RANDOM ACCESS PROTOCOL AND ITS IMPACTS ON AIR INTERFACE IN CELLULAR NETWORKS

Muhammad Jahangir Hossain Sarker

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S4 at Helsinki University of Technology (Espoo, Finland) on the 18th, May, 2005, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Communications Laboratory

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Tietoliikennelaboratorio

HELSINKI UNIVERSITY OF TECHNOLOGY	ABSTRACT OF DOCTORAL DISSERTATION

P.O. BOX 1000, FIN-02015 HUT
http://www.hut.fi

Author   Muhammad Jahangir Hossain Sarker

Name of the dissertation
Random Access Protocol and its Impacts on Air Interface in Cellular Networks

Date of Manuscript: April 5, 2005	Date of dissertation: May 18, 2005
☐ Monograph	☑ Article dissertation (summary + original articles)

Department   Electrical and Communications Engineering
Laboratory   Communications Laboratory
Field of Research   Communications Engineering
Opponent(s)   Professor Jens Zander
Supervisor   Professor Sven-Gustav Häggman
(Instructor)

Abstract

Cellular technology has already evolved through its first and second generations. Currently it is on the verge of the third generation, characterized by the addition of different types of data transmission techniques. In the near future, optimization of radio resource allocation with different kinds of traffic sources having different quality of services will be one of the main challenging problems. This thesis intends to find few solutions to optimize the air interface of current cellular systems with both kinds of traffic, current voice and low speed data traffic and future different types of data traffics.

Nowadays, GSM, TDMA 136, IS-95 and PDC are the most successful second-generation cellular systems. The random access part and the traffic channel part are utilized, respectively, for call reaching in the base transceiver station and information transmission in the uplink direction of a wireless cellular network such as the Global System for Mobile communications (GSM). Optimization of the random access part and the traffic channel part are studied, respectively, in the first part and the second part of the thesis. Overall radio interface optimization considering these two parts together is studied in the third part of the thesis. Suitable cases are confirmed by simulations.

Some form of Slotted ALOHA is used in all current digital cellular networks. The multiple power level transmission system in the original Slotted ALOHA concept, has one inherent problem. It exhibits a throughput-collapse phenomenon whereby the throughput decreases exponentially to zero as the mean number of aggregate transmission attempts increases beyond optimum. Few throughput-collapse preventing algorithms are presented first in this thesis. The random access protocol with the retransmission cut-off scheme is used in the access channel for initial access into the TDMA based traffic channel. The analysis of the random access channel accompanied by retransmission cut-off and its optimization is devised. The number of slots occupied by a terminal can be changed dynamically in the newly developed General Packet Radio Services (GPRS) system. An analysis of the GPRS transmission system with and without RLC/MAC block (radio packet) dropping is carried out.

A call is either rejected or blocked depending on its inability to succeed either in the random access part or in the traffic channel part. It is shown that the number of random access slots provisioned in the wireless access system has a significant impact on the overall call set up probability and traffic channel efficiency. The number of random access slots at optimum performance is derived considering four parameters: average call arrival rate, number of retransmissions, number of traffic channels and average call holding time. The results show that the number of random access slots at optimum performance depends on the average call arrival rate and the number of retransmissions. Moreover, it is independent of the number of traffic channels and average call holding time.

The models presented in this thesis could be used as a guideline for the optimization of radio resource allocation in the future evolution of cellular networks.

PREFACE

The work has been carried out during the years 2000-2005 at Communications Laboratory of Helsinki University of Technology and Espoo-Vantaa Institute of Technology.

I confess my profound indebtedness to my supervisor Prof. Sven-Gustav Häggman for his guidance, constructive suggestions, sincere cooperation, and encouragement. I wish to thank late Prof. Seppo J. Halme, for providing me the opportunity to work for more than eight years in a relatively independent manner and for his support to let me put my boldest ideas into practice. I express my deep gratitude to Jan Hubach (associated with the Communications Laboratory) for improving the linguistic structure of this thesis. I acknowledge the encouragement and help from all friends at Helsinki University of Technology, especially Edward Mutafungwa and Michael Hall. In addition, I also appreciate the support of the financers, particularly the Graduate School in Electronics Telecommunications and Automation (GETA). Prof. Iiro Hartimo, head of GETA, gave me the opportunity to become a GETA student and subsequently a GETA teacher. I would like to thank Prof. Riku Jäntti of the University of Vaasa for his financial support under one project, when the thesis was already near to completion.

Finally, I concede the support and encouragement of my father, late Md. Abul Hossain Sarker, mother Jahida Hossain and relatives. I appreciate the understanding of my family members: my wife, Mahruba Siddika and my daughter Shifa Hossain Sarker. Without my wife's patience and love this final version of the thesis would not have been accomplished.

Jahangir H. Sarker
Espoo, Finland
April 2005

This work is dedicated to the loving memory of my father Md. Abul Hossain Sarker (1932-2004).

Contents

ABBREVIATIONS

3GPP	Third Generation Partnership Project
ATDMA	Advanced Time Division Multiple Access
ATM	Asynchronous Transfer Mode
BTS	Base Transceiver Station
CDMA	Code Division Multiple Access
EDGE	Enhanced Data Rates for GSM Evolution
ETSI	European Telecommunications Standard Institute
FDMA	Frequency Division Multiple Access
GERAN	GPRS/EDGE Radio Access Network
GPRS	General Packet Radio Services
GSM	Global System for Mobile communications
HSCSD	High Speed Circuit Switched Data
IMT-2000	International Mobile Telecommunications by the year 2000
IP	Internet Protocol
LLC	Logical Link Control
PDC	Personal Digital Communications
PDMA	Power level Division Multiple Access
PRACH	Packet Random Access Channel
PRMA	Packet Reservation Multiple Access
PTCH	Packet Traffic Channel
QoS	Quality of Service
RACH	Random Access Channel
RAN	Radio Access Network
RLC/MAC	Radio Link Control/Medium Access Control
SGSN	Serving GPRS Support Node
SNDCP	Subnetwork Dependent Convergence Protocol
TBF	Temporary Block Flow
TCH	Traffic Channel
TDMA	Time Division Multiple Access
UMTS	Universal Mobile Telecommunications System
UTRAN	Universal Terrestrial Radio Access Network
WCDMA	Wideband Code Division Multiple Access

SYMBOLS

a	Traffic activity factor
B	IP call blocking probability
D	IP packet dropping probability
E	Channel occupancy with IP traffic
G	Aggregate packet generation rate per time slot
N	Number of power levels
q	Maximum number of slot occupation by one GPRS terminal
r	Number of retransmissions
R	Rejection probability per time slot
R_G	GSM call rejection probability per random access slot
S	Throughput per slot
S_G	GSM call success probability per random access slot
T	GSM call set up probability
U	Maximum number of allowed IP calls
V	Maximum number of available slots for all IP calls
x	Optimum number of random access slots
λ	New packet generation rate
λ_d	IP call arrival rate
λ_G	GSM call arrival rate
μ_d	IP call departure rate
μ_G	GSM call departure rate
ρ_d	IP traffic generation rate
ρ_G	GSM traffic generation rate

LIST OF PUBLICATIONS

The thesis contains an overview with the following publications:

P1. Jahngir H. Sarker and Seppo J. Halme, "Slowly Decreasing Throughput Algorithms for Random Access Channel", *17th International Teletraffic Congress (ITC-17), 2001*, Salvador da Bahia, Brazil, December 2-7, 2001, pp. 335-346.

P2. Jahangir H. Sarker and Seppo J. Halme, "Auto Controlled Algorithm for Slotted ALOHA", *Communications, IEE Proceedings-,* Volume 150, Issue: 1, February 2003, pp. 53 –58.

P3. Jahangir H. Sarker and Seppo J. Halme, "An Optimum Retransmission Cut-off Scheme for Slotted ALOHA, *Wireless Personal Communications*, May 2000, Volume 13, Nos. 1-2, pp. 185-202.

P4. Jahangir H. Sarker, Seppo J. Halme and Mika Rinne, "Performance Analysis of GSM Traffic Channel Capacity With(out) High Speed Circuit Switched Data", *The IEEE Semiannual Vehicular Technology Conference, IEEE (VTC2000 Fall)*, Boston, USA, September 24-28, 2000, pp. 1603-1609.

P5. Jahangir H. Sarker and Seppo J. Halme, "Efficiency of GSM and GPRS Air Interface for Real-Time IP Traffic Flow With and Without Packet Dropping", *Wireless Personal Communications*, Volume 21, No 1, April 2002, pp. 125-140.

P6. Jahangir H. Sarker and Seppo J. Halme, "Optimizing the use of Random Access Channels in GSM-GPRS", *Wireless Personal Communications*, 2002, Volume 22, No 3, September 2002, pp. 387-408.

P7. Jahangir H. Sarker and Seppo J. Halme, "Single Channel Occupied Voice and Multi-Channel Occupied HSCSD over GSM", *Computer Communications*, Volume 25, Issue 5, 15 March 2002, pp. 522-533.

AUTHOR'S CONTRIBUTION

The author has technically accomplished the research work presented in paper 1 through paper 7. His contributions include setting the objectives, creating the ideas, theoretical analyses, data analyses and writing the manuscripts.

Co-author Prof. Seppo J. Halme's contributions include scientific advice, valuable technical comments and revising the papers.

In paper 4, the co-author M. Rinne contributed to the choice of the parameters for drawing the graphs and the revision of the paper.

Prof. Sven-Gustav Häggman is the supervisor of the author. His contributions to this thesis work include continuous guidance, scientific advice, technical comments, constructive suggestions, sincere cooperation, constant encouragement and revision of the thesis.

1. INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

There has been an evolutionary change in mobile communication systems in every decade of recent history. The first generation cellular technology was an analog system designed for voice transmission; the service started in the beginning of the 1980's. The cellular system entered the second generation by a conversion from analog to digital ten years later in the beginning of the 1990's. According to one of the latest statistics (March 2004), the most successful second generation digital cellular system is the Global System for Mobile communications (GSM) based on FDMA/TDMA, occupying 73.52% of the total 'digital cellular world market'. In addition, other second generation cellular systems such as CDMA covers 13.98%, US TDMA 7.81% and the PDC system 4.38% of the world's digital cellular networks. The rest 0.3% of the digital cellular world market is occupied by the third generation WCDMA system [1].

From 1999 through 2010, voice oriented services (which were the most important services in the first and second generation cellular systems) are expected to grow by 1.5 times per year in the number of subscribers and double the amount of traffic annually. It is expected that the voice services will level off after 2010. The future mobile communications systems should accommodate increased multimedia traffic in 2010's. Those systems should not only consist of high-speed networks, but also high capacity, with a low cost per bit [2].

The traffic flow situation has already changed for wired communication networks. Due to the mega-trend of the rapid growth of Internet type services the packet oriented traffic now exceeds circuit switched traffic. Therefore, communication systems will be developed and optimized for packet-oriented traffic. Within the next few years the bulk of traffic will consist of packet data of which voice traffic will be just a fractional part. One possible future solution is the *transparent transport network*, which will support packet- and circuit-oriented protocols by direct photonic connections of edge nodes [3].

The initial implementation of GSM was intended mainly for voice communication. An efficient packet based data transmission protocol called General Packet Radio Services (GPRS) over the GSM network was designed and is already used in the mobile business. Currently, 3rd Generation Partnership Project (3GPP) is responsible for standardization of Enhanced Data Rates for GSM Evolution (EDGE) [4, 5]. The new EDGE deploys 8-PSK modulation in addition to GMSK, thus enabling higher data throughput per slot [6].

A common third generation mobile communication system called Universal Mobile Telecommunications System (UMTS) is being standardized in the 3rd Generation Partnership Project (3GPP) [5]. Release R99 of the UMTS only supported a Wideband Code Division Multiple Access (WCDMA) based Radio Access Network (RAN) called Universal Terrestrial Radio Access Network (UTRAN) technology using Asynchronous Transfer Mode (ATM) based transport. The UMTS Release 5 (released in 2000) defines two RAN technologies, a GPRS/EDGE radio access network (GERAN) and the previous one: UTRAN. These two access networks and possibly other types of access networks (e.g., WLAN, BRAN) will be connected to the same packet-switched core network, which is the evolution of GPRS [7].

The Internet Protocol (IP) is a network layer protocol and it is independent of link/physical layers. Therefore, it has very good capabilities to allow a wide selection of lower layer technologies including Synchronous Optical Networks (SONET) or Wavelength Division Multiplexing (WDM). IP is not only becoming the basis for packetization of voice, data, signaling, and operation, but it will also form the core network in 3G [7, 8].

From the end-to-end connectivity point of view, GSM offers a circuit switched connection between a mobile host and either a Public Switched Telephone Network (PSTN) or Integrated Services Digital Network (ISDN). On the contrary, a network based on packet switching provides the connectivity between the UMTS terminal and an Internet host [9].

In 2000, more PCs than mobile terminals were connected to the Internet. It was expected during that time that after 2004, the number of mobile portable handsets will exceed the number of PCs connected to the Internet, due to the dominating role of mobile radio access [10]. Therefore, the Internet traffic over the current and future radio interface plays an important role. The newly developed radio access system (a part of UMTS) and IP based core network will handle this situation [10].

From the above discussion it is clear that the packet-oriented traffic now exceeds circuit switched traffic for wired communications networks, and for the future cellular systems the scenario will be the same. In a cellular system, the bandwidth of the air-interface part is limited and needs to be optimized every time when a new service is introduced. The optimization of new kinds of services within the limited air-interface is a difficult problem. Therefore, extensive research is required for the future Internet services (traffic classes) transmitted over the existing GSM and GPRS networks. The present study intends to fill in a small part of the big gap in this area.

The operators and vendors can optimize their networks by setting the parameters according to specifications, when new services are added. The proposed models in this thesis would help the operators and vendors to dimension the uplink radio resources of the evaluated second generation system GPRS/UMTS, especially for future discontinuous traffic sources.

1.2 RESEARCH CONTRIBUTIONS

The research contributions of this thesis can be subdivided into three parts.

The first part covers the random access protocol. Among the protocols for random access through a common wireless channel, Slotted ALOHA is one of the most attractive protocols due to its simplicity and low delay (under light load) for bursty traffic. Slotted ALOHA is frequently used as a component of more complex protocols in wireless systems, e.g. in the capacity/channel request phase of wireless ATM [11] and GSM cellular networks [12]. The main problem of the standard Slotted ALOHA is that the throughput decreases abruptly at higher traffic load conditions, and consequently it drops almost to zero, thus leading to unstable conditions. A multiple power level transmission system can significantly boost the capacity (maximum throughput) of Slotted ALOHA but the throughput collapse phenomenon at high system load remains a serious problem. Three schemes, that can prevent the throughput collapse at high traffic load, have been proposed in Publication 1 [P1].

Throughput collapse prevention algorithms can solve the stability problem at a high retransmission probability as well: those are presented in Publication 2 [P2]. Slotted ALOHA with retransmission cut-off is used in some modern mobile communications systems when requesting a dedicated mobile channel. Most of the earlier works considered the case with no retransmission cut-off. Extended research work is required considering retransmission cut-off in various communications systems. An optimum retransmission cut-off scheme for Slotted ALOHA is discussed in Publication 3 [P3].

Like several other technologies the second-generation GSM system is evolving via GPRS, High Speed Circuit Switched Data (HSCSD) and EDGE towards UMTS [13]. One HSCSD terminal can occupy multiple traffic channels continuously at a given time. On the other hand the number of channels occupied by a GPRS terminal changes dynamically from frame to frame. In the second part of the thesis, the traffic channel occupancy is evaluated, considering both types of cases, HSCSD and GPRS [P4, P5]. In the existing GSM system, the number of traffic channels is limited, in a given base transceiver station. A HSCSD call occupies multiple traffic channels during the entire data call session. If multiple traffic channels are occupied by each HSCSD call from a limited number of traffic channels in a base transceiver station then the number of traffic channels for voice calls is reduced, thus increasing the existing voice call blocking probability. The main purpose of the Publication 4 [P4] is to find the advantages and disadvantages for High Speed Circuit-Switched Data (HSCSD) traffic co-existing with the voice traffic in a GSM base transceiver station. It is found that for the same voice call blocking probability, the overall channel occupancy with continuous multichannel occupied HSCSD calls is higher than in the case without HSCSD calls. The higher channel occupancy with the same voice call blocking probability is obtained by reducing voice traffic intensity. The same result can also be obtained by increasing the number of traffic channels in a given base transceiver station. The overall channel occupancy increases with increasing number of HSCSD calls pertaining to the same voice call blocking probability. The possibility of real-time traffic over the GPRS air-interface is studied in Publication 5 [P5]. It is found that the channel occupancy is extremely low with real-time traffic over GPRS, causing wastage of radio resources. The RLC/MAC block (radio packet) dropping strategies can enhance the channel occupancy and thus increases the radio resource utilization with a negligible degradation of the prescribed QoS.

Most of the researchers have analyzed either the traffic channel part [56-62, P4, P5] or the random access part [25-47, P1-P3] of a TDMA based cellular system. The last two papers fill this gap by combining the two parts together [P6, P7]. In the third and final part of the thesis, the impact of the random access channels on the traffic channel part is investigated [P6, P7]. The random access channels and traffic channels are utilized, respectively, for call reaching in the base transceiver station and information transmission in the uplink direction (from mobile to base transceiver station) of the GSM networks. A call is either rejected or blocked due to its inability to succeed either in the random access channels or in the traffic channels. The overall call setup probability is devised considering call rejection and call blocking probabilities in Publication 6 [P6] and Publication 7 [P7].

1.3 ORGANISATION OF THE THESIS

The intention of the thesis is to aim for a two level approach: the expansion of generic theory and application to GSM-technology.

All the major digital wireless cellular networks use some form of ALOHA. A multiple power level transmission system augments the maximum throughput of Slotted ALOHA, exhibiting the similar behavior to the standard Slotted ALOHA, where the throughput decreasing abruptly at higher traffic load conditions and consequently, drops to almost zero, leading to unstable condition. The slowly decreasing throughput algorithms for Slotted ALOHA are developed to prevent that throughput-collapse and thus increasing the stability is described in Chapter 2. The algorithm to limit the number of retransmissions to ensure stable operation of ALOHA is well known. It is shown in Chapter 3 (as well as in P3) that retransmission cut-off is not needed for the stable operation of Slotted ALOHA. A detailed of random access protocol with the retransmission cut-off scheme and its application to GSM network is also presented in Chapter 3. A HSCSD call occupies multiple traffic channels continuously during the entire data call session. On the other hand the number of channels occupied by a GPRS terminal changes dynamically from frame to frame. In Chapter 4 the traffic channel occupancy - where one mobile terminal can occupy multiple radio resources at a time - is discussed. The random access part and traffic channel part are utilized, respectively, for call reaching in the base transceiver station and information transmission in the uplink direction of a wireless cellular network. In Chapter 5 a more theoretically sound explanation of radio resource management considering random access signaling channels and traffic channels is presented. Finally, the conclusion of the thesis is presented in Chapter 6.

2. SLOWLY DECREASING THROUGHPUT ALGORITHMS FOR SLOTTED ALOHA UNDER HEAVY LOAD CONDITIONS

2.1 THE FEASIBILITY OF THE POISSON PROCESS

The input traffic process of an ALOHA based channel is typically modeled as a Poisson process because of its attractive theoretical properties. It has been a long-standing tradition to investigate the performance behavior of an ALOHA channel loaded by a non-Poisson process. The throughput performance of ALOHA channels with an arbitrary packet interval time distribution is investigated in [14]. It is found that the throughput behavior of an ALOHA channel with a very large number of users is approaching the Poisson arrival model [14]. This information is more than twenty years old and the traffic behavior has changed a lot within the last ten years because of the appearance of many new types of high-speed communications networks.

Fractal phenomena, also known as self-similarity is abound in real-world applications, e.g., geography, hydrology, economics and biology [15]. Fractal or self-similarity in traffic engineering describes that the resulting traffic flows look very similar to each other on differing time scales, i.e., bursty over a wide time scale. Extensive recent studies indicate that the traffic in high-speed communications networks exhibits self-similar (fractal) [16-18] and impulsive character [19, 20]. It is shown that in case of wide area traffic, Poisson arrival processes are valid only for modeling the user sessions (TELNET connections, FTP control connections) arrival process. The self-similar model is more appropriate for other WAN packet arrival processes [21]. Poisson modeling is inappropriate for bulk transfer such as machine-generated SMTP (email) and NNTP (network news) [21]. Clearly, the self-similar model should be used where the packets are communicated as a sequential packet stream [22].

It is interesting to see the ALOHA channel performance under the self-similar arrival model and its comparison when loaded under the Poisson arrival model. If any arrival model shows worsened network performance, then the system needs a higher number of network elements such as usable capacity or transmission resources, buffer size etc. For a stable system design, it is more desirable to obtain worsened system performance results comparing with the better system performance results. The reason is that the worsened performance results will encourage the system designer to overestimate the resources such as transmission resources, buffer size etc. These overestimated resources will make the system more stable. For achieving the system stability by the expense of more resources is logical.

This chapter and next chapter will consider the ALOHA channel, especially the system stability. Therefore, we will select an arrival model, which will show worsened performance of the ALOHA channel.

The ALOHA channel performance (such as normalized throughput and inverse of packet collision probability) under self-similar input shows a significant increase in comparison with the Poissonian case [23]. Clearly, if the ALOHA channel's performance is worse under the Poisson arrival model, then the required number of network elements is higher.

It is shown in [24] that the ALOHA channels show better performance under bursty Pareto distributed packet inter-arrival times compared with the Poisson arrival model [24]. The findings presented in [23] and [24] support that the Poisson arrival model shows worsened performance for ALOHA channels. If the Poisson arrival model is used by a system designer, he will obtain worsened performance results. More resources will be used, because of these worsened performance results, to increase the system stability. Thus, the choice of the Poisson arrival model is appropriate for a stable ALOHA system design, when stability is the main concern.

Furthermore, another 'bursty Poisson' arrival model is discussed next for the throughput performance of Slotted ALOHA. Results show that 'throughput' and 'probability of success' performances of Slotted ALOHA, under a pure Poisson arrival model is worse than that of the proposed 'bursty Poisson' arrival model.

2.1.1   Another bursty Poisson arrival model

In the late 1960's a number of efforts were made in wireless data packet communications. Multiple access is one of the main challenges in data packet communications. Multiple access is a process where a number of data packet sources transmit their digital data to a central receiver using common radio resources. The simple solution for this problem is to divide the whole radio resource by time-partitioning (known as time division multiple access, TDMA) or by frequency-partitioning (known as frequency division multiple access, FDMA). Then each fixed part of this whole radio resource can be allocated to each data packet generator to send their data packets. Fixed allocation multiple access such as FDMA or TDMA is suitable for a configuration with a small number of data packet generators with a stable traffic pattern [25].

Currently, in most of the digital networks, the basic traffic of the network is organized into packets and the traffic arrival is fundamentally bursty and random in nature. Fixed allocated multiple access such as FDMA or TDMA is inadequate for data packet communications and because of that ALOHA was invented. The ALOHA system provides an excellent multiple access for a large number of distributed packet sources with a very high peak- to- average traffic pattern and a time- varying demand.

In the first version of the ALOHA system, the whole radio resource was given to all distributed data packet sources. Any packet source can transmit its any data packets randomly into the ALOHA channel, without the knowledge of activities of other packet sources. A transmitted packet can be received incorrectly for the following two reasons: (1) an occasional noise error from inside the transmission medium and (2) overlapping of another packet transmitted by another packet source. The first type of error is not serious because of its very low possibility of happening due to adding some extra bits to correct the packets. On the other hand, the second type of error is gaining importance when a large number of packet sources are trying to transmit packets at the same time. Later on, a 'slotted' version of ALOHA called Slotted ALOHA was invented; it came to reduce the collision of packets. In the Slotted ALOHA, the common high-speed radio resource is separated in time frame, with equal interval of time, called 'slots'. Each data packet source can start transmitting its packets at the beginning of any slot and the packet must be fit into that slot. The central receiver can decode a

packet from a slot if only one packet is transmitted into that slot. In the Slotted ALOHA channel, the rate of successful transmission per time slot is called 'throughput'.

Traditionally the packet arrival at Slotted ALOHA from an infinite number of bursty terminals is modeled as a Poisson process. If the average arrival rate is G packets per time slot, then the Poisson arrival model can be shown as:

$$a = \frac{G^n}{n!} e^{-G} \qquad\qquad 0 \leq n < \infty \qquad\qquad\qquad (2.1)$$

Here a, defines Poisson distribution with average value of G. In Slotted ALOHA, a packet is successfully transmitted into a given slot, if only one packet is transmitted. The probability of only one packet transmission at a given slot (from Eq. 2.1) is Ge^{-G}, and known as throughput. Traditionally, throughput of Slotted ALOHA with the Poisson arrival model $S = Ge^{-G}$ is considered.

It is shown through with Shannon's theory that the capacity of a 'low duty cycle' ALOHA channel without capture can not be exceeded by any other multiple access methods [25]. Recent investigation [26] indicates that a single-power-level system in which all sources transmit at the maximum allowable power level and shortest packet length achieves both optimal throughput and efficiency of used energy.

A low duty cycle and a shorter packet length will make the arrival more bursty compared with the current "bursty equation" given in Eq. (2.1). Modern high -speed - optical, satellite, broadband – systems etc. generate shorter packets with a low duty cycle, and thus the future packet arrival process will be burstier. There are an infinite number of components in Eq. (2.1). If we contemplate more burstyness, from Eq. (2.1), then it can be said that the first two components will exist. In other words, the probability for zero packets and the probability of exactly one packet will exist, and the probability of more than one packet will be very small or will not exist at all. In other words, the probability of zero packets and the probability of exactly one packet will exists or packet communicated as a sequential packet stream [22]. And thus the probability of more than one packet will be very small or will not exist at all, if the average arrival rate is not very high. For a larger value of the average packet arrival rate, the probability of more than one packet will be high, if burstyness is also high.

For doing this, we can use an extra parameter β, ($\beta \geq 1$) to reduce the arrival rate G, when the number of arrival components is more than one packet per time slot. Thus the modified 'bursty Poisson' arrival model is:

$$a' = \begin{cases} \dfrac{G^n}{n!} e^{-G} & \text{if} \qquad n = 0 \text{ and } 1 \\[4mm] \dfrac{(G/\beta)^n}{n!} e^{-(G/\beta)} & \text{if} \qquad 2 \leq n < \infty \end{cases} \qquad\qquad (2.2)$$

After normalization, the 'bursty Poisson' arrival model is:

$$b_n = \begin{cases} \dfrac{\dfrac{G^n}{n!}e^{-G}}{1+e^{-G}-e^{-G/\beta}+Ge^{-G}-\left(G/\beta\right)e^{-G/\beta}} & \text{if} \qquad n = 0 \text{ and } 1 \\[4ex] \dfrac{\dfrac{\left(G/\beta\right)^n}{n!}e^{-(G/\beta)}}{1+e^{-G}-e^{-G/\beta}+Ge^{-G}-\left(G/\beta\right)e^{-G/\beta}} & \text{if} \qquad 2 \le n < \infty \end{cases} \qquad (2.3)$$

The 'bursty Poisson' arrival model becomes a pure Poisson arrival model when β is equals to one. Burstiness increases with the increased value of β. There are two groups in the 'bursty Poisson' distribution. First group contains transmission of zero packets and one packet ($n = 0$ and 1). The second group contains two or more packets ($n \ge 2$). If the value of β increases, then the deviation between these two groups increases. If the value of β is less than one, then we can assume that the traffic arrival is less bursty and we are not interested about in that region.

The Slotted ALOHA throughput performance with this 'bursty Poisson' arrival model can be written as:

$$Sb = \dfrac{Ge^{-G}}{1+e^{-G}-e^{-G/\beta}+Ge^{-G}-\left(G/\beta\right)e^{-G/\beta}} \qquad (2.4)$$

For calculating the average number of arrival packets per time slot or the modified aggregate packet generation rate with the 'bursty Poisson' model is

$$Gb = \sum_{n=0}^{\infty} nb_n = \dfrac{\left(G/\beta\right)+Ge^{-G}-\left(G/\beta\right)e^{-G/\beta}}{1+e^{-G}-e^{-G/\beta}+Ge^{-G}-\left(G/\beta\right)e^{-G/\beta}} \qquad (2.5)$$

Finally, the ratio of successful slots to the average arrival rate per slot, or probability of success with the 'bursty Poisson' arrival model is:

$$Pb = \dfrac{Sb}{Gb} = \dfrac{Ge^{-G}}{\left(G/\beta\right)+Ge^{-G}-\left(G/\beta\right)e^{-G/\beta}} \qquad (2.6)$$

The numerical representation of Eqs. (2.4) and (2.6) is depicted in Figure 2.1. Most of the time the Poisson arrival model shows the worst throughput performance when compared with the other more bursty arrival traffic. The Poisson arrival shows better performance if the traffic load is high and the burstiness factor is not very high.

Figure 2.1 also shows the discontinuity, especially with the higher values of β, when the arrival rate Gb is also higher. The reason is that we calculated Gb from the corresponding value of G. Thus, if the value of G is equal to 4, then the corresponding value of Gb is equal to 0.8, when $\beta = 100$.

At this stage, from Figure 2.1, it is sufficient to say that the Poisson arrival model shows worsened performance most of the time.

The work here differs from 'very bursty, in [27] in at least two aspects. Firstly, the Slotted ALOHA throughput was considered here *without* blocking, while throughput of Slotted ALOHA with blocking is considered in [27]. Secondly, the definition of burstiness is defined here differently compared to the burstiness in Ref. [27].
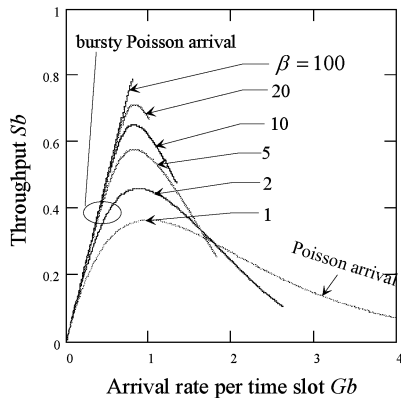
(a) Throughput vs. arrival rate
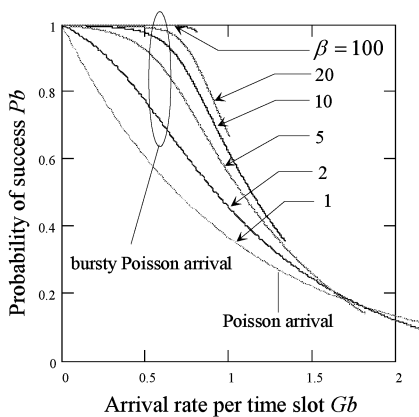
(b) Probability of success vs. arrival rate

Figure 2.1: Throughput and probability of success with bursty Poisson arrival model.

From the above discussion it is clear that the pure Poisson arrival model shows worsened performance results compared to the non-Poisson arrival models. Worsened performance could mean overestimation of needed resources such as transmission resources, buffer size etc. in the system. Physical realization of overestimated resources based on overestimation will provide extra means to make the system more stable.

Finally, we can conclude that the ALOHA channel is more stable under the Poissonian arrival model than under the non-Poissonian arrival models. Therefore, choosing the Poisson arrival model for random access ALOHA channels seems to be appropriate.

2.2 THROUGHPUT COLLAPSE OF SLOTTED ALOHA UNDER HEAVY LOAD

The original Slotted ALOHA concept, which uses a single power level for transmission, has two inherent problems. Firstly, it has a very low capacity (maximum throughput) of only 0.37 which is achieved at a system load of one transmission attempt per slot on average ($G = 1$). Secondly, it exhibits a throughput-collapse phenomenon whereby the throughput decreases exponentially to zero as the mean number of aggregate transmission attempts increases beyond one. In the literature, the throughput collapse is referred to as the stability problem of Slotted ALOHA. There are several algorithms [28-30] to stabilize Slotted ALOHA. These algorithms try to dynamically adjust transmissions to keep the system load at the optimal operating point of $G = 1$. Such adjustments add complexity to the system, increase delay for bursty traffic, and detract them from the original appeal (simplicity and low delay) of Slotted ALOHA.

2.3 BACKGROUND OF MULTIPLE POWER LEVELS TRANSMISSION SYSTEMS

In some popular versions of Slotted ALOHA, users can be separated by transmitting their packets at different power levels. The receiver is able to receive the packet with the highest received power level, even when packets with lower power levels are transmitted into the same slot, and thus the overall throughput increases. The possibility of transmitting packets at multiple power levels and of receiving the packet with the highest power level, was first introduced and analyzed by Metzner [31]. Shacham [32] derived the detailed throughput and delay performance analysis. In that approach, higher classes have advantage over lower classes, which is unfair. To make the system fair and efficient, the user transmits the packets with pre-selected power levels by random selection, which implies that all users belong to the same class. The performance analysis of multiple random power levels by uniformly distributed power level selections with capture effect is developed in [33].

For a given number of power levels, the probability of more than one packet transmitted with the highest power level increases when the traffic load is very high. Because more than one packet is transmitted at the highest power level, the receiver is unable to receive any correct packets, and thus the throughput decreases [34-36].

In the above mentioned system model we have assumed that the power of a given packet on the receiving side is the same as on the transmitting side, which is not a realistic assumption for a typical cellular system. However, this initial work may bring some new results and may thus be useful in a real cellular environment. Therefore, let us initially concentrate on the multiple power level transmission system with this assumption of identical power levels on the receiving and transmitting side.

Although it is now established that by using multiple power levels one can significantly boost the capacity (maximum throughput) of Slotted ALOHA, the throughput collapse phenomenon at high system load remains a serious problem, especially for uniformly distributed power level selections. The objective of this study is to devise simple distributed algorithms for selecting power levels that prevent throughput collapse at high system loads.

One solution to this problem is to introduce Power level Division Multiple Access (PDMA) [36], where a sequentially assigned power level based multiple access scheme is used (Figure 2.2). Figure 2.2 shows the PDMA scheme with 12 terminals and 3 power levels. In this scheme, the maximum number of terminals that can transmit a packet in a given time slot is the same as the number of power levels. Thus only one packet is transmitted at each power level. So the throughput is kept at the highest value, i.e. one. One frame consists of a number of slots and power levels. Every terminal is assigned a slot and a power level in every frame. The allocated power level should be one level higher than the previous power level assigned in the previous frame as shown in Figure 2.2.
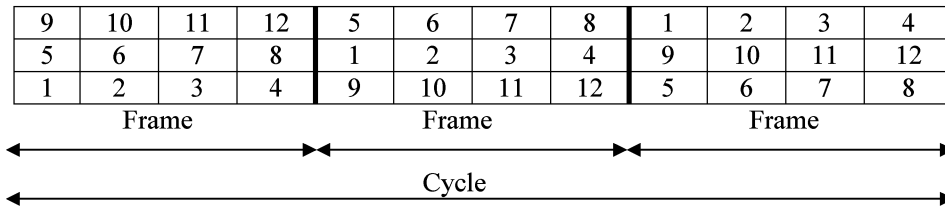
9	10	11	12	5	6	7	8	1	2	3	4
5	6	7	8	1	2	3	4	9	10	11	12
1	2	3	4	9	10	11	12	5	6	7	8

Frame   Frame   Frame

◄──────────────►◄──────────────►◄──────────────►

Cycle

◄──►

Figure 2.2: PDMA with 12 terminals and 3 power levels.

In the original version of PDMA, the power level allocation among the different terminals is unfair. In this scheme, one terminal has advantage over another terminal at a greater number of times in a given cycle (block). A complete cycle is shown in Figure 2.2. In this cycle, terminal 5 has advantage over terminal 1 twice. On the other hand, terminal 1 has advantage over terminal 5 only once. A fair PDMA scheme is introduced [37] using the *permutation* method to allocate the power levels and time slots to every mobile terminal so that each terminal has advantage over another terminal on an equal number of occasions.

A generalized multicopy ALOHA [38] scheme is also introduced to increase the throughput at higher traffic load conditions. In this scheme, two enhancement methods of the multicopy ALOHA are developed. First, a relaxed multicopy transmission policy, which does not decrease the successful transmission probability especially during higher traffic load conditions, and secondly, an artificial modified capture model is used.

In the above mentioned throughput collapse preventing algorithms [36-38], a central entity (e.g., base transceiver station) assigns some information to the mobile terminals to maintain very high throughput under heavy load. However, the involvement of a central entity no longer classifies it as a true distributed protocol and hence is not so attractive from the implementation point of view.

2.4 SLOWLY DECREASING THROUGHPUT ALGORITHMS UNDER HIGH LOAD CONDITIONS

To prevent the throughput collapse at high load, we propose three truly distributed algorithms in a system with multiple power levels. Herein, a mobile terminal can transmit any packet at a higher probability at the lower power levels that increases the probability of only one packet at the highest power levels and thereby augmenting the capture probability. If the traffic load is high, the probability of interfering packets in the same slot increases.

Consequently, the transmission probability at the lower power levels increases and the probability of exactly one packet at the highest power levels remain unaltered. As a result, the system is automatically controlled, since it does not need any extra information about the network. All our algorithms have the following attractive features:

- they prevent throughput-collapse under heavy load (G>>1)
- they are truly distributed in the sense that they do not have any central component and they do not require the mobile hosts to know the current network size
- very simple, no complex estimation
- no need for dynamically adjustment of transmissions to keep the system load at the optimal operating point of $G = 1$.

In our model, we therefore do not consider the power roll-off effect with distance. This assumption can be justified for the emerging pico-cell environments, such as in densely populated urban areas, for low radius cells for personal area networks, and so on, where all mobile terminals are located close to the receiver. Also for large cells in conventional cellular networks, there are now sophisticated power control technologies which can smooth out the power roll-off effect.

All three schemes are based on the following intuitive conjecture:

Conjecture: If higher power levels are selected with lower probability, the probability of receiving multiple packets at the highest power level received will decrease, which in turn will increase the probability of successful transmission (in other words this will increase the system throughput).

2.4.1   Linear approach (LA) scheme

In the linear approach (LA) scheme, we decrease the selection probability of higher power levels *linearly*. To achieve this, we first construct a probability-height diagram as shown in Figure 2.3, where the height of a power level determines its selection probability. Note that for a uniformly distributed selection, each of the N power levels is selected with probability $1/N$ and hence all heights are equal (the middle diagram in Figure 2.3) resulting in a horizontal roof-line. Here 1 is the highest power level and N is the lowest power level.

Now all we have to do to decrease the selection probability of higher power levels linearly is to tilt the roof-line into the direction shown in the top diagram in Figure 2.3. The increase of probability height for the lowest power level (N) is denoted by positive h. Therefore, depending on the sign of h, we have the following three cases:

1.   If $h > 0$, the probability of transmission at lower power levels is *higher* than that of at higher power levels. We consider this case for our linear approach (LA) scheme.

2.   If $h = 0$, the probability of transmission of packets at all N power levels is uniformly distributed. The power level selection is random [33, 34, 35].

3.   If $h < 0$, the probability of transmission at lower power levels is *lower* than that of at higher power levels. This case is against our conjecture stated above.
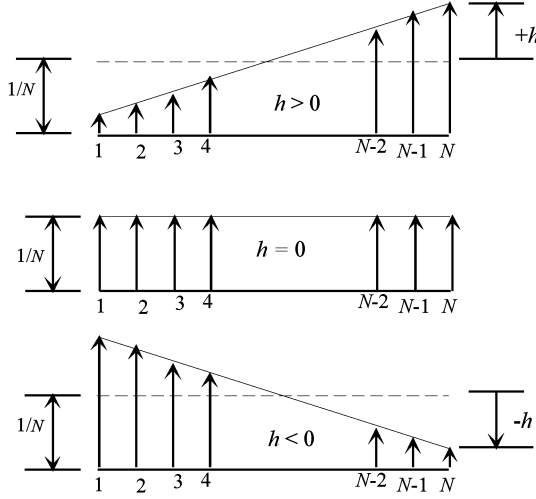

Figure 2.3: Probability distribution of packet transmission into different power levels.

Note that the absolute value of h is bounded by $1/N$, which occurs when either end of the roof-line hits the floor. Let us define α_i^L as the probability of transmitting a packet to the ith power level using the linear approach (LA) scheme. This probability can be formulated using simple geometry from Figure 2.3 as:

$$\alpha_i^L = \frac{2h}{N-1}(i-1) + \frac{1}{N} - h = \frac{h}{N-1}(2i - N - 1) + \frac{1}{N} \qquad |h| \leq 1/N \qquad (2.1)$$

while the throughput of the improved stability scheme with the linear approach is [P1]:

$$S_{LA} = G \sum_{j=1}^{N} \left[\left\{ \frac{h}{N-1}(2j - N - 1) + \frac{1}{N} \right\} \exp \left\{ \frac{Gjh(N-j)}{N-1} - \frac{Gj}{N} \right\} \right] \qquad (2.2)$$

The detailed derivation can be found in Publication 1 [P1].

The random selection (RS) power level transmission system [33, 34, 35] is a truly distributed system, but it can not prevent throughput-collapse under heavy load ($G \gg 1$). That is why we will compare our slowly decreasing throughput (prevent throughput-collapse) algorithms with that random selection (RS). The performance comparison of LA and RS are depicted in Figure 2.4. Clearly, performance improvement is only possible with positive values of h. The linear approach (LA) becomes random selection (RS), when $h = 0$.

In the simulation, the packet generations from all active users transmit identical packets either newly generated or retransmitted. The aggregated packet arrival is modeled as Poisson with an average rate of G packets per time slot. These G packets are transmitted into N power levels and the selection of each power level is the main concern.

We used Eq. (2.1) for the power levels selection probabilities in Figure 2.4. The power levels selection probabilities for next two Figures are given in Publication 1 [P1]. In the simulation, concentrating on i^{th} power level varying from 1 to N, the receiver can receive the packet transmitted at the ith power level if only one packet is transmitted at that power level and all other interfering packets are at *lower* power levels than that of ith power level.
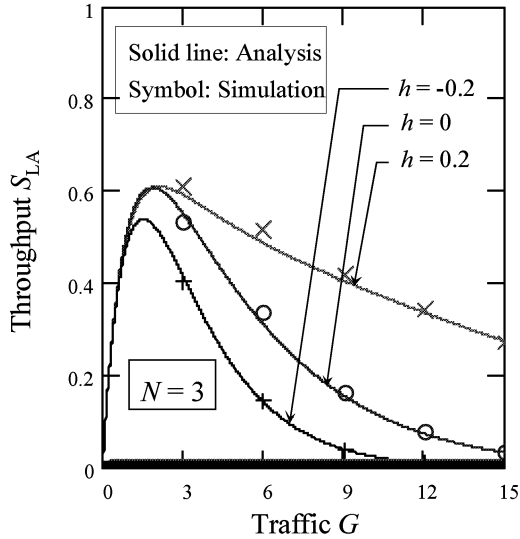
Figure 2.4: Performance comparison of RS and LA.

2.4.2   Two other schemes

We consider two more different power level selection schemes where the packet transmission probabilities of higher power levels are decreased according to *power law*. Here power law means, for example the selection of ith power level is proportional to i^2, i^3 etc. The packet transmission probabilities into higher power levels decreases according to power laws. The power law decrease is motivated by the fact that the throughput of Slotted ALOHA decreases exponentially at higher load. Therefore, if the higher power levels are selected with power law decreasing probability, the throughput loss may be addressed somewhat evenly.

Two other power levels selection approaches are annular approach (AA) and circular shell (CS) approach. The detailed descriptions and derivations of those can be found in Publication 1. The performance comparison of AA and RS are depicted in Figure 2.5. Similarly, the higher throughput of the circular shell (CS) scheme is evidenced in Figure 2.6, especially under heavy load. Note that the proposed slowly decreasing throughput algorithms are not optimized, only the proposal. And many other similar probability distributions can solve the stability problem.

Figure 2.5: Performance comparison of AA and RS.

Figure 2.6: Performance of CS and traditional RS approaches.

2.5 STABILITY CONSIDERATION

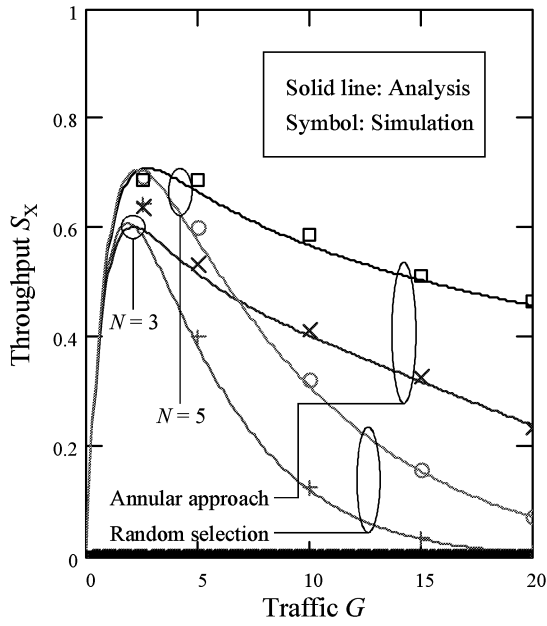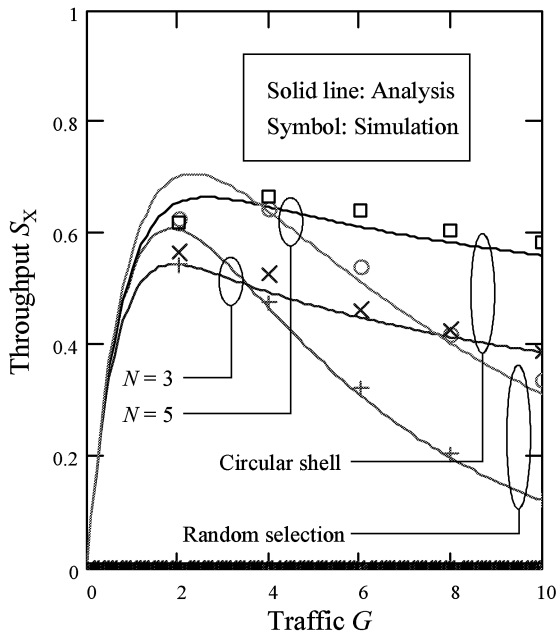In a stable channel, equilibrium throughput-delay results are achievable over an *infinite time* horizon. However, if the channel is unstable, the equilibrium throughput-results are achieved only for a *finite time* before the channel goes into saturation and succumbs to throughput-collapse discussed earlier. The throughput collapse is also referred to as the stability problem.

The stability of Slotted ALOHA can be achieved by preventing throughput-collapse. Three throughput-collapse preventing algorithms are discussed in Publication 1 [P1]. We have also shown in Publication 1 [P1] that delay can be reduced by proposed power level selection schemes.

The purpose of the Publication 2 is to mathematically analyze the stability of the Slotted ALOHA channel operated particularly with a very high retransmission probability. Through numerical examples, it can be shown that the proposed power selection schemes pass some selective stability tests, whereas the uniform power selection (RS) scheme fails those tests [P2].

2.6 SUMMARY OF THE CHAPTER

Randomly selecting a power level (with uniform probability distribution) in multiple power level Slotted ALOHA leads to instability and throughput collapse under heavy load. We have proposed three power selection schemes (in publication 1) where the probability of selecting a higher power level is lower than the probability of selecting a lower power level. We have shown that all three schemes can improve the system throughput and thus stability of the Slotted ALOHA under heavy load. All proposed schemes are distributive in nature and can be easily implemented in mobile communication systems without requiring any central control.

It is found in Publication 2 that the auto-controlled algorithm demands remarkable throughput performance when compare with the traditional random power level selection algorithm particularly with a very high retransmission probability.

We did not consider the real cellular environment, i.e., we did not include the near-far effect, fading and shadowing to avoid complexity. We provide only the ideas to prevent the throughput collapse of Slotted ALOHA. These ideas along with the cellular environment might give interesting results.

3. A RETRANSMISSION CUT-OFF SCHEME FOR RANDOM ACCESS CHANNELS IN CELLULAR SYSTEMS

The slotted ALOHA is used mainly for two purposes: (1) the mobile data transmission and (2) the request for a dedicated mobile channel. For the first purpose each packet should be received successfully. So, the retransmissions take place until the successful reception of each data packet. On the other hand, for the request of a dedicated mobile channel, the successful reception of each data packet is not needed. Therefore, the retransmission cut-off occurs after a certain number of retransmissions trials. Note that slotted ALOHA with retransmission cut-off is used in some modern mobile communication systems. Some potential applications of the transmission number cut-off policy can be found in [39]. Slotted ALOHA with retransmission cut-off is considered in this chapter.

3.1 RETRANSMISSION CUT-OFF SCHEME

Let us reconsider the retransmission cut-off procedure as shown in Figure 3.1. The new packet generation rate from all active users is Poisson distributed with a mean of λ packet per time slot and all users are in the active mode [40]. A user re-enters the active mode after a successful transmission in its first attempt. If this packet is not transmitted successfully, the user enters into the retransmission 1 mode (RT_1). The user will retransmit the same packet in any of the next slots with the probability of q, where q =1/(randomized retransmission delay). We will show in this section (along with the other factors) that the basic analytical result is independent of retransmission probability q.

The *new* packet generation rate is λ packets per time slot from all active users (Figure 3.1). Including r times retransmissions, packet generation rate is G packets per time slot. Assuming Poisson arrival of newly generated and retransmitted packets, the probability of successful transmission of any packet is $\exp(-G)$.

The probability of failure of any packet is $\{1 - \exp(-G)\}$. The failure rate of first time transmission is $\lambda\{1 - \exp(-G)\}$. This unsuccessful part $\lambda\{1 - \exp(-G)\}$ will be entered into the retransmission mode 1 (RT_1). Let n_1 be the number of users in the RT_1 mode. The transmission probability of any user from the RT_1 mode is q. Therefore transmission rate from RT_1 mode is $n_1 q$. The probability of failure for any packet is the same $\{1 - \exp(-G)\}$. So the failure rate of 1st time retransmission is $n_1 q\{1 - \exp(-G)\}$ and this part will be entered in the RT_2 mode, and so on. In general, a user-packet involved in a collision in the RT_l mode enters into the RT_{l+1} mode.
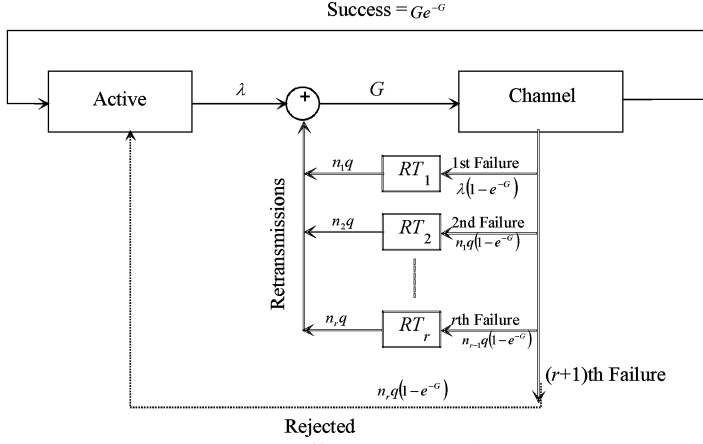
Figure 3.1: Retransmissions cut-off system considering retransmission modes.

If a maximum of r retransmissions is allowed, including the first transmission, the total failure is $(r+1)$ times. After the $(r+1)^{th}$ failure, the rate is $n_r q\{1 - \exp(-G)\}$ and this part is rejected. If δ's define the balance function of each mode, then according to [40]

- Active mode: $\qquad \delta_{Active} = G \exp(-G) + n_r q\{1 - \exp(-G)\} - \lambda$

- RT_1 mode: $\qquad \delta_1 = \lambda\{1 - \exp(-G)\} - n_1 q$

- RT_l mode ($2 \le l \le r$): $\quad \delta_l = n_{l-1} q\{1 - \exp(-G)\} - n_l q$

The solution can be obtained considering equilibrium points $\delta_{Active} = \delta_1 = \delta_2 = \delta_3 = \ldots = \delta_r = 0$. At the equilibrium points, we obtained $G \exp(-G) - \lambda\left[1 - \{1 - \exp(-G)\}^{r+1}\right] = 0$, or

$$G = \frac{\lambda\left[1 - \{1 - \exp(-G)\}^{r+1}\right]}{\exp(-G)} \qquad\qquad (3.1)$$

This is the basic Equation of Slotted ALOHA with retransmission cut-off, which contains neither n_l nor q.

Clearly, the basic Equation of Slotted ALOHA with retransmission cut-off can be obtained without considering the number of retransmission mode users ($n_1 + n_2 + \ldots + n_r$) and randomized retransmission delay ($1/q$).

To obtain the above basic Equation, we assumed retransmitted packet arrivals as a Poisson process. It is obvious that the Poisson arrival model is not valid for the reattempt with a very low randomized retransmission delay. Fortunately, it is generally assumed that if the random retransmission delay is larger than five slots, the combined new and retransmitted packet arrivals can be approximated as a Poisson process [41, 42].

According to Publication 3 [P3], the same Equation (3.1) can be obtained by an easier way, which is described next.

3.2 A SIMPLE FORMULATION FOR RETRANSMISSION CUT-OFF SCHEME

In this section we consider a simple formulation for the Slotted ALOHA protocol with retransmission cut-off as shown in Figure 3.2. Consider the *new* packet generation rate from all active users as λ packets per time slot. Including r times retransmissions, packet generation rate is G packets per time slot. Assume that the resultant of new and retransmitted packet arrivals is a Poisson process.
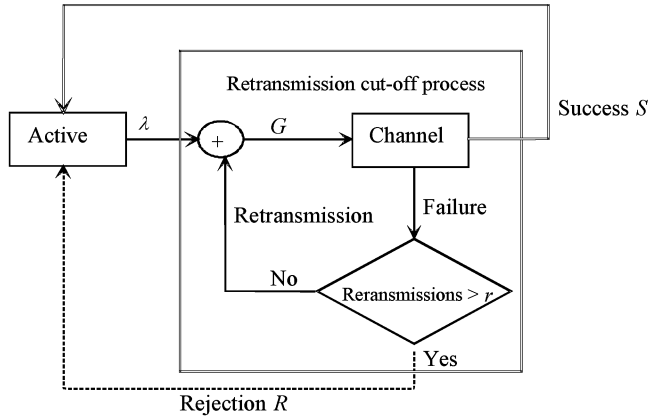
Figure 3.2: Retransmission cut-off scheme.

According to the Poisson arrival model of Slotted ALOHA, the probability of success of any packet without capture is $\exp(-G)$. The probability of failure of any packet is $1 - \exp(-G)$. The unsuccessful fraction of traffic $\lambda\{1 - \exp(-G)\}$ is transmitted in a first time retransmission. Assuming independent failure and success probability each time, the probability of two successive failures is $\{1 - \exp(-G)\}^2$. Hence, the traffic submitted for the second time retransmission is $\lambda\{1 - \exp(-G)\}^2$, and so on. In general, the amount of traffic submitted for the ith time retransmission is $\lambda\{1 - \exp(-G)\}^i$. Let the maximum allowable number of retransmissions of a packet be r, then the mean aggregate offered traffic from all active users is given by

$$G = \lambda \sum_{i=0}^{r} \{1 - \exp(-G)\}^i ,$$

which can be recast as

$$\lambda = \frac{G\exp(-G)}{1 - \{1 - \exp(-G)\}^{r+1}} \tag{3.2}$$

which is same as Eq. (3.1).

This Equation and system model signifying that if the *new* packet generation rate from all active users λ is high, the number of retransmissions r should be low for a given value of the total offered traffic load G and vice versa. The role of aggregate traffic generation rate G is important to make the system 'stable'.

3.3 STABILITY OF THE RETRANSMISSION CUT-OFF SCHEME

It is well known that in Slotted ALOHA, the channel throughput S increases initially with an increase of the aggregate packet generation rate G as shown in Figure 3.3. The throughput reaches its maximum value at a certain point of the aggregate packet generation rate, and it starts decreasing with a further increase of the aggregate packet generation rate. The *optimum* packet generation rate is the aggregate packet rate that maximizes the throughput. According to Figure 3.3, the optimum packet generation rate is one.
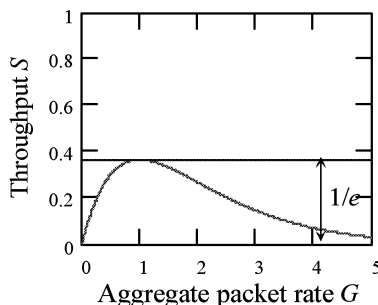

Figure 3.3: Throughput vs. aggregate packet generation rate.

Stability is also studied in [43] for the retransmission cut-off scheme of Slotted ALOHA. It is shown in [43] that the Slotted ALOHA system is stable if the aggregate packet generation rate is less than the optimum packet generation rate. Therefore, according to Figure 3.3, it can be said that the system is 'stable' if the aggregate packet generation rate from all active users is less than one packet per time slot.

Note that, this is dissimilar with the findings in the previous chapter, where we were also interested after the 'optimum point'. In this chapter we consider a Slotted ALOHA system as stable, if its aggregate packet generation rate from all active users is less than the optimum packet generation rate.

The number of retransmissions r should be controlled to make the aggregate packet generation rate G limited to one, to achieve system stability. The main purpose of the retransmission number control is to make the system stable.

Under the above 'stability' definition, the system stability can be achieved by controlling the number of retransmissions r, if the new packet generation rate is less than one. Note that the system performance described in Figure 3.3 is the simplest one. A more general system model is described in Publication 3 [P3].

It is shown in Publication 3 that the aggregate packet generation rate, G never reaches its optimum value if the new packet generation rate, λ is less than a critical limit, irrespective of the number of retransmissions r. Therefore, the system is always stable. Detailed derivation of this critical limit is given in Publication 3. According to Publication 3, the aggregate traffic load, G, never reaches 1, if the new packet generation rate, λ, is less than e^{-1} packets per time slot. Thus, the system is always stable, if the new packet generation rate is less than e^{-1} packets per time slot irrespective of the number of retransmissions.

Publication 3 deals with the following aspects:- (1) a critical limit of the new packet generation rate, below which the number of retransmission control is not needed for a stable operation of Slotted ALOHA, (2) the exact number of retransmissions for a given new packet generation rate exceeding the critical limit, and (3) the maximum value of the new packet generation rate, where no retransmission is allowable for a stable operation. Furthermore, the packet rejection probability is derived and a stable operating region is discussed in Publication 3.

A few retransmission number cut-off policies are discussed in [44-47].

3.4 APPLICATION OF THE RETRANSMISSION CUT-OFF SCHEME

Slotted ALOHA with retransmission cut-off is used in some new communications systems. In this section, we discuss the application of this feature in GSM networks. Consider the RACH of the GSM Systems, where Slotted ALOHA with retransmission cut-off is used. According to specifications [12], the parameter r can be set for four different possible values 1, 2, 4 or 7. Including the first transmission, the maximum allowed transmission number is $(r+1)$. The retransmission cut-off scheme is studied in Publication 3 [P3] and accordingly, the throughput per slot is redefined as

$$S = \lambda \left[1 - \{1 - \exp(-G)\}^{r+1} \right] \qquad (3.3)$$

where the new packet arrival rate per random access slot is λ and the aggregate arrival rate in *each* access slot G is related to

$$\lambda = \frac{G \exp(-G)}{1 - \{1 - \exp(-G)\}^{r+1}} \qquad (3.2)$$

The throughput or the successful access rate per random access slot is S packets per time slot. The random access throughput per slot S for different values ascribed to the number of retransmissions r, is depicted in Figure 3.4. The next section (Section 3.5) describes the simulation model used in Figures 3.4 to 3.6.
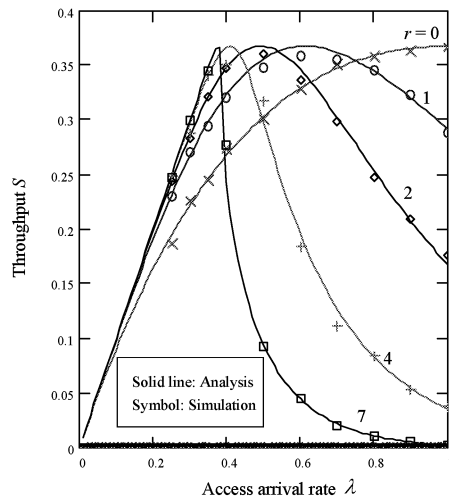

Figure 3.4: Random access throughput S vs. average access arrival rate λ.

3.5 SIMULATION MODEL

The new packet generation rate of λ packet per time slot from all active users is modeled as a Poisson process. The capture effect is not considered. Thus, if more than one packet is transmitted in the same slot, they will destroy each other and all colliding packets need retransmission(s) for a successful transmission.

In a practical situation, the retransmissions are not Poisson distributed and depend on the retransmission probability (randomized retransmission delay). If a given packet is unsuccessful, then it is retransmitted into any next slot with a given probability called retransmission probability. Throughout this chapter the retransmission probability is set to 0.05 in the simulation. The selection of a slot is considered successful if only one packet is transmitted into that slot. A given packet will terminate its retransmission(s) for the following two reasons. Firstly, if the packet is successfully transmitted. Secondly, if the same packet has already been retransmitted a maximum of r times. We considered 10 000 slots for simulation. For finding throughput, we calculated the ratio of successful slots to the total number of simulated slots, i.e., 10 000.

3.6 VALIDITY OF THE INDEPENDENT RETRANSMISSIONS ASSUMPTION

It was indicated in earlier sections (Section 3.1 and 3.2) that the aggregate packet arrival on the Slotted ALOHA channel is consisting of newly transmitted packets plus retransmitted packets. We have assumed that the total arrival rate is Poisson, with parameter G. For doing this, we have made the additional assumption that the retransmitted packets are Poisson distributed as well. This is obviously not true, since retransmissions depend on collisions. The Poisson arrival model is ideal for independent arrivals, but not for dependent arrivals. However, the computer simulation shows that the assumption of independent retransmitted packets arrivals is valid if the random retransmission delay is comparatively large [48, p. 425].

The retransmitted packet arrival process can be assumed as a Poisson process. Computer simulation also shows that the final result does not give much difference compared to the independent retransmissions analytical results, if the randomized retransmission delay is large and the number of retransmissions is not too large [49]. Therefore, we can conclude that the arrivals of the new and retransmitted packets are independent in nature and the total arrival process can be assumed as a Poisson process. This assumption is valid if each retransmission is memoryless and independent of newly generated packets as well as retransmitted packets [50, p. 272]

3.7 AVERAGE NUMBER OF RETRANSMISSIONS

The collapse of the Slotted ALOHA is controlled by different parameters; the retransmission number cut-off [P3] is one of them [12].

In the retransmission cut-off scheme, the retransmissions terminate before last rth retransmission because of successful transmission after the kth retransmission, and the probability of that is [47, P6]

$$P_k = \{1 - \exp(-G)\}^k \exp(-G) \qquad\qquad 0 \le k \le r-1 \qquad\qquad (3.4a)$$

In case of last rth retransmission, two possibilities exist. Firstly, it may transmit successfully and the probability of that is

$$P_r = \{1 - \exp(-G)\}^r \exp(-G) \qquad\qquad (3.4b1)$$

Secondly, the access packet is rejected and it will not be retransmitted due to exceeding of the maximum retransmission count and its probability is

$$P_r = P[r\text{th time retransmission}] * P[\text{probability of failure}] = \{1 - \exp(-G)\}^{r+1} \qquad (3.4b2)$$

Therefore, the distribution is

$$P_k = \begin{bmatrix} \{1 - \exp(-G)\}^k \exp(-G) & 0 \le k \le r-1 \\ \\ \{1 - \exp(-G)\}^k \exp(-G) + \{1 - \exp(-G)\}^{r+1} & k = r \end{bmatrix} \qquad (3.5)$$

To make the Eq. (3.5) as absolute distribution for the access attempt k ($0 \le k \le r$), we do not have to normalize this, since $\sum_{k=0}^{r} P_k = 1$.

Finally, the expected number of retransmissions of an access packet is

$$Rx = \sum_{k=0}^{r} kP_k = \exp(-G) \sum_{k=0}^{r} k\{1 - \exp(-G)\}^k + r\{1 - \exp(-G)\}^{r+1}$$

$$= \frac{\{1 - \exp(-G)\}\left[1 - (r+1)\{1 - \exp(-G)\}^r + r\{1 - \exp(-G)\}^{r+1}\right]}{\exp(-G)} + r\{1 - \exp(-G)\}^{r+1}$$

$$= \frac{\{1 - \exp(-G)\}\left[1 - (r+1)\{1 - \exp(-G)\}^r\right] + r\{1 - \exp(-G)\}^{r+1}}{\exp(-G)} \qquad (3.6)$$

where the relationship between the aggregate packet arrival rate G packets per time slot and the average access packet arrival rate λ packets per time slot can be obtained from Eq. (3.2).

Let us apply the average number of retransmissions for the Random Access Channel (RACH) or a Packet Random Access Channel (PRACH) in a GSM-GPRS network. A GSM-GPRS call tries to access the base transceiver station immediately, if it has any information for transmitting. If the first access transmission is unsuccessful, it retransmits the access packet using a special algorithm. The time between the two transmissions, measured in units of RACH slots, is an integer random variable, uniformly distributed between 1 and a parameter *Tx-integer*, where *Tx-integer* can be set to some values ranging from 3 to 50 slots [12]. In GSM a maximum of r (r = 1, 2, 4, or 7) retransmissions are allowed. The average number of retransmissions for GSM-GPRS is depicted in Figure 3.5.

Figure 3.5 shows a slight difference of analytical and simulated results only when the number of retransmissions is set to seven. The reason might be that the independent retransmission assumption is only valid if the number of retransmissions is not very large and the new packet generation rate is low. Clearly, the simulated and analytical outcomes show very similar results for the smaller number of retransmissions.
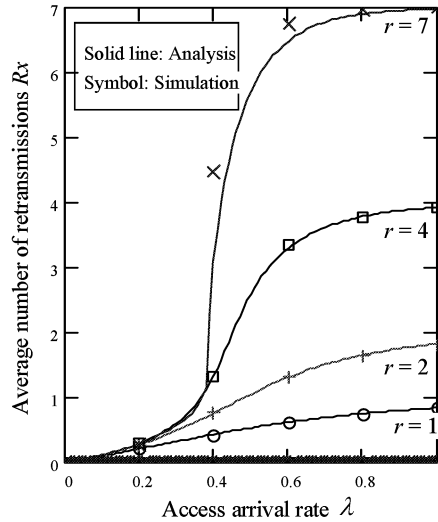
Figure 3.5: Average number of retransmissions.

3.8 REJECTION PROBABILITY

The interesting parameter in the random access channel part is the packet rejection probability. It defines the probability that an access packet is rejected. The packet rejection probability is [P3]

$$R = \{1 - \exp(-G)\}^{r+1} \qquad (3.7)$$

Eq. (3.7) is the general equation for packet rejection probability. We applied these results for the case of GSM-GPRS, when $r = 1, 2, 4$ or 7 are used. The numerical and simulation results considering these four cases is depicted in Figure 3.6. In the simulation, for finding the rejection probability, we calculated the ratio of packets that are leaving system without successful transmission after r retransmissions, and the total number of new arriving packets.

Clearly, simulation results are well matched with the analysis, however we did not include the retransmission probability term in the analysis. The reason is discussed in the Section 3.1.

Throughout this chapter we simulate the performance by setting the randomized retransmission delay to twenty slots (retransmission probability 0.05) and the results show very good agreement with aggregate new and retransmitted packet Poisson arrival analytical model.
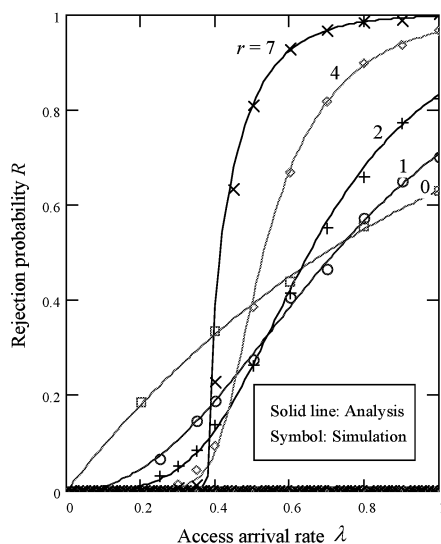
Figure 3.6: Call rejection probability.

Thus, from the sections presented in this chapter, we can conclude that the collisions dependent retransmissions can be assumed independent and can be modeled as Poisson, if the randomized retransmission delay is large, new packet generation rate is low and the number of retransmissions is small.

3.9 CHAPTER SUMMARY

An analytical approach for the performance evaluation of a scheme comprising of Slotted ALOHA with retransmission cut-off scheme has been studied, and has been validated by simulations.

It is found that for a given critical limit of the new packet generation rate, limiting the number of retransmissions is not needed for stable operation. The reason is that if the new packet generation rate is below a certain limit, the probability of success is very high due to the lower aggregate packet generation rate. This lower aggregate packet generation rate is not sufficient to make the system unstable. Publication 3 shows that the number of transmissions has to decrease sharply to keep the system operation stable if the new packet generation rate exceeds this critical limit.

Finally, we can conclude that the Poisson analytical model for new and retransmitted packet arrivals shows very good agreement with the simulation if the simulation follows these three conditions.

(1) New packet generation rate is not very high ($\leq$ one packet per time slot).
(2) The average randomized retransmission delay is large compared to the slot duration (retransmission probability was set to 0.05).
(3) The number of retransmissions is not too large (maximally seven retransmissions were allowed).

4. CHANNEL OCCUPANCY BY HSCSD AND GPRS

The random access channels and traffic channels are utilized, respectively, for call reaching in the Base Transceiver Station (BTS) and information transmission in the uplink direction (from mobile to BTS) of the GSM and GPRS systems. In GSM and GPRS every call request has to go past the random access part before it can reach the BTS and compete for traffic channel allocations. The random access part is contention based and is already discussed in Chapter 2 and Chapter 3. The calls successfully past the random access part compete for traffic channels (slots) in the traffic channel part. This is the second part of the thesis where the call admission control only in the traffic channel part is investigated. The call set up (call admission) probability based on these two parts (random access part and traffic channel part) together is devised in Chapter 5.

In a cellular system, such as GSM, each mobile voice call can occupy a single traffic channel for the entire call duration. A High Speed Circuit-Switched Data (HSCSD) call can occupy multiple traffic channels continuously up to completing its all data transmission. On the other hand, the number of traffic channels allocated for a given GPRS call can be changed dynamically. This chapter deals with the second part of the thesis, where HSCSD and GPRS multichannel occupancy is discussed. Therefore, the traffic channel part is the main subject in this chapter. The handover of the cellular system is not considered. In general, a given Base Transceiver Station (BTS) is studied, where basic radio resources are available to serve the system. In Section 4.1, we consider given basic radio resources of a BTS that serves High Speed Circuit-Switched Data (HSCSD) co-existing with the voice services in a GSM system. From Section 4.2 to Section 4.7 we consider a fixed amount of radio resources, which are available in the GPRS air interface for serving only calls of real-time traffic (e.g., call of real-time IP traffic).

4.1 BLOCKING BASED DIMENSIONING FOR HSCSD CO-EXISTING WITH VOICE SERVICES

The European Telecommunications Standard Institute (ETSI) has defined a multiple time slot technology for data service, High Speed Circuit-Switched Data (HSCSD), over the existing GSM system. In a given BTS the number of traffic channels is limited. In the existing GSM system, free traffic channels are given to a circuit-mode voice call on a reservation scheduling basis for the whole conversation period on a 'one traffic channel for one voice call' basis. A voice call is *blocked* when all traffic channels in a BTS are occupied.

A HSCSD call occupies multiple traffic channels at one time up to the moment it has completed its call. If multiple traffic channels are occupied by each HSCSD call from a limited number of traffic channels in a BTS then the number of traffic channels for voice calls is reduced, thus increasing the existing voice call blocking probability [58-62].

The imminent question is whether or not an operator is willing to implement HSCSD to co-exist with the voice services in a GSM system? Traffic 'channel occupancy' is an important measure for the operators, since the users might pay their bills partially based on that (traffic channel occupancy). On the other hand, the existing voice user will *not* be unsatisfied, if they find new services based on HSCSD within the same network without

hampering the existing facilities. Publication 4 [P4] offers useful information to help the operators to make the decision.

It is shown in Publication 4 [P4] that for the same voice call blocking probability, the overall channel occupancy with continuous multichannel HSCSD calls is higher than in the case without HSCSD calls. The overall channel occupancy increases with increasing number of HSCSD calls pertaining to the same voice call blocking probability. Note that this higher overall channel occupancy is achieved by reducing the voice traffic intensity. If reducing the voice traffic intensity is not possible then the same result can be obtained by a higher number of traffic channels in that BTS. It means that the BTS should contain a higher number of basic radio resources. Such a result encourages the operators to provide different kinds of services using HSCSD in addition to the existing voice services.

4.2 GPRS ACCESS PROTOCOL AND ASSUMPTIONS

The original GSM system is developed for real-time voice traffic transmission, which is circuit-switched transmission. On the other hand, packet-switched non real-time data transmissions have increased a lot in the fixed network. Therefore, a packet based transmission system called GPRS is developed for non real-time packet data transmission system by the evolution of existing GSM network. It is expected that in a near future most of the transmission will be non real-time type of transmission and almost all the networks will be packet based. A fractional part of the transmission will be real-time type of transmission [3]. Therefore, it is interesting to find the difficulties of real-time traffic transmission over the GPRS system, since the GPRS is specified for non real-time type of traffic transmission. May be, the next interesting matter is to overcome the difficulties of real-time traffic transmission over the GPRS system. Real time traffic, such as voice over GPRS called 'Push to talk over Cellular (PoC) is already in the market [71]. Therefore, discussion about the real-time traffic over GPRS is not only an interesting also an important issue.

In a given BTS the number of traffic channels is limited. The circuit mode voice calls will not occupy all traffic channels all the time and some unused traffic channel(s) are available during those times. The packet-mode data transmission technology, GPRS, can utilize the unused traffic channels. Some fixed radio resource can also be allocated only for GPRS. Therefore, the sum of radio resources for GPRS in a given BTS can vary depending on circuit-mode traffic. For simplicity, we assumed in [P5], and in this chapter, that a fixed amount of radio resources are available in a GPRS based BTS, for serving the calls of real-time traffic only.

In the GSM, handover occurs when one mobile call transfers from one BTS to another BTS. In GPRS, cell reselection occurs when one mobile call transfer from one BTS to another BTS. In GPRS, cell reselection is time consuming and not considered here. Only one BTS without any cell reselection is considered here where a limited number of Packet Data Traffic Channels (PDTCH) are allocated to serve calls of real-time traffic using GPRS system. In GSM or GPRS, each carrier frequency is divided (time multiplexed) into TDMA frames and each TDMA frame is divided into eight time slots. The subsequent time slots, one from each subsequent TDMA frames, form one full rate Packet Data Traffic Channels (PDTCH) in GPRS.

In the GPRS data flow (Figure 4.1), the Subnetwork Dependent Convergence Protocol (SNDCP) layer is responsible for formatting the network packets. The SNDCP also does header compression and multiplexing of data coming from different sources. The current GPRS specifications contain a header compression scheme for real-time IP packets. The Logical Link Control (LLC) layer is responsible for reliable logical links between the mobile station and the Serving GPRS Support Node (SGSN). A Temporary Block Flow (TBF) is used in the RLC/MAC layer and it represents a physical connection used by the two radio resource (RR) peer entities to support the unidirectional transfer of LLC PDUs. A TBF is temporary and is maintained until a session of the RLC/MAC blocks transmission is completed over the air interface. A given call consist of several sessions of RLC/MAC blocks flow. The radio resource allocation methods for a given session are discussed later in this section.

The IP packets are variable in size. Each IP packet is mapped into a LLC frame as long as the size of IP packet does not cross the maximum LLC frame size (1520 octets). The LLC frames are passed through Radio Link Control/Multiple Access Control (RLC/MAC) layer. In the RLC/MAC layer, LLC blocks are segmented into fixed length RLC/MAC blocks. Each RLC/MAC block are divided into four GPRS bursts. The Physical Link Layer is responsible to transmit the GPRS bursts over the air interface. Four GPRS bursts of one RLC/MAC block are transferred over four subsequent time slots, one from each four subsequent TDMA frames. It is reasonable to assume that one time slot is allocated from one TDMA frame to transmit one RLC/MAC block.

The MAC layer is responsible to distribute the radio resources among different calls. A limited number of calls are accepted in a given GPRS BTS. When a mobile terminal has LLC PDUs to transmit, it makes requests for resource allocation. Two broad class of resource allocation methods are supported by GPRS [70, pp.49-51].

In **fixed** allocation a call requests for sufficient radio resources to transmit its ready RLC/MAC blocks. The network then allocates a detailed fixed uplink resource to that call. The call can request the continuation of the radio resource allocation, if needed.

The **dynamic** channel allocation provides a more flexible use of the radio resource with traffic channels being 'open' and 'closed'. In the **extended dynamic** allocation, a single GPRS call can occupy up to eight traffic channels, in the same frequency [67, 70]. Therefore, one call can generate up to eight RLC/MAC blocks at a time. The RLC/MAC layer is considered for GPRS performance analysis in this chapter.

According to current GPRS specifications, a RLC/MAC block can only be transferred over the air interface, if four subsequent time slots, one from each four subsequent TDMA frames, are free. For simplicity, we assumed that a RLC/MAC block is dropped or lost if all slots are full during one TDMA frame duration.

From the above discussion it is clear that one call can generate up to eight RLC/MAC blocks at the same time. Each call generates RLC/MAC blocks when it is 'active' with some probability, which is explained later. Thus each call can generate up to eight streams of RLC/MAC blocks flow. Assume that each stream of RLC/MAC block flow is independent of the others.
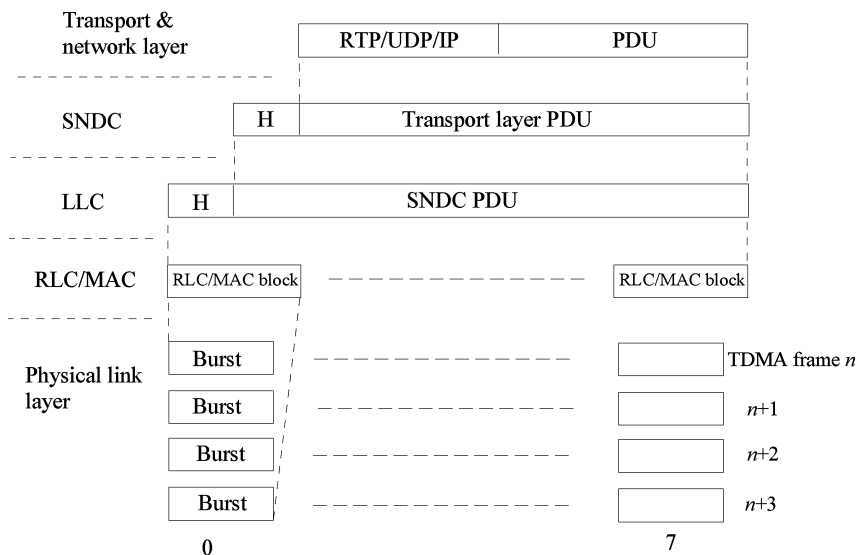
Figure 4.1 GPRS data flow diagram.

According to GPRS specifications [70, p. 49], the channel reservation (holding time) algorithm can also be implemented on assignment basis. This allows individual calls to transmit a predetermined amount of time without interruptions. A given call may be allowed to use the uplink resources as long as there is queued data on the RLC/MAC layer to be sent from the call. It can comprise a number of LLC frames. In that sense the radio resources are assigned on the initially "unlimited" time basis. Alternatively, the uplink assignment for each assignment may be limited to a number of radio blocks (e.g. in order to offer more fair access to the medium at higher loads).

In the analysis we assumed that the duration (channel holding time) of each stream of RLC/MAC blocks flow is exponentially distributed. Note that, this kind of assumption is valid if one GPRS call contains multiple independent services. Therefore, the mathematical model presented here is not valid where one call transmits high data rate. If one call transmits high data rate using multiple streams of RLC/MAC block flow, then each stream of RLC/MAC block flows is dependent of the other one.

Now the question is how many calls can be accepted in a given BTS? Because of the limitation of radio resources, the number of accepted calls per BTS is also limited. Consider a BTS where the calls with real-time traffic (e.g. real-time IP calls or PoC calls) arrive at the BTS with a Poisson arrival process and with a mean rate λ_d. The service time of each call is independent of each other and exponentially distributed with a mean duration $1/\mu_d$ and thus the traffic intensity is $\rho_d = \lambda_d / \mu_d$. If maximum U calls are allowed to transmit the multislot (multichannel) or singleslot (singlechannel) GPRS calls of real-time traffic, the probability that n calls are active in that BTS is obeying the Erlang-B model.

$$p(n) = \frac{(\rho_d)^n / n!}{\sum_{i=0}^{U} (\rho_d)^i / i!} \qquad 0 \le n \le U, \qquad (4.1)$$

and the average number of calls in a given BTS where maximum U calls are allowed is

$$C = \sum_{n=0}^{U} n p(n) = \rho_d \left\{ 1 - p(U) \right\} = \rho_d \left(1 - B \right), \qquad (4.2)$$

where $B = \dfrac{\left(\rho_d \right)^U / U!}{\sum_{i=0}^{U} \left(\rho_d \right)^i / i!}$ and defines the probability that U calls are already admitted and it is

known as the *call blocking probability* or *call loss* probability and any further attempted call is lost.

Now let us concentrate on the mathematical modeling of RLC/MAC block or radio packet or burst generation. Before considering the multiple streams of RLC/MAC block generation from a single call, it is better to consider a simplified case, where each call generates a single stream of RLC/MAC blocks. In other words, let us consider first, when one call can occupy one slot from subsequent TDMA frames, when it is 'active'.

4.3 GPRS MODELING WHEN ONE CALL OCCUPIES ONE SLOT

Consider the GPRS case where one IP call can occupy maximum one slot per TDMA frame. A given call can generate only one stream of real-time RLC/MAC blocks or radio packets. A call (source) generates RLC/MAC blocks while it is in active periods and does not produce any RLC/MAC blocks or radio packets during silence periods.

It is logical to assume real-time voice such as 'push to talk over cellular (PoC)' for this case. A PoC call generates RLC/MAC blocks while it is in talk spurt and does not produce any RLC/MAC blocks or radio packets during silence periods. PoC is a half-duplex system where users receiving the transmission hear the sender's voice in real-time without having to answer the sender. The delay in PoC is not that crucial, since it is unidirectional and listeners do not reply using this service. PoC provides operators an opportunity to develop their new voice services offering without changing conventional voice services. It is a new real-time voice service over GPRS [71].

The probability of n calls are active was already given in Eq. (4.1). Of these n voice calls, all are not in the talk spurt in a given TDMA frame. It is well known that a talker does not talk continuously. Talker activity can be modeled with a two state Markov model (Figure 4.2 a) [73]. Each of the n sources either is idle for an exponential length of average $1/\lambda_2$ or generates RLC/MAC blocks with exponential holding time of average length $1/\mu_2$.

The push to talk service uses cellular access and radio resources more efficiently, reserving network resources only for the duration of talk spurts instead of for entire call duration [74]. An RLC/MAC block is a basic data unit transferred on a GPRS Packet Data Channel (PDCH). It is interleaved over four TDMA bursts and transmitted over one 20ms block period [70]. Obtained by the measurements of conversational speech: the mean talk spurt period is 1s and the duration of mean silence period is 1.35s [73].

We assume here that each source occupies one slot from one TDMA frame and the duration of slots occupation from subsequent TDMA frames is exponentially distributed with

a mean $1/\mu_2$. The distribution is discrete, but the actual active periods can be approximated with continuous exponential distribution [73].

In push to talk, the Media Resource Function (MRF) is responsible to prevent two or more users from sending media at the same time. The MRF controls the talk spurts from different users by employing a request/response mechanism to control transmission right. A user who wants to transmit talk spurt must wait until their request have been granted. If a call is already connected, then the time needed for granting a new talk spurt session is almost zero [71]. The MRF can also cancel transmission right when a user is not using its given radio resources [74].

(a) Model for a single call.

(b)The model for n calls
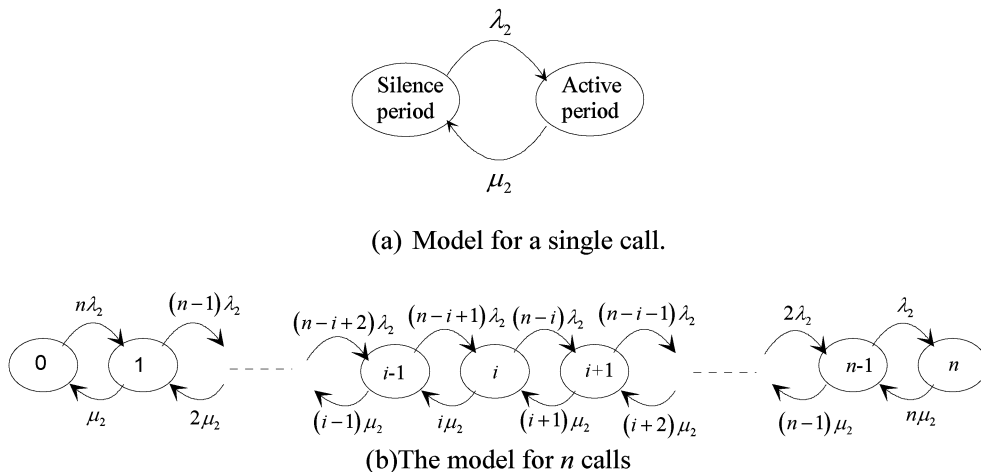
Figure 4.2 Source activity model.

Assume that the source active state and silence state of each call are uncorrelated over the GPRS RLC/MAC block periods. Consider a special case, where the number of available slots (channels) is equal or greater than n active sources. The state probabilities of such a system can be shown as in Figure 4.2 b, which is a special case of Engset distribution [48]. The conditional probability that i of the n calls (sources) are in the active period is given by [48]

$$p_{ac}(i|n) = \frac{\binom{n}{i}(\lambda_2/\mu_2)^i}{\sum_{i=0}^{n}\binom{n}{i}(\lambda_2/\mu_2)^i} \qquad 0 \le i \le n \qquad (4.3)$$

This is known as the modified Engset formula. Note that this modified Engset formula can also be expressed as a binomial distribution:

$$p_{ac}(i|n) = \binom{n}{i}a^i(1-a)^{n-i} \qquad 0 \le i \le n, \qquad (4.4)$$

where $a = \dfrac{\lambda_2}{\lambda_2 + \mu_2} = \dfrac{1/\mu_2}{1/\mu_2 + 1/\lambda_2} =$ (mean active period)/(mean active period + mean silence period). The parameter a defines the probability that a source is generating a RLC/MAC block and occupies one slot from one TDMA frame.

Let us define $p_{ac}(i)$ as the probability of i slots being occupied by all U calls of real-time traffic, when they are in the active period. If a maximum of U calls can be accepted, then applying law of total probability yields

$$p_{ac}(i) = \sum_{n=0}^{U} p_{ac}(i|n) p(n) \tag{4.5}$$

Let us define $\sum_{n=0}^{U} p_{ac}(i|n) p(n)$ more clearly. Mathematically, it is clear from Eqs. (4.3) and (4.4) that $p_{ac}(i|n)$ have some real values, when $n \geq i$. There is no actual value of $p_{ac}(i|n)$, when $n < i$. Thus, the lower limit of n in Eq. (4.5) will be i and not zero. Therefore, Eq. (4.5) can be rewritten as

$$
\begin{aligned}
p_{ac}(i) &= \sum_{n=0}^{U} p_{ac}(i|n) p(n) \\
&= \sum_{n=i}^{U} p_{ac}(i|n) p(n) \qquad \text{with } p_{ac}(i|n) = 0 \qquad \text{for} \quad n < i \\
&= \frac{\displaystyle\sum_{n=i}^{U} \frac{(\rho_d)^n}{n!} \binom{n}{i} a^i (1-a)^{n-i}}{\displaystyle\sum_{n=0}^{U} \frac{(\rho_d)^n}{n!}} \qquad\qquad 0 \leq i \leq U
\end{aligned}
\tag{4.6}
$$

Consider practical consequence of Eq. (4.6), by investigating an example. Let the maximum number of allowed calls $U = 10$ and each call can occupy a maximum of *one* slot at a time. So the value of i varies from 0 to 10. Let $i = 7$, then $p_{ac}(i|n) = 0$ with $n < 7$. It means, at least 7 calls are needed to occupy 7 slots. Less than 7 calls can *not* occupy 7 slots, since one call occupies maximum one slot, according to current model. Of course more than 7 calls can also occupy the same 7 slots. Therefore, in this example the lower limit of n is 7 and not zero. This is why, in $\sum_{n=i}^{U} p_{ac}(i|n) p(n)$, n varies from i to U, not from 0 to U.

Eq. (4.6) shows the slot occupation distribution, when one call occupies one slot discontinuously from subsequent TDMA frames. Eq. (4.6) can be considered as the push to talk slot occupation modeling for GPRS if the radio resource is sufficient to serve in real-time. The average number of slot occupation by PoC sources without any burst (RLC/MAC block) loss or delay can be shown as [75]:

$$CD = \sum_{i=0}^{U} i p_{ac}(i) = a\rho_d (1-B) \tag{4.7}$$

If one call occupies one slot continuously from TDMA frames, then slot occupation probability $a = 1$, and Eq. (4.7) will become same as Eq. (4.2) as expected.

Maximum U calls are accepted to transmit push to talk at a time, where each call can occupy a maximum of one slot from a TDMA frame. Therefore, without any RLC/MAC block dropping, the minimum requirement of the slots V is at least U (i.e., $V \geq U$). Finally,

the traffic channel occupancy with calls of real-time traffic over the GPRS air interface is given by [P5]

$$E_1 = \frac{CD}{V} = \frac{\rho_d a}{V}(1-B)$$ (4.8)

Consider a case where the number of available channels (slots during one TDMA frame duration) for push to talk calls in a BTS is 30 (V = 30). Traffic channel occupancy and the call loss (blocking) probabilities are shown in Figure 4.3, without RLC/MAC block dropping. Real-time transmission is considered here, where each RLC/MAC block must be transmitted just after generation, without any delay. The solid lines curves shown in Figure 4.3 are the traffic channel occupancy without RLC/MAC block dropping. Three dotted curves from left to right show the call blocking probabilities, U =10, U = 20 and U = 30 respectively. The horizontal dashed line show the 2% call blocking probability. Considering the same call blocking probabilities three perpendicular arrows show the traffic channel occupancy for the same call blocking probabilities. Figure 4.4 shows more clearly the traffic channel occupancy for real-time voice (push to talk) traffic without RLC/MAC block dropping. Current non real-time traffic over GPRS uses a delay to achieve higher traffic channel occupancy (efficiency) and is not interesting issue here. The analysis for *non* real-time traffic with delay over GPRS can be found in [76].

Figures 4.3 and 4.4 show that the channel occupancy of the air interface with real-time traffic is unacceptably low especially with a lower value of *a*. One possible solution of the lower channel occupancy problem is to activate a RLC/MAC block *dropper* in each mobile terminal as well as in the BTS. The main function of this RLC/MAC block dropper is to drop a real-time RLC/MAC block, when all radio resources are full.
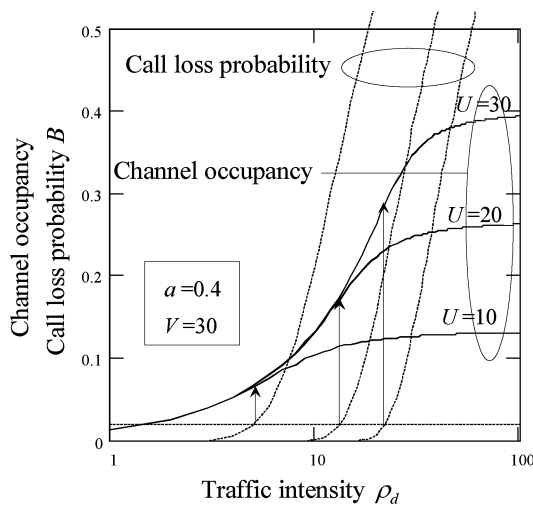
Figure 4.3 Traffic channel occupancy and call blocking probability of push to talk without RLC/MAC block dropping.
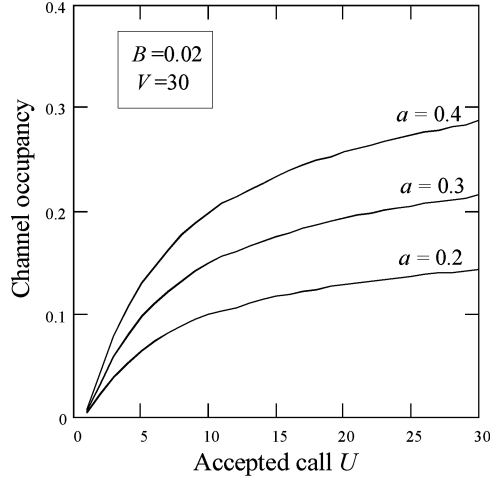
Figure 4.4 Traffic channel occupancy for the same call blocking probability of push to talk without RLC/MAC block dropping.

4.4 ENHANCED CHANNEL OCCUPANCY FOR PUSH TO TALK OVER GPRS

In the case of real-time push to talk traffic, each burst must be transmitted immediately. If maximum V (i.e., $V < U$) slots are available over all the GPRS carriers available at the given BTS, at any time if more than V bursts are generated they will be rejected. Thus we need a modified distribution of equation (4.6).

$$p_{ac}'(i) = p_{ac}(i) \frac{\sum_{i=0}^{U} p_{ac}(i)}{\sum_{k=0}^{V} p_{ac}(i)}$$

$$= \frac{\sum_{n=i}^{U} \frac{(\rho_1)^n}{n!} \binom{n}{i} a^i (1-a)^{n-i}}{\sum_{i=0}^{V} \sum_{n=i}^{U} \frac{(\rho_1)^n}{n!} \binom{n}{i} a^i (1-a)^{n-i}} \qquad 0 \le i \le V , \qquad (4.9)$$

where $\sum_{i=0}^{V} p_{ac}'(i) = 1$ \qquad (4.10)

Thus the improved traffic channel efficiency for push to talk transmission through GPRS air interface is shown as

$$EE_1 = \frac{1}{V} \sum_{i=1}^{V} i p_{ac}'(i) \qquad (4.11)$$

34

By accepting more calls, the traffic channel occupancy is not only increased, the call blocking probability is also reduced. The improvement of the traffic channel occupancy and reduced call blocking probability is obtained at the expense of some RLC/MAC block dropping.

According to current specifications, one RLC/MAC block is dropped if four subsequent slots, one form each in four subsequent TDMA frames, are *not* available. For simplicity, we assumed that one RLC/MAC block is dropped if all slots (channels) are full during one TDMA frame duration

The RLC/MAC block dropping occurs if all V slots are occupied [68, 69]. We have assumed that when $p_{ac}(k \leq V)$, no dropping occurs and when $p_{ac}(k > V)$, only the surplus RLC/MAC blocks are dropped. The average number of dropped RLC/MAC blocks is given by:

$$H_1 = \sum_{k=V+1}^{U} (k-V) p_{ac}(k) \tag{4.12}$$

The average slot occupation per TDMA frame duration is derived in Eq. (4.7). So, the average number of RLC/MAC blocks transmitted during one TDMA frame duration is the same as Eq. (4.7). Finally, the RLC/MAC block dropping probability is the ratio of dropped RLC/MAC blocks to the offered RLC/MAC blocks and this probability is:

$$D_1 = \frac{H_1}{CD} = \frac{\sum_{k=V+1}^{U} (k-V) p_{ac}(k)}{\rho_d a(1-B)} \tag{4.13}$$

For 2% call blocking probability, the enhanced traffic channel occupancy and the corresponding RLC/MAC block dropping probability is shown in Table 4.1. This table may help the operators for optimum push to talk network design, if the RLC/MAC block dropping is considered.

Table 4.1: Improved traffic channel occupancy and corresponding RLC/MAC block dropping probability.

| U | B = 0.02 | | V = 30 | | | |
| | a = 0.2 | | a = 0.3 | | a = 0.4 | |
	EE_1	D_1	EE_1	D_1	EE_1	D_1
30	0.143	0	0.215	0	0.287	0
35	0.173	0	0.26	$2.84*10^{-15}$	0.346	$9.46*10^{-12}$
40	0.203	0	0.304	$3.15*10^{-12}$	0.406	$5.53*10^{-9}$
45	0.233	$1.06*10^{-14}$	0.35	$4.13*10^{-10}$	0.466	$3.91*10^{-7}$
50	0.263	$7.31*10^{-13}$	0.395	$1.63*10^{-8}$	0.527	$8.57*10^{-6}$
55	0.294	$2.27*10^{-11}$	0.441	$2.95*10^{-7}$	0.587	$8.75*10^{-5}$
60	0.325	$3.98*10^{-10}$	0.487	$3.04*10^{-6}$	0.647	$5.24*10^{-4}$
65	0.356	$4.51*10^{-9}$	0.533	$2.06*10^{-5}$	0.703	$2.11*10^{-3}$
70	0.387	$3.64*10^{-8}$	0.579	$1*10^{-4}$	0.754	$6.34*10^{-3}$
75	0.418	$2.23*10^{-7}$	0.625	$3.75*10^{-4}$	0.797	0.0151
80	0.449	$1.09*10^{-6}$	0.669	$1.13*10^{-3}$	0.832	0.03
85	0.48	$4.4*10^{-6}$	0.711	$2.88*10^{-3}$	0.859	0.0518
90	0.512	$1.51*10^{-5}$	0.749	$6.32*10^{-3}$	0.881	0.08
95	0.543	$4.52*10^{-5}$	0.783	0.0122	0.898	0.1129
100	0.574	$1.2*10^{-4}$	0.812	0.0214	0.911	0.1487
105	0.605	$2.86*10^{-4}$	0.836	0.0344	0.922	0.1854
110	0.636	$6.2*10^{-4}$	0.857	0.0513	0.931	0.2215
115	0.666	$1.24*10^{-3}$	0.874	0.0718	0.938	0.2561
120	0.694	$2.31*10^{-3}$	0.889	0.0953	0.944	0.2887
125	0.721	$4.03*10^{-3}$	0.901	0.1211	0.949	0.3191
130	0.746	$6.64*10^{-3}$	0.911	0.1483	0.953	0.3472
135	0.77	0.01	0.919	0.176	0.957	0.3733
140	0.791	0.015	0.926	0.2036	0.96	0.3974

4.5 MULTIPLE PUSH SERVICES OVER GPRS AIR INTERFACE

Presently there is a trend to bind everything in IP. Real time voice over few packet transport technologies are discussed in [64]. The multislot real-time IP traffic flow over GPRS is studied here. We mainly consider the air-interface of GPRS technology.

A GPRS call can occupy a maximum of 8 slots in a TDMA frame and the number of occupied slots in the following TDMA frames can be changed dynamically [65-67]. As mentioned earlier, the model presented in this chapter is not applicable for higher rate data transmissions over GPRS. This mathematical model is more applicable where multiple independent sources are producing traffic from the same mobile call. Different kind of multimedia services over GPRS were already discussed [77, 78, 79]. This mathematical model may show best results when multiple independent push to talks are generated from the same call. More precisely, this model may be used for a future service, called for example: 'multiple push services'.

Consider the case where one call can occupy a maximum of q slots per TDMA frame. The probability of n calls are active was already given in Eq. (4.1). Therefore, all n calls can occupy a maximum of nq slots. Assuming that the occupation of each slot is independent of others having probability a as described in Section 4.3. Therefore, the probability that n calls occupy k slots is given by

$$p_{m\_ac}\left(k\mid n\right)=\binom{nq}{k}a^{k}\left(1-a\right)^{nq-k} \qquad (4.14)$$

The probability that all U calls occupy k slots is given as

$$p_{m\_ac}\left(k\right)=\sum_{n=\lceil k/q\rceil}^{U}p_{m\_ac}\left(k\mid n\right)p\left(n\right)$$

$$=\sum_{n=\lceil k/q\rceil}^{U}\frac{\left(\rho_{d}\right)^{n}/n!}{\sum_{i=0}^{U}\left(\rho_{d}\right)^{i}/i!}\binom{nq}{k}a^{k}\left(1-a\right)^{nq-k} \qquad 0\le k\le Uq \qquad (4.15)$$

where the notation $\lceil z\rceil$ defines the smallest integer $\ge z$.

Let us illustrate Eq. (4.15) by an example. Consider a case, where the maximum number of allowed calls is $U=10$ and each call can occupy a maximum of $q=8$ slots at a time. So the value of k varies from 0 to 10*8=80. Let $k=7$, then $\lceil k/q\rceil=\lceil 7/8\rceil=1$. It means $n=1$ to 10, that is: at least 1 call is needed to occupy 7 slots. Of course more than one call can also occupy the same 7 slots. Similarly, if $k=8$, then $\lceil 8/8\rceil=1$, and hence $n=1$ to 10, that is at least 1 call is needed to fill 8 slots. Obviously, more calls can also produce 8 slots in a TDMA frame. If we consider next $k=9$, then $\lceil 9/8\rceil=2$. This implies $n=2$ to 10, that is: at least 2 calls are needed to occupy 9 slots, because one call will not occupy more than 8 slots. Of course more than two calls can also occupy 9 slots.

The average number of slots occupied by all calls of real-time traffic can be obtained from [P5, P6]

$$A=\sum_{k=0}^{Uq}kp_{m\_ac}\left(k\right)=\sum_{k=1}^{Uq}k\sum_{n=\lceil k/q\rceil}^{U}\frac{\left(\rho_{d}\right)^{n}/n!}{\sum_{i=0}^{U}\left(\rho_{d}\right)^{i}/i!}\binom{nq}{k}a^{k}\left(1-a\right)^{nq-k}$$

$$=\rho_{d}qa\left(1-\frac{\left(\rho_{d}\right)^{U}/U!}{\sum_{i=0}^{U}\left(\rho_{d}\right)^{i}/i!}\right)=\rho_{d}qa\left(1-B\right) \qquad (4.16)$$

Note that the traditional system average slot occupation can be obtained by considering the maximum allowed slots per call $q=1$ and the probability of occupation of each slot $a=1$.

Maximum U calls are accepted to transmit real-time traffic at a time where each call can occupy a maximum of q slots in a TDMA frame duration. Therefore, the minimum

requirement of the slots V without any RLC/MAC block dropping is at least Uq (i.e., $V \geq Uq$). Finally, the traffic channel occupancy with real-time IP traffic over the GPRS air interface is given by [P5]

$$E_2 = \frac{A}{V} = \frac{\rho_d q a}{V}(1 - B),$$
(4.17)

where B is the call blocking probability.

Consider a case where the number of available channels (slots during one TDMA frame duration) for GPRS calls in a BTS is 40 ($V = 40$). One call can occupy maximum of 8 slots at a time. Or one call can transmits up to 8 RLC/MAC blocks at a time. The maximum number of accepted calls is 5 (=40/8) and each RLC/MAC blocks must be transmitted in real-time. The traffic channel occupancy with 5, 4, 3 and 2 accepted calls for the real-time traffic over the GPRS air interface is shown in Figure 4.5. The corresponding call loss probability is high and it is not included in Figure 4.5. Clearly, the channel occupancy of the air interface with calls of real-time traffic is unacceptably low especially with a lower value of a.

Channel occupancy is the main consideration and we choose the traffic intensity ρ_d as a variable, which is the ratio of call arrival rate λ_d and service rate μ_d. For the value of ρ_d less than one, the channel occupancy is so low that we are not interested to observe the system behavior in that range.
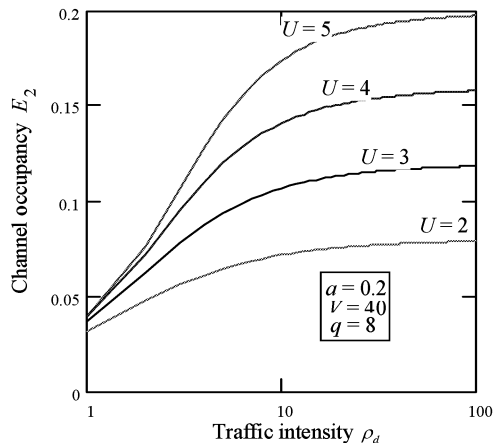
Figure 4.5: Channel occupancy without RLC/MAC block dropping.

The traffic channel occupancy is an important issue for operators, since some parts of the users' bills are paid based on this. There are at least two possibilities for improving this traffic channel occupancy. Firstly, accepting delay, this is more applicable for non real-time type of traffic [76]. This is not considered here. Without thinking of transmission delay, another possible solution of the lower channel occupancy problem is to activate a RLC/MAC block *dropper* in each mobile terminal as well as in the BTS. The main function of this RLC/MAC block dropper is to drop a real-time RLC/MAC block, when all radio resources are full.

4.6 CALLS OF MULTISLOT REAL-TIME TRAFFIC WITH RLC/MAC BLOCK DROPPING

Consider the case with a higher number of accepted calls, where the maximum number of required packet transmissions at a time is higher than what can be handled by the available radio resources. This approach is only possible if users accept the RLC/MAC block dropping possibility. On the other hand, the operators will imply that because of higher channel occupancy. Channel occupancy is an important factor for operators as mentioned earlier.

Let U calls be accepted and thus those calls can produce a maximum of Uq RLC/MAC blocks. These Uq RLC/MAC blocks can occupy same amount of slots within one TDMA frame duration. If maximum of V ($V < Uq$) slots are available for all GPRS calls in a given BTS, the radio resources are not able to serve immediately and the RLC/MAC blocks are dropped. It should be noted that in this mathematical model there is no buffer to absorb the RLC/MAC block (packet) scale fluctuations. An additional assumption is that dropped RLC/MAC blocks are cleared. The steady state probability $p_{m\_ac}{}'(k)$ that there are k packets in the system is given by

$$p_{m\_ac}{}'(k) = \frac{1}{Z(V,Uq)} \sum_{n=\lceil k/q \rceil}^{U} \frac{(\rho_d)^n / n!}{\sum_{i=0}^{U} (\rho_d)^i / i!} \binom{nq}{k} a^k (1-a)^{nq-k} \qquad 0 \le k \le V \qquad (4.18)$$

where $Z(V,Uq)$ is a normalization constant or partition function. By definition, we have

$$Z(V,Uq) = \sum_{k=0}^{V} \sum_{n=\lceil k/q \rceil}^{U} \frac{(\rho_d)^n / n!}{\sum_{i=0}^{U} (\rho_d)^i / i!} \binom{nq}{k} a^k (1-a)^{nq-k} \qquad (4.19)$$

and $p_{m\_ac}{}'(k) = 0$, for $k > V$. The distribution Eq. (4.18) is called a truncated distribution of Eq. (4.15). The traffic channel occupancy for calls of real-time traffic with RLC/MAC block or radio packet dropping is [P5]:

$$E = \frac{1}{V} \sum_{k=0}^{V} k p_{m\_ac}{}'(k) \qquad (4.20)$$

In Figure 4.5, the number of available slots for GPRS is considered as 40. If one call can occupy maximum of 8 slots at a time, without RLC/MAC block dropping, a maximum of 5 calls can be accepted. In the case described in Figure 4.5, the channel occupancy is maximum if a maximum of 5 calls is accepted. It is also obvious that the call blocking probability is at a minimum in that case. Thus, this is the best case of without RLC/MAC block dropping. We want to compare this best case (without RLC/MAC block dropping) with two other cases where RLC/MAC block dropping is considered.

Traffic channel occupancy and the call loss (blocking) probabilities are shown in Figure 4.6, with and without RLC/MAC block dropping. Therefore, the curves shown in Figure 4.6 indicated with $U=5$ are without RLC/MAC block dropping as described in the previous paragraph. If the number of calls is more than 5, then there are some possibilities of

RLC/MAC block dropping. When RLC/MAC block dropping is acceptable, then more than 5 calls can be allowed. Thus, the curves shown in Figure 4.6 indicated with U=10 and 20 are with RLC/MAC block dropping. Three dotted curves from left to right show the call blocking probabilities, U =5, U = 10 and U = 20 respectively. The horizontal dashed line show the 2% call blocking probability. Considering the same call blocking probabilities three perpendicular arrows show the traffic channel occupancy for the same call blocking probabilities. Figure 4.6 shows clearly that the traffic channel occupancy with an acceptable call blocking probability increases significantly with RLC/MAC block dropping.

Consider again the occupancy curves of Fig. 4.6. Take for instance the middle curve with U = 10. If the traffic intensity ρ_d, is very high (e.g., 100 Erlang) and the maximum number of allowed calls U = 10, then there is a very high probability that 10 calls are supported by the subsystem. Each call can occupy 8 slots and the occupation probability of each slot is 0.2. Therefore each call occupies 8*0.2 slots and all 10 calls occupy 10*8*0.2=16 slots. Thus the channel occupancy is 16/40=0.4 as shown in Figure 4.6.
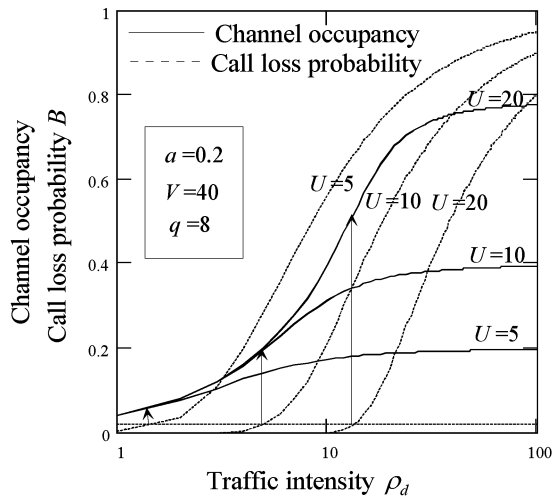
Figure 4.6: Traffic channel occupancy and call blocking probabilities as a function of traffic intensity.

From the above discussion, it is clear that by allowing some RLC/MAC block dropping, the channel occupancy can be increased. But this channel occupancy increment is not without a price. The price is 'RLC/MAC block dropping'. Since the available number of slots is less than the maximum number of occupied slot, some RLC/MAC blocks are dropped. There is a corresponding probability that a RLC/MAC block is dropped because 'all V serving slots are busy'.

4.7 RLC/MAC BLOCK DROPPING PROBABILITY

According to the previous five sections, the service time of each call is independent of the others and exponentially distributed. The traffic intensity of all calls is ρ_d. Maximally U calls are allowed to transmit the multislot GPRS real-time traffic; we considered a case where n calls are active. Consider the case where one call can occupy a maximum of q slots per TDMA frame duration. Therefore, all n calls can occupy a maximum of nq slots during one TDMA frame duration.

According to Eq. (4.15), the probability that all calls occupy k slots is given as

$$p_{m\_ac}(k) = \sum_{n=\lceil k/q \rceil}^{U} p_{m\_ac}(k|n) p(n)$$

$$= \sum_{n=\lceil k/q \rceil}^{U} \frac{(\rho_d)^n / n!}{\sum_{i=0}^{U} (\rho_d)^i / i!} \binom{nq}{k} a^k (1-a)^{nq-k} \qquad 0 \le k \le Uq \qquad (4.21)$$

where the notation $\lceil z \rceil$ defines the smallest integer $\ge z$.

Eq. (4.21) shows the probability that k slots are occupied in a given BTS when n calls are generating RLC/MAC blocks. According to current specifications, one RLC/MAC block is dropped if four subsequent slots, one from each in four subsequent TDMA frames, are *not* available. For simplicity, we assumed that one RLC/MAC block is dropped if all slots are full during one TDMA frame duration.

If maximum V ($V < Uq$) slots are available during a TDMA frame duration, in a given BTS, then the radio resources are not able to serve more than V RLC/MAC blocks during a TDMA slot duration. Therefore, RLC/MAC blocks in excess of V are dropped. The RLC/MAC blocks dropping occurs if all V slots are occupied [68, 69]. We have assumed that when $p_{m\_ac}(k \le V)$, no dropping occurs and when $p_{m\_ac}(k > V)$, only the surplus RLC/MAC blocks are dropped. The average number of dropped RLC/MAC blocks is given by:

$$H_2 = \sum_{k=V+1}^{Uq} (k-V) p_{m\_ac}(k) \qquad (4.22)$$

According to the GPRS specifications, each RLC/MAC block occupies four subsequent slots, one from each four subsequent TDMA frames. Thus, one RLC/MAC block occupies one slot during one TDMA frame duration. Therefore, the average number of slot occupation during one TDMA frame duration can be same as the average number RLC/MAC blocks transmitted during one TDMA frame duration. The average number of slot occupation per TDMA frame duration is derived in Eq. (4.16). So, the average number of RLC/MAC blocks transmitted during one TDMA frame duration is same as Eq. (4.16). Finally, the RLC/MAC block dropping probability is the ratio of dropped RLC/MAC blocks to the offered RLC/MAC blocks and this probability is:

$$D_2 = \frac{H_2}{A} = \frac{\sum_{k=V+1}^{Uq} (k-V) p_{m\_ac}(k)}{\rho_d q a (1-B)} \qquad (4.23)$$

It should be noted that the RLC/MAC block dropping Eq. (4.23) presented here is not same as the Eq. (21) presented in Publication 5. GPRS is not designed for real-time traffic and RLC/MAC block dropping probability evaluation approach presented in Eq. (4.23) is more appropriate approach comparing to the simple approach presented in Eq. (21) of Publication 5. The numerical calculations show that the results obtained by Eq. (4.23) and Eq. (21) in Publication 5 do not differ in practice.

The RLC/MAC block dropping probability is depicted in Figure 4.7.

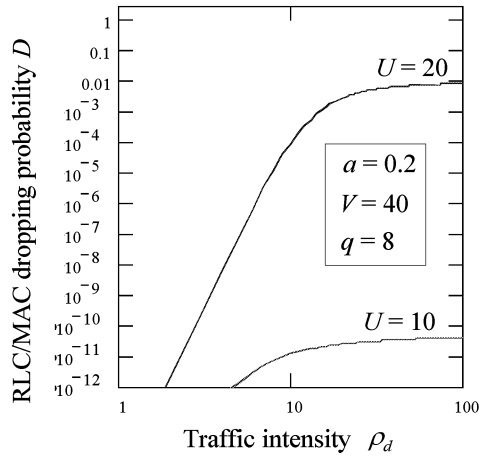*Figure 4.7: RLC/MAC block dropping probability as a function of traffic intensity.*

The traffic channel occupancy improvement with accepting RLC/MAC block dropping is shown in Table 4.2 when one call can transmit up to 8 independent RLC/MAC blocks at a time and call blocking probability is 2%. Comparing, Table 4.1 and Table 4.2, the RLC/MAC block dropping is more significant with multislot (multichannel) operation.

Table 4.2: Traffic channel occupancy improvement and corresponding RLC/MAC block dropping probability.

	$B = 0.02$		$q = 8$		$V = 40$	
	$a = 0.2$		$a = 0.3$		$a = 0.4$	
U	E	D_2	E	D_2	E	D_2
5	0.009	0	0.041	0	0.108	0
10	0.077	$1.64*10^{-12}$	0.245	$1.6*10^{-7}$	0.383	$8.88*10^{-5}$
15	0.306	$6.54*10^{-7}$	0.517	$1.1*10^{-3}$	0.647	0.0215
20	0.51	$2.9*10^{-4}$	0.705	0.028	0.787	0.1307
25	0.664	$6.92*10^{-3}$	0.809	0.1178	0.854	0.2863
30	0.773	0.0407	0.866	0.2461	0.891	0.4227
35	0.837	0.11	0.897	0.3611	0.911	0.5188
40	0.877	0.2002	0.916	0.4528	0.925	0.5893
45	0.903	0.2899	0.929	0.5231	0.935	0.6423
50	0.92	0.3681	0.939	0.578	0.942	0.6835

4.8 CHAPTER SUMMARY

Results in Publication [P4] show that GSM can extend its services using HSCSD technology. An analytical approach was carried out in Publication 4 [P4] to show the channel occupancy (utilization) with and without HSCSD. The results show that the overall traffic channel occupancy increases with the increase of HSCSD calls pertaining to the same voice call blocking probability under the condition of lower voice traffic intensity. Therefore, depending on these results, operators can implement HSCSD alongside GPRS to approach the 3rd generation capability.

The traffic channel occupancy of the GSM-GPRS air interface for calls of real-time traffic flow was studied in this chapter. Calls of real-time traffic without RLC/MAC block dropping keeps a lot of slots unoccupied. Thus, the traffic channel occupancy and the call acceptance probability are low. Increasing the number of accepted calls could augment traffic channel occupancy due to a decreasing number of unoccupied slots. The channel occupancy increases with increasing traffic intensity and further enhancement is achieved with more maximum number of admitted calls. The trend of RLC/MAC block dropping probability resembles the trend of traffic channel occupancy. The traffic channel occupation enhancement is more significant when one call can transmit multiple RLC/MAC blocks at a time.

The form of dropping probability equations presented in this chapter is different compared to the dropping probability equations presented in Publication 5. In this chapter, for evaluating dropping probability, standard technique usually followed in packet-switched technology literature is followed. In Publication 5, a comparatively simpler approach was used for calculating the same dropping probability. A comparison of numerical results obtained from these two approaches shows that they differ very slightly in practice. Hence, the results in Publication 5 are still usable and supportive for this chapter.

5. THE BRIDGE BETWEEN THE RANDOM ACCESS CHANNEL AND THE TRAFFIC CHANNEL

This is the third part of the thesis where the impact of the random access channels on the traffic channel is investigated. 'Call set up probability' is typically assumed to be one minus call blocking probability. Such assumptions are not always valid in wireless cellular networks, such as GSM, where every call request has to go past the random access part first before it can reach the base transceiver station and compete for traffic channel allocations. We provide an analysis of call set up probability in GSM networks by explicitly accounting for the fate of call requests in the random access part. Our analytical model enables us to study the actual call set up probability in GSM networks under a given traffic load and to compute the optimum number of random access slots in the wireless access network to maximize call set up probability. In this chapter we present a more theoretically sound explanation of radio resource management while combining the random access signaling channel and the traffic channel.

5.1 THE APPLICABILITY OF THE POISSON PROCESS

The random access channels and traffic channels are utilized, respectively, for call reaching in the base transceiver station and information transmission in the Global System for Mobile communications (GSM). We have assumed that 'the random access channels' is the first part and its successfully transmitted (output) traffic enters as the input call in the traffic channel part. Therefore it is interesting to know the characteristics of the output traffic of the random access part. In case of Poisson arrival, the departure of packets from the servers is also Poisson with the same rate as arrival, if the service time is exponentially distributed and the arrival rate is less than the total service rate [80, pp. 516-517].

In a random access channel, the service time is fixed and packet transmission occurs based on contention. The output process in a packet broadcasting system with contention was studied in [81]. In this study, it was assumed that the throughput (output) of the random access channel could be modeled as a Poisson process. The output of the random access part is same as the successful call arrival in a GSM cellular base transceiver station. Now we have to show the validity of the Poisson model for the arrival of successful calls in a cellular system like GSM.

Claims have been made about the appropriateness or otherwise of the Poisson call arrival model in a cellular system where blocked calls may be reattempted. The call arrival measurements of the air interface of a "real" cellular network are presented in [82]. In [82] the mean of measured data was compared with that of the Poisson call arrival model invoking the Chi-square test. The Chi-square test results show that the difference between the measured arrival process and the Poisson arrival model is *rejected* at a level of significance of $\alpha = 0.05$. This is true when the number of traffic channels is greater than 12 and experienced blocking probability is greater than 1%. In the measured arrival process in [82], reattempted calls were also included. From this result, it can be said that reattempted calls for an acceptable blocking probability does not hamper the appropriateness of the Poisson call arrival model in a cellular network.

From the above discussion, we can conclude that the throughput (output) of the random access part or the successful arrival to the traffic channel part in a cellular system is following the Poisson rule.

5.2 RANDOM ACCESS CHANNEL AND TRAFFIC CHANNEL IN GSM

A single base transceiver station is considered where maximum M voice calls are allowed to simultaneously transmit information. Only voice calls are considered for simplicity. Assume that the arrival of all mobile calls is Poisson distributed with a mean λ_G. The number of random access slots is x^1. The selection of a random access slot is random in nature. Therefore, the Poisson arrival calls are distributed uniformly over the random access slots. Thus, the call arrival into each random access slot is also Poisson distributed. The output process of different packet broadcasting systems with contention is studied in [81]. We have assumed that the output of *each* random access slot is Poisson with an average value amounting to the Slotted ALOHA throughput. Since the sum of Poisson arrivals is also Poisson, the overall output of the random access part (also the input of the traffic channel part) is Poisson where the mean is the sum of multiple output mean values corresponding to each random access slot. The system model is shown in Fig. 5.1. A more detailed Figure is presented in Publication 6 [P6].
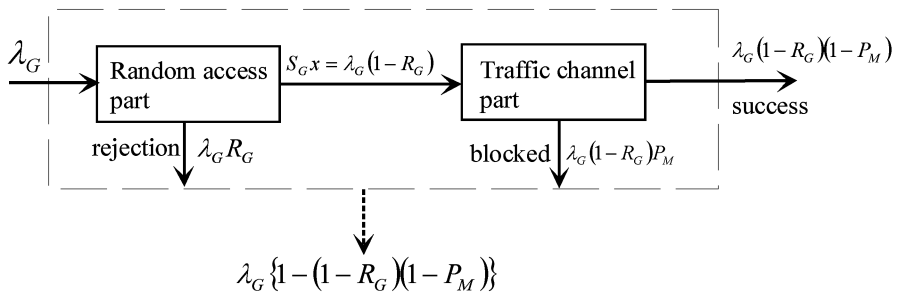

$$\lambda_G \qquad \text{Random access part} \qquad S_G x = \lambda_G (1 - R_G) \qquad \text{Traffic channel part} \qquad \lambda_G (1 - R_G)(1 - P_M)$$

$$\text{rejection} \quad \lambda_G R_G \qquad\qquad \text{blocked} \quad \lambda_G (1 - R_G) P_M$$

$$\lambda_G \{ 1 - (1 - R_G)(1 - P_M) \}$$

Figure 5.1: Random access part and traffic channel part.

After a successful reception of an access packet through the random access part, the base transceiver station allocates a Traffic CHannel (TCH) for that mobile call to transmit its voice bursts if the network has that TCH available. The traffic channel occupancy for GSM voice users pertaining to the random access channel is discussed in this chapter.

A GPRS call transmits its 'bursty' packets occupying 0 to 8 slots from a single time frame if needed and if the base transceiver station has the free slots available. If the slots are not available during that time, the packets are buffered and transmitted in the next available slots. For a given QoS (in terms of outage probability) of existing voice services, the number of slots occupied in each time frame should be limited [83]. The traffic channel occupancy of

---

[1] x is basically the random access slot rate. In the GSM and GPRS specifications, x is always defined as an integer number of slots, within a given time, which is nothing but the number of slots per unit time. This is why slot rate is defined as the smallest integer of the slot rate. Throughout the thesis we will call x as 'the number of random access slots' instead of 'the random access slot rate'.

GPRS for the multislot occupied real-time traffic pertaining to the random access channel is discussed in Publication 6 [P6]. In this chapter, we consider only the voice calls for a limited radio resource in a base transceiver station. We will not consider GPRS in this chapter. For this reason we did not mentioned GPRS in the beginning of this chapter. Interested readers investigating the bridge between the random access channel and traffic channel for GPRS are referred to Publication 6 [P6].

5.3 THE SIMULATION MODEL

In the simulation, we have developed two separate parts: the random access part and the traffic channel part. The input process of the random access part is the same as described in Chapter 3. In Figure 5.1, a random access system that consists of x parallel slots are considered. If the throughput of each random access slot is S_G, then the output rate of all x slots is $S_G x$.

We have simulated a system whose random access part consists of 20 000 slots. To obtain the output, we selected x number of slots (x is integer) out of 20 000 slots by random selection. We have repeated the selection process 10 000 times with the same number of selected slots (x), out of the same 20 000 slots, but by choosing them each time at random. Then we have taken the average of successful slots out of x slots. That is, how many slots on average are successful out of x slots? Precisely, it provides the value of $S_G x$. Therefore, there are x random access slots in each time frame. The time frame is an important parameter which will become clearer in the traffic channel part. Obviously, the output average number, $S_G x$, might be a fractional number and it was extrapolated to an integer so as to consider it as the call arrival to the traffic channel part.

Therefore, in the simulation, we consider the successful output from the random access part as the call arrival to traffic channel part. This might be one of the reasons of a slight difference between simulation and analytical results. It should be noted that we will assume the Poisson arrival model for call arrival to the traffic channel part in the analysis (Section 5.5).

The reader can immediately imagine why we need to take the average successful slots? x slots should be chosen by random selection from the random access part and the successful output slots should be considered as call arrivals to the traffic channel part. Our experiences show that the variation of the successful slots is very high if x is not very large. If we simulate the output of the *traffic channel part* with these various results, some final results are exactly similar to the analytical results and a few of them are quite different. The basic reason for this is, that a significant percentage of access packets (those are showing the results) might not be available, if the value of x is not very large. This is why we considered the average successful rounded integer value of slots as the representative call arrival to the traffic channel part.

The call arrival to the traffic channel part is considered as the successfully (integer number of) transmitted packets through the random access part as described above. Please note that we assumed the Poisson arrival model in the traffic channel part for this analysis.

Forty traffic channels are considered in the traffic channel part. A time frame was considered in the simulation. If this time frame duration is a minute, then the unit of time will be a minute.

The average successful integer number of slots out of x slots is considered as the call arrival to the traffic channel part in each time frame. Therefore, it is an analogy that there exists x number of random access slots in each time frame.

An exponentially distributed call holding time was considered. Two different call holding times were considered in the simulations and analysis. Average values of these two different exponentially distributed times are 5 time frame and 10 time frame respectively. If all 40 channels are occupied, then any new call arriving during that time is lost. Traffic channel occupancy is the ratio of occupied channels in all time frames to the total number of channels in all time frames. Blocking probability is defined as the ratio of the total number of lost calls to the total number of arrival calls. We consider 10 000 time frames in the simulation.

Please note that, we used 'unit time' only in Figures 5.2 to 5.6 to describe the time frame duration as mentioned above. We ignored the 'unit time' words in the text to avoid the haze.

5.4 THE NUMBER OF RANDOM ACCESS SLOTS

There are five different structures of the RACH [46, 54, 84] with approximately 400,000 and n*780,000 RACH slots per hour (n= 1, 2, 3, 4). It is well known (also shown in Section 3.3) that the Slotted ALOHA shows maximum throughput when the aggregate packet generation rate from all active GSM calls per each random access slot is $G = 1$. If the aggregate arrival rate is more than this value, the throughput decreases sharply [P3]. Considering this case in the random access part, the relation between the access arrival rate and the number of slots at optimum performance becomes [85, P7]

$$x = \left\lceil \lambda_G e \left\{ 1 - \left(1 - e^{-1} \right)^{r+1} \right\} \right\rceil = \lceil w \rceil \tag{5.1}$$

where r is the number of retransmissions and $\lceil w \rceil$ is the smallest integer $\geq w$. The numerical representation of Eq. (5.1) is depicted in Figure 5.2. Since this equation is derived based on optimum performance, we can call this term as 'the minimum required number of slots'. Evidently, the minimum required number of slots x is higher if the number of retransmissions is high. Remarkably, this higher slot rate implementation stimulates lower rejection probability and thus enhances the output traffic of the random access part only with the higher number of retransmission number r. In the case of lower retransmission number r, the rejection probability is high. With a lower retransmission number r, and the excess random slot implementation can not reduce the rejection probability. This is why, implementation of excess random access slots with a lower retransmission number interpreted as a waste of radio resources.

The overall successful arrival rate into the traffic channel part from all active calls is $S_G x$ (Figure 5.1). Note that the value of S_G should always be less than one, but the value of

$S_G x$ might be more than one; of course its value should be less than or equal to λ_G. Figure 5.3 depicts the random access carried traffic (throughput rate) $S_G x$ and the random access rejection traffic (packet rejection rate) $\lambda_G R_G$ with the variation of call arrival rate λ_G, considering the number of random access slot $x = 20$. Clearly, the call arrival λ_G is the sum of two different parts: the access part carried traffic $S_G x$ and rejection traffic $\lambda_G R_G$.
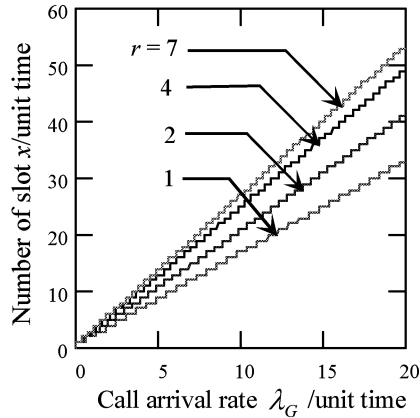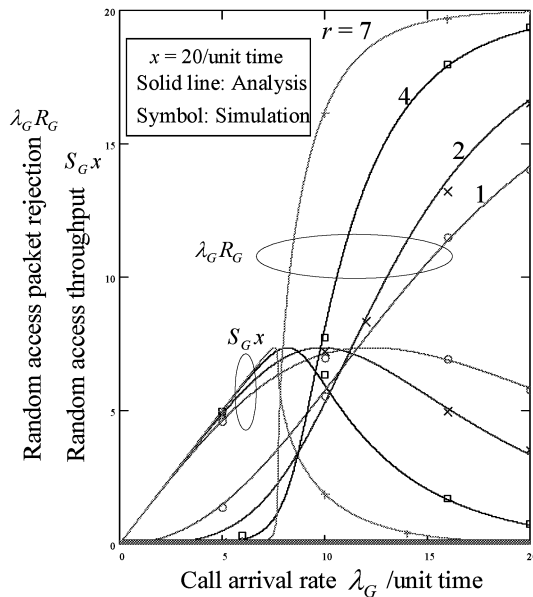

Figure 5.2: Number of random access slots x.

Figure 5.3: Call arrival rate, throughput rate and rejection rate of random access part.

5.5 THE BRIDGE BETWEEN THE RANDOM ACCESS CHANNEL AND THE TRAFFIC CHANNEL

According to GSM specifications, a maximum of r retransmissions is allowed for each mobile call during the access period. The parameter r can be set to four different possible values 1, 2, 4 or 7. A higher retransmission number r reduces the access rejection probability, and hence $r = 7$ is a widely used setting parameter [12, 65]. Considering this particular case, including the first transmission, the maximum allowed transmission number is 8. The retransmission cut-off scheme is studied in Chapter 3 (and Publication 3) and accordingly, the throughput of the random access part is redefined as $S_G x = \lambda_G \left[1 - \left\{ 1 - \exp(-G) \right\}^8 \right]$.

Consider only voice calls, where a voice user can hold one traffic channel at a time. An assumption is made in this chapter that the random access slot rate is x slots per unit time and those slots are exclusively used for voice calls. If the output of each random access slot is S_G, then the average successful call arrival rate to the traffic channels is $S_G x$ (Figure 5.1). If a voice call is successfully set up (neither rejected nor blocked), it holds a traffic channel for the entire conversation period of its user. Let us further assume that the traffic channel (call) holding time of each user is independent and exponentially distributed with a mean $1/\mu_G$. If a maximum of M voice calls are allowed to transmit voice at a time, the probability that voice calls occupy m traffic channels is

$$P_m = \frac{\left(S_G x / \mu_G \right)^m / m!}{\sum_{i=0}^{M} \left(S_G x / \mu_G \right)^i / i!} \tag{5.2}$$

The traffic channel occupancy is the ratio of the average number of occupied traffic channels to the total number of traffic channels [P6],

$$U = \frac{1}{M} \sum_{m=0}^{M} m P_m = \frac{\left(S_G x / \mu_G \right)}{M} \left[1 - P_M \right] \tag{5.3}$$

where P_M is the probability that a voice call is blocked because of traffic channel saturation and it is given by:

$$P_M = \frac{\left(S_G x / \mu_G \right)^M / M!}{\sum_{i=0}^{M} \left(S_G x / \mu_G \right)^i / i!} \tag{5.4}$$

The variation of traffic channel occupancy with the average call arrival rate for different number of random access slots x and different call holding time $1/\mu_G$ is shown in Figure 5.4. The utilization is deteriorating substantially, if *less* random access slots are involved.

Another important parameter is the call blocking probability. A blocked call is defined as that call is being successful through the random access part, but the network (base transceiver station) has no traffic channel to serve the call. Thus, the probability of a call successful through random access channel, but can not served by traffic channels, is known as call blocking probability. Eq. (5.4) shows the analytical result of the call blocking probability and Figure 5.5 shows the graphical representation of the same (call blocking probability).
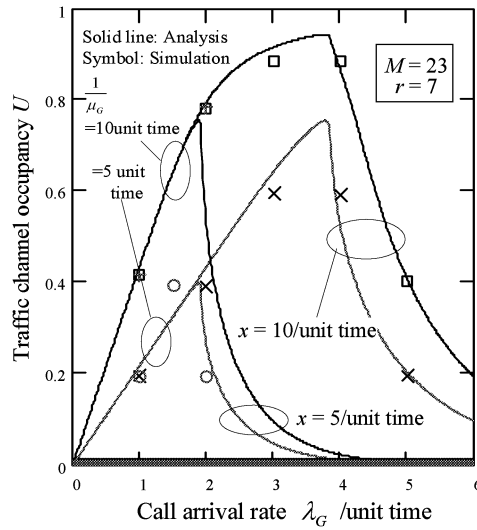

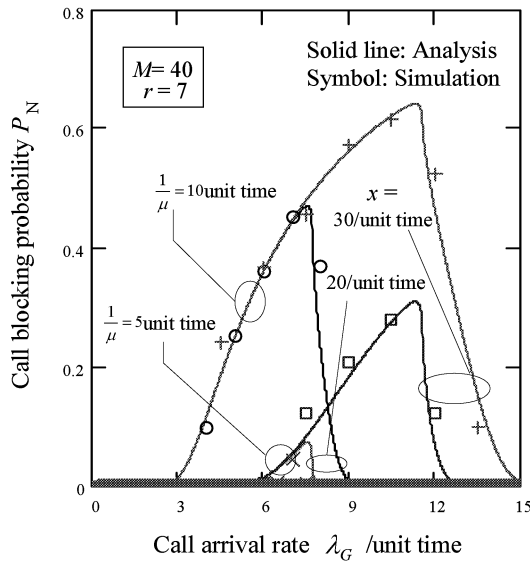Figure 5.4: Traffic channel utilization vs. the average access arrival rate.

Figure 5.5: Call blocking probability.

5.6 CALL SET UP PROBABILITY

In a given base transceiver station the number of traffic channels is limited. In the existing GSM system, free traffic channels are allocated to a circuit-mode voice call on a reservation basis for the whole conversation period: 'one traffic channel for one voice call'. A new voice call into a given area is *blocked* when all traffic channels in the base transceiver station serving that area are occupied by existing voice calls.

Call blocking probability therefore defines the probability that a new call will be blocked due to lack of unoccupied traffic channels in the base transceiver station. The call set up probability, i.e., the probability that a new call will be successfully assigned a traffic channel can be derived using an intuitive approach. This can be simply taken as one minus call blocking probability. This assumption is valid for fixed networks where all new call requests compete for a channel allocation. However, for wireless cellular networks, such as GSM, new call requests have to "break the barrier" of the random access part before they can be processed any further by the base transceiver station.

With random access, call request packets from several mobile calls may collide in the random access slots resulting in loss of packets. Unlimited retransmission attempts of lost packets would guarantee a one hundred percent success rate in the random access part, in which case the "one minus blocking probability" would correctly define the call set up probability in GSM networks. However, for practical purposes, a retransmission cut-off is enforced whereby a mobile call terminates the access attempt mostly after seven unsuccessful retransmissions. With retransmission cut-off in place, there is a non-zero probability that a new call request may not reach the base transceiver station to compete for a traffic channel allocation. Call rejection probability is the term often used to refer to the failing probability of a call request packet through the random access part.

It is now clear that for an accurate analysis of the call set up probability in GSM networks, one has to account for both call rejection probability in the random access part and the call blocking probability at the traffic channel part. Earlier work in this area either concentrated on call blocking probability at the base transceiver station without any reference to call rejection probability [56-62, P4, P5], or they analyzed the random access part in isolation to study the dynamics of random access protocols and systems [25-47, P1-P3]. This chapter consolidates these isolated works with the objective of defining a more accurate model to study the call set up probability in wireless cellular networks in general and GSM networks, to be specific.

Considering that a call is rejected in the random access part with a probability R_G, the input probability of the traffic channel part is $(1 - R_G)$. A call is blocked in the traffic channel part with a probability P_M. Therefore, the probability that a call is successfully set up (neither rejected nor blocked) is given by

$$T = (1 - R_G)(1 - P_M) \tag{5.5}$$

We are more interested in the access state to give full availability for all kinds of GSM users. In the second stage, if all traffic channels are full and the network is unable to provide services, the network informs the mobile terminal of its *blocked* call. On the contrary,

if the call is *rejected* in the access stage, it cannot receive any message from the base transceiver station. Physically, it is unknown to the network that one terminal tries to access to transmit its information. In this case, the mobile terminal may receive an automatic "access rejection" signal, which may request the mobile user to try again. The mobile user may think that the network is full and attempt to call later. This is true for the case of call blocking but not for call rejection. Considering this situation for call blocking and call rejection, it can be said that *call blocking* is a better situation than that of *call rejection*.

A numerical representation of Eq. (5.5) is depicted in Figure 5.6 for a few sets of specific values of different parameters that constitute two separate regions:

1. If $\lambda_G < xe^{-1}\left\{1-\left(1-e^{-1}\right)^8\right\}^{-1}$:

   - The call rejection probability is almost zero.
   - The call blocking probability increases linearly and thus the call set up probability decreases linearly.
   - The steepness of these two curves is proportional to the average call holding time $1/\mu_G$, and inversely proportional to the number of traffic channel M.

2. If $\lambda_G \geq xe^{-1}\left\{1-\left(1-e^{-1}\right)^8\right\}^{-1}$:

   - The call rejection probability increases abruptly. Thus, most of the calls are rejected in the random access part.
   - The arrivals to the traffic channel part decreases and thus the call blocking probability also decreases.
   - The overall call set up probability T, decreases abruptly because of a sharp increase in call rejection probability R_G.

Finally, it must be emphasized that the overall call set up probability can be definitely calculated using the traditional Erlang call blocking formula if and only if the number of random access slots is more than $\left\lceil \lambda_G e\left\{1-\left(1-e^{-1}\right)^8\right\}\right\rceil$.

Solid line: Analysis
Symbol: Simulation

$\dfrac{1}{\mu_G} = 10$ unit time

$x = 20$/unit time

$M = 40$

$r = 7$

Set up

Rejection

Blocking

Y-axis: Probability

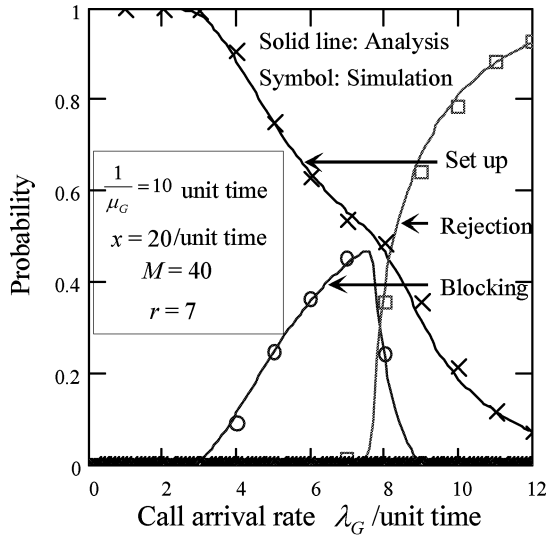X-axis: Call arrival rate λ_G /unit time

Figure 5.6: Rejection, blocking and set up probabilities.

5.7 CHAPTER SUMMARY

We have developed a more accurate model for analyzing the call set up probability in wireless cellular networks. Using this model, we have presented the analysis of call set up probability in GSM networks. Our analysis shows that the number of random access slots provisioned in the wireless access system has a significant impact on the overall call set up probability for a given call arrival rate. To be more precise, there exists an optimal choice for the number of random access slots, below which the call set up probability drops drastically. Simulation results validate theoretical findings. We have derived a closed form expression to compute this optimum number of random access slots. Our results will help wireless cellular network operators and vendors to properly dimension their wireless access systems.

6. Conclusions

6.1 Conclusions

This thesis optimized the radio interface protocols of a TDMA based cellular system like GSM. A few analytical methods were developed in [P1-P7]. The verification of most of those analytical methods is presented in this thesis by means of simulations.

In the first part of the thesis (Chapter 2), three throughput collapse preventing algorithms are presented. In the implementation of these algorithms, a central entity (e.g., base transceiver station) is not needed. The central entity will assign some feedback information about the network load for maintaining very high throughput under heavy load. It has also been shown in Publication 2 [P2] that the auto-controlled algorithm demands remarkable throughput performance when compared to the traditional random power level selection algorithm particularly with a very high retransmission probability. Thus, the auto-controlled algorithm prevents the throughput collapse in that high retransmission probability region and consequently increases the system stability.

In the retransmission cut-off scheme, if the first transmission is unsuccessful the mobile call may retransmit the same request within the next few slots by random selection. It is obvious that the retransmission attempts are not independent and it can *not* be regarded as a Poisson arrival model. In this thesis and in [P3], an analytical model is proposed, where the first transmissions and the subsequent dependent retransmissions are assumed to be Poisson arrival processes. Fortunately the analytical results are almost the same as the simulated results and thus justify the assumption. People usually believe that the stability of Slotted ALOHA can be achieved by limiting the number of retransmissions. It is proved in Publication 3 that the system stability can be achieved if the generation rate of new packets is less than a critical limit, irrespective of the number of retransmissions. The close form analytical result of that critical limit is shown in the thesis and more details about this critical limit can be found in [P3].

The presence of continuous multislot occupied HSCSD services increases the blocking probability of the voice calls. It is found that, in a given base transceiver station, for the same voice blocking probability, the overall channel occupancy increases with the higher number of allowed HSCSD calls, as well as with the higher number of channels to be allocated for each HSCSD call. This achievement can be obtained either by reducing the voice traffic intensity or by increasing the number of traffic channels in a given base transceiver station. Bursty real-time traffic over the packet switched GPRS transmission system is inefficient in terms of channel occupancy. It is shown in the thesis and in [P5] that the channel occupancy can be increased at the expense of RLC/MAC block (radio packet) dropping.

It is interesting to know the output (carried traffic) of the Slotted ALOHA with retransmission cut-off. It is assumed that the output process of the Slotted ALOHA with retransmission cut-off is Poisson distributed and the analytical model is developed as described in Chapter 5. Due to the assumption, the analysis is approximate and thus computer simulation is used to validate it. In the simulation, we insert the output traffic of the Slotted

ALOHA with retransmission cut-off directly into the traffic channel part. The simulation results match well with the analytical results and validate the assumption.

The random access part and traffic channel part are utilized, respectively, for call reaching in the base transceiver station and information transmission in the uplink direction (from mobile to base transceiver station) of a wireless cellular network such as the Global System for Mobile communication (GSM). A call is either rejected or blocked depending on its inability to succeed either in the random access part or in the traffic channel part. It is shown that the number of random access slots provisioned in the wireless access system has a significant impact on the overall call set up probability for a given call arrival rate. To be more precise, there exists an optimal choice for the number of random access slots, below which the call set up probability drops drastically. It is shown that the optimum number of random access slots is directly proportional to the average call arrival rate, being independent of the average channel holding time and the number of traffic channels. A complete analysis is executed for the traffic channel utilization and call blocking probability with an exact number of random access slots that provide almost zero call rejection probability. The overall call set up probability is devised considering call rejection and call blocking probabilities.

6.2 FURTHER RESEARCH

In a cellular system, a mobile terminal close to the base station has always an advantage over a mobile terminal far from the base station to transmit information. A terminal closer to the base station does not require perfect power control. Path-loss, fading or shadowing is not considered in the present study. The multiple power level transmission system including path-loss, fading and shadowing might be one interesting topic for future study. The multiple power level transmission system can nullify (at least partially) the advantages of the mobiles close to the base station. Definitely, the multiple power level transmission system is a simpler technology than that of perfect power control. Therefore one interesting future topic is how this simpler technology can be improved to approach the capability of perfect power control.

In Slotted ALOHA with a retransmission cut-off scheme, it is found that for a given critical limit of the generation rate of new packets, a limited number of retransmissions is not needed for stable operation. The exact number of retransmissions for a given new packet generation rate exceeding the critical limit, is derived. Therefore, a central entity (e.g. base transceiver station) needs to account for the fact that the mobile terminals crossed that critical limit and to request the mobile terminal to change the number of retransmissions. Hence, it is not so attractive from the implementation point of view. To solve this problem, throughput collapse preventing algorithms can be included with the retransmission cut-off scheme.

In this work, we considered the traffic channel occupancy without considering path-loss, fading and shadowing. Traffic channel occupancy with those factors is one of the interesting topics for further study. The radio resources for GPRS in a given base transceiver station can vary depending on circuit-mode traffic. It is considered here for simplicity that a fixed amount of radio resources are available for serving real-time traffic. The combined channel occupancy is considered in [P4, P7], where single channel occupied voice calls and

multichannel occupied HSCSD calls are working together. It is also interesting to find the channel occupancy, where single channel occupied voice calls and dynamically changing multiple slots occupied GPRS calls will work together.

Finally, the random access channel and traffic channel are considered together for a successful call set up procedure in a cellular system like GSM. For simplicity, only Rayleigh fading channels in the random access part (in Publication 6), are considered. The exact number of random access channels that provides almost zero call rejection probability was found. It is interesting and necessary to include path-loss and shadowing in the traffic channel part to incorporate in the random access part. After including path-loss and shadowing also in the traffic channel part, the obtained value of the optimum number of random access slots will be more precise and it will provide motivation for other future work.

REFERENCES

[1]   http://www.gsmworld.com/news/statistics/substats.shtml

[2]   S. Ohmori, Y. Yamao and N. Nakajima, "The future generations of mobile communications based on broadband access methods", *Wireless Personal Communications*, Vol. 17, No. 2-3, June 2001, pp. 175-190.

[3]   A. Jajszczyk, "What is the future of telecommunications networking" *IEEE Communications Magazine*, Vol. 37, No. 6, June 1999, pp. 12 –20.

[4]   S. Nanda, K. Balachandran, S. Kumar, "Adaptation techniques in wireless packet data services", *IEEE Communications Magazine*, Vol. 38, No. 1, Jan. 2000, pp. 54 –64.

[5]   3GPP pages: http://www.3gpp.org/

[6]   J.-H. Park, "Wireless Internet access for mobile subscribers based on the GPRS/UMTS network", *IEEE Communications Magazine*, Vol. 40, Issue: 4, Apr. 2002, pp. 38 –49.

[7]   S. Dixit, G. Yile and Z. Antoniou, "Resource management and quality of service in third generation wireless networks", *IEEE Communications Magazine,* Vol. 39, No. 2, Feb. 2001, pp. 125 –133.

[8]   G. H. Sasaki, Ching-Fong Su, "The interface between IP and WDM and its effect on the cost of survivability", *IEEE Communications Magazine*, Vol. 41, No. 1, Jan. 2003, pp. 74-79.

[9]   A. Gyasi-Agyei, S. J. Halme and J. H. Sarker, "GPRS-features and packet random access channel performance analysis", *IEEE International Conference on Networks*, 2000, (ICON 2000), pp. 13 –17.

[10]   W. Mohr and W. Konhäuser, "Access network evolution beyond third generation mobile communications", *IEEE Communications Magazine*, Vol. 38, No. 12, Dec. 2000, pp. 122-133.

[11]   M. Ivanovich and M. Zukerman, "A study of deadlock models for a multiservice medium access protocol employing a slotted signalling channel," *IEEE/ACM Transactions on Networking*, Vol. 8, No. 6, December 2000.

[12]   M. Mouly and M.-B. Pautet, "The GSM system for mobile communications", Published by authors, 1992.

[13]   W. Mohr, "Development of mobile communications systems beyond third generation", *Wireless Personal Communications*, Vol. 17, No. 2-3, June 2001, pp. 191-207.

[14]   D. Sant, "Throughput of unslotted ALOHA channels with arbitrary packet interarrival time distributions", *IEEE Transactions on Communications*, Vol. 28, Issue 8, Aug. 1980, pp. 1422 –1425.

[15]   B. B. Mandelbrot, *"The fractal geometry of nature"*, W. H. Freeman, Ed. San Francisco, CA, 1982.

[16]   W.E. Leland, M. S. Taqqu, W. Willinger and D.V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)" *IEEE/ACM Transactions on Networking*, Vol. 2 Issue 1, Feb. 1994, pp. 1 –15.

[17]   M. E. Crovella, A. Bestavros, "Self-similarity in World Wide Web traffic: evidence and possible causes", *IEEE/ACM Transactions on Networking*, Vol. 5, Issue 6, Dec. 1997, pp. 835 –846.

[18]   W. Willinger, M. S. Taqqu, R. Sherman and D. V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level", *IEEE/ACM Transactions on Networking*, Vol. 5, Issue 1, Feb. 1997, pp. 71-86.

[19]   A. T. Andersen and B. F. Nielsen, "An application of superpositions of two state Markovian source to the modelling of self-similar behaviour" *INFOCOM '97*, Vol. 1, 1997, pp. 196 –204.

[20]   X. Yang and A. P. Petropulu, "The extended alternating fractal renewal process for modeling traffic in high-speed communication networks" *IEEE Transactions on Signal Processing*, Vol. 49, Issue 7, July 2001, pp. 1349 –1363.

[21]   V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling", *IEEE/ACM Transactions on Networking*, Vol. 3, Issue 3, June 1995, pp. 226 –244.

[22]   R. Jain and S. A. Routhier, "Packet trains: measurements and a new model for computer network traffic", *IEEE Journal of Selected Areas in Communications*, Vol. 4, 1986, pp. 986-995.

[23]   J. Aracil, L. Munoz, "Performance of aloha channels under self-similar input" *Electronics Letters*, Vol. 33, Issue 8, 10 April 1997, pp. 659 –660.

[24]   Z. Harpantidou, M. Paterakis, "Random multiple access of broadcast channels with Pareto distributed packet interarrival times", *IEEE Personal Communications* [see also *IEEE Wireless Communications*], Vol. 5, Issue 2, April 1998, pp. 48 –55.

[25]   N. Abramson, "Fundamentals of packet multiple access for satellite networks", *IEEE Journal on Selected Areas in Communications*, Volume: 10, Issue: 2, Feb. 1992, pp. 309 – 316.

[26]   Wei Luo and A. Ephremides, "Power levels and packet lengths in random multiple access", *IEEE Transactions on Information Theory*, Volume: 48, Issue: 1, Jan. 2002, pp. 46 – 58.

[27]   Jianbo Gao and I. Rubin, "Analysis of a random-access protocol under long-range-dependent traffic", *IEEE Transactions on Vehicular Technology*, Vol. 52, Issue 3, May 2003, pp. 693 –700.

[28]   D. Raychaudhuri and K. Joseph, "Performance evaluation of slotted ALOHA with generalized retransmission backoff", *IEEE Transactions on Communications*, Vol. 38, No. 1, Jan. 1990, pp. 117 –122.

[29]   D. G. Jeong and W. S. Jeon "Performance of an exponential backoff scheme for slotted-ALOHA protocol in local wireless environment", *IEEE Transactions on Vehicular Technology*, Vol. 44, Aug. 1995, pp. 470 –479.

[30]   J. Alcober and V. Casares, "Throughput and delay analysis of the ideal window stabilized ALOHA", *European Transactions on Telecommunications,* Vol. 10, No.5, Sept.-Oct. 1999, pp. 505-511.

[31]   J. J. Metzner, "On improving utilization in ALOHA networks", *IEEE Transactions on Communications* April 1976, pp. 447-448.

[32]   N. Shacham, "Throughput-delay performance of packet switching multiple-access channel with power capture", *Performance Evaluation*, 4 (1984), pp. 153-170.

[33]   C. C. Lee, "Random signal levels for channel access in packet broadcast networks", *IEEE Journal of Selected Areas in Communications,* Vol. SAC-5, No. 6, July 1987.

[34]   J. H. Sarker and M. Hall, "A novel (Multi)[3] random access model in packet broadcast networks", (*MDMC '96) The 2nd International Workshop on Multi-Dimensional Mobile Communications*, Seoul, Korea, July 18-20 1996, pp. 711-715.

[35]   J. H. Sarker, "A random access system with multiple copies and random power levels selection", *IEEE International Symposium on Wireless Communications (ISWC'98)*, Montreal, Quebec, Canada, (May 22-23, 1998) p. 3.

[36]   S. Shimamoto, Y. Onozato and S. Teshigawara, "Performance evaluation of power level division multiple access (PDMA) scheme", *IEEE International Conference on Communications 92 (ICC'92)*, 1992, pp. 1333-1337.

[37]   Y.-W. Leung, "Fair power level division multiple access", *IEEE Communications Letters*, Vol. 1, No. 3, May, 1997, pp. 61-63.

[38]   Y.-W. Leung, "Generalised multicopy ALOHA", *Electronics Letters*, vol. 31, No. 2, Jan. 9, 1995, pp. 82-83.

[39]   J. H. Sarker and S. J. Halme, "The prudence transmission method I (PTM I): a retransmission cut-off method for contention based multiple access communication systems", *IEEE Vehicular Technology Conference 97, (VTC'97)*, USA, May 4-7, 1997, pp. 397-401.

[40]   K. Sakakibara, H. Muta, and Y. Yuba, "The effect of limiting the number of retransmission trials on the stability of slotted ALOHA systems", *IEEE Transactions on Vehicular Technology*, Vol. 49, No. 4, July 2000, pp. 1449 –1453.

[41]   E. W. M. Wong and T.-S.P. Yum, "The optimal multicopy Aloha", *IEEE Transactions on Automatic Control*, Vol. 39 Issue: 6, June 1994, pp.1233 –1236.

[42]   L. Kleinrock and S. S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation", *IEEE Transactions on Communications*, Vol. 23, April 1975, pp.410-423.

[43] Sung Hyun Cho and S. H. Park, "Stabilised random access protocol for voice/data integrated WCDMA system", *Electronics Letters*, Vol. 37, Issue 19, 13 Sept. 2001, pp. 1197 –1199.

[44]   K. Sakakibara, "Performance approximation of a multi-base station slotted ALOHA for wireless LAN's", *IEEE Transactions on Vehicular Technology*, Vol. 41, No. 4, Nov. 1992, pp. 448 –454.

[45]   S. W. Kim, "Frequency-hopped speed-spectrum random access with retransmission cutoff and code rate adjustment", *IEEE Journal of Selected Areas in Communications*, vol. 10, No. 2, Feb. 1992, pp. 344-349.

[46]   C. Luders and R. Haferbeck, "The performance of the GSM random access procedure", *IEEE Vehicular Technology Conference 94, (VTC' 94)*, Stockholm, Sweden, June 8-10, 1994, pp. 1165-1169.

[47]   J. H. Sarker and S. J. Halme, "Performance analysis and optimization of the GSM-GPRS random access channel in a Rayleigh fading channel", *5th International Conference on Information Systems Analysis and Synthesis (ISAS'99)*, Orlando, Florida, USA, July 31-August 4, 1999, Volume 7, pp. 172-176.

[48]   M. Schwartz, "Telecommunication networks: protocols, modeling and analysis", Addison-Wesley, 1987.

[49]   N. Abramson, "The ALOHA system-another alternative for computer communications", *Fall Joint Computer Conference, (AFIPS Conf. Proc.), 1970*, vol. 37, AFIPS Press, 1970, pp. 281-285.

[50]   J. Zander and S.-L. Kim, "Radio resource management for wireless networks", Artech House, 2001.

[51]   Davenport, "Probability and random processes", McGraw-Hill.

[52]   E. Parzen, "Stochastic Processes", Holden-Day, Inc., Publishers, San Francisco, 1962.

[53]   ETSI-GSM Technical Specification, GSM 04.08, Version 4.4.0, "Mobile radio interface-layer 3 specification", April 1993.

[54]   ETSI EN 300 940, GSM 04.08 Version 7.1.2, "Digital cellular telecommunications system (Phase 2+), Mobile radio interface layer 3 specification", 1998.

[55]   GSM 05.02 version 8.2.0, "Digital cellular telecommunications systems (Phase 2+); multiplexing and multiple access on the radio path", Release 1999.

[56]   J. H. Sarker, "A consistent time division multiple access (CTDMA) for integrated voice and data wireless networks", *International Conference on Telecommunications (ICT'97),* 2-4 April, 1997, pp. 1211-1216.

[57]   D. A. Menasce and V. A. F. Almeida, "Capacity planning for Web performance: matrics, models, and methods", *Prentice-Hall*, 1998.

[58]   D. Calin and D. Zeghlache, "Analysis of GSM HSCSD service with channel allocation constraints", *Multiaccess, Mobility, and Teletraffic for Wireless Communications: Volume 3*, Ed. By K. Leung and B. Vojcic, Kluwer Academic Publishers, 1999, pp. 237-252.

[59]   A. A. Nilsson, M. Perry, A. Gersht and V. B. Iversen, "On multi-rate Erlang-B computations", 16th *International Teletraffic Congress, (ITC 16)*, 1999, pp. 1051-1060.

[60]   J. S. Kaufman, "Blocking in a shared resource environment", *IEEE Transactions on Communications*, Vol. COM-29, No. 10, October 1981, pp. 1474-1481.

[61]   J. M. Aein, "A multi-class, blocked-calls-cleared, demand access model", *IEEE Transactions on Communications*, Vol. COM-26, No. 3, March 1978, pp. 374-385.

[62]   M. Stasiak, "Combinatorial considerations for switching systems carrying multi-channel traffic streams", 14th *International Teletraffic Congress (ITC 14)*, 1994, pp. 1229-1238.

[63]   J. W. Roberts, "Traffic theory and the Internet", *IEEE Communications Magazine*, Vol. 39, Issue 1, Jan. 2001, pp. 94 –99.

[64]   D. Wright, "Voice over MPLS compared to voice over other packet transport technologies", *IEEE Communications Magazine*, Vol. 40, No. 11, Nov. 2002, pp. 124 –132.

[65]   G. Brasche and B. Walke, "Concept, services, and protocols of the new phase 2+ general packet radio services", *IEEE Communications Magazine*, August 1997, pp. 94-104.

[66]   J. Cai and D. J. Goodman, "General packet radio service in GSM", *IEEE Communications Magazine,* October 1997, pp. 122-131.

[67]   GSM 03.64, V5.1.0, "Digital cellular telecommunication system (Phase 2+); general packet radio service (GPRS); overall description of the GPRS radio interface (Stage 2)," November 1997.

[68]   J. Roberts, U. Mocci and J. Virtamo (Eds.), "Burst scale loss systems", *Broadband Network Teletraffic, Final Report of Action, COST 242.*

[69]   R. L. Freeman, "Telecommunication system engineering", Third Edition, John Wiley &Sons, 1996.

[70]   3GPP TS 03.64 V8.12.0 (2004-04), "Radio access network; general packet radio service (GPRS); Overall description of the GPRS radio interface; Stage 2", (Release 1999).

[71]   D. Lenton, "Big push is on [mobile phones]" *IEE Review*, Vol. 50, Issue: 1, Jan. 2004, pp. 28 – 29.

[72]   S. N. Fabri, S. Worrall, A. Sadka and A. Kondoz, "Real-time video communications over GPRS", *First International Conference on 3G Mobile Communication Technologies, 2000.* (IEE Conf. Publ. No. 471), 27-29 March 2000, pp. 426 – 430.

[73]   S. Nanda, D. J. Goodman and U. Tomor, "Performance of PRMA: a packet voice protocol for cellular systems", *IEEE Transactions on Vehicular Technology*, Vol. 40, Issue: 3, Aug. 1991, pp. 584 – 598.

[74]   N. Medman, K. Svanbro and P. Synnergren, "Ericsson instant talk", "Ericsson Review", No. 1, 2004, pp. 16-19.

[75]   J. H. Sarker, "Voice and data transmission over the TDMA based networks" *Master's thesis*, 1996, Helsinki University of Technology, Finland.

[76]   J. H. Sarker, "IP traffic transmission over the general packet radio services", *International Conference on Communications Technology, ICCT 2003*, Beijing, China, April 9-11, 2003, pp. 263-266.

[77]   D. Soldani, R. Cuny, "On the deployment of multimedia services in wireless networks: radio dimensioning aspects for UTRAN FDD", *First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, 2004, QSHINE 2004*, 18-20 Oct. 2004, pp. 189 – 196.

[78]   K. Murai, S. Nakamura, "Real time face detection for multimodal speech recognition", *IEEE International Conference on Multimedia and Expo, 2002*, 26-29 Aug., 2002, pp. 373 – 376.

[79]   G. Niklfeld, M. Pucher, R. Finan, W. Eckhart, "Mobile multi-modal data services for GPRS phones and beyond", *IEEE International Conference on Multimodal Interfaces*, 2002, 14-16 Oct., 2002, pp. 337 – 342.

[80]   R. Jain, "The art of computer systems performance analysis", John Wiley & Sons, Inc., 1991.

[81]   H. Takagi and L. Kleinrock, "Output process in contention packet broadcasting systems", *IEEE Transactions on Communications*, Vol. COM-33, No. 11, Nov. 1985, pp. 1191-1199.

[82]   G. W. Tunnicliffe, A. R. Murch, A. Sathyendran and P. J. Smith, "Analysis of traffic distribution in cellular networks", *IEEE Vehicular Technology Conference 98 (VTC 98)*, pp. 1984 –1988.

[83]   S. Ni, Y. Liang and S.-G. Häggman, "Outage probability in GSM-GPRS cellular systems with and without frequency hopping", *Wireless Personal Communication*, September 2000, pp. 215-234.

[84]   ETSI-GSM Technical Specification, GSM 05.02, Version 4.2.0, "Physical layer on the radio path, multiplexing and multiple access on the radio path", April 1993.

[85]   J. H. Sarker and S. J. Halme, "Estimation of random access slots for GSM evolution", *The Third International Symposium on Wireless Personal Multimedia Communications, WPMC '00*, Bangkok, Thailand, November 12-15, 2000, pp. 1021-1026.

PUBLICATIONS 1-7