay than circuit switched connections. In [29], the performance of Internet applications such as http and ftp over GPRS is studied through simulations using OPNET. The results show that RLC/MAC retransmissions play a crucial role in the throughput performance for Internet traffic.

On QoS: In GPRS, QoS profiles for a number of service classes (e.g., reliability class, delay class) has been specified. In practice, QoS management can be provided by means of traffic scheduling, traffic shaping, and connection admission control.

A performance study of resource sharing by circuit-switched GSM connections and packet-switched GPRS sessions under a static and a dynamic channel allocation scheme is presented in [20]. Three different QoS profiles modeled by a weighted fair queueing scheme are considered. It is shown that the dynamic allocation of resources is preferred. In [21], several traffic scheduling

[3]We, in this thesis, provide a performance analysis of the slot level retransmission at RLC.

methods, including FIFO[4], static priority scheduling (SPS) and earliest deadline first (EDF), have been evaluated through simulations. It is shown that EDF is able to meet the delay requirements at a much higher channel utilization compared to FIFO and SPS. In [22],[23], the throughput, delay and loss rate performance of different scheduling algorithms, including FCFS with and without priority, fluid fair queueing (FFQ), and FCFS with windows, are evaluated through simulations for different traffic types.

In [33],[34], an admission control policy is proposed for multiservice GSM/GPRS networks. The performance of three different strategies of introducing GPRS traffic to GSM networks have been evaluated and compared. It is shown that with the use of the 'partial sharing' or 'complete sharing' strategy, better utilization of the GSM system is achieved. In [40], the influence of Internet traffic volume and characteristics on performance measures like new and handover blocking probabilities, system utilization and mean data rates of the services in multiservice GPRS network is evaluated. It is shown that by using efficient call admission and giving higher priority to voice service, $a)$ the increase of the volume of Internet traffic will not influence the blocking probability for voice, $b)$ the average bit rates achievable by the Internet connections will decrease, and $c)$ the utilization of the network will increase. In [31], the throughput and packet loss probability on the downlink (base station-to-mobile link), considering QoS, are evaluated using a continuous-time Markov chain model. This study assumes a dynamic channel allocation scheme. The employed traffic model constitutes a Markov modulated Poisson process.

1.7 Contributions

The key contributions in this thesis are as follows.

- We provide a modeling and *analytical* evaluation of the performance of the RLC/MAC protocols in GPRS, considering the uplink *request-reservation mechanism*. In the analysis, we use the theory of Markov chains to derive expressions for the uplink throughput and delay

[4] In our study in this thesis, we adopt the FIFO discipline of traffic scheduling.

performance of the GPRS MAC protocol. Our analysis quantifies the throughput-delay performance as a function of system traffic load, number of random access channels (to carry resource requests from mobiles) and number of traffic channels (to carry data traffic) on the uplink. The results can be of use in choosing the optimal mix of number of random access channels and traffic channels for a given traffic load.

- We propose a *slot level retransmission* at the RLC layer to enhance performance. We provide an analysis of the performance of slot level retransmission (SLR) at the RLC. We provide a performance comparison of SLR versus block level retransmission (BLR) at the RLC. We show that SLR improves the RLC layer performance compared to BLR, particularly when the channel error rates are high. In addition, SLR enables resource allocation at the individual time-slot level to achieve a finer resource allocation granularity and greater flexibility.
- We present a simulation study of the interaction between the ARQ schemes at the RLC and the LLC layers. Our performance results show that it is more beneficial to do error recovery by allowing more retransmission attempts at the RLC layer than at the LLC layer.
- We carry out a simulation study of the TCP performance on the uplink in GPRS. The difference of our TCP on GPRS study from others reported earlier in the literature is that we evaluate the TCP performance with slot level retransmission and compare it with TCP performance with block level retransmission. Also, we investigate the effect of various TCP parameters like fast retransmit threshold, and maximum window size on the TCP throughput performance.

1.8 Thesis Organization

The rest of this thesis is organized as follows. An overview of the GPRS system concept, architecture and protocol stack are presented in Chapter 2. In Chapter 3, we present the Markov chain analysis of the throughput and delay performance of the GPRS MAC protocol. We also provide the performance analysis of slot level retransmission at the RLC layer. Performance at the LLC layer is also presented. Chapter 4 presents the performance of TCP over GPRS using slot level retransmission versus block level retransmission at the RLC layer. Chapter 5 provides the conclusions and potential areas of further extensions to this research. A list of acronyms are provided in Chapter 7.

Chapter 2

Overview of GPRS

General Packet Radio Service (GPRS) is a *packet mode* wireless data system that has been standardized to operate on GSM infrastructure, by introducing new packet support nodes and associated protocol stacks [10],[11],[12]. A portion of the radio resources (channel frequencies) in an existing GSM system may be dedicated for packet data services using GPRS. Alternatively, GPRS and GSM services may dynamically share the same radio resources. Thus, GSM voice services and GPRS packet data services can co-exist in the same GSM system.

GPRS provides IP connectivity to the mobile users. The maximum data rate supported is 170 Kbps, realized through statistical multiplexing of traffic from different mobile users over-the-air. Point-to-point (e.g., e-mail, Internet access, remote access, road toll) as well as point-to-multipoint (e.g., multicast, financial updates, fleet management, traffic information) applications are envisaged. In this Chapter, we provide a brief overview of GPRS.

2.1 GPRS System Architecture

In order to integrate GPRS into the existing GSM architecture, a new class of network nodes, called GPRS support nodes (GSN), has been introduced. GSNs are responsible for the delivery and routing of data packets between the mobile stations and the external packet data networks (PDN). There are two types of GPRS support nodes (GSNs), namely, Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). The SGSN and GGSN enable the mobile data users to connect to the Internet, as shown in Fig. 2.1. The SGSN and GGSN essentially play analogous roles of the Serving MSC and Gateway MSC in a GSM system that connects voice users to the PSTN.

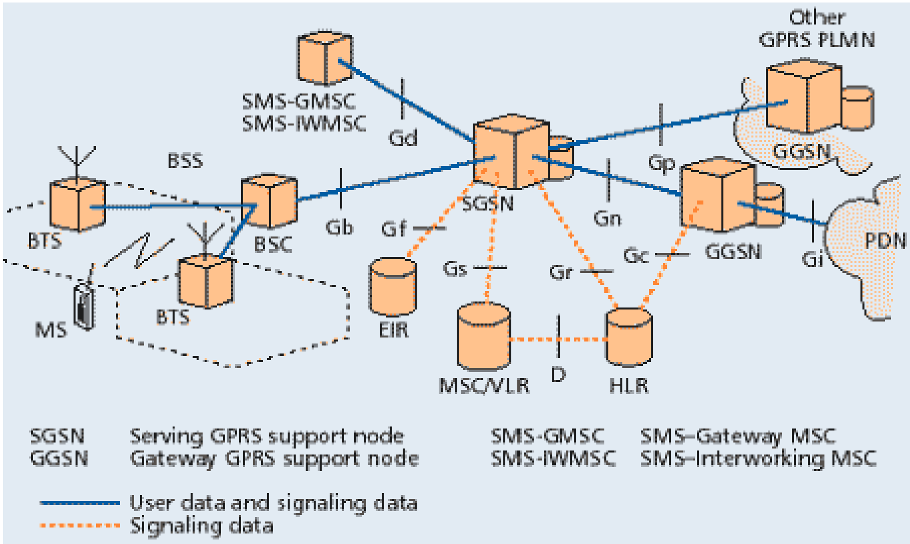


Figure 2.1: GPRS system architecture. (Ref: [18])

2.1.1 GPRS Support Nodes

Serving GPRS Support Node (SGSN): A serving GPRS support node (SGSN) is responsible for the delivery of data packets from and to the mobile stations within its service area. Its tasks include packet routing and transfer, mobility management, logical link management, authentication, and charging functions. The location register in the SGSN stores location information (e.g., current cell, current VLR) and user profiles (e.g., IMSI, address(es) used in the packet data network) of all GPRS users registered with this SGSN.

Gateway GPRS Support Node (GGSN): A gateway GPRS support node (GGSN) acts as an interface between the GPRS backbone network and the external packet data networks. It is mainly responsible for packet routing and transfer, and mobility management. It converts the GPRS packets coming from the SGSN into the appropriate packet data protocol (PDP) format (e.g., IP or X.25) and sends them out on the corresponding packet data network. In the other direction, PDP addresses of incoming data packets are converted to the GSM address of the destination user. The readdressed packets are sent to the responsible SGSN. This encapsulation of protocol data units (PDUs) at originating GSN and decapsulation at the receiving GSN is called *tunneling*. In between the GSNs, the Internet protocol (IP) is used to transfer the PDUs.

In general, there is a many-to-many relationship between the SGSNs and the GGSNs: a GGSN is the interface to external packet data networks for several SGSNs; a SGSN may route its packets over different GGSNs to reach different packet data networks.

The HLR stores the user profile, the current SGSN address, and the PDP address(es) for each GPRS user in the PLMN. The Gr interface is used to exchange this information between HLR and SGSN. For example, the SGSN informs the HLR about the current location of the MS. When the MS registers with a new SGSN, the HLR will send the user profile to the new SGSN. The signaling path between GGSN and HLR (Gc interface) may be used by the GGSN to query a user's location and profile in order to update its location register.

In addition, the MSC/VLR may be extended with functions and register entries that allow efficient coordination between packet switched (GPRS) and circuit switched (conventional GSM) services. Examples of this are combined GPRS and non-GPRS location updates and combined attachment procedures. Moreover, paging requests of circuit switched GSM calls can be performed via the SGSN. For this purpose, the Gs interface connects the data bases of SGSN and MSC/VLR.

2.1.2 A Routing Example

Fig. 2.2 illustrates an example of how packets are routed in GPRS. We assume that the packet data network is an IP network. A GPRS mobile station located in PLMN1 sends IP packets to a host connected to the IP network, e.g., to a web server connected to the Internet. The SGSN that the mobile station is registered with encapsulates the IP packets coming from the mobile station, examines the PDP context, and routes them through the intra-PLMN GPRS backbone to the appropriate GGSN. The GGSN decapsulates the packets and sends them out on the IP network, where IP routing mechanisms are used to transfer the packets to the access router of the destination network. The latter delivers the IP packets to the host.

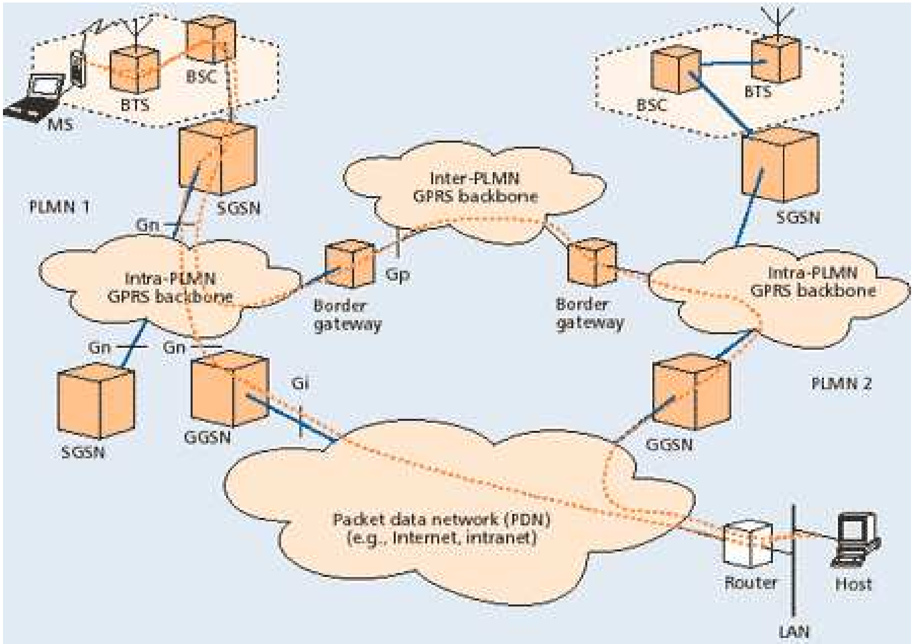


Figure 2.2: GPRS routing example. (Ref: [18])

Assume that the home-PLMN of the mobile station is PLMN2, but the MS roaming in PLMN1. In this case, an IP address has been assigned to the MS by the GGSN of PLMN2. Thus, the MS's IP address has the same network prefix as the IP address of the GGSN in PLMN2. The correspondent host is now sending IP packets to the MS. The packets are sent out onto the IP network and are routed to the GGSN of PLMN2 (the home-GGSN of the MS). The latter queries the HLR and obtains the information that the MS is currently located in PLMN1. It encapsulates the incoming IP packets and tunnels them through the inter-PLMN GPRS backbone to the appropriate SGSN in PLMN1. The SGSN decapsulates the packets and delivers them to the MS.

2.2 Protocol Architecture

Figure 2.3 illustrates the protocol architecture of the GPRS transmission plane [12] providing transmission of user data and its associated signaling.

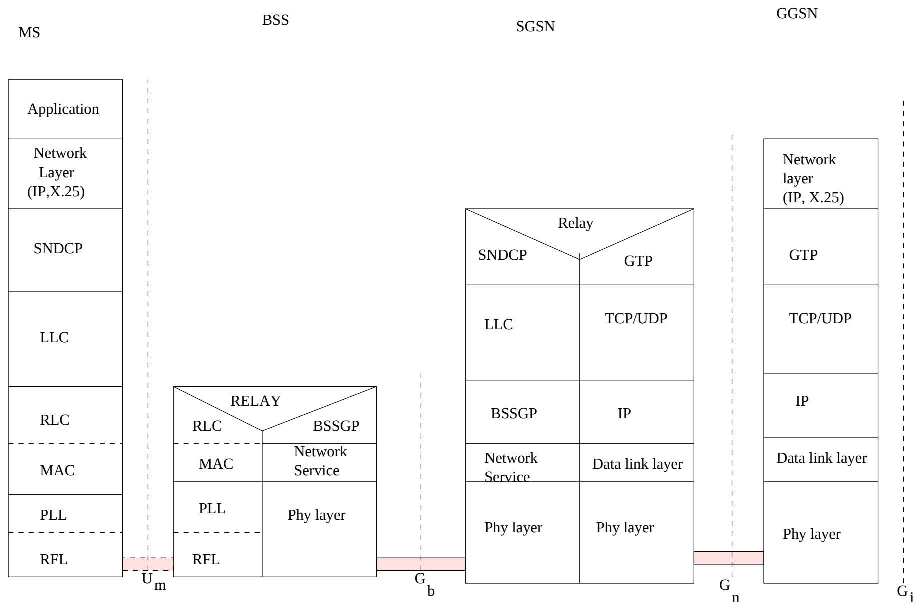


Figure 2.3: GPRS protocol stack

2.2.1 GPRS Backbone: SGSN–GGSN

As mentioned earlier, user data packets are encapsulated within the GPRS backbone network. The GPRS Tunneling Protocol (GTP) tunnels the user data packets and related signaling information between the GSNs. The protocol is defined both between GSNs within one PLMN (Gn interface) and between GSNs of different PLMNs (Gp interface). In the transmission plane, GTP employs a tunnel mechanism to transfer user data packets. In the signaling plane, GTP specifies a tunnel control and management protocol. The signaling is used to create, modify, and delete tunnels.

GTP packets carry the user's IP or X.25 packets. Below GTP, standard protocols like TCP or UDP are employed to transport the GTP packets within the backbone network. X.25 expects a reliable data link, thus TCP is used. UDP is used for access to IP-based packet data networks, which do not expect reliability in the network layer or below. IP is employed in the network layer to route packets through the backbone. Ethernet, ISDN, or ATM-based protocols may be used below IP.

To summarize, in the GPRS backbone we have an IP/X.25-over-GTP-over-UDP/TCP-over-IP transport architecture as shown in Fig. 2.3.

2.2.2 Subnetwork Dependent Convergence Protocol

The Subnetwork Dependent Convergence Protocol (SNDCP) is used to transfer data packets between SGSN and MS. Its functionality includes:

- Multiplexing of several connections of the network layer onto one virtual logical connection of the underlying LLC layer.
- Compression and decompression of user data and redundant header information.

2.2.3 Data Link Layer

The over-the-air communication between a MS and the GPRS network is defined by the physical layer and the data link layer. The physical layer functions involve modulation, demodulation, channel encoding/decoding, etc. The data link layer consists of two sublayers, namely, Logical Link Control (LLC) layer and the Radio Link Control/Media Access Control (RLC/MAC) layer.

The LLC layer [14] operates between the MS and the SGSN, and provides a logical link between them. The LLC layer supports two modes of operation, namely, Asynchronous Disconnected Mode (ADM) and Asynchronous Balanced Mode (ABM). In ADM, the LLC performs 'unacknowledged' operation and does not provide any error recovery procedure to guarantee in-order delivery. Thus, ADM mode of LLC operation may be applicable for delay sensitive, error tolerant applications (e.g., voice over IP). In ABM, the LLC performs 'acknowledged' operation which includes functionalities like error recovery through ARQ, in-order delivery and flow control (these functionalities are essentially based on the well known HDLC protocol). Also, LLC ensures data confidentiality through ciphering functions.

The RLC/MAC layers are primarily responsible for the efficient sharing of common radio resources by several mobiles. The RLC/MAC peers are at the mobile and the base station. The main purpose of the RLC layer is to establish a reliable link between the MS and the BSS. Each LLC frame is segmented into several RLC data blocks of fixed size. Each RLC data block occupies four time slots. The RLC layer can operate either in acknowledged or unacknowledged mode. In acknowledged mode, RLC provides for the selective retransmission of erroneous RLC data blocks (block level retransmission).

The MAC protocol operates on a slotted-ALOHA based reservation protocol [9]. The MAC layer requests/reserves resources in terms of number of data slots. The MAC function provides arbitration between multiple mobiles attempting to transmit simultaneously, and provides collision detection and recovery procedures[1].

2.2.4 Physical Layer

The physical layer between MS and BSS is divided into the two sublayers: the physical link layer (PLL) and the physical RF Layer (RFL). The PLL provides a physical channel between the MS and the BSS. Its tasks include channel coding (detection of transmission errors, forward error correction, indication of uncorrectable codewords), interleaving, and detection of physical link congestion. The RFL operates below the PLL. Among other things, it includes modulation and demodulation.

GPRS uses the same physical layer as defined in GSM. GSM uses a combination of FDMA and TDMA for multiple access. Two frequency bands 45 MHz apart are used for GSM operation: 890 - 915 MHz for transmission from the mobile station to the BS (i.e., uplink), and 935 - 960 MHz for transmission from the BS to the MS (i.e., downlink). Each of these bands of 25 MHz width is divided into 124 single carrier channels of 200 kHz width.

Each of the 200 kHz wide frequency channels is partitioned into eight TDMA time slots which constitute one TDMA frame. Each frequency-time slot pair constitute a physical channel. Each time slot can carry 156.25 bits of information at a channel transmission rate of 270.833 Kbps. Accordingly, each time slot occupies 576.9 μsec and each TDMA frame occupies 4.613 msec. The modulation used is GMSK with a bandwidth-delay product of 0.3.

The resource allocation in GSM is done in such a way that each voice call is assigned one time slot in a TDMA frame. Also, the slot assignment in GSM is symmetrical on both links

[1] Chapter 3 provides a more detailed description of the MAC protocol, its modeling and performance analysis.

(i.e., one time slot on the the uplink and one time slot on the downlink are alloted per call). On the other hand, in GPRS, each call can be assigned either one slot per TDMA frame (single slot operation) or multiple slots per TDMA frame (multislot operation). Also, the resource allocation in GPRS is done in units of RLC blocks each occupying four time slots. In addition, resources can be alloted in an assymetrical fashion in GPRS. For example, in typical web based applications, a short query from a mobile on the uplink could result in a large download from the BS side on the downlink, which would require a higher data rate on the downlink. In such scenarios, GPRS allows allocation of more slots on the downlink than on the uplink. Both fixed channel allocation (FCA) and dynamic channel allocation (DCA) are allowed in GPRS.

In GSM, a physical channel (time slot) is permanently allocated to a user for the entire duration of the call (whether data is transmitted or not). In contrast to this, in GPRS, the slots are allocated only when data packets need to be sent, and the assigned slots are released after the transmission. For bursty traffic, this results in a much more efficient usage of the scarce radio resources.

2.3 Logical Channels

On top of the physical channels, a series of logical channels are defined to perform a multiplicity of functions, e.g., signaling, broadcast of general system information, synchronization, channel assignment, paging, or payload transport. As with conventional GSM, they can be divided into two categories: traffic channels and signaling (control) channels. For a complete listing of the logical channels in GPRS, refer [11]. Here, we define those logical channels which are relevant to our study.

The Packet Data Traffic CHannel (PDTCH) is employed for the transfer of user data. These PDTCH slots are the resources that the mobiles request for and get assigned from the BS for user data transfer.

The Packet Common Control CHannel (PCCCH) is a bidirectional point-to-multipoint signaling channel that transports signaling information for network access management, e.g., for allocation of radio resources and paging. It consists of four sub-channels:

- Packet Random Access Channel (PRACH) is used by the mobiles to send their resource requests to the BS. PRACH is a slotted ALOHA channel on which mobile users can send their resource request packets in an uncoordinated manner. Hence, collision among request packets from multiple users can result, which is resolved through suitable collision resolution mechanisms. The system operator may choose to define one or more PRACH depending on the number of users and traffic load in the system.

- Packet Access Grant Channel (PAGCH) is used by the BS to convey the allocated channel information to the mobiles.

- Packet Paging CHannel (PPCH) is used by the BS to convey an incoming call to a mobile.

- Packet Notification Channel (PNCH) is used by the BS to inform a mobile of incoming PTM messages (multicast or group call).

In addition to common control channels, there are dedicated control channels, including Packet Associated Control CHannel (PACCH). The PACCH is always allocated in associated with one or more PDTCH that are assigned to one MS. PACCH transports signaling information specific to one specific MS (e.g, ACKs).

The coordination between circuit switched GSM and packet switched GPRS logical channels is important. If the PCCCH is not available in a cell, a mobile station can use the common control channel (CCCH) of GSM to initiate the packet transfer. For example, if PRACH is not available, packet channel requests can be sent on RACH.

In the following subsections, we describe typical mobile originated packet transfer and mobile terminated packet transfer in GPRS.

2.4 Mobile Originated Packet Transfer

A mobile station originates a packet transfer by sending a Packet Channel Request to the network on the PRACH. The network assigns the requested resources (if available), and informs the assignment information to the MS on the PAGCH. This request/assignment procedure can be done either in one-phase or in two-phases. In the one-phase access, the network responds to the Packet Channel Request with the Packet Immediate Assignment, reserving the requested resources on PDTCHs (in units of number of radio blocks) for uplink transfer. In the two-phase assignment, on the other hand, the network responds to the Packet Channel Request with the Packet Immediate Assignment, which reserves the uplink resources ONLY for sending a Packet Resource Request. The MS then sends the Packet Resource Request message on the PACCH indicating the complete details of the requested resources for uplink transfer. Thereafter, the network responds with the Packet Resource Assignment on the PACCH, reserving the requested resources for the uplink transfer. The mobile then sends its data frames on the assigned PDTCH. Channel errors in the data frames are handled through ARQ mechanisms at the RLC and LLC layers. While one-phase access is mandatory procedure, two-phase access is an optional procedure[2]. Figure 2.4 illustrates a two-phase resource request/assignment procedure for uplink data transfer.

It is possible that the Packet Channel Requests from several mobiles can cause the BS to queue up these requests before allocation of resources. A Packet Queuing Notification is sent to the resource requesting mobiles. This notification includes information that the Packet Channel Request message is correctly received and Packet Immediate Assignment may be transmitted later.

2.5 Mobile Terminated Packet Transfer

When a packet call arrives from the external PDN to a mobile station, the network initiates a packet transfer by paging the mobile. It sends a Packet Paging Request on the Packet Paging Channel(PPCH) on the downlink. The MS responds to the page by sending a Packet Channel Request on the PRACH to the BS. The BS then sends a Packet Immediate Assignment message

[2]In our study in this thesis, we consider one-phase access.

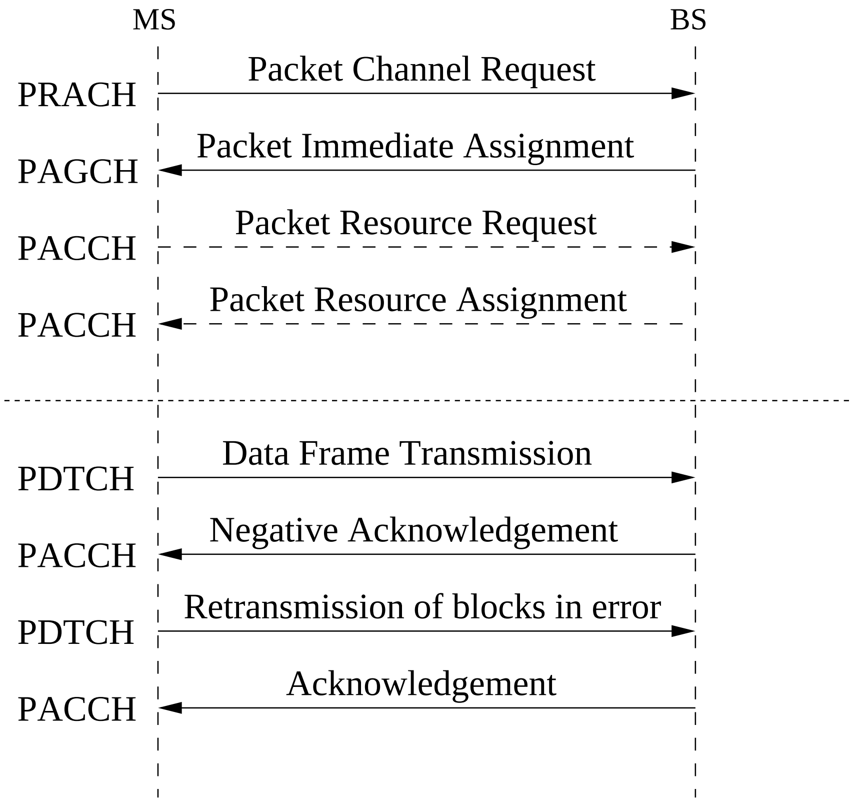


Figure 2.4: Mobile originated data transfer procedure

on the PAGCH, indicating that the BS is ready to allocate resources. The MS responds to this message with a Packet Paging Response on the PACCH asking the identity of the alloted resources. A Packet Resource Assignment message from the BS provides the assigned resources information to the MS. The BS then starts transmitting date frames on the assigned PDTCHs. The MS listens to these PDTCHs and sends the appropriate ACK/NACK messages to the BS. In the case of a NACK, only those radio blocks listed as erroneous are retransmitted. Figure 2.5 illustrates the paging and packet transfer procedures.

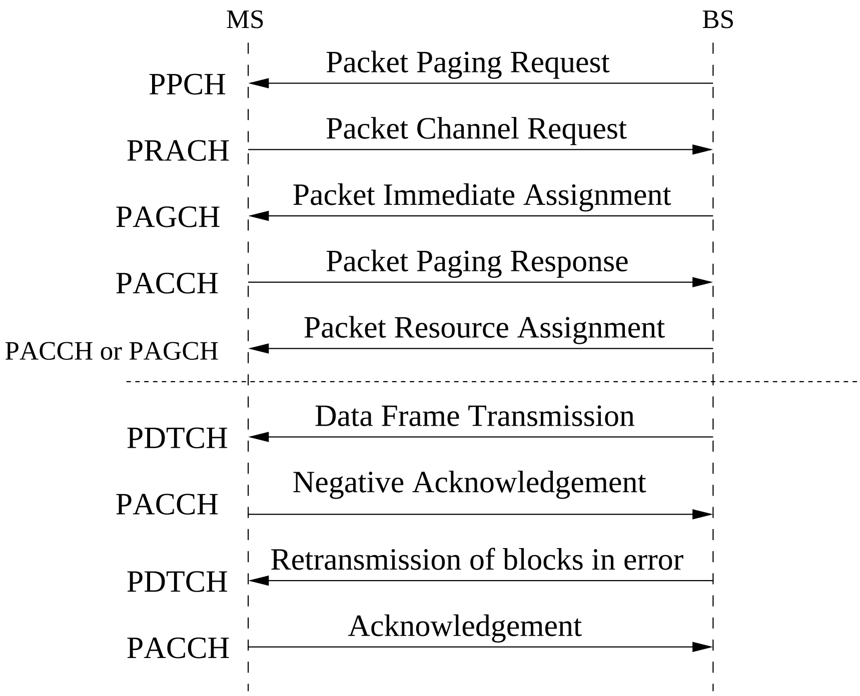


Figure 2.5: Mobile terminated data transfer procedure

Chapter 3

Performance Analysis of LLC/RLC/MAC Layers

The over-the-air communications between the mobile station (MS) and the GPRS network is defined by the physical layer and the data link layer functionalities. The physical layer functions involve modulation, demodulation, channel encoding/decoding, etc. The data link layer consists of two sublayers, namely, Logical Link Control (LLC) layer, and the Radio Link Control/Media Access Control (RLC/MAC) layer. In this Chapter, we present the functionalities, the modeling and performance analysis of the LLC, RLC, and MAC layers in the GPRS protocol stack. Refer Fig. 2.3 for the GPRS protocol stack architecture.

3.1 LLC/RLC/MAC Layers

In this Section, we present functionalities of the LLC, RLC and MAC layers that are relevant to our performance analysis.

LLC Layer: The LLC layer operates between the MS and the SGSN, and provides a logical link between them. Packet data units (PDUs) from higher layers (IP layer) are segmented into variable size LLC frames (see Figs. 3.1, 3.2). The functions of LLC layer include link level flow control and ciphering. The LLC layer can operate either in an unacknowledged mode or in an acknowledged mode. In the unacknowledged mode of operation, the LLC layer does not attempt recovery of erroneous frames. LLC frames, erroneously received or otherwise, are passed on to the higher layers. In the acknowledged mode, the LLC layer provides an ARQ mechanism to retransmit erroneous LLC frames. A Frame Check Sequence (FCS) is used provided in each LLC frame to detect frame errors. A retransmission count variable $N200$ is defined [14]. The LLC is reset and error recovery is passed on to higher layers (e.g., TCP) if frames errors could not be recovered within $N200$ retransmission attempts.

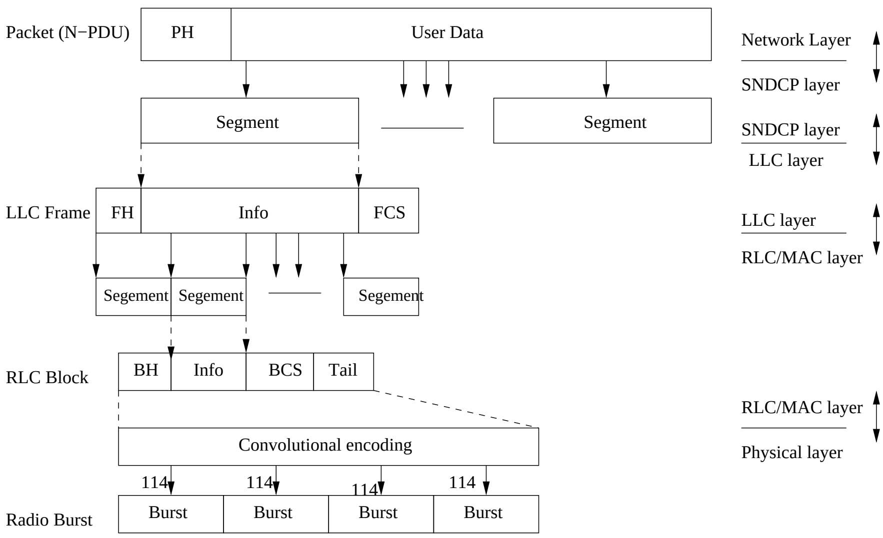
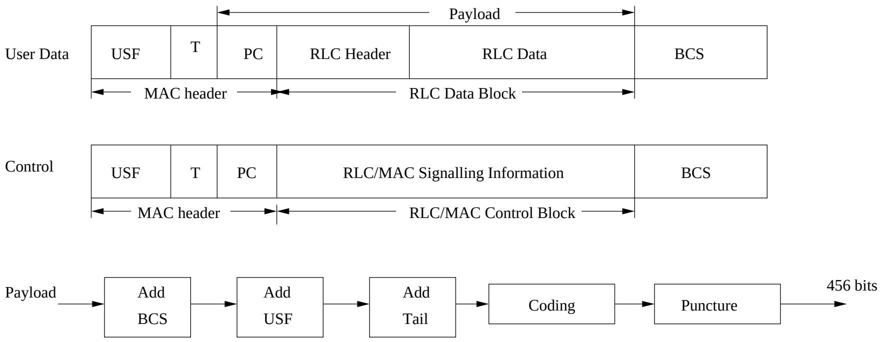


Figure 3.1: Network layer PDU segmentation into LLC frames, RLC blocks, and MAC bursts.

RLC Layer: The RLC layer is provided below the LLC layer and above the MAC layer. The RLC specifications are given in [9]. The RLC peers are at the MS and the BS as shown in Fig. 2.3. On the transmit side, the RLC layer segments each LLC frames into several RLC data blocks (see Fig. 3.1). Each RLC data block occupies four time slots, irrespective of the type of channel coding scheme used. Coding schemes CS-1, CS-2, CS-3, CS-4 are defined with rate 1/2, rate 2/3, rate 3/4, and rate 1 (no coding), respectively [9]. In the case of coding scheme CS-1, each RLC block consists of 181 information bits, 40 block check sequence (BCS) bits, and 7 tail/control bits. With single slot operation, the capacity of coding scheme CS-1 is 9.05 Kbps (i.e., 181 bits in four TDMA frames, where each TDMA frame occupies 4.615 ms). Similarly, the maximum information rates possible using other coding schemes are as follows: CS-2: 13.4 Kbps, CS-3: 15.6 Kbps, CS-4: 21.4 Kbps. With multislot operation (allocation of upto 8 slots in a TDMA



Scheme	Code rate	Payload	BCS	USF	Tail	Coded bits	Puntured	Rate (Kbps)
CS−1	1/2	181	40	3	4	456	0	9.05
CS−2	~2/3	268	16	6	4	588	132	13.4
CS−3	~3/4	312	16	6	4	676	220	15.6
CS−4	1	428	16	12	0	456	0	21.4

Figure 3.2: RLC/MAC block structure and coding

frame to an user), these maximum possible information rates are increased by eight fold.

Like LLC, RLC too can operate either in an acknowledged mode or in an unacknowledged mode. In acknowledged mode, RLC provides a selective repeat ARQ mechanism to recover erroneous RLC data blocks. A block check sequence (BCS) is provided in each RLC data block to enable error detection. A RLC retransmission counter N3104 is defined [9] which keeps track of the number of times RLC blocks are retransmitted in case of error. The RLC layer is allowed to attempt a maximum of N3104_MAX retransmissions to recover blocks in error. If an erroneous block is not recovered within N3104_MAX retransmission attempts, then control is passed on to LLC to recover errors at the frame level. In the unacknowledged mode, there is no retransmission of erroneous RLC blocks. On the receive side, the RLC performs reassembly of LLC frames.

MAC Layer: The GPRS MAC protocol operates on a slotted-ALOHA based reservation protocol. The MAC layer peers are at the MS and the BS. The MAC layer requests/reserves resources in number of traffic data slots. The MAC function provides arbitration between multiple mobiles attempting to transmit simultaneously, and provides collision detection and recovery procedures.

The Packet Random Access CHannel (PRACH) is used by all the mobiles, on a contention basis, for the purpose of sending resource request packets. Typically, TS0 slot in a GSM frame of 8 slots can be used as PRACH. All mobiles are allowed to transmit on PRACH slots, following slotted-ALOHA protocol [6]. Depending on the system load, the number of PRACHs can be increased. The Packet Data Traffic CHannels (PDTCH), on the other hand, are used for the transfer of data packets. Resource requests are made by the mobiles in terms of number of uplink PDTCH slots required. Based on these requests, PDTCH slots are dynamically assigned to the mobiles by the base station. Allocation can be done on a one time slot per GSM TDMA frame basis (called *single slot operation*) or multiple time slots per GSM TDMA frame basis (called *multislot operation*).

When the MAC at the mobile side receives RLC data blocks to be transferred to the base station, it sends a request packet on the immediately following PRACH slot. The request packet indicates K, the number of PDTCH slots required. If the base station receives the request packet without collision or channel errors, and if PDTCH slots are available to honor the request, the base station informs the reservation information to the mobile on the downlink Packet Access Grant CHannel (PAGCH) channel. The reservation information include the PDTCH frequency-time slots that can be used by the mobile for data transfer. The mobile then sends data in those K reserved slots. On the other hand, if the request packet is lost (due to collision or channel errors) or if PDTCH slots are not available, then the mobile will not get the reservation. The mobile will then reschedule its request packet retransmission attempt to a later time (typically, after a random backoff time).

The MAC control parameters include MAX_RETRANS, PERSISTENCE_LEVEL as defined in [9]. The MAC layer can send channel requests on the PRACH slots upto a maximum of MAX_RETRANS+1 retransmission attempts in the event of loss due to collision or channel errors. The delay between retransmission attempts is defined by the PERSISTENCE_LEVEL.

Thus, in summary, in terms of error recovery at different layers,

- MAC layer attempts to resolve collision of request packets,
- RLC layer attempts to recover RLC data block errors through a selective repeat ARQ mechanism, and

- LLC attempts recovery of erroneous LLC frames through another ARQ mechanism.

Link errors unresolved at LLC layer are passed on to higher layers (e.g., transport layer) to resolve.

In the following sections, we provide the system model and the performance analysis of the MAC, RLC, and LLC layer.

3.2 System Model

Consider a single cell GPRS system with M, $M \geq 2$ uplink channels and N mobile users. Each channel corresponds to a frequency-time slot pair in the mobile-to-base station direction. Out of M channels, L, $1 \leq L < M$, channels are used as packet random access channels (PRACH), and the remaining $M - L$ channels are used as packet data traffic channels (PDTCH). Typically, slot TS0 in all GSM TDMA frames on a given frequency can form a PRACH. Likewise, on a given frequency, slot TS1 in all GSM TDMA frames can form PDTCH-1, slot TS2 can form PDTCH-2, and so on.

We consider the *single slot operation*, where only one slot per GSM TDMA frame is assigned to a user on PDTCH. For example, TS1 slots in consecutive frames n, $n+1$,...,$n+K$ being assigned to a mobile for data transfer is a typical illustration of single slot operation.

Considering single slot operation, all M uplink channels can be modeled as synchronized slotted channels as shown in Figure 3.3. One request packet is one slot in size. One network layer packet data unit (PDU), including LLC/RLC headers and checksums, occupies several slots. Between the successful transmission of a request packet on a PRACH slot and the corresponding data transmission on the assigned PDTCH slots, some finite time gets elapsed because of the propagation and processing delays involved. This delay is typically of the order of a few slots.

We are interested in analyzing the throughput and delay performance of the GPRS MAC protocol modeled as above. In order to carry out the performance analysis, we assume the following:

1. The network layer PDU arrival process (hence the new request packet generation process)

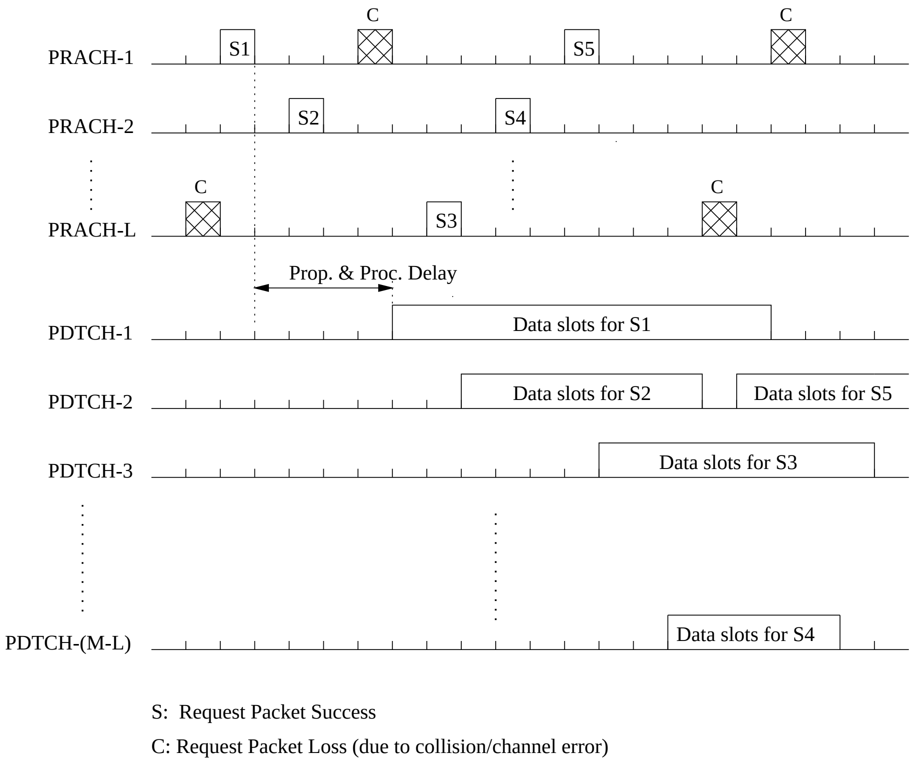


Figure 3.3: GPRS-MAC protocol operation

is Bernoulli with arrival probability, λ, in each slot. A new network layer PDU is accepted only after the completion of the transfer of the previously accepted PDU.

2. The length of the PDU (including LLC/RLC headers and checksums), measured in number of slots, is geometric with parameter g_d, $0 < g_d < 1$.

3. Loss of request packets on PRACH is only due to collision.

4. Retransmission attempts of request packets following a collision on PRACH (or non-availability of PDTCH) is geometrically delayed with parameter g_r, $0 < g_r < 1$. This parameter g_r essentially models the backoff delay (in the event of request packet loss due to collision/channel errors) which is characterized by the standards defined parameter PERSISTENCE_LEVEL.

5. Propagation and processing delays are assumed to be negligible. This assumption can be valid in our considered system of *single slot* operation, where the response from the BS can come within one TDMA frame time itself.

3.3 MAC Layer Throughput-Delay Analysis

As per the GPRS MAC protocol and the system model described in the previous sections, the mobile can be in any one of the following states, namely,

- *Idle* state,
- *Backlogged* state,
- *Data-Tx Success* state, and
- *Data-Tx Failure* state,

as shown in Figure 3.4.

In *idle* state, a mobile remains idle with probability (1-λ) and generates a PDU with probability λ. If n requests come in a slot, then at most one request is received correctly (capture) with probability $cap(n, 1)$. The remaining $n-1$ mobiles go to *backlogged* state. If a mobile's request succeeds, but there are no available PDTCHs to serve the request, then also the mobile goes to the *backlogged* state. On the other hand, if there are available PDTCHs to serve a request, then the mobile goes to the *data-tx success* state or *data-tx failure* state, where it sends data on the assigned PDTCH slots.

3.3.1 Per Channel Throughput

Let $\{\mathbf{D_{ft}}; \mathbf{t} \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}, ...\}\}$ represent the number of mobiles in *data-tx failure* state at the beginning of slot t, $\{\mathbf{D_{st}}; \mathbf{t} \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}, ...\}\}$ be the number of mobiles in the *data-tx success* state, and $\{\mathbf{B_t}; \mathbf{t} \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}, ...\}\}$ be the number of mobiles in *backlogged* state. The three dimensional process $\{\mathbf{Z_t} = (\mathbf{D_{ft}}, \mathbf{D_{st}}, \mathbf{B_t}); \mathbf{t} \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}, ...\}\}$ is a Markov chain because of the assumptions made above. The one-step state transition probability, $\tilde{P}_{z_1,z_2}$, that the system moves from state $\mathbf{Z_t} = \mathbf{z_1} = (\mathbf{i_1}, \mathbf{j_1}, \mathbf{k_1})$ at time t, to state $\mathbf{Z_{t+1}} = \mathbf{z_2} = (\mathbf{i_2}, \mathbf{j_2}, \mathbf{k_2})$, $0 \leq i_1 \leq M-L$, $0 \leq j_1 \leq M-L-i_1$, $0 \leq k_1 \leq N-i_1-j_1$, $0 \leq i_2 \leq M-L$, $0 \leq j_2 \leq M-L-i_2$, $0 \leq k_2 \leq N-i_2-j_2$,

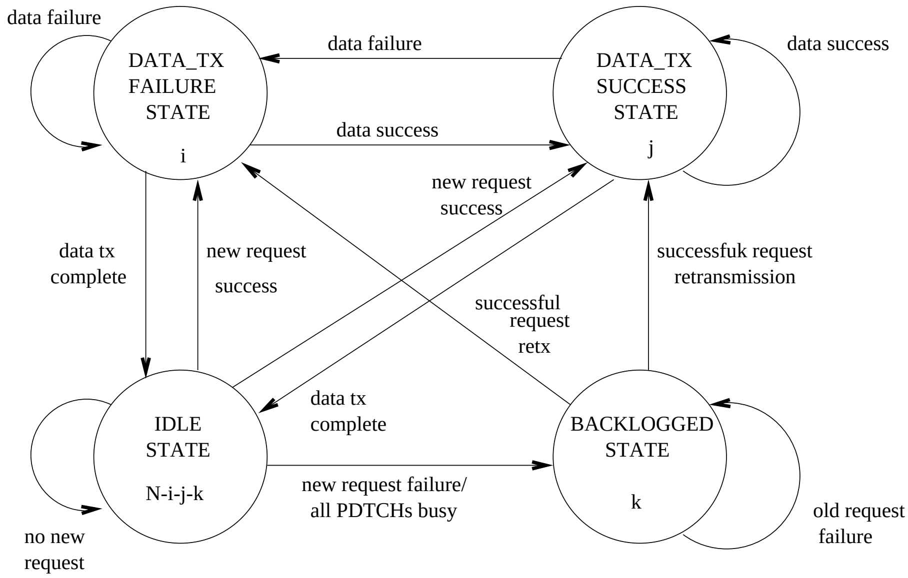


Figure 3.4: Mobile station state transition diagram

at time $t+1$ is given by [1]

$$\begin{aligned}
\tilde{P}_{z_1,z_2} \quad = \quad & \sum_{n=0}^{N-i_1-j_1} \sum_{c_s=0}^{L} \sum_{s_j=0}^{c_s} \sum_{s_f=0}^{j_1} \sum_{f_s=0}^{i_1} \qquad (3.1)\\
& \cdot \binom{N-i_1-j_1-k_1}{a} \lambda^a (1-\lambda)^{N-i_1-j_1-k_1-a} \\
& \cdot \binom{k_1}{n-a} {g_r}^{n-a} (1-g_r)^{k_1-n+a} \\
& \cdot \binom{i_1}{b_i} {g_d}^{i_1-b_i} (1-g_d)^{b_i} \\
& \cdot \binom{b_i}{f_s} (1-P_s)^{f_s} P_s^{b_i-f_s} \\
& \cdot \binom{j_1}{b_j} {g_d}^{j_1-b_j} (1-g_d)^{b_j} \\
& \cdot \binom{b_j}{s_f} {P_s}^{s_f} (1-P_s)^{b_j-s_f} \\
& \cdot \binom{c_s}{s_j} (1-P_s)^{s_j} P_s^{c_s-s_j} . f(c_s, n, L),
\end{aligned}$$

[1]The derivation of this transition probability expression is given in Appendix A.1.

where $a = k_2 - k_1 + c_s$, n is the number of requests sent, c_s is the number of successful requests in the t^{th} slot, $b_j = j_2 - s_j - f_s + s_f$, $b_i = i_2 - c_s + s_j - s_f + f_s$, $f_s \leq b_i$, $s_f \leq b_j$, P_s is the average probability of slot error, and

$f(c_s, n, x)$ = Prob(c_s successes given that n requests are sent and x PRACHs are available). Then,

$$f(c_s, n, x) = \sum_{l=0}^{n} \sum_{c=0}^{1} \binom{n}{l} \left(\frac{1}{x}\right)^l \left(1 - \frac{1}{x}\right)^{n-l} \cdot cap(l, c).f(c_s - c, n - l, x - 1), \tag{3.2}$$

where $cap(l, c)$ is the probability of capturing c out of l colliding request packets.

Eqn. (3.3) is terminated by fixing $f(c, n, 1)$ = $cap(n, c)$. As in [15], we set $cap(0, 1) = 0$, $cap(1, 1) = 1$, $cap(2, 1) = 0.67$, $cap(3, 1) = 0.48$, $cap(4, 1) = 0.40$, $cap(5, 1) = 0.35$, $cap(n, 1) = 0$ for $n > 5$, and $cap(n, 0) = 1 - cap(n, 1)$.

The above analysis assumes a large number of PDTCHs, so that a successful request always gets an assignment. If the number of PDTCHs is small compared to the number of users N, (i.e., $M - L < N$), then a successful request may not get an assignment as all the PDTCHs may be busy. This event occurs if $M-L-(L-1) \leq (i_1+j_1) \leq M-L$ and $M-L-(L-1) \leq (i_2+j_2) \leq (i_1+j_1)$. This implies that the total number of *idle* channels $M - L - (i_1 + j_1) < L$. In principle, $y \leq L$ requests succeed in any time slot and the number of idle channels $M - L - (i_1 + j_1)$ is less than y then all y requests are backlogged. For those z_1 and z_2 satisfying the above condition, compute the probability $\hat{P}_{z_1,z_2}$ and add it to the corresponding $\tilde{P}_{z_1,z_2}$ term calculated using Eqn. (3.2).

Let x represents the number of idle channels which is given by $M - L - (i_1 + j_1)$, and y represents the number of sessions ended in that time slot which is given by $(i_1 + j_1) - (i_2 + j_2)$ and $\mu = k_2 - k_1$ the new arrivals which are backlogged due to non availability of channels. Therefore, the probability $\hat{P}_{z_1,z_2}$ for the transition $z_1 = (i_1, j_1, k_1)$ to $z_2 = (i_2, j_2, k_2)$ transition is given by,

$$\hat{P}_{z_1,z_2} = \sum_{\theta=0}^{k_1} \sum_{c_s=y+x+1}^{L} \sum_{s_f=0}^{i_2} \sum_{f_s=0}^{j_2} \tag{3.3}$$

$$
\begin{aligned}
&\cdot \binom{N-i_1-j_1-k_1}{\mu} \lambda^{\mu}(1-\lambda)^{N-i_1-j_1-k_1-\mu} \\
&\cdot \binom{k_1}{\theta} {g_r}^{\theta}(1-g_r)^{k_1-\theta} \\
&\cdot \binom{i_1}{b_i} {g_d}^{i_1-b_i}(1-g_d)^{b_i} \\
&\cdot \binom{b_i}{f_s} (1-P_s)^{f_s} P_s^{b_i-f_s} \\
&\cdot \binom{j_1}{b_j} {g_d}^{j_1-b_j}(1-g_d)^{b_j} \\
&\cdot \binom{b_j}{s_f} {P_s}^{s_f}(1-P_s)^{b_j-s_f} \\
&\cdot f(c_s, \mu+\theta, L)
\end{aligned}
$$

where $b_j = j_2 - f_s + s_f$, $b_i = i_2 - s_f + f_s$, $f_s \le b_i$, $s_f \le b_j$.

Thus,

$$P_{z_1,z_2} \;=\; \tilde{P}_{z_1,z_2} + \hat{P}_{z_1,z_2}. \tag{3.4}$$

The Markov chain $\{\mathbf{Z_t}; \mathbf{t} \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}, ...\}\}$ has a finite number of states and is positive recurrent [30]. Hence, it has a stationary steady state distribution and is found by solving

$$\Pi = \Pi\mathbf{P}, \tag{3.5}$$

where $\Pi = [\pi_{ijk}], 0 \le i \le M-L, 0 \le j \le M-L-i, 0 \le k \le N-i-j$, is the steady state probability vector. The total system throughput, η, is defined as

$$\eta \;=\; \sum_{i=0}^{M-L} \sum_{j=0}^{M-L-i} \sum_{k=0}^{N-i-j} j\pi_{ijk}. \tag{3.6}$$

The per channel throughput, η_c, is then given by [2]

$$\eta_c = \eta/M. \tag{3.7}$$

3.3.2 Mean Delay

Next, we derive the mean PDU transfer delay performance. The mean PDU transfer delay, D, is defined as the the average number of slots elapsed from the slot where a PDU arrived to the slot where the PDU transmission is complete. The number of users in the *non-idle* state (i.e., *data-tx failure*, *data-tx success*, and $backlogged$) contribute to the mean delay. There are $\nu = (i + j + k)$ non $idle$ users in the system and averaging it over steady state distribution gives

$$\mathrm{E}(\nu) = \sum_{i=0}^{M-L} \sum_{j=0}^{M-L-i} \sum_{k=0}^{N-i-j} (i + j + k)\pi_{ijk}. \tag{3.8}$$

There are $N - i - j - k$ $idle$ users each will generate requests with probability λ in each slot. The average arrival rate to the system is given by

$$\Lambda = \lambda(N - E(\nu)). \tag{3.9}$$

¿From Little's theorem, the average time an user spends in the system is given by the ratio between the number of users in the system to the average arrival rate. Hence,

$$\mathrm{D} = 1 + \frac{E(\nu)}{\Lambda}. \tag{3.10}$$

Note that the one in Equation (3.10) is added to ensure that there is one slot delay for the mobiles to enter into the *non-idle* state.

[2]The limiting behavior of η_c for $\lambda \to 0$ and $\lambda \to 1$ are discussed in Appendix A.2.

3.4 RLC Layer Performance

Note that the analysis in the previous section corresponds to the MAC protocol operation with RLC in the unacknowledged mode (i.e., there is no ARQ at the RLC). In the acknowledged mode, however, RLC retransmits erroneous blocks using a selective repeat ARQ mechanism. Each RLC block consists of four slots. It is noted that slot level retransmission in RLC is proposed as an alternative to block level retransmission [38].

In our RLC analysis here, we consider a slot level retransmission mechanism by which a slot in error is repeatedly retransmitted until it succeeds. This implies that the number of retransmission attempts at the RLC is infinity. We derive the throughput-delay performance of this slot level retransmission scheme as follows.

Let the random variable T represent the number of slots per PDU and the random variable Y_i represent the number of transmission attempts of i^{th} slot until success. Thus, the total number of slots required for successfully transmitting T slots is given by,

$$\mathrm{X} = \sum_{i=1}^{T} Y_i. \tag{3.11}$$

The distribution of X is given by,[3]

$$\mathrm{Pr(X}{=}x) = g_d(1 - P_s)[1 - g_d(1 - P_s)]^{x-1}, x = 1, 2, 3... \tag{3.12}$$

where P_s is the slot error rate.

Now, in order to obtain the throughput and mean delay for the RLC (acknowledged mode) with slot level retransmission, we need to just change the parameter g_d to $g_d(1 - P_s)$ in Eqn. (3.2) and Eqn. (3.4).

Note that the above analysis is for infinite slot retransmission attempts at the RLC. The effect of finite number of RLC retransmission attempts is evaluated through simulations. Likewise, the performance using block level retransmission at RLC is evaluated through simulations. In the

[3]The derivation of the distribution of X is given in Appendix A.3.

following section, we provide the results and discussions of the RLC/MAC performance obtained through the analysis and simulations.

3.4.1 RLC/MAC Performance Results and Discussion

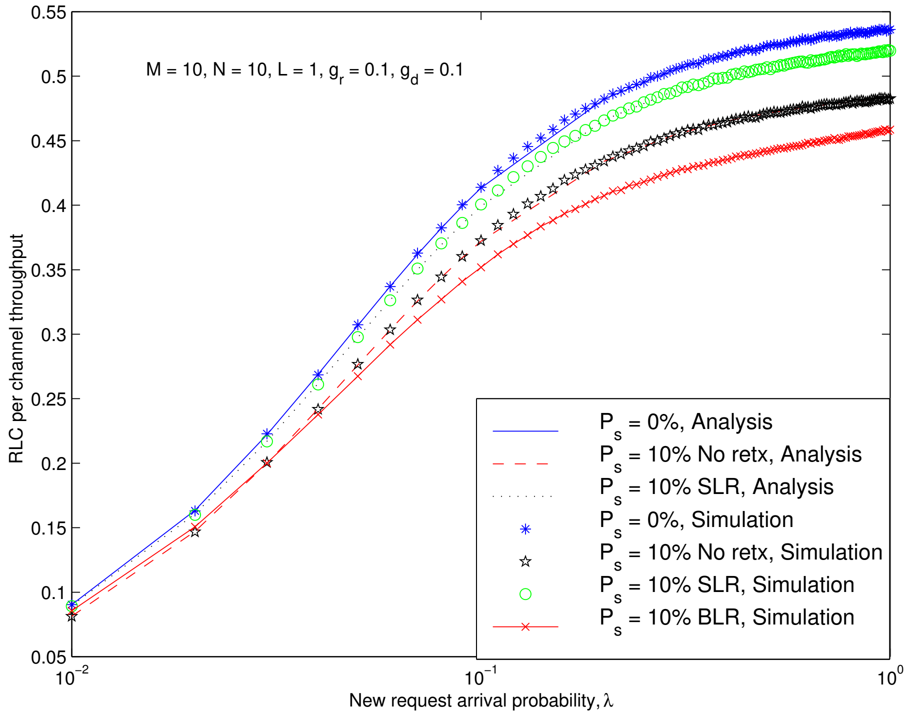


Figure 3.5: RLC/MAC average per channel throughput, η_c, versus new request arrival probability, λ. $N = 10$, $M = 10$, $L = 1$, $g_r = 0.1$, $g_d = 0.1$.

In Figure 3.5, numerical results for the average per channel throughput of the GPRS RLC/MAC protocol, obtained from Eqn. (3.7), for N=10, M=10, L=1, g_r = 0.1, and g_d = 0.1 are plotted as a function of new request arrival probability, λ. It is noted that g_r value of 0.1 implies that the average backoff delay after collision is 10 slots. Likewise, g_d value of 0.1 implies that the average length of the PDU measured in number of slots is 10. According to the throughput computation in Eqn. (3.7) the PRACH slots do not contribute to the effective throughput. In other words, for M channels out of which L channels are PRACH, the maximum capacity is given by $(M - L)/M$. The effect of slot errors with/without RLC slot level retransmission is also plotted for a slot error rate of 10% (which corresponds to a poor channel condition). These results are compared with the corresponding RLC block level retransmission performance obtained through simulation.

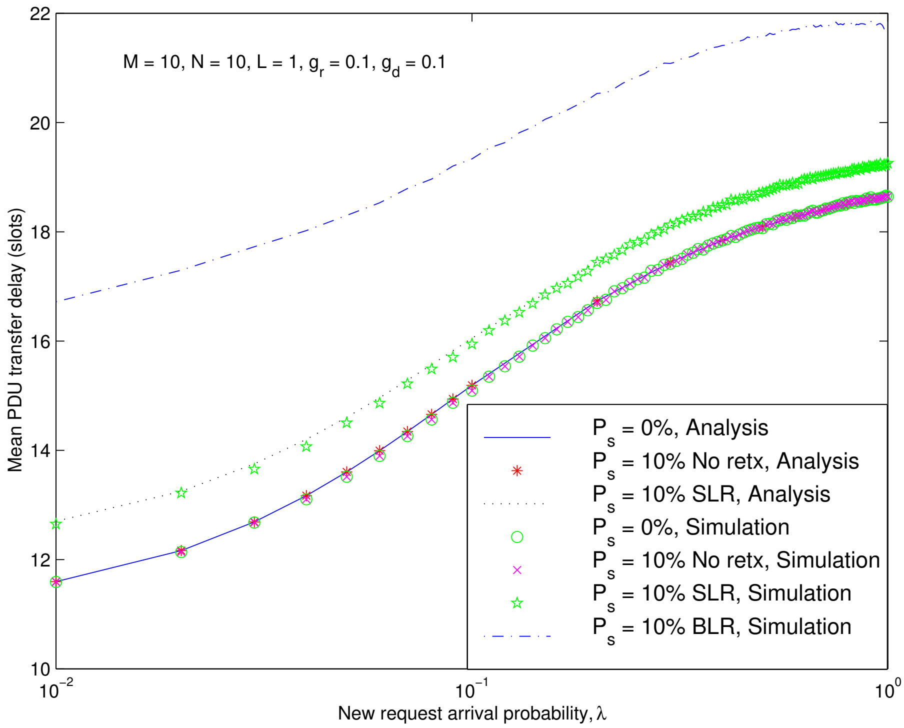


Figure 3.6: Mean PDU transfer delay (in number of slots) versus new request arrival probability, λ. $N = 10$, $M = 10$, $L = 1$, $g_r = 0.1$, $g_d = 0.1$.

Note that the normalized average per channel throughput in Figure 3.5 gives the average number of successful slots and the average number of successful blocks for slot level and block retransmissions, respectively. In order to convert this normalized throughput to an equivalent data rate in Kbps, the number of information and control bits (including BCS and SCS bits) defined in the various coding schemes CS-1, CS-2, CS-3, CS-4 needs to be taken into account. For example, when CS-1 is used, a normalized per channel throughput of, say, 0.45 corresponds to an effective data rate of 4.07 Kbps (i.e., 0.45×9.05 Kbps) for block level retransmission. For slot level retransmission, assuming 10 bits of slot check sequence (SCS) in each slot, the effective data rate becomes 3.9 Kbps (i.e., 0.45×8.7 Kbps). Note that the maximum information rate of 8.7 kbps in the above is obtained as follows: Each slot carries 114 (rate 1/2) coded bits, so that 40 information bits, 10 SCS bits and 7 control bits constitute one slot before coding. 40 information bits sent every 4.615 ms results in an effective information rate of 8.7 Kbps. It is noted that when the slot error rate is zero (or very low values of P_s) block level retransmission performs marginally better than slot level retransmission (i.e., maximum information rate of 9.05 Kbps versus 8.7 Kbps), which is due to the slightly increased overheads due to the tail/control bits in slot level retransmission. However, as we will see subsequently, slot level retransmission performs much better than block level retransmission when the channel error rates are high (high values of P_s). The effective data rates for other coding schemes CS-2, CS-3, and CS-4 can be obtained likewise.

¿From Figure 3.5 we observe that the throughput results obtained from Eqn. (3.7) closely match with the results obtained through simulation. This validates our analysis in Section 3.3. Also, it is observed that for low values of new request arrival probability, λ, the system spends most time in the $idle$ state, resulting in very low throughputs. As λ increases, the fraction of time the system spends in $idle$ state decreases, and this results in increased throughput. When there are no slot errors (i.e., $P_s = 0\%$), all the slots carrying data traffic are successful, which represents the best possible performance. For example, for $\lambda = 0.01$ the per channel throughput achieved is 0.09 (i.e., an effective data rate of 666 bps using coding scheme CS-1), and for λ values closer to unity the per channel throughput increases to 0.54 (4 Kbps using CS-1).

It is noted that the 'no retx' plot in Figure 3.5 corresponds to MAC with RLC in the unacknowledged mode of operation. For high slot error rates (say, $P_s = 10\%$), when there are no RLC retransmissions, the fraction of successful slots decreases, and hence the throughput decreases. On the other hand, for the same slot error rate (of 10%), slot level retransmission at the RLC improves the throughput performance. This is because, the fraction of time the channel is left $idle$ is reduced due to the retransmission attempts, and as long as the slot error rate is reasonably good, this would result in increased throughput. Another observation in Figure 3.5 is that the throughput achieved with block level retransmission at the RLC is much lower than the slot level retransmission. This is because, in block level retransmission, even if one slot in a block is in error, the entire block (of 4 slots) will be retransmitted, and this considerably reduces the throughput. For example, for high arrival rates ($\lambda = 1$) the per channel throughput achieved using block level retransmission is 0.46 (effective data rate of 4.16 Kbps using CS-1), whereas slot level retransmission achieves a per channel throughput of 0.52 (effective data rate of 4.52 Kbps using CS-1). The effective data rates for CS-2, CS-3, CS-4 for the above per channel throughput values are given in Table 3.1.

¿From Table 3.1, it is observed that gains in throughput are possible with slot level retransmission compared to block level retransmission.

Coding scheme	FEC	BLR (Kbps)	SLR (Kbps)
CS-1	1/2	4.16	4.52
CS-2	2/3	6.16	6.97
CS-3	3/4	7.17	8.06
CS-4	1(No FEC)	9.84	11.02

Table 3.1: Effective data rates (in Kbps) for various coding schemes. $\lambda = 1$. (Ref: Fig. 3.5).

The mean PDU transfer delay performance of the GPRS RLC/MAC protocol is evaluated using Eqn. (3.10) for the same set of parameters used in Figure 3.5. The mean PDU transfer delay in number of slots is plotted in Figure 3.6 as a function of λ. In the case of 'no retx' (i.e., RLC unacknowledged mode), it takes the same number of slots to carry the traffic as in the no error case. Hence, the delay is the same for both no error case as well as error case with no retransmission. The delay for slot level retransmission increases as it takes more slots to successfully deliver the data slots. In block level transmission, since the entire block gets retransmitted even if one slot in a block is in error, the delay performance is worse than slot level retransmission. For example, for λ values near unity the mean delay is 19 slots for slot level retransmission and 23 slots for block level retransmission.

The effect of the number of channels, M, on the throughput characteristics of the GPRS RLC/MAC protocol is shown in the Figure 3.7, for N=10, L = 1, g_r = 0.1, g_d = 0.1, and for λ = 1. The total number of channels, M, is varied in the range 2 to 10. ¿From Figure 3.7, we observe the following. The per channel throughput increases as M increases, up to a certain a value of M, beyond which the throughput decreases. This is because, at low values of M, requests are $backlogged$ due to unavailability of PDTCHs, whereas at high values of M, PDTCHs are $idle$ most of the time. More interestingly, around the optimum value of M for the chosen set of parameters and traffic load, slot level retransmission significantly outperforms block level retransmission. For example, when $M = 5$, block level retransmission gives a per channel throughput of 0.47, whereas slot level retransmission gives 0.64. These normalized per channel throughputs converted to effective data rates (in Kbps) for different coding schemes are given in the Table 3.2. The mean PDU transfer delay performance for the same set of parameters used in Figure 3.7 is illustrated in Figure 3.8.

¿From Figures 3.5 to 3.8 and Tables 3.1 and 3.2, we observe that significant improvement

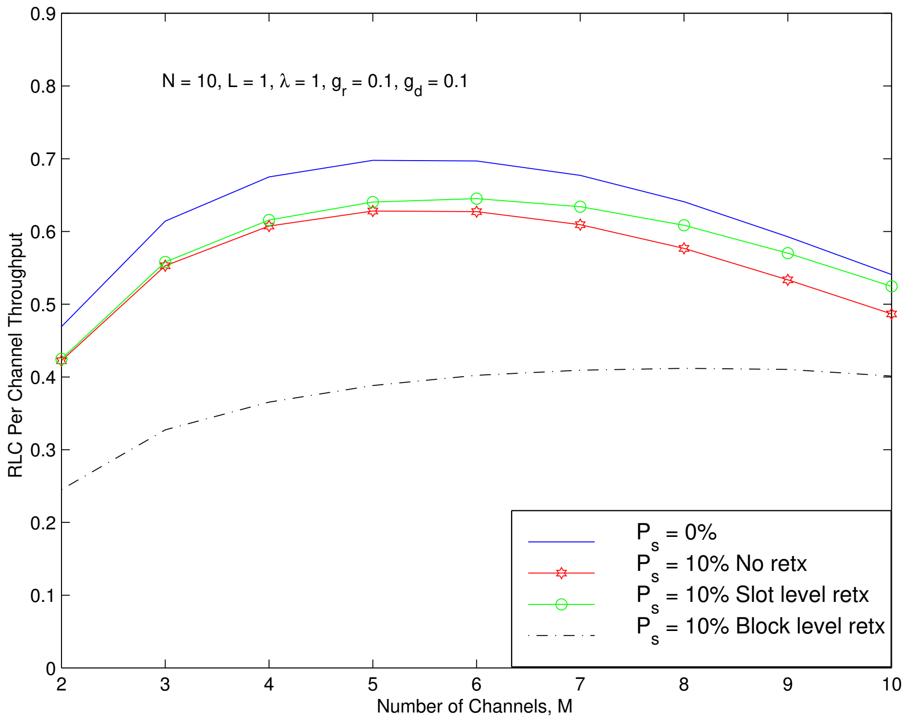


Figure 3.7: RLC/MAC average per channel throughput, η_c, versus number of channels, M. $N = 10$, $L = 1$, $\lambda = 1$, $g_r = 0.1$, $g_d = 0.1$.

in throughput and delay performance is possible using slot level retransmission instead of block level retransmission.

The effect of slot error rate on the per channel throughput and delay performance is shown in Figures 3.9 and 3.10, for N=10, L = 1, g_r = 0.1, g_d = 0.1, and λ = 1. The plot shows the performance of RLC in unacknowledged mode ('no retx') as well as the RLC in acknowledged mode with slot level and block level retransmissions. As expected, the per channel throughput decreases as slot error rate increases. As mentioned before, when the slot error rate is small (good channel

Coding scheme	FEC	BLR (Kbps)	SLR (Kbps)
CS-1	1/2	4.25	5.57
CS-2	2/3	6.30	8.57
CS-3	3/4	7.33	9.92
CS-4	1(No FEC)	10.06	13.57

Table 3.2: Effective data rates for various coding schemes. $M = 5$, $P_s = 10\%$. (Ref: Fig. 3.7).

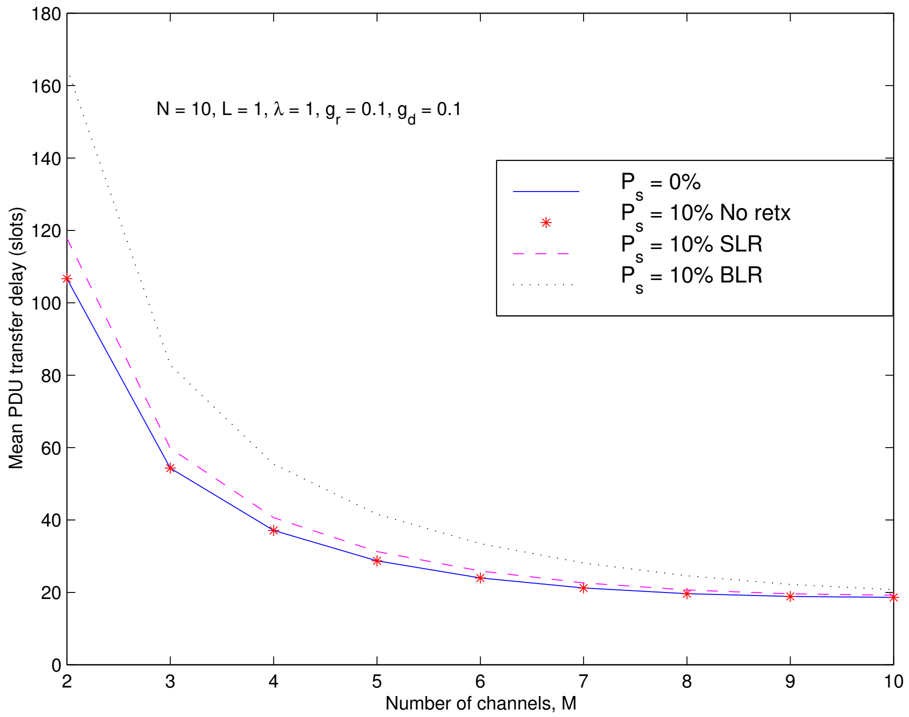


Figure 3.8: Mean PDU transfer delay versus number of channels, M. $N = 10$, $L = 1$, $\lambda = 1$, $g_r = 0.1$, $g_d = 0.1$.

condition), block level retransmission marginally performs better than slot level retransmission due to more number of overhead bits in SLR. However, as the slot error rate increases, block level retransmission performs poorer than even the unacknowledged mode. This is because even if one slot goes into error all the four slots of a block are retransmitted. Also, slot level retransmission is found to significantly outperform both unacknowledged mode and block level retransmission, particularly when the slot error rate is high.

3.5 LLC Layer Performance

We evaluate the throughput of the LLC layer with the RLC and MAC layers below, through simulations. We consider the LLC acknowledged mode of operation. Of particular interest here is the effect and the optimum choice of the maximum LLC retransmission count (R_{llc}) and the maximum RLC retransmission count (R_{rlc}). We use a similar system model as described in Section

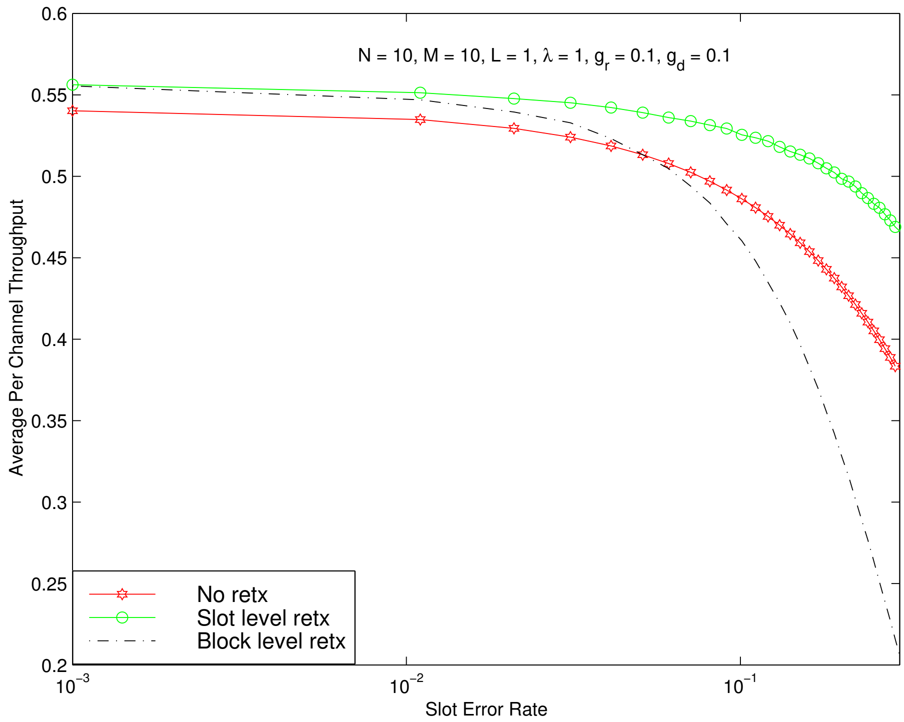


Figure 3.9: RLC/MAC average per channel throughput, η_c, versus slot error rate, P_s. $N = 10$, $M = 10$, $L = 1$, $\lambda = 1$, $g_r = 0.1$, $g_d = 0.1$.

3.2, except that here we assume the message length measured in *number of LLC frames* is geometrically distributed with parameter g_f. We consider selective acknowledgement (SACK) of LLC frames [14]. Also, we assume that each LLC frame consists of 5 RLC blocks (20 slots). If an erroneous LLC frame is not recovered within R_{llc} retransmission attempts, the link layer is reset and re-established. We assume that this reset and re-establishment delay is RESET_DELAY in number of slots.

The per channel throughput at the LLC layer as a function of maximum number of RLC retransmission count R_{rlc} is plotted in Figure 3.11, for $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $g_r = 0.1$, $g_f = 0.2$, and RESET_DELAY = 200 slots. Both block as well as slot level retransmissions at a slot error rate of 10% is considered. The R_{llc} values considered are 1, 2, and 3. We define the throughput at the LLC layer as the average number of successful LLC frames. We observe from Figure 3.11 that, for a fixed R_{llc}, as R_{rlc} increases throughput increases. By increasing R_{rlc} we try to recover erroneous blocks in the RLC layer itself rather than giving up the entire frame (that contains the erroneous blocks) to the LLC. We also observe that as R_{llc} is increased, throughput

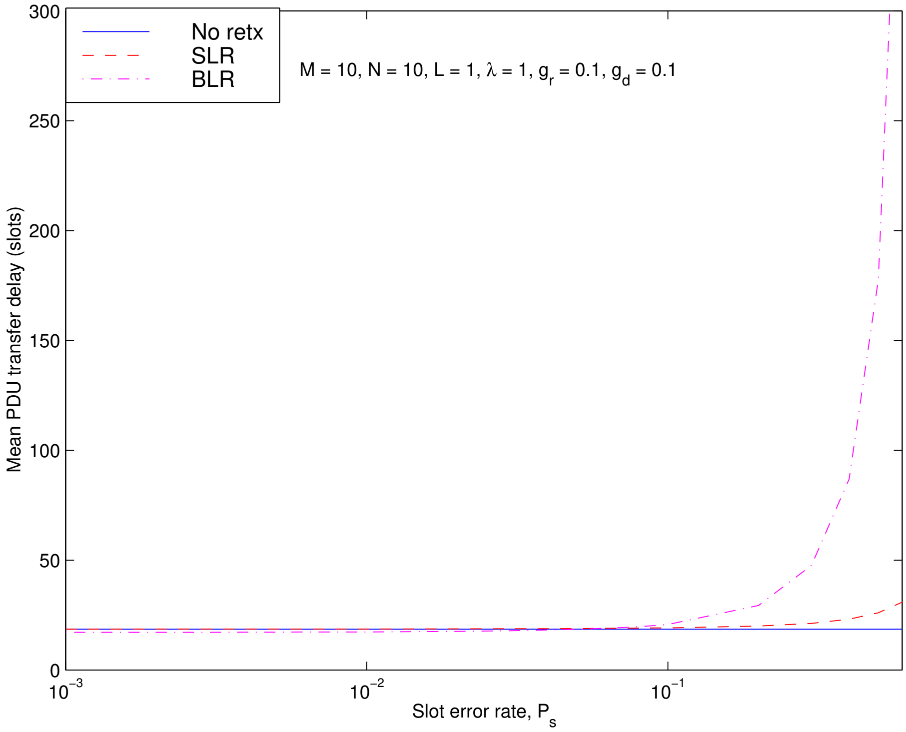


Figure 3.10: Mean PDU transfer delay versus slot error rate, P_s. $N = 10$, $M = 10$, $L = 1$, $\lambda = 1$, $g_r = 0.1$, $g_d = 0.1$.

increases. When R_{llc} is small, LLC *resets* occur frequently. A reset makes the channel to stay idle for RESET_DELAY slots. Hence, the per channel throughput at the LLC is less for small values of R_{llc}. By increasing R_{llc}, we try to avoid too many resets. This gives a better performance in throughput. We also note that slot level retransmission (SLR) performs better than block level retransmission (BLR).

The delay performance at the LLC layer is computed as a function of R_{rlc} and plotted in Figure 3.12, for $M = 10$, $N = 10$, $L = 1$, $\lambda = 0.1$, $g_r = 0.1$, $g_f = 0.2$, $P_s = 10\%$ and RESET_DELAY = 200 slots. ¿From Figure 3.12, we observe that, for a fixed R_{llc}, as R_{rlc} increases delay decreases. Also, for a fixed R_{rlc}, increasing R_{llc} increases throughput and decreases delay as seen in Figures 3.11 and 3.12. If R_{rlc} is too small to recover an erroneous block, then the entire LLC frame that contains the erroneous block is retransmitted. When R_{rlc} and R_{llc} are both small, then LLC resets will occur frequently, which will increase the PDU transfer delay. Thus, for small R_{rlc}, the delay performance improves by increasing R_{llc}. When R_{rlc} is large enough (e.g., > 10), the delay performance for R_{llc} = 2 or 3 is approximately the same. This is because almost all the

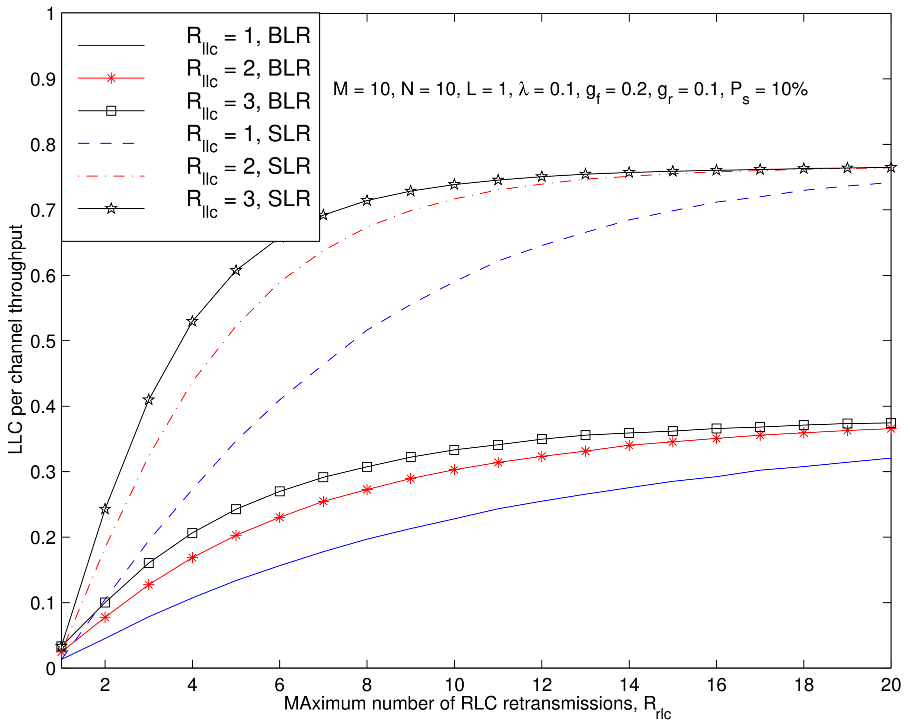


Figure 3.11: LLC average per channel throughput versus maximum number of RLC retransmissions, R_{rlc}. $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $P_s = 10\%$, $g_r = 0.1$, $g_f = 0.2$.

erronoeus RLC blocks are recovered in the RLC layer itself and there may not be much need for LLC retransmission.

The effect of slot error rate on the per channel throughput and delay performance at the LLC is shown in Figures 3.13 and 3.14, for M=10, N=10, L = 1, λ=0.1, g_r = 0.1, g_f = 0.2, and for various combinations of R_{llc} and R_{rlc}. We consider the block level retransmission here. In Figures 3.13 and 3.14, ($R_{llc} \to \infty$, $R_{rlc} = 1$) corresponds to the case where the stack has no RLC and has a persist-until-success LLC, and $R_{rlc} \to \infty$ corresponds to persist-until-success RLC and no LLC. Note that the absence of RLC ($R_{rlc} = 1$) gives the worst case performance even if R_{llc} is taken to ∞. Also, complete recovery at the RLC itself ($R_{rlc} \to \infty$) gives the best performance. This indicates a larger value of R_{rlc} than R_{llc} is beneficial in terms of performance.

A comparison of slot level retransmission and block level retransmission for LLC throughput and delay performance is shown in Figures 3.15 and 3.16 for $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $g_r = 0.1$, $g_f = 0.2$, and RESET_DELAY = 200 slots. Figure 3.15 shows the throughput performance for $R_{llc} \to \infty$ and for R_{rlc} taking values one and ∞ for both BLR and SLR. Note that

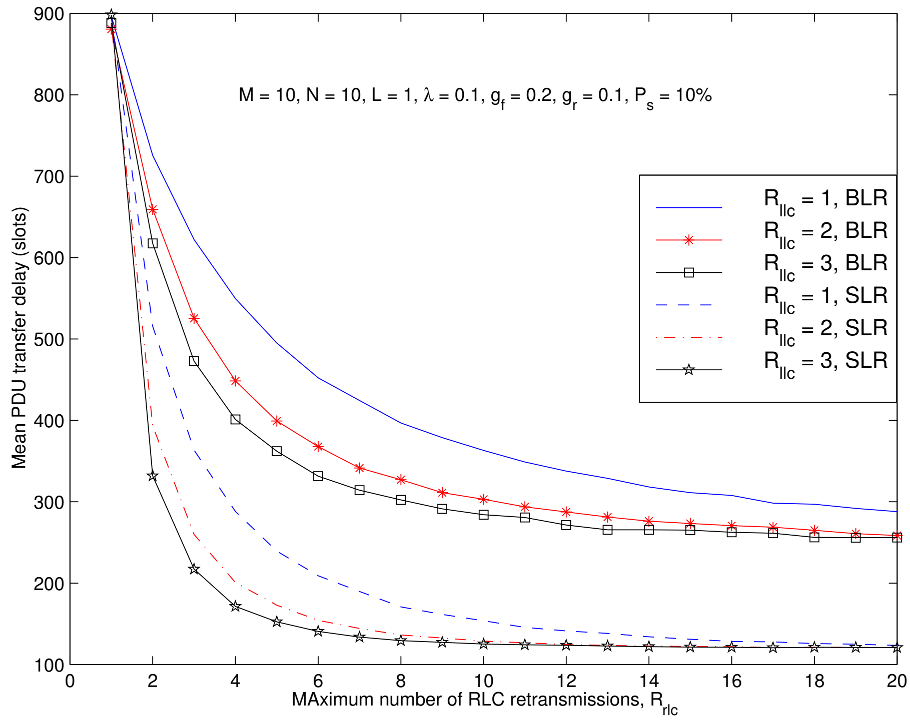


Figure 3.12: LLC mean PDU transfer delay versus maximum number of RLC retransmissions, R_{rlc}. $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $P_s = 10\%$, $g_r = 0.1$, $g_f = 0.2$.

$R_{rlc} = 1$ and $R_{llc} \to \infty$ corresponds to no error recovery at RLC at complete recovery at LLC. Also $R_{rlc} \to \infty$ corresponds to complete recovery at RLC (and no recovery at LLC). ¿From Figures 3.15 and 3.16 we can infer that slot level retransmission performs significantly better than block level retransmission in terms of both throughput and delay.

We can compute the effective data rate at LLC in Kbps from the normalized LLC throughput, using the relation

$$\eta_{c,llc}(\text{Kbps}) = \eta_{c,llc} \cdot \frac{n_{inf} * FL - OH}{FL * BL * 4.615} \text{ Kbps},$$

where

$\eta_{c,llc}$ represents the normalized per channel throughput at the LLC layer,

$\eta_{c,llc}$(Kbps) represents the effective data rate in Kbps at the LLC layer,

n_{inf} represents the number of information bits in each RLC block including the LLC header and checksum,

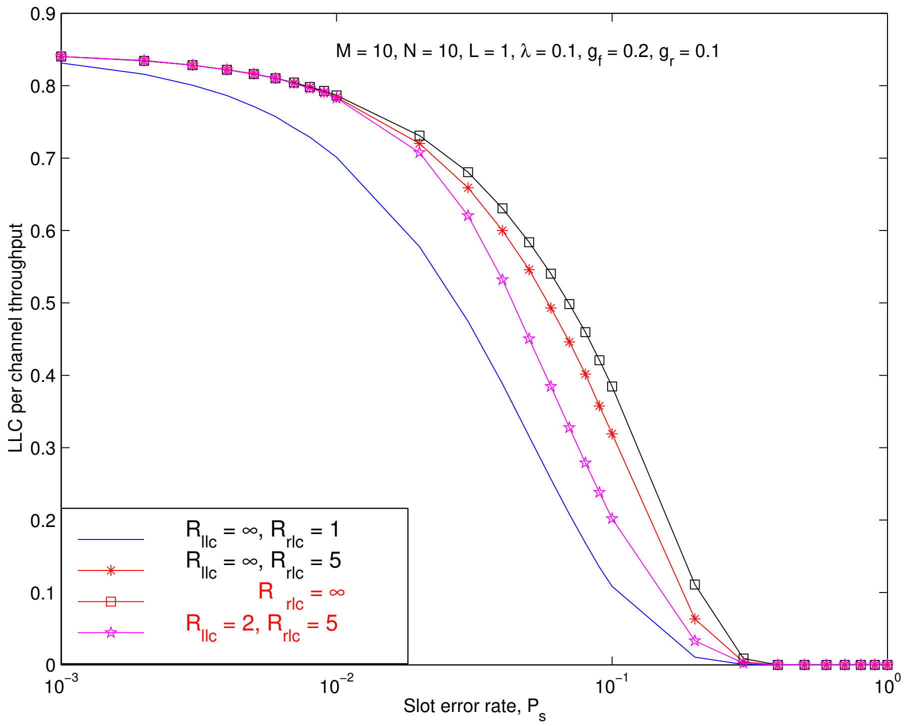


Figure 3.13: LLC average per channel throughput versus slot error rate, P_s. $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $g_r = 0.1$, $g_f = 0.2$. BLR at RLC.

FL represents the LLC frame length in terms of number of RLC blocks,

BL represents the RLC block length in terms of number of GSM slots, and

OH represents the number of overhead bits per LLC frame.

Note that the 4.615 in the denominator of the above expression accounts for the one slot duration which is equivalent to one TDMA frame length of 4.615 ms. For the coding scheme CS-1, n_{inf} = 181 bits for block level retransmission and n_{inf} = 40 bits for slot level retransmission. In our simulations, we set FL = 5 and BL = 4 for block level retransmission and FL = 20 and BL = 1 for slot level retransmission. The overhead bits, $OH = 56$ corresponds to LLC header and checksum bits.

Note that $R_{rlc} \to \infty$ in Figures 3.13, 3.15 corresponds to full recovery in the RLC layer. This means that the normalized per channel throughput is the same at both RLC and LLC layers, but the effective data rates in Kbps are different. For slot level retransmission (see Fig. 3.15), the normalized per channel throughput at $P_s = 10\%$ is 0.77, which corresponds to an effective

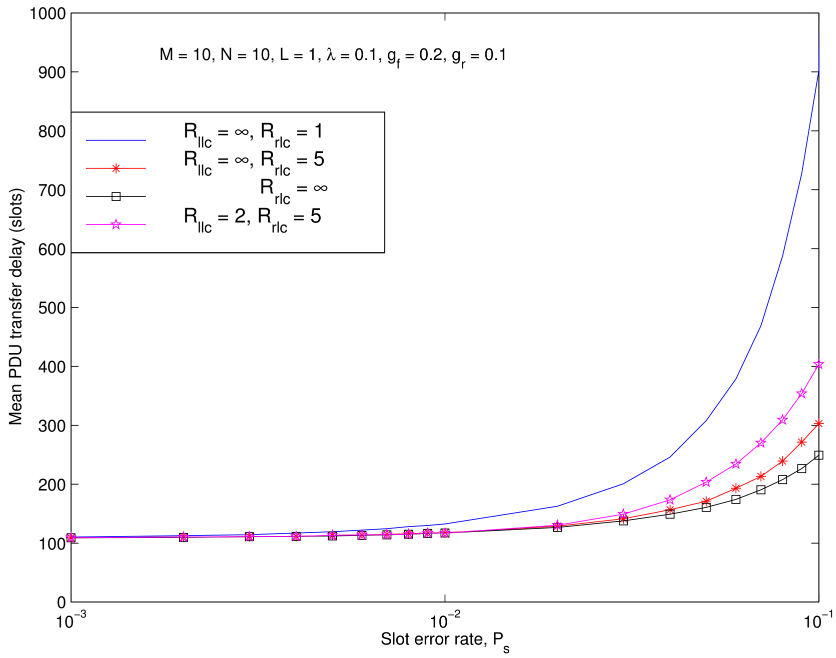


Figure 3.14: Mean PDU transfer delay versus slot error rate, P_s. $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $g_r = 0.1$, $g_f = 0.2$. BLR at RLC.

data rate of 6.19 Kbps at the LLC layer. The corresponding effective data rate at the RLC layer is given by 6.68 Kbps. This reduced effective data rate at the LLC compared to RLC is due to the LLC overheads due to checksum and header. For block level retransmission (see Fig. 3.13), the normalized per channel throughput for the same system parameters is 0.38. This corresponds to an effective data rate of 3.42 Kbps at the LLC layer and 3.44 kbps at the RLC layer. When $R_{rlc} = 1$ and $R_{llc} \to \infty$, and for the same system parameters with block level retransmission, the normalized per channel throughput at the LLC layer is 0.1085. This corresponds to an effective data rate of 0.99 Kbps at the LLC layer.

We summarize the points to be noted for better channel utilization.

- Slot level retransmission is preferable than block level retransmission.
- Large values of R_{rlc} (e.g., $R_{rlc} > 10$) can be used.
- For a particular R_{rlc}, increasing R_{llc} can result in some marginal improvement in performance. For example, taking a R_{rlc} value larger than 10, R_{llc} value of 2 or 3 is adequate to

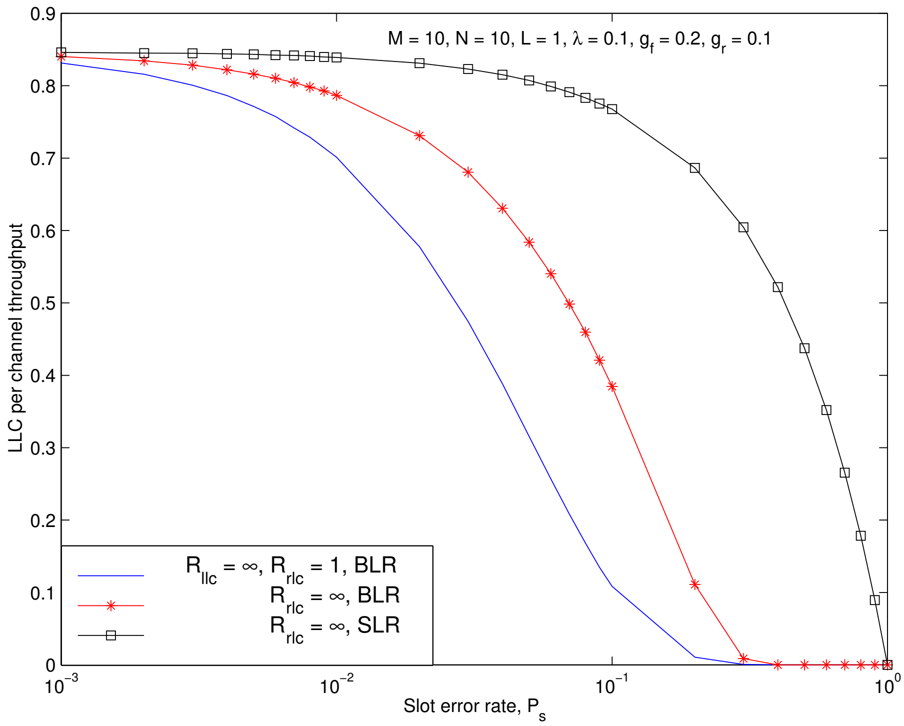


Figure 3.15: LLC average per channel throughput versus slot error rate, P_s. $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $g_r = 0.1$, $g_f = 0.2$.

achieve a good performance.

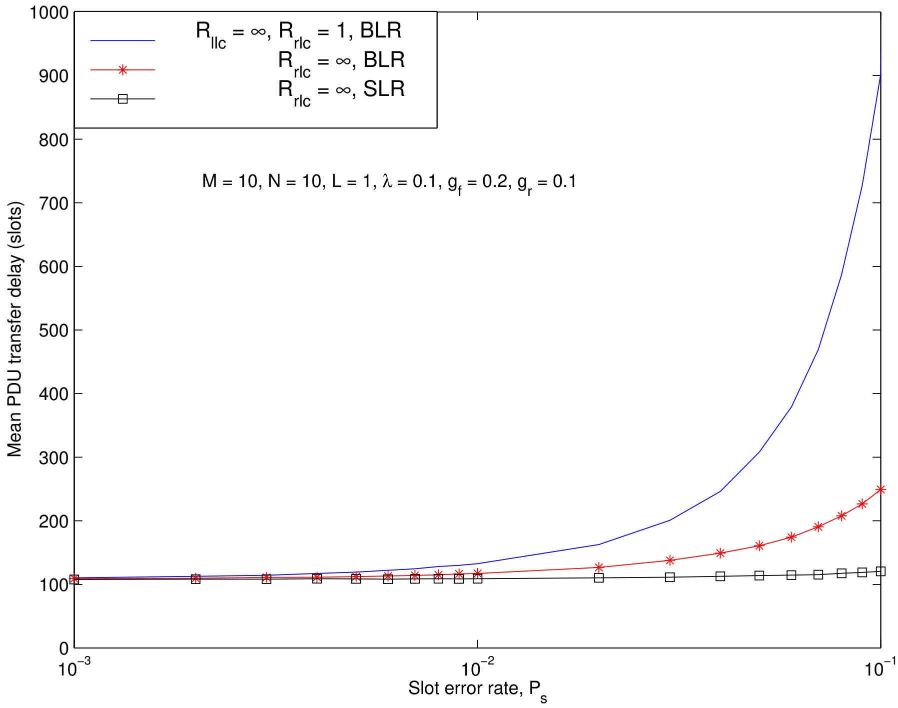


Figure 3.16: Mean PDU transfer delay versus slot error rate, P_s. $N = 10$, $M = 10$, $L = 1$, $\lambda = 0.1$, $g_r = 0.1$, $g_f = 0.2$.

Chapter 4

TCP Performance in GPRS

Transmission Control Protocol (TCP) is a well known transport layer protocol in the Internet. TCP is a reliable, connection-oriented protocol which is widely used in popular applications like http, ftp, telnet, etc. Several studies have analyzed the performance of TCP on wireless, but without considering ARQ in the link layer [42],[43],[44],[45]. In this Chapter, we estimate throughput performance of TCP on the uplink in GPRS with the associated LLC/RLC/MAC layers.

Being a connection oriented protocol, TCP has a call setup phase, data transfer phase, and a call clear phase. The bulk throughput performance of TCP is determined primarily by the data transfer phase. We, in this study, are interested in evaluating the TCP throughput in the data transfer phase. In Section 4.1, we present a brief description of the basic data transfer phase of TCP. In Section 4.2, we present the simulation model considered. We consider both the RLC block level retransmission (as defined in current GPRS standards) and the RLC slot level retransmission we propose. In Section 4.3, we present the results and discussions on the TCP performance on GPRS.

4.1 The TCP Protocol

Several studies have modeled the data transfer part of TCP [44], [45]. The following description follows from [44].

TCP receiver: The TCP receiver has a finite resequencing buffer. It advertises a maximum window size, W_{max} equal to the buffer size at connection setup time, and the transmitter ensures that the amount of unacknowledged data is never more than this value. The receiver accepts packets out of sequence number, buffers them in the TCP buffer, and delivers them to its TCP user in sequence. The receiver returns an *acknowledgement (ack)* for every good packet that it receives. An ack packet that acknowledges the first receipt of an error-free, in-sequence packet will be called a *first ack*. The acks are *cumulative*, i.e., an ack carrying the sequence number n acknowledges all data up to, and including, the sequence number $n - 1$. If there is data in the resequencing buffer, the acks from the receiver will carry the *next expected* packet number, which is the first among the packets required to complete the sequence of packets in the sequencing buffer. Thus, if a packet is lost (after a long sequence of good packets), then the transmitter keeps getting acks with the sequence number of the first packet lost, if some packets transmitted after the lost packet do succeed in reaching the receiver. These are called *duplicate acks*.

TCP transmitter: At all times t the TCP transmitter maintains the following variables for each connection:

$A(t)$ = the *lower window edge*; all data numbered upto and including $A(t)-1$ has been transmitted and acked. $A(t)$ is nondecreasing; the receipt of an ack with sequence number $n > A(t)$ causes $A(t)$ to jump to n.

$W(t)$ = the *congestion window*. The transmitter can send packets with the sequence numbers $n, A(t) \leq n < A(t) + W(t)$. $W(t) \leq W_{max}$; $W(t)$ increases or decreases as described below.

$W_{th}(t)$ = the *slow-start threshold*; $W_{th}(t)$ controls the increments in $W(t)$ as described below.

Retransmission time-out: The transmitter measures the round-trip times of *some* of the packets that it has transmitted and received acknowledgements for. These measurements are used to obtain a running estimate of the packet *round-trip time (rtt)* on the connection. Each time a new packet is transmitted, the transmitter starts a timer and resets the already running timer, if any; i.e., there is a timeout only for the last transmitted packet. The timer is set for a *retransmission* *time-out (rto)* value that is derived from the rtt estimation procedure. The TCP transmitter process measures time and sets timeouts only in multiples of a *timer granularity*; for example, BSD based systems have a timer granularity of 500ms. Further, there is a minimum timeout duration in most implementations; e.g., 2 timer ticks in BSD, implying an average minimum timeout of 750ms.

Window adaptation and time-out based recovery: The basic algorithm is common to all TCP versions, and was originally developed by Van Jacobson. The normal evolution of the processes $A(t), W(t)$, and $W_{th}(t)$ is triggered by first acks (see definition above) and timeouts as follows.

1. *Slow start:* If $W(t) < W_{th}(t)$, each first ack causes $W(t)$ to be incremented by 1.

2. *Congestion Avoidance:* If $W(t) \geq W_{th}(t)$, each first ack causes $W(t)$ to be incremented by $\frac{1}{W(t)}$.

3. *Timeout* at epoch t causes $W(t^+)$ to be set to 1, $W_{th}(t^+)$ to be set to $\lceil \frac{W(t)}{2} \rceil$, and retransmissions to begin from $A(t)$.

Packet loss recovery: $A(t)$ and $W(t)$ will be incremented as and when first acks come. The evolution then stops once a packet goes into error. For a particular loss instance, let the values at the instant a packet is lost be denoted by A and M, respectively; We will call M a *loss window*. Then the transmitter will continue to send packets upto the sequence number $A+M-1$. If some of the packets sent after the lost packet get through, they will result in duplicate acks, all carrying the same sequence number A. The last packet transmitted (i.e., $A+M-1$) will have an rto associated with it.

Fast Retransmit: There is a transmitter parameter K, a small positive integer. Typically $K = 3$. If the transmitter receives the Kth duplicate ack at time t (before the timer expires), then the transmitter behaves as if a timeout has occurred and begins retransmission, with $W(t^+)$ and $W_{th}(t^+)$ as given by the basic algorithm.

4.2 Simulation Model

Internet traffic forms a new class of traffic that is difficult to be described using classical traffic models. The reason for this is a significant probability for very long sessions, very long interarrival times between sessions and packets or very large size. This leads to heavy tailed (long tailed) complementary cumulative distribution functions (CCDF) for these typical measures [46]. Much work is done to find a suitable description of Internet traffic. A common approach is to approximate the empirical heavy tailed CCDFs with *Pareto*, *Weibull*, *hyperexponential* or *power law* distributions. Such a theoretical distribution allows easy implementation of a generator which produces values of the described measure with a similar characteristic. In our simulation to evaluate the performance of TCP over GPRS, we consider each TCP source to generate a traffic pattern where the inter-arrival time between packets follows a Pareto distribution.

The classical Pareto distribution with shape parameter β and location parameter a has the CDF

$$F(x) = P\{X \leq x\} = 1 - (a/x)^{\beta}, a, \beta \geq 0, x \geq a, \tag{4.1}$$

with the corresponding probability density function:

$$f(x) = \beta a^{\beta} x^{-\beta-1}. \tag{4.2}$$

An ON-OFF traffic model, such as the web and e-mail traffic, is considered. The OFF period distribution is modeled to be a Pareto distribution, for the reasons given above. We assume one TCP packet to consist of five LLC frames, and each LLC frame contains 536 bytes. So the TCP packet size is 536x5 bytes. An LLC frame is segmented into 25 RLC blocks, each of size 4 GSM slots (see Figure 3.1 for PDU segmentation). An erroneous RLC block can be retransmitted upto R_{rlc} times by a selective ARQ mechanism in the RLC layer (see Chapter 3 for detailed description). If the RLC layer can not recover the erroneous blocks, they are passed on to the LLC layer. The LLC layer has a selective ARQ mechanism that tries to recover erroneous LLC frames by retransmission upto R_{llc} times. TCP takes care of the unrecovered erroneous frames. We consider 20 TCP sessions each sending TCP packets to some hosts in an external Packet Data Network

(PDN). The simulation programs of all the protocol layers, including TCP layer are written in C. The simulation is carried out for one million slots.

4.3 Results and Discussion

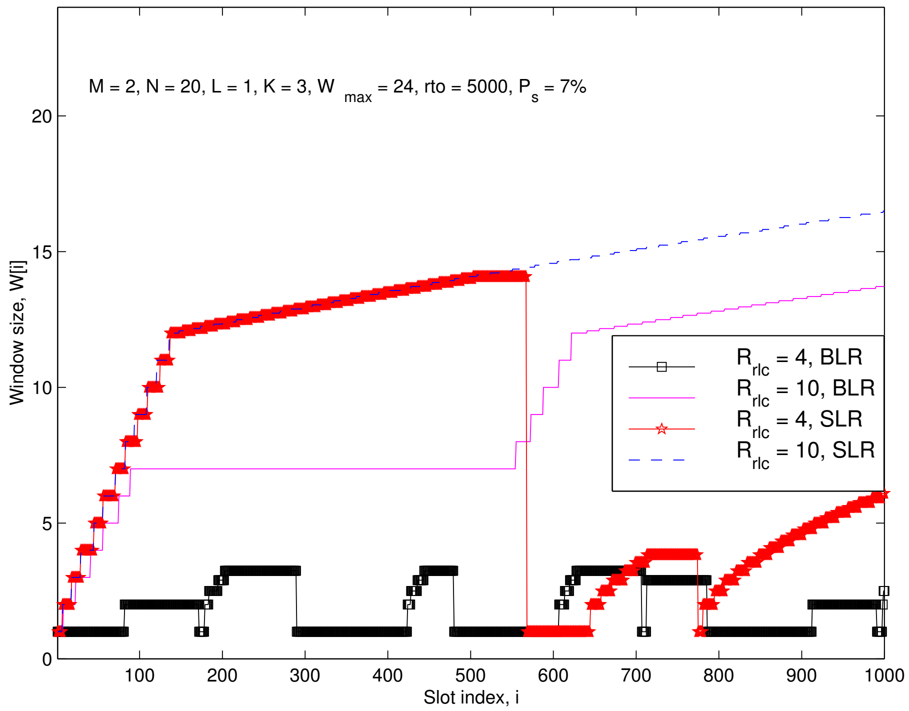


Figure 4.1: Evolution of TCP window size, $W[i]$, versus slot index, i. $M = 2$, $N = 20$, $L = 1$, $R_{llc} = 3$, $K = 3$, $P_s = 7\%$, $W_{max} = 24$, $rto = 5000$.

Figure 4.1 shows the evolution of window size, $W[t]$, as a function of time for a slot error rate (P_s) of 7%, $N = 20$, $L = 1$, $K = 3$, $R_{llc} = 3$, W_{max} of 24 TCP packets, and a round trip timeout value of 5000 slots. The window evolution of four different cases are plotted: $a)$ BLR with $R_{rlc} = 4$, $b)$ BLR with $R_{rlc} = 10$, $c)$ SLR with $R_{rlc} = 4$, and $d)$ SLR with $R_{rlc} = 10$. In Figure 4.1, a comparison between $R_{rlc} = 4$ versus $R_{rlc} = 10$ for both BLR and SLR indicates that the window size is more open for $R_{rlc} = 10$ than $R_{rlc} = 4$. This is because for $R_{rlc} = 4$ the recovery of erroneous blocks can be incomplete and this can result in more TCP timeouts and fast retransmits, which shrinks the window size to 1. Since larger instantaneous window sizes are good for achieving high throughput, the choice of parameter value $R_{rlc} = 10$ is preferred over $R_{rlc} = 4$.

Also a comparison between BLR and SLR for $R_{rlc} = 10$ reveals that SLR results in a significantly better performance at the TCP layer. The window evolution behaviour in Figure 4.1 results in a TCP throughput performance shown in Figure 4.2.

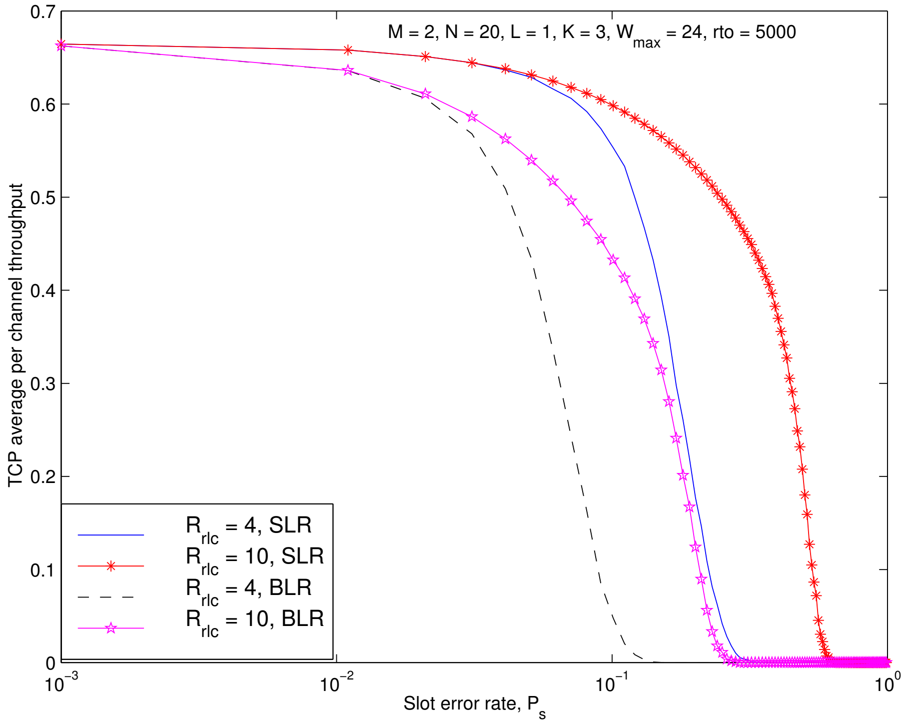


Figure 4.2: TCP throughput performance as a function of P_s and R_{rlc}. $M = 2$, $N = 20$, $L = 1$, $R_{llc} = 3$, $K = 3$, $W_{max} = 24$, $rto = 5000$.

Figure 4.2 shows the throughput performance of TCP as a function of slot error probability, P_s, for different values of R_{rlc}. The following system parameters are taken: $M = 8$ PDTCHs, $N = 20$ users, $L = 1$ PRACH, $R_{llc} = 3$, fast retransmit parameter (K) = 3, maximum advertised window size (W_{max}) = 24 TCP packets, and a round trip timeout value of 5000 slots. We observe that as R_{rlc} is increased, throughput increases, as expected. When R_{rlc} is large, more erroneous blocks are retransmitted and recovered in the RLC layer itself rather than leaving them to LLC or TCP to recover by retransmitting the LLC frame or entire TCP packet. We also observe that the RLC slot level retransmission offers a better throughput than the block level retransmission.

We can compute the effective TCP data rate in Kbps from the TCP normalized throughput, using the relation

$$\eta_{c,TCP}(\text{Kbps}) = \eta_{c,TCP} \cdot \frac{n_{inf} * PL * FL - OH}{PL * FL * BL * 4.615} \text{ Kbps},$$

where

$\eta_{c,TCP}$(Kbps) represents the effective data rate at the TCP layer in Kbps,

$\eta_{c,TCP}$ represents the normalized per channel throughput at the TCP layer,

n_{inf} represents the number of information bits in each RLC block including the IP/LLC headers and checksum,

PL represents the TCP packet length in terms of number of LLC frames.

FL represents the LLC frame length in terms of number of RLC blocks,

BL represents the RLC block length in terms of number of GSM slots, and

OH represents the total number of overhead bits per TCP packet (IP header, LLC header/checksum).

Note that 4.615 in the denominator of the above expression accounts for the one slot duration which is equivalent to one TDMA frame length of 4.615 ms. For coding scheme CS-1, n_{inf} = 181 for block level retransmission and n_{inf} = 40 slot level retransmission. In our simulations, we set PL = 5, FL = 25, and BL = 4 for block level retransmission and PL = 5, FL = 100, and BL = 1 for slot level retransmission. The number of OH bits per TCP packet is 376 (320 IP header bits + 56 LLC overhead bits).

In Figure 4.2, for $P_s = 20\%$ the normalized per channel throughput at the TCP layer is 0.1021 for block level retransmission which corresponds to an effective date rate of 0.93 Kbps at the TCP layer. For the same system parameters, for slot level retransmission, the normalized per channel throughput at the TCP layer is 0.7087 which corresponds to an effective date rate of 6.02 kbps at the TCP layer. Thus, there is a substantial improvement in the effective data rate as we go from block level to slot level retransmission at the RLC.

Figure 4.3 shows the throughput performance for different values of M, for $R_{rlc} = 10$, $N = 20$, $L = 1$, $K = 3$, $W_{max} = 24$, $R_{llc} = 3$, and a round trip timeout value of 5000 slots. We find that as M is increased throughput increases. If the number of PDTCHs, M is less, a successful request may not get an available PDTCH and will be backlogged. As M is increased the successful

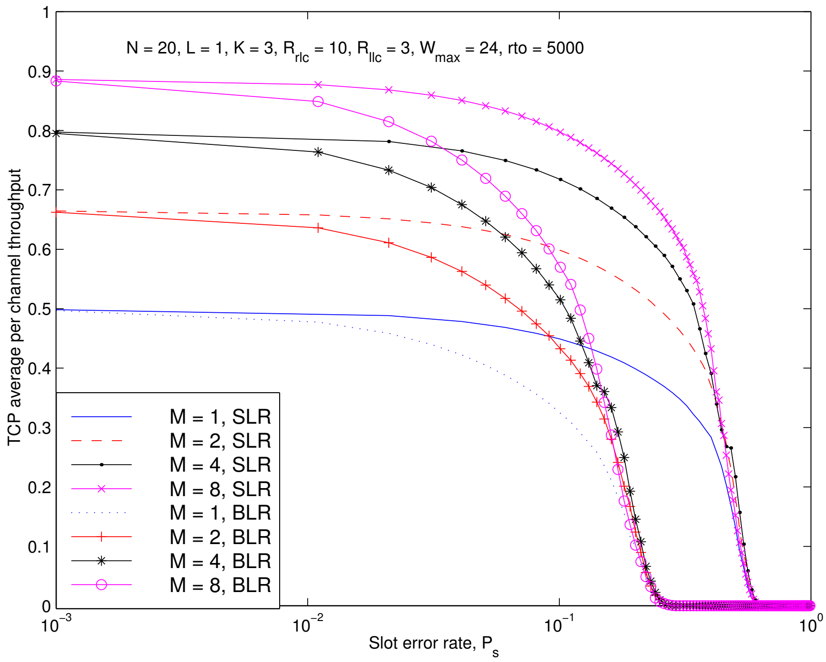


Figure 4.3: TCP throughput performance as a function of P_s and M. $N = 20$, $L = 1$, $R_{rlc} = 10$, $R_{llc} = 3$, $K = 3$, $W_{max} = 24$, $rto = 5000$.

requests will see an available PDTCH with high probability and hence the throughput increases as M increases. For a given arrival rate, if M is increased beyond a certain value then the throughput decreases. This is because, the channels remain idle most of the time. This behaviour is illustrated in Figure 4.4.

Figure 4.5 gives the throughput as a function of fast retransmit parameter, K, for $M = 8$ PDTCHs, $N = 20$ users, $L = 1$ PRACHs, maximum advertised window size (W_{max}) of 24 packets, and a round trip timeout value of 5000 slots. We observe a small performance improvement as K is decreased from 3 to 1. This is because, if the error rate is high or errors are bursty, then the receiver may not get $K = 3$ *duplicate acks* before TCP timeout. The receiver in that case *times out* and initiates the *error recovery* mechanism. Since the channel remains idle from the point where the packet in the upper edge of the TCP transmitter window is sent to the point where timeout occurs, the throughput is less. By reducing K we try to initiate *error recovery* sooner. The reduction of fast retransmit threshold is expected to improve significantly performance when the errors are bursty [45].

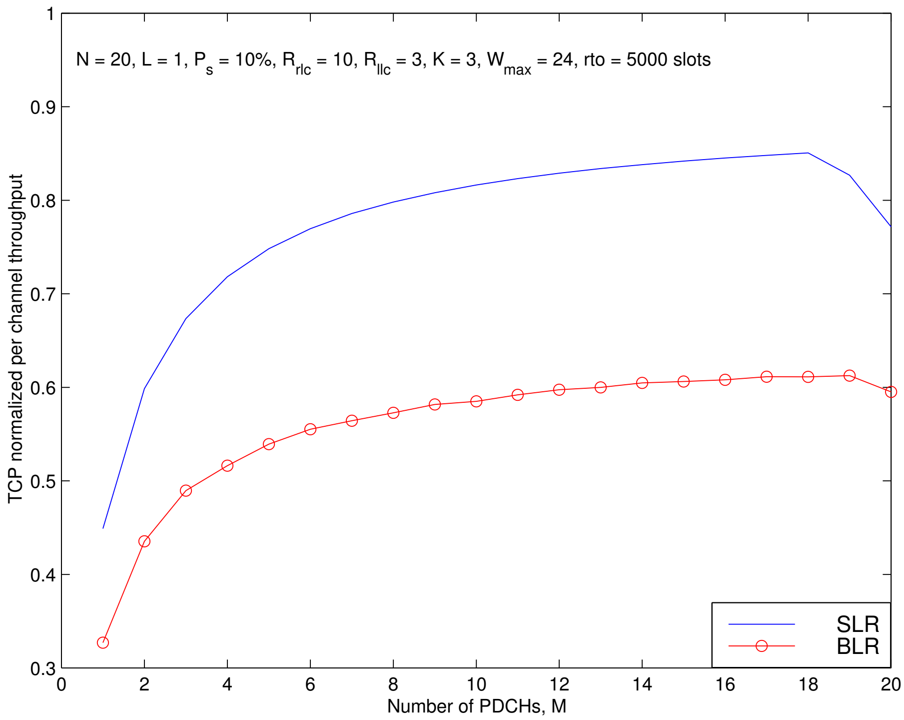


Figure 4.4: TCP throughput performance as a function of M. $N = 20$, $L = 1$, $R_{rlc} = 10$, $P_s = 10\%$, $R_{llc} = 3$, $K = 3$, $W_{max} = 24$, $rto = 5000$.

Figures 4.5, 4.6 show that as R_{rlc} is increased the TCP performance becomes independent of the value of K (i.e., almost same performance for K = 1,2,3). This is because, as R_{rlc} increases, the erroneous blocks are recovered in the RLC layer itself. Hence, the effective packet error rate as seen by the TCP receiver decreases, and the chances of a fast retransmiti, whether K = 3 or 1, is high.

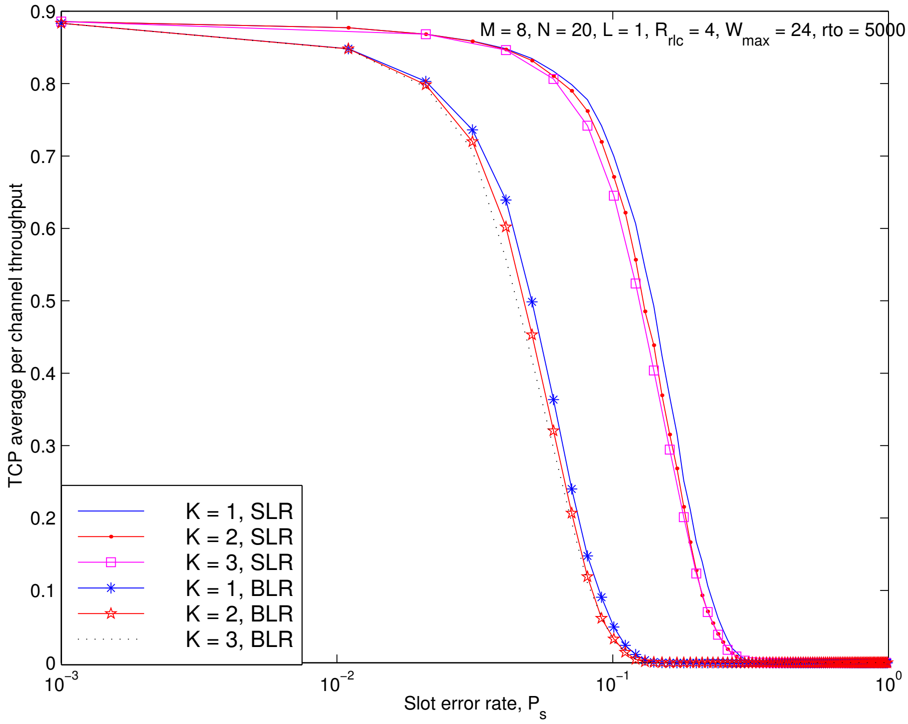


Figure 4.5: TCP throughput performance as a function of P_s and K. $M = 8$, $N = 20$, $L = 1$, $R_{rlc} = 4$, $R_{llc} = 3$, $W_{max} = 24$, $rto = 5000$.

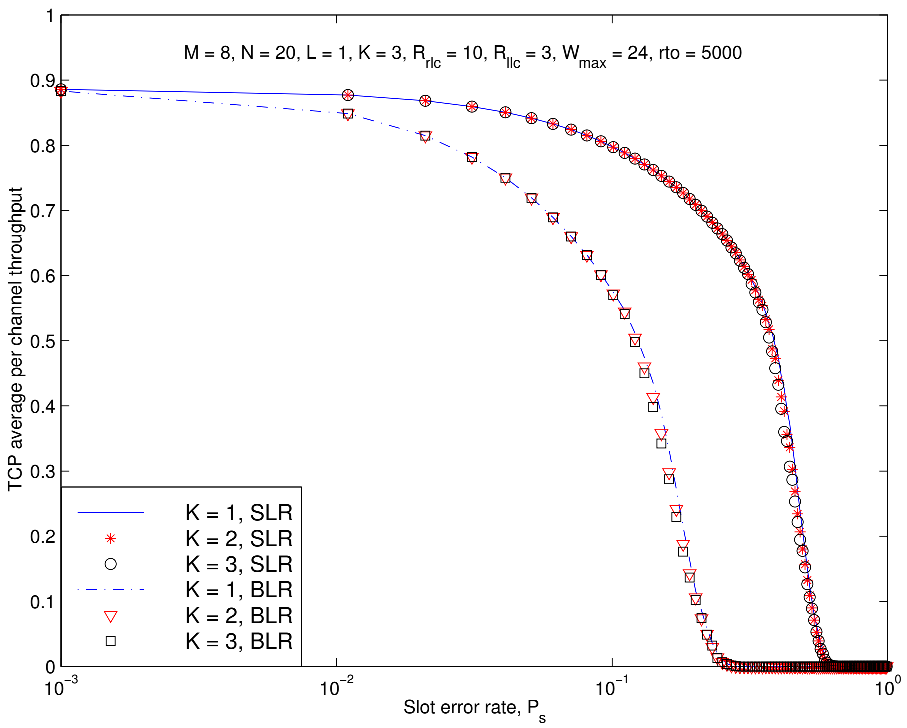


Figure 4.6: TCP throughput performance as a function of P_s and K. $M = 8$, $N = 20$, $L = 1$, $R_{rlc} = 10$, $R_{llc} = 3$, $W_{max} = 24$, $rto = 5000$.

Chapter 5

Conclusion

5.1 Summary

In this thesis, we analyzed the performance of various layers (TCP/IP/LLC/RLC/MAC) of the GPRS protocol stack. Using theory of Markov chains, we derived the throughput and delay performance of the GPRS MAC protocol, considering the uplink request-reservation mechanism. We proposed a slot level retransmission scheme in the RLC layer and compared its performance with that of block level retransmission as defined in the current GPRS standards. We showed that slot level retransmission can give substantial performance gains compared to block level retransmission, particularly when the channel error rate is high. We studied the impact of LLC ARQ on the throughput-delay performance. Specifically, we addressed the question of the choice of number of retransmission attempts at the RLC and at the LLC layers. We also studied the performance of TCP on the uplink in GPRS with both BLR and SLR at the RLC layer.

The following observations were made in the study:

- Slot level retransmission performs better than block level retransmission both in terms of throughput and delay.
- For a better utilization of the channel, maximum number of retransmissions at RLC (R_{rlc}) can be as large as possible.

- For a large enough R_{rlc} (> 10), maximum number of retransmissions at LLC, R_{llc}, of 2 or 3 is adequate to achieve a good performance.

- Substantial increase in TCP throughput is possible by increasing the maximum allowable number of retransmissions (R_{rlc}) at the RLC layer. Given a large value of R_{rlc}, the fast retransmit parameter, K, does not have much impact on TCP throughput performance.

5.2 Scope of Future Work

There are further extensions possible to this work.

- We analyzed the GPRS protocol stack performance for the single slot operation. Extension of the analysis for the multislot operation can be done.

- We considered an i.i.d channel model, in this study. The evaluation of performance in a correlated fading channel will be useful.

- We analyzed the performance for fixed resource allocation. This work can be extended to include dynamic allocation and extended dynamic allocation.

Chapter 6

List of Publications Accepted/Submitted out of this thesis

1. K. Premkumar and A. Chockalingam, "Performance Analysis of RLC/MAC Protocol in General Packet Radio Service," *Proc.* 7^{th} *NCC-2001,* IIT Kanpur, pp. 173–177, January 2001.

2. K. Premkumar and A. Chockalingam, "Performance of LLC and TCP on GPRS Uplink with RLC Slot Level Retransmission," being submitted to 8^{th} *NCC-2002, IIT Bombay*.

3. K. Premkumar and A. Chockalingam, "Performance Analysis of General Packet Radio Service Protocol Stack," journal paper under preparation.

Appendix A

A.1 Derivation of $\tilde{P}_{z_1,z_2}$

Here, we present the derivation of the transition probability expression (3.1) for $\tilde{P}_{z_1,z_2}$. The notation followed here is the same as in Sec. 3.3.1. Refer Fig. A.1 which illustrates the state transitions from slot t to slot $t+1$.

At the beginning of slot t, the system is in state $\mathbf{z_1} = (\mathbf{i_1}, \mathbf{j_1}, \mathbf{k_1})$, i.e., there are i_1 mobiles in *Data_Tx failure* state, j_1 mobiles in *Data_Tx success* state, k_1 mobiles in *Backlogged* state, and $(N - i_1 - j_1 - k_1)$ mobiles in *Idle* state.

During the slot t, b_i mobiles $(0 \leq b_i \leq i_1)$ out of i_1 mobiles in *Data_Tx failure* state have data to transmit and the remaining $(i_1 - b_i)$ mobiles go to *Idle* state. Out of those b_i mobiles, f_s mobiles, $0 \leq f_s \leq b_i$, go to *Data_Tx success* state, and the remaining $(b_i - f_s)$ mobiles stay in the *Data_Tx failure* state.

Also, b_j mobiles $(0 \leq b_j \leq j_1)$ out of j_1 mobiles in *Data_Tx success* state will continue to transmit and the remaining $(j_1 - b_j)$ mobiles go to *Idle* state. Out of those b_j mobiles, s_f mobiles, $0 \leq s_f \leq b_j$, go to *Data_Tx failure* state, and the remaining $(b_j - s_f)$ mobiles stay in the *Data_Tx success* state.

Out of k_1 mobiles in *Backlogged* state, $(n - a)$ mobiles attempt a packet request and the remaining $(k_1 - (n - a))$ mobiles stay in the *Backlogged* state. Also, 'a' *Idle* mobiles send a resource request packets giving a total of n requests out of which c_s, $0 \leq c_s \leq n$ succeeds . The remaining $(n - c_s)$ mobiles go to *Backlogged* state. Out of c_s succeeding mobiles, s_j mobiles go to *Data_Tx success* state and the remaining $(c_s - s_j)$ go to *Data_Tx failure* state.

So, at the end of slot t (or at the beginning of slot $t + 1$), the total number of mobiles in *Data_Tx failure* state is $(b_i - f_s + s_f + c_s - s_j)$, in *Data_Tx success* state is $(f_s + b_j - s_f + s_j)$, and in *Backlogged* state is $(k_1 + a - c_s)$. All the above transitions follow Binomial distribution and are given by

$$\begin{aligned}
\Pr\{b_i \text{ out of } i_1 \text{ is still active}\} &= \binom{i_1}{b_i} {g_d}^{i_1 - b_i}(1 - g_d)^{b_i}, \\
\Pr\{f_s \text{ out of } b_i \text{ go to success}\} &= \binom{b_i}{f_s}(1 - P_s)^{f_s} P_s^{b_i - f_s}, \\
\Pr\{b_j \text{ out of } j_1 \text{ is still active}\} &= \binom{j_1}{b_j} {g_d}^{j_1 - b_j}(1 - g_d)^{b_j}, \\
\Pr\{s_f \text{ out of } b_j \text{ go to failure}\} &= \binom{b_j}{s_f} {P_s}^{s_f}(1 - P_s)^{b_j - s_f},
\end{aligned}$$

where P_s is the probability of slot error. Therefore, the transition probability $\tilde{P}_{z_1,z_2}$ is given by

$$\begin{aligned}
\tilde{P}_{z_1,z_2} &= \Pr\left(\mathbf{z_1} = (\mathbf{i_1}, \mathbf{j_1}, \mathbf{k_1}) | \mathbf{z_2} = (\mathbf{i_2}, \mathbf{j_2}, \mathbf{k_2})\right) \qquad \text{(A.1)} \\
&= \sum_{n=0}^{N-i_1-j_1} \sum_{c_s=0}^{L} \sum_{s_j=0}^{c_s} \sum_{s_f=0}^{j_1} \sum_{f_s=0}^{i_1} \Pr\Big\{n, c_s, s_j, s_f, f_s, \mathbf{z_1} = (\mathbf{i_1}, \mathbf{j_1}, \mathbf{k_1}) | \mathbf{z_2} = (\mathbf{i_2}, \mathbf{j_2}, \mathbf{k_2})\Big\} \\
&= \sum_{n=0}^{N-i_1-j_1} \sum_{c_s=0}^{L} \sum_{s_j=0}^{c_s} \sum_{s_f=0}^{j_1} \sum_{f_s=0}^{i_1} \Pr\Big\{'a' \text{ new arrivals,} \\
&\qquad (n - a) \text{ arrivals from backlogged mobiles,} \\
&\qquad b_i \text{ out of } i_1 \text{ is still active, } f_s \text{ out of } b_i \text{ go to success,} \\
&\qquad b_j \text{ out of } j_1 \text{ is still active, } s_f \text{ out of } b_j \text{ go to failure,} \\
&\qquad s_j \text{ out of } c_s \text{ go to success, } c_s \text{ out of } n \text{ requests go to success}\Big\} \\
&= \sum_{n=0}^{N-i_1-j_1} \sum_{c_s=0}^{L} \sum_{s_j=0}^{c_s} \sum_{s_f=0}^{j_1} \sum_{f_s=0}^{i_1} \Pr\{'a' \text{ new arrivals}\} \\
&\qquad \cdot \Pr\{(n - a) \text{ arrivals from backlogged mobiles}\} \\
&\qquad \cdot \Pr\{b_i \text{ out of } i_1 \text{ is still active}\} \cdot \Pr\{f_s \text{ out of } b_i \text{ go to success}\} \\
&\qquad \cdot \Pr\{b_j \text{ out of } j_1 \text{ is still active}\} \cdot \Pr\{s_f \text{ out of } b_j \text{ go to failure}\} \\
&\qquad \cdot \Pr\{s_j \text{ out of } c_s \text{ go to success}\} \cdot \Pr\{c_s \text{ out of } n \text{ requests go to success}\}
\end{aligned}$$

$$
\begin{aligned}
= & \sum_{n=0}^{N-i_1-j_1} \sum_{c_s=0}^{L} \sum_{s_j=0}^{c_s} \sum_{s_f=0}^{j_1} \sum_{f_s=0}^{i_1} \binom{N-i_1-j_1-k_1}{a} \lambda^a (1-\lambda)^{N-i_1-j_1-k_1-a} \\
& \cdot \binom{k_1}{n-a} {g_r}^{n-a} (1-g_r)^{k_1-n+a} \\
& \cdot \binom{i_1}{b_i} {g_d}^{i_1-b_i} (1-g_d)^{b_i} \cdot \binom{b_i}{f_s} (1-P_s)^{f_s} P_s^{b_i-f_s} \\
& \cdot \binom{j_1}{b_j} {g_d}^{j_1-b_j} (1-g_d)^{b_j} \cdot \binom{b_j}{s_f} {P_s}^{s_f} (1-P_s)^{b_j-s_f} \\
& \cdot \binom{c_s}{s_j} (1-P_s)^{s_j} P_s^{c_s-s_j} . f(c_s, n, L).
\end{aligned}
$$

A.2 Limiting Behaviour of η_c as $\lambda \to 0$ and $\lambda \to 1$

If $\lambda \to 0$, then the average time duration between arrivals is given by $1/\lambda$ and during this time interval, the average number of successful slots are given by $1/g_d$. Thus, the average throughput is given by

$$\eta_c = (1-P_s) \cdot \frac{1/g_d}{1/\lambda} \cdot \frac{M-L}{M}$$

For $\lambda = 0.01$ and P_s = 0, this corresponds to 0.09 which is close to the value shown in Fig. (3.5). For the case of $\lambda \to 0$, the mean RLC PDU transfer delay is given by $(1/g_d)$ + 1. For g_d = 0.1, the mean delay is 11 slots, as observed in Fig. (3.6).

When $\lambda \to 1$, then all *Idle* mobiles will transmit a resource request packets with probability tending to 1. This means that in the above Eqn. (A.1), the transition probability $\tilde{P}_{z_1,z_2}$ will be zero if $a \neq N - i_1 - j_1 - k_1$. Therefore, $a = N - i_1 - j_1 - k_1$. Also, $a = k_2 - k_1 + c_s$, since we know that $k_2 = k_1 + a - c_s$ which gives $a = k_2 - k_1 + c_s$. This, in turn, implies $c_s = N - i_1 - j_1 - k_2$. If $c_s \geq 0$, $\tilde{P}_{z_1,z_2}$ is given by

$$
\begin{aligned}
\tilde{P}_{z_1,z_2} = & \sum_{n=0}^{N-i_1-j_1} \sum_{s_j=0}^{c_s} \sum_{s_f=0}^{j_1} \sum_{f_s=0}^{i_1} \qquad (A.2) \\
& \binom{k_1}{n-(N-i_1-j_1-k_1)} {g_r}^{n-(N-i_1-j_1-k_1)} (1-g_r)^{k_1-(n-(N-i_1-j_1-k_1))}
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \binom{j_1}{b_j} {g_d}^{j_1-b_j}(1-g_d)^{b_j} \cdot \binom{b_j}{s_f} {P_s}^{s_f}(1-P_s)^{b_j-s_f} \\
&\cdot \binom{i_1}{b_i} {g_d}^{i_1-b_i}(1-g_d)^{b_i} \cdot \binom{b_i}{f_s} (1-P_s)^{f_s} P_s^{b_i-f_s} \\
&\cdot \binom{c_s}{s_j} (1-P_s)^{s_j} P_s^{c_s-s_j} . f(c_s, n, L).
\end{aligned}
$$

We know that the new arrivals during the slot t are given by $\mu = k_2 - k_1$[1]. But, for $\lambda \to 1$, $\mu = N - i_1 - j_1 - k_1$. This, in turn, implies $k_2 = N - i_1 - j_1$. Therefore, if $k_2 = N - i_1 - j_1$, then $\hat{P}_{z_1,z_2}$ is given by

$$
\begin{aligned}
\hat{P}_{z_1,z_2} = & \sum_{\theta=0}^{k_1} \sum_{c_s=x+y+1}^{L} \sum_{s_f=0}^{i_2} \sum_{f_s=0}^{j_2} \\
& \binom{k_1}{\theta} {g_r}^{\theta}(1-g_r)^{k_1-\theta} \\
& \cdot \binom{j_1}{b_j} {g_d}^{j_1-b_j}(1-g_d)^{b_j} \cdot \binom{b_j}{s_f} {P_s}^{s_f}(1-P_s)^{b_j-s_f} \\
& \cdot \binom{i_1}{b_i} {g_d}^{i_1-b_i}(1-g_d)^{b_i} \cdot \binom{b_i}{f_s} (1-P_s)^{f_s} P_s^{b_i-f_s} \\
& \cdot f(c_s, \mu+\theta, L)
\end{aligned} \tag{A.3}
$$

A.3 Derivation of the distribution of X

Let the random variable T represent the length of PDU in number of slots and the random variable Y_i represent the number of transmission attempts of i^{th} slot until success. The distribution of T is *geometric* (by assumption) and is given by,

$$
\Pr(T = t) = g_d(1-g_d)^{t-1}, t = 1, 2, 3, \ldots \tag{A.4}
$$

[1]Refer page 36 for the definition of μ.

and the distribution of Y_i (Y_i's being i.i.d. since we consider i.i.d. slot errors) is given by

$$\Pr(Y_i{=}y) \;=\; (1-P_s)(P_s)^{y-1}, y = 1, 2, 3, \ldots \tag{A.5}$$

where P_s is the slot error rate.Thus, the total number of slots required for successfully transmitting T slots is given by

$$X \;=\; \sum_{i=1}^{T} Y_i. \tag{A.6}$$

The distribution of X is evaluated using Transform techniques as follows.

$$\begin{aligned}
\phi(z) &= E[z^X] \\
&= E\left[z^{\sum_{i=1}^{T} Y_i}\right] \\
&= E\left[\prod_{i=1}^{T} z^{\,Y_i}\right] \\
&= E_T\left[E_{Y_i}\left[\prod_{i=1}^{t} z^{\,Y_i} | T=t\right]\right] \qquad (\because T \amalg Y_i) \\
&= E_T\left[\prod_{i=1}^{t} E_{Y_i}\left[z^{\,Y_i} | T=t\right]\right] \qquad (\because Y_i \amalg Y_j, \forall i \neq j) \\
&= E_T\left[\left(\frac{z(1-P_s)}{1-zP_s}\right)^T\right] \\
&= \sum_{t=1}^{\infty}\left(\frac{z(1-P_s)}{1-zP_s}\right)^t \cdot g_d(1-g_d)^{t-1} \\
&= \frac{zg_d(1-P_s)}{1-z(1-g_d(1-P_s))} \\
&= g_d(1-P_s)\cdot z\left(1-z(1-g_d(1-P_s))\right)^{-1} \\
&= g_d(1-P_s)\cdot \sum_{j=1}^{\infty}\left((1-g_d(1-P_s)\right)^{j-1} z^j.
\end{aligned} \tag{A.7}$$

Therefore,

$$\text{Pr(X=}x) \quad = \quad g_d(1 - P_s)[1 - g_d(1 - P_s)]^{x-1}, x = 1, 2, 3, ... \tag{A.8}$$

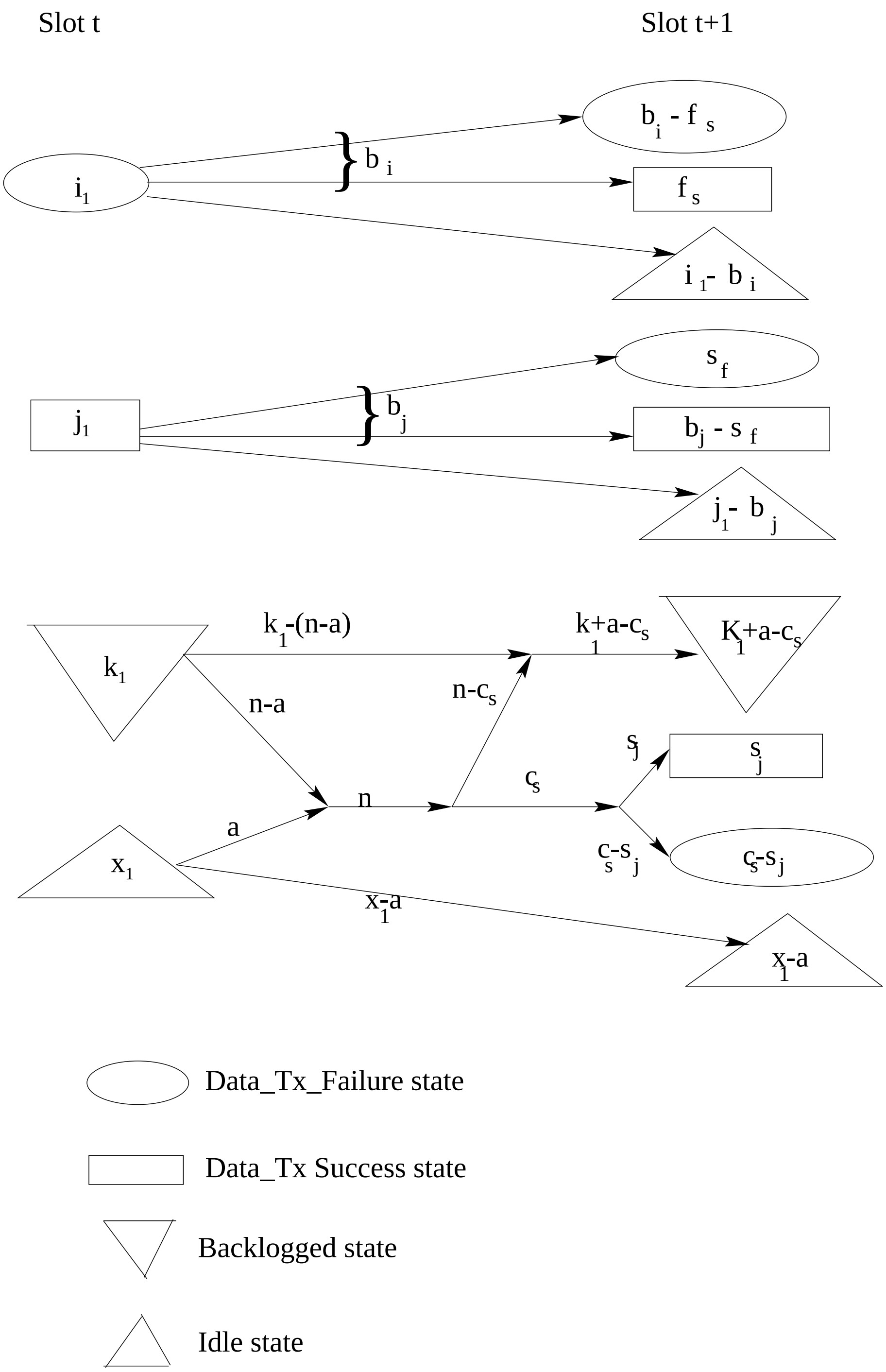


Figure A.1: State transition from slot t to slot $t + 1$

Bibliography

[1] M. Zorzi and A. Chockalingam, *Wireless Communications, Encyclopedia of Physical Science and Technology*, Academic Press, 2001.

[2] W. C. Jakes, Jr., *Microwave Mobile Communications,* New York: John-Wiley & Sons, 1974.

[3] T. S. Rappaport, *Wireless Communications: Principles & Practice*, Prentice Hall, 1996.

[4] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communications*, Addison-Wesley, 1995.

[5] W. R. Stevens, *TCP/IP Illustrated, Volume 1,* Addison-Wesley, Nov. 1994.

[6] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[7] G. Brasche and B. Walke, "Concepts, Services, and Protocols of the New GSM Phase 2+ General Packet Radio Service," *IEEE Commun. Mag.,* pp. 94–104, August 1997.

[8] ETSI TC-SMG GPRS Ad hoc, "Digital Cellular Telecommunications System (Phase 2+); Physical Layer on the Radio Path; General Description (GSM 05.01)." http://www.etsi.org/

[9] ETSI TC-SMG GPRS Ad hoc, "Digital Cellular Telecommunications System (Phase 2+); General Packet Radio Service(GPRS); Mobile Station (MS)-Base Station System (BSS) interface; Radio Link Control/Medium Access Control (RLC/MAC) protocol (GSM 04.60)." http://www.etsi.org/

[10] ETSI TC-SMG GPRS Ad hoc, "Digital Cellular Telecommunications System (Phase 2+); General Packet Radio Service(GPRS); Requirement Specifications of GPRS (GSM 01.60)." http://www.etsi.org/

[11] ETSI TC-SMG GPRS Ad hoc, "Digital Cellular Telecommunications System (Phase 2+); General Packet Radio Service(GPRS); Overall Description of the GPRS Radio Interface (GSM 03.64)." http://www.etsi.org/

[12] ETSI TC-SMG GPRS Ad hoc, "Digital Cellular Telecommunications System (Phase 2+); General Packet Radio Service(GPRS); Service Description (GSM 03.60)." http://www.etsi.org/

[13] ETSI TC-SMG GPRS Ad hoc, "Digital Cellular Telecommunications System (Phase 2+); General Packet Radio Service(GPRS); Multiplexing and Multiple Access on the Radio Path (GSM 05.02)." http://www.etsi.org/

[14] ETSI TC-SMG GPRS Ad hoc, "Digital Cellular Telecommunications System (Phase 2+); General Packet Radio Service(GPRS); Mobile Station - Serving GPRS Support Node (MS-SGSN) Logical Link Control (LLC) Layer Specification (GSM 04.64)." http://www.etsi.org/

[15] ETSI SMG2 GPRS Ad hoc, "Evaluation Criteria for the GPRS Radio Channel," Tdoc 56/95.

[16] D. J. Goodman, "Cellular Packet Communications," *IEEE Trans. on Commun.,* pp. 1272–1280, August 1990.

[17] A. K. Salkintzis, "A Survey of Mobile Data Networks," *IEEE Commun. Surveys,* pp. 2–18, vol. 2, no. 3, 3^{rd} Quarter, 1999. http://www.comsoc.org/pubs/surveys

[18] C. Bettstetter, H. J. Vögel, and J. Eberspächer, "GSM Phase 2+ General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface," *IEEE Commun. Surveys,* pp. 2–14, vol. 2, no. 3, 3^{rd} Quarter, 1999. http://www.comsoc.org/livepubs/surveys/public/3q99issue/bettstetter.html

[19] J. Cai and D. J. Goodman, "General Packet Radio Service in GSM," *IEEE Commun. Mag.,* pp. 122–131, October 1997.

[20] C. Lindemann and A. Thümmler, "Evaluating the GPRS Radio Interface for Different Quality of Service Profiles," *Proc.* 12^{th} *GI/ITG Fachtagung Kommunikation in Verteilten Systemen,* Hamburg, Germany, February 2001.

[21] Q. Pang, A. Bigloo, V.C.M. Leung and C. Scholefield, "Service Scheduling for General Packet Radio Service Classes," *Proc. WCNC'99,* New Orleans, LA, September 1999.

[22] W. Ajib and P. Godlewski, "Service Disciplines Performance for Best-Effort Policies in Packet-Switching Wireless Cellular Networks," *Proc. IEEE VTC Spring 2000,* Tokyo, May 2000.

[23] W. Ajib and P. Godlewski, "Service Disciplines Performance for WWW Traffic in GPRS System," *Proc. IEE 3G2000,* London, March 2000.

[24] G. Brasche, "Evaluation of a MAC Protocol proposed for a General Packet Radio Service in GSM," *Proc. IEEE VTC'96,* pp. 668–672, 1996.

[25] G. Brasche and B. Walke, "Analysis of Multi-Slot MAC Protocols proposed for the GSM Phase 2+ General Packet Radio Service," *Proc. IEEE VTC'97,* pp. 1295–1300, 1997.

[26] W. Ajib and P. Godlewski, "Acknowledgment Procedures at Radio Link Control Level in GPRS,", *ACM/Baltzer Mobile Networks and Applications Journal // Wireless Networks Journal (MONET/WINET)* - Special Issue Fall 2000.

[27] Y. Y. Kim and S. Q. Li, "Performance Evaluation of Packet Data Services over Cellular Voice Networks," *Proc. Conference on Computer Communications (IEEE Infocom)*, Japan, pp. 1022, April 1997.

[28] C. Demetrescu, "LLC-MAC Analysis of General Packet Radio Service in GSM," *Bell Labs Tech. Jl.,* pp. 37–50, July-September 1999

[29] Q. Pang, A. Bigloo, V. C. M. Leung, and C. Scholefield, "Performance Evaluation of Retransmission Mechanisms in GPRS Networks," *Proc. WCNC' 2000*.

[30] R. W. Wolff, *Stochastic Modeling and the Theory of Queues,* Prentice Hall, 1989.

[31] C. Lindemann and A. Thümmler, "Performance Analysis of the General Packet Radio Service," *Proc.* 21^{th} *Int. Conference on Distributed Computing Systems,* Phoenix, Arizona, April 2001.

[32] Y. Argyropoulos, A. Bigloo, and J. van Peursem, "GPRS Delay and Capacity Analysis for Web Browsing Application," *Proc. International Conference on Telecommunications,* Vol. 2, pp. 32–36, 1998.

[33] M. Ermel, K. Begain, T. Müller, J. Schüler, and M.Schweigel, "Analytical Comparison of Different GPRS Introduction Strategies," *Proc.* 3^{rd} *ACM International Workshop on Modeling, Analysis and Mobile Systems (ACM MSWiM' 2000)*, pp. 3–10, Boston, August 2000.

[34] K. Begain, M. Ermel, T. Müller, J. Schüler, and M.Schweigel, "Analytical Call Level Model of a GSM/GPRS Network," *Proc. SPECTS 2000 (Symposium on Performance and Evaluation of Computer & Telecommunication Systems),* Vancouver, pp. 444–451, July 2000.

[35] W. Ajib and P. Godlewski, "Effect of Circuit Switched Transmission over GPRS Performance," *ACM International Workshop on Modeling and Simulation of Wireless and Mobile Systems (ACM MSWiM'00)* in conjunction with MobiCom'00, Boston, August, 2000.

[36] R. V. Nobelen and N. Seshadri, "An Adaptive Radio Link Protocol with Enhanced Data Rates for GSM Evolution," *IEEE Personal Commun. Magazine,* pp. 54–65, February 1999.

[37] S.S. Chakraborty and S. Wager, "A New Approach to Medium-Access Control for Data Traffic and its Adaptation to the GSM General Packet Radio Services," *IEEE Trans. on Veh. Tech.,* vol. 48, no. 1, pp. 240–248, January 1999.

[38] X. Qiu, K. Chawla, L. F. Chang, J. Chuang, N. Sollenberger, and J. Whitehead, "RLC/MAC Design Alternatives for Supporting Integrated Services over EGPRS," *IEEE Personal Commun. Mag.,* August 1999.

[39] R. Kalden, I. Meirick, and M. Meyer, "Wireless Internet Access Based on GPRS," *IEEE Personal Commun. Mag.,* pp. 8–18, April 2000.

[40] K. Begain, M.Ermel, T. Mueller, J. Schueller, and M. Schweigel "Performance of Internet Traffic over GPRS," *8th IFIP Workshop on Performance and Evaluation of ATM & IP Networks,* pp. 16–20, Ilkley, UK, July 2000.

[41] G. Liebl, T. Stockhammer, and F. Burkert, "Modeling and Simulation of Wireless Packet Erasure Channels,“ *Proc. Virginia Tech's* 10^{th} *Symposium on Wireless Personal Communications,* Blacksburg, VA, June 2000.

[42] T. V. Lakshman and U. Madhow, "The performance of TCP/IP for Networks with High Bandwidth-Delay Products and Random Loss," *IEEE/ACM Trans. Networking,* June 1997.

[43] H. Chaskar, T. V. Lakshman, and U. Madhow, "TCP Over Wireless with Link Level Error Control: Analysis and Design Methodology," *IEEE/ACM Trans. Networking,* October 1999.

[44] A. Kumar, "Comparative Performance Analysis of Versions of TCP in a Local Network with a Lossy Link," *ACM/IEEE Trans. Networking,* August 1998.

[45] M. Zorzi, A. Chockalingam, and R. R. Rao, "Throughput Analysis of TCP on Channels with Memory," *IEEE Jl. Sel. Areas Commun. (JSAC)- Wireless Comm. Series,* vol. 18, no. 7, pp. 1289-1300, July 2000.

[46] V. Paxson and S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Trans. Networking,* pp. 226–244, June 1995.

Appendix B

Glossary

AMPS	Advanced Mobile Phone Service
AuC	Authentication Center
ARQ	Automatic Repeat Request
BCS	Block Check Sequence
BLR	Block Level Retransmission
BSC	Base Station Controller
BSS	Base Station Subsystem
CDMA	Code Division Multiple Access
CS	Coding Scheme
EIR	Equipment Identity Register
FCFS	First Come First Serve
FCS	Frame Check Sequence
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FH	Frame Header
ftp	file transfer protocol
GGSN	Gateway GPRS Support Node
GMSK	Gaussian Minimum Shift Keying
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications

GSN	GPRS Support Node
GTP	GPRS Tunneling Protocol
HDLC	High Level Data Link Control
http	hypertext transfer protocol
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
LL-PDU	LLC Packet Data Unit
LLC	Logical Link Control
MAC	Medium Access Control
MSC	Mobile Switching Center
N-PDU	Network Protocol Data Unit
PACCH	Packet Associate Control CHannel
PAGCH	Packet Access Grant CHannel
PBCCH	Packet Broadcast Control CHannel
PC	Power Control
PCCCH	Packet Common Control CHannel
PDCH	Packet Data CHannel
PDTCH	Packet Data Traffic CHannel
PDN	Packet Data Network
PDP	Packet Data Protocol (e.g., IP or X.25)
PDU	Protocol Data Unit
PLMN	Public Land Mobile Network
PNCH	Packet Notification CHannel
PPCH	Packet Paging CHannel
PRACH	Packet Random Access CHannel
PTM	Point-To-Multipoint
PTP	Point-To-Point
QoS	Quality of Service
RLC	Radio Link Control
SCS	Slot Check Sequence
SGSN	Serving GPRS Support Node
SLR	Slot Level Retransmission
SMTP	Simple Mail Transfer Protocol

SNDC	SubNetwork Dependent Convergence
TACS	Total Access Communications System
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
UDP	User Datagram Protocol