PERFORMANCE ANALYSIS AND ENHANCEMENTS OF RADIO RESOURCE CONTROL WITH APPLICATIONS TO LTE ADVANCED

by

Ali Taha Koç

APPROVED BY SUPERVISORY COMMITTEE:

Professor Murat Torlak, Chair

Professor Naofal Al-Dhahir

Professor Mohammad Saquib

Professor John P. Fonseka

PERFORMANCE ANALYSIS AND ENHANCEMENTS OF RADIO RESOURCE CONTROL WITH APPLICATIONS TO LTE ADVANCED

by

ALI TAHA KOÇ, BSE, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

May 2013

Dedicated to my mom and dad.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Professor Murat Torlak for his invaluable support, guidance and encouragement during my PhD study. My time in the PhD program at the Electrical Engineering Department of University of Texas at Dallas has been one of the most valuable experiences in my life. I would also like to thank my committee members, Professor Al-Dhahir, Professor Saquib and Professor Fonseka, for reviewing my thesis and serving in my committee. Their comments and insights have made my research more fruitful.

I would also like to thank my colleagues Satish C. Jha, Maruti Gupta and Rath Vannithamby in Mobile Wireless Platform group at Intel labs for their support during my PhD study.

Finally, I would like to express my eternal gratitude to my mother Şadiye Koç, my father Atilla Koç and my sisters Sırma and Zehra for their love and support. Throughout this long and difficult process, their love and encouragement has kept me going. There have been times when I almost lost hope, but their wonderful faith in me gave me the strength to finish.

December 2012

PREFACE

This dissertation was produced in accordance with guidelines which permit the inclusion as part of the dissertation the text of an original paper or papers submitted for publication. The dissertation must still conform to all other requirements explained in the "Guide for Preparation of Master's Theses and Doctoral Dissertations at The University of Texas at Dallas." It must include a comprehensive abstract, a full introduction and literature review, and a final overall conclusion. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported.

It is acceptable for this dissertation to include as chapters authentic copies of papers already published, provided these meet type size, margin and legibility requirements. In such cases, connecting texts which provide logical bridges between different manuscripts are mandatory. Where the student is not the sole author of a manuscript, the student is required to make an explicit statement in the introductory material to that manuscript describing the student's contribution to the work and acknowledging the contribution of the other author(s). The signatures of the Supervising Committee which precede all other material in the dissertation attest to the accuracy of this statement.

PERFORMANCE ANALYSIS AND ENHANCEMENTS OF RADIO RESOURCE CONTROL WITH APPLICATIONS TO LTE ADVANCED

Publication No. _____

Ali Taha Koç, PhD
The University of Texas at Dallas, 2013

Supervising Professor: Professor Murat Torlak

With all the developments and advances in wireless communications, we are faced with the challenge of securing scarce resources including frequency spectrum and device battery power. The objective of efficiently utilizing the scarce radio spectrum under acceptable delay and limited battery power has become even more challenging especially after the proliferation of smart phones. Scarce radio spectrum is utilized by radio resource control (RRC) which is the part of protocol stack in advanced wireless systems which handles signalling between user terminals and wireless networks. The configuration of RRC has significant impact on both efficient utilization of the spectrum and battery life.

The radio resource control mechanisms in wireless networks include Automatic Repeat Request (ARQ) and Discontinuous Reception (DRX) operations. In this dissertation, we develop performance analysis techniques for ARQ over different channel conditions and explore analytical models for DRX mechanism in advanced wireless networks.

ARQ is designed to improve the reliability of wireless networks under fading and collision via retransmitting erroneous packets. We first investigate the fading issues and present the

system model for collision free ARQ systems to optimize the number of retransmissions. Statistical analysis of collision free ARQ systems rely on the product distributions of fading channels. Existing methods to evaluate the statistical properties of product distributions either use numerical approximation or special functions in the form of infinite series. We propose a simple and efficient analytical approximation to the probability distribution of product of independent random variables. We extend our analysis to ARQ collision channels with a unified approach for throughput calculation for both single antenna and multi antenna systems.

An important part of radio resource control is device battery power saving. Discontinuous reception (DRX) is one of the key mechanisms used by current wireless networks. We present an analytical model for DRX mechanism as Semi-Markov process. Using this analytical model, tradeoff between power saving and latency of DRX mechanism is investigated. Findings are verified with analytical and simulation results over LTE advanced networks with mobile application traffic. Since the requirements of various mobile applications are getting more diverse, our analysis and corresponding algorithm development need to be tested under a variety of channel and network conditions. Thus, we present a common network simulation platform that implements realistic communication links and data traffic for advanced wireless systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The demand for wireless communication has been growing rapidly. Wireless communication services grew from 11 million subscribers worldwide in 1990 to more than 2 billion in 2005. Moreover, The International Telecommunication Union (ITU) said the number of mobile broadband subscriptions exceeded one billion in 2010 [1]. To meet the demands of these increasing number of mobile broadband users, wireless communication systems need to support reliable high data rate transmissions. Providing high data rates and reliable communications are challenging problems due to harsh and time varying channel conditions that the mobile users experience. In particular, mobility and multipath propagation add more variety into the received signal which is called fading. Fading can be overcome with increase in transmission power. Due to restrictions on maximum transmit power and also limited battery capacity of the mobile devices, increasing transmission power may not be a good solution.

Also the usage of smart phones, tablets etc. that offer internet applications like social networking, email and web browsing is increasing drastically. As such mobile devices become commonplace, user expectations also rise in terms of higher data rates and anytime-anywhere connectivity. These new expectations create power consumption issues for the mobile devices.

Wireless communication systems use multiple radio resource control (RRC) mechanisms to overcome fading and power consumption issues. In the next section, we provide an overview of these radio resource control mechanisms.

1.1 Radio Resource Control

Radio resource control which is the part of protocol stack in advanced wireless systems that handles signalling between user terminals and wireless network. The configuration of RRC has a significant impact on battery life and efficient utilization of the spectrum. The radio resource control mechanisms in wireless networks include Automatic Repeat Request (ARQ) and Discontinuous Reception (DRX) operations. In this dissertation, we develop performance analysis techniques for ARQ over different channel conditions and explore analytical models for DRX mechanism in advanced wireless networks.

1.1.1 Automatic Repeat Request (ARQ)

Automatic repeat request (ARQ) is a feedback based communication over different layers of the network stack. The general concept of ARQ is to detect packets with errors at the receiver and then to request the transmitter to repeat those packets using a feedback mechanism. Basic ACK (successful) or NACK (unsuccessful) signalling is used for feedback. The basic automatic repeat request for multi-access was developed around 1970 to provide radio communication between the central computer and various data terminals at the campuses of the University of Hawaii and called Aloha [2]. Aloha scheme is a simple mechanism whose throughput depends on the arrival rate and failures due to collisions. However, in wireless communication failure due to channel conditions are also important. In order to resolve failures due to channel conditions, ARQ schemes can be combined with the lower layer of the stack and called Hybrid-ARQ (H-ARQ). H-ARQ is a combination of channel coding and ARQ. In H-ARQ schemes, retransmissions are designed according to first transmission coding rate and success. Moreover, packets are stored even if they are unsuccessfully received and used to increase the probability of correct decoding. Two basic H-ARQ schemes are chase combining (CC) and incremental redundancy (INR). In chase combining, a NACK leads to retransmission of the same packet in the second round

of H-ARQ. In incremental redundancy, a NACK leads to a packet with more information and redundancy bits instead of the same packet. Decoders are designed to decode each round alone in simple ARQ, or using all rounds in CC and INR.

Aloha scheme exhibits fairly poor performance due to absence of centralized controller. Whenever a user has a packet to transmit, user does so without considering other users. Figure 1.1 shows the performance of the Aloha scheme vs. offered load. Offered load G lower then one causes the system being under utilized and creates unused empty slots. Offered load G bigger then one causes excessive collisions and performance degradation. With Aloha, optimum capacity is achieved only when one user accesses to the medium. However, in wireless communication systems, we have limited resources and multiple users

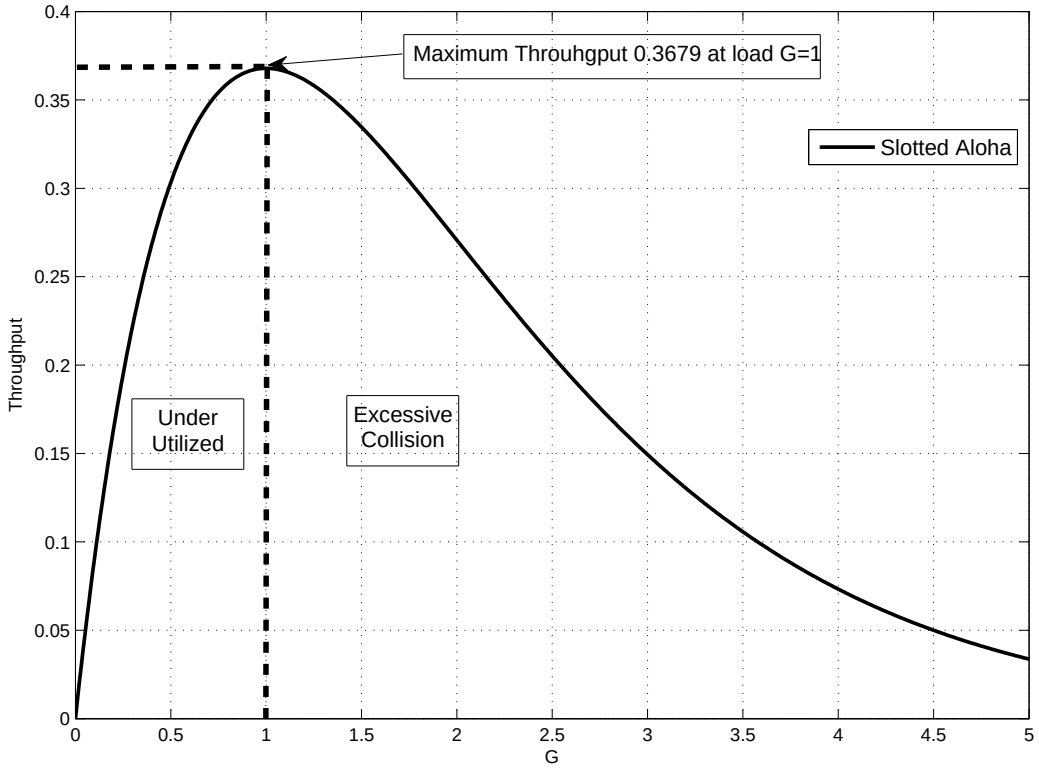

Figure 1.1. Slotted Aloha transmission

to access the same medium. Our motivation is to improve the capacity of ARQ for wireless communication systems over different channel conditions. The product of independent random variables (RVs) has an important role in studying the capacity of ARQ systems of modern wireless communication systems. The product of independent random variables (RVs) are used to calculate outage probability and ϵ-outage capacity or rate offset under various fading conditions. The capacity analysis is discussed in more detail in Chapter 2.

We also investigate the capacity of Aloha, chase combining and incremental redundancy with the information theoretic approach and improve their throughput by adding threshold and multiple antennas. Adding a predetermined threshold leads to selective access to the medium which will decrease the collision as well as the level of interference. Adding multiple antennas to the base station adds more degrees of freedom and improvements on received signal strengths. Moreover, Aloha scheme assumes that a decoding failure occurs whenever a collision happens. In wireless systems, user's packets may be decoded even if there is a collision. On the other hand, the packets may be dropped due to insufficient received signal strength even if there is no collision. Received signal strength of a user may change due to fading, shadowing, and path loss. Instead of focusing on collisions, adding information theoretic capacity to Aloha, CC and INR schemes and designing medium access control according to the achieved rate will be more effective for wireless systems. These enhancements are discussed in more detail in Chapter 3.

1.1.2 Discontinues Reception (DRX)

Long Term Evolution Advanced (LTE-A) is designed to keep up with the today's rapidly growing data traffic, especially with the proliferation of smart phones and mobile applications. It has become a necessity to enhance LTE system in order to achieve higher data rates with less power consumption. LTE-A is targeted to support 1Gbps peak data rate in downlink (DL) in addition to the existing LTE capabilities. The increased data rate is

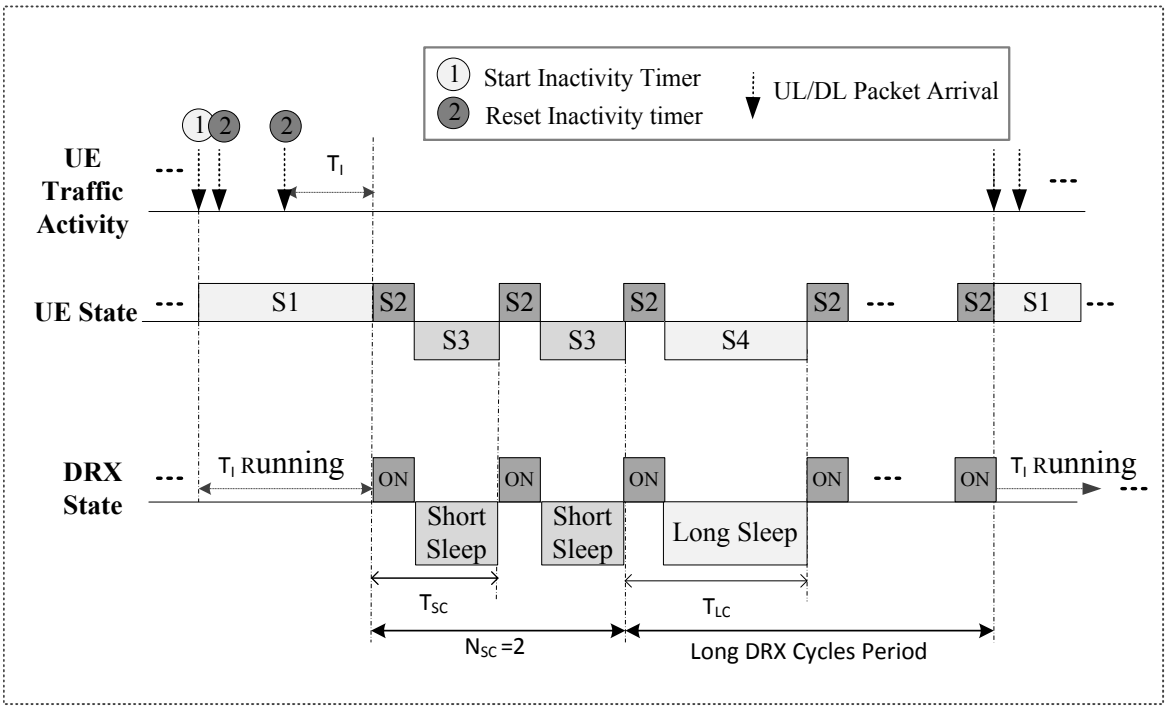

Figure 1.2. A typical scenario of DRX mechanism in LTE-A

achieved by adopting higher transmission bandwidth with carrier aggregation, higher order modulation, advanced coding techniques and multiple advanced antenna schemes [3][4][5]. Support of these features and diverse data traffic in LTE-A devices have resulted in battery hungry devices.

Discontinuous Reception (DRX) has been adopted as an effective power saving mechanism, which largely addresses the power saving and unnecessary state transition issues [5][6]. DRX saves UE battery power by allowing UE to monitor downlink control channel less frequently and go to sleep whenever there is no packet activity for the UE [7] [8]. Since power saving in DRX usually comes at the expense of increased end-to-end delay, DRX may create serious latency issues for active traffic where latency may be of prime focus. A typical scenario of DRX mechanism is shown in Figure 1.2. DRX mechanism in LTE is discussed in more detail in Chapter 4.

1.2 Existing Work

ARQ and H-ARQ is widely used in wireless systems such as WiMAX [9], HSPA [10], LTE [7]. Several types of H-ARQ protocols have been proposed in [11]. However, majority of the research is focused on code design while relatively little research has focused on performance analysis of H-ARQ. Also, most throughput analysis considers a symmetric system and studies the behaviour of the ARQ schemes as a Markov chain instead of focusing on information theoretic capacity. Such an analysis of Hybrid-ARQ protocols in terms of throughput and delay can be found in [12] and [13]. Most relevant work in [14] has established a relationship between H-ARQ throughput and mutual information in classical Shannon capacity sense. Moreover, in [14] a closed-form throughput formula under a delay and code rate constraint has been provided. In [15], the performance of H-ARQ in block fading channels has been investigated using a fixed outage transmission over a single antenna system. For multiple antenna systems, the diversity-multiplexing-delay trade-off of H-ARQ has been studied in [16]. Since adaptive DRX mechanism in LTE is relatively a new research area, only a few work has been published in literature. A tradeoff relationship between power saving and packet waiting delay has been investigated for LTE DRX mechanism with adjustable DRX cycles in [17]. The work in [17] has also shown that LTE DRX achieves better power saving than that of Universal Mobile Telecommunications System (UMTS) DRX at the cost of longer packet waiting delay in the former. Authors in [6] have pointed out the fact that power saving by DRX comes at the cost of degraded system utility. The authors proposed an adaptive DRX inactivity timer based on the channel condition between UE and enhanced nodeB (eNB), so that UE with bad channel quality will go to sleep for lesser time. In this way UE with bad channel condition can increase its transmission opportunity to increase average throughput at the cost of reduced power saving. Note that inactivity timer defines the period during which UE must monitor downlink control channel (PDCCH) after the previous successful decoding of PDCCH message. The effect

of DRX functionality on the traditional LTE scheduling is discussed in [18]. Authors in [18] have proposed a DRX-aware scheduling scheme to reduce the packet loss.

1.3 Contributions

In the dissertation, we first propose an analytical approximation to the probability density function of the product of independent random variables (RVs) based on moment matching. The product distributions arise in studying the performance of modern wireless communications systems. Existing methods to evaluate the statistical properties of product distributions are limited to numerical approximations based on distribution fitting tools, Gaussian approximation, or special functions in terms of infinite series. For more general cases that involve the Rician distribution, the distribution function exists only for the product of two RVs as the sum of infinite series of special functions [19]. The proposed method is sufficiently general to approximate the probability distribution of the products of Rician RVs. The accuracy of the new approximation is evaluated in computing the rate offset of the automatic repeat request (ARQ) transmission. Performance of new approximation is also compared with Gaussian approximation for both Rayleigh and Rician RVs.

Second, we develop a generalized throughput formula for multi-access schemes depending on a predetermined threshold value to exploit the multi-user diversity. We also exploit the multi antenna support for different H-ARQ schemes and derive the throughput expression with respect to the number of users and the number of antennas at the base station.

Third, we focus on the power saving and latency impact of Discontinuous Reception (DRX) on LTE-Advanced. We formulate an analytical model for power saving as well as latency for a given DRX parameter set. We then solve a multi-objective optimization problem which balance the trade off between power saving and latency. In order to validate the proposed enhancements, we propose a common simulation framework which can model the

proposed radio resource control mechanisms in advanced wireless communication systems.

1.4 Organization of Dissertation

The organization of the dissertation is as follows:

In Chapter 2, we present the system model and assumptions for collision free ARQ systems. Statistical analysis of ARQ channels rely on the product distributions of fading channels. In Chapter 2, we propose a simple and efficient analytical approximation to the probability distribution of product of independent random variables. We apply the approximated probability distribution to analyze and optimize ARQ protocols. In Chapter 3, we present a unified approach for throughput calculation in H-ARQ collision channels. We also expand our analysis of H-ARQ schemes to multi antenna systems. An important part of radio resource control is device battery power saving. Discontinuous Reception (DRX) is one of the key mechanisms which is used by current wireless networks. In Chapter 4, we present analytical model for DRX mechanism in advanced wireless systems. More specifically, we investigate the tradeoff between power saving and latency with both analytical and simulation results for LTE advanced networks. In the context of radio resource control, this dissertation develops analytical performance analysis techniques and explores analytical models to improve throughput and delay. Since the requirements of various mobile applications are getting more diverse, our analysis and corresponding algorithm development need to be tested under a variety of channel and network conditions. Thus, in Chapter 5 we present a common network simulation framework that implements realistic communication links for advanced wireless systems. This part of dissertation serves the purpose of validating the proposed models and algorithms in the presence of realistic physical layer and mobile applications. In Chapter 6, we provide future research directions and conclude the dissertation.

CHAPTER 2
COLLISION FREE ARQ SYSTEMS

Automatic repeat request (ARQ) is used to improve the data rate and reliability over the cellular networks. For the optimization of the ARQ parameters, the distribution of the product of independent random variables is needed. Besides ARQ, the product of independent random variables (RVs) has an important role in studying the performance metrics of modern wireless communication systems such as outage probability and ϵ-outage capacity or rate offset under various fading conditions. For example, in a multiple input multiple output keyhole system, signals propagate through multiple keyholes such that the overall channel gain can be approximated as the product of the random variables [20]. This modelling was applied to independent Rayleigh channels in [21] where the distribution of the product of Rayleigh random variables was derived in terms of Meijer-G function. Also for ARQ systems, the rate offset can be approximated as the logarithm of the product of the channel gains [22]. Also in [22], expression for the rate offset is given for Rayleigh channels using the Meijer G-function. All the aforementioned works result in either infinite series or special functions. However, the special functions such as Meijer G-function are essentially contour integrals which are not easy to compute. For the infinite series expressions, truncation is needed for numerical calculations. But the number of terms truncated which highly depends on the distribution parameters needs to be found. In [23], a numerical approximation to the probability density function of the product of independent Rayleigh random variables is proposed in terms of a more tractable distribution, i.e., the Nakagami-m distribution. In [24], the same approximation is extended to several commonly used distributions in wireless communication. In both [23] and [24], the Nakagami-m distribution parameters are heuristically found using MATLAB distribution fitting tool. However,

applying numerical approximation in [24] to the product of RVs with multiple distribution parameters such Rician RVs may be prohibitive due to extended search space. In this chapter, we also explore approximating the distribution of the product of the independent random variables as a Nakagami-m distribution. However, as an alternative approach we obtain one-to-one mapping between the fractional moments of the original random variables and the parameters of the Nakagami-m distribution to closely approximate the product of both Rayleigh and Rician RVs [25]. In our approximation, the Nakagami-m parameters are found as a function of the original distribution parameters to allow the use of the closed-form expressions. Numerical results show that the proposed approximation based on moment matching yields sufficient accuracy for the product distributions of Rayleigh and Rician RVs, respectively. We also compare the performance of proposed approximation with Gaussian approximation. Our results show that Nakagami-m approximation follows the exact values closer than Gaussian approximation for the range of N values that we are interested in for ARQ systems.

2.1 System Model

The system consists of a single user that accesses the network. We consider a block-fading channel where the channel remains constant but various independently from one slot to another. At slot n, the received signal is given by:

$$y_n = h_n x_n + w_n \tag{2.1}$$

where h_n, x_n, w_n are the channel conditions of the user in slot n, the transmitted signal of the user in slot n and AWGN noise in slot n, respectively.

2.2 Analytical Approximation to Distribution of Product of Rayleigh RVs Using Moment Matching Method

Product of n independent Rayleigh variables can be expressed as

$$Y = \prod_{i=1}^{n} X_i \tag{2.2}$$

where X_i, $i = 1, 2, \ldots, n$ follow the normalized Rayleigh distribution of

$$f_{X_i}(x) = 2xe^{-x^2} \tag{2.3}$$

The pdf of Y is derived using Meijer G-function as [21]

$$f_Y(y) = 2G_{0,n}^{n,0}\left(y^2|_{\frac{1}{2},\cdots,\frac{1}{2}}^{-}\right) \tag{2.4}$$

where $G_{m,n}^{p,q}\left(z|_{b_1,b_2,\cdots,b_q}^{a_1,a_2,\cdots,a_p}\right)$ is the Meijer G-function

$$G_{m,n}^{p,q}\left(z|_{b_1,b_2,\cdots,b_q}^{a_1,a_2,\cdots,a_p}\right) = \frac{1}{2\pi i}\int \frac{\prod_{j=1}^{m}\Gamma(b_j - s)\prod_{j=1}^{n}\Gamma(1 - a_j + s)}{\prod_{j=m+1}^{q}\Gamma(1 - b_j + s)\prod_{j=n+1}^{p}\Gamma(a_j - s)} z^s ds \tag{2.5}$$

where $\Gamma(.)$ is the complete Gamma function. It can be seen from (2.4) that a numerical integration is needed to solve the contour integral. For larger n values, the computational complexity of the probability distribution in (2.4) is very high. An approximation which does not have any special functions or infinite series would be helpful for applications in wireless communication.

Let us define random variable W as

$$W = Y^{\frac{1}{n}} = \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}} \tag{2.6}$$

While the moment based techniques are widely used in literature, algebraic expression obtained to match the first two moments of the candidate distribution lacks insights. The raw moments of $X_i^{(1/n)}$ can be written as

$$\mu_h = E[X_i^{h/n}] = \int_0^{\infty} 2x^{1+h/n}e^{-x^2}dx = \Gamma\left(1 + \frac{h}{2n}\right) \tag{2.7}$$

It was found that W can be well approximated as a Nakagami-m random variable with the pdf [26]

$$f_W(w) \approx 2\left(\frac{m}{\Omega}\right)^m \frac{1}{\Gamma(m)} w^{2m-1} e^{-\frac{m}{\Omega}w^2}, \qquad w \geq 0 \tag{2.8}$$

where fading parameter $m = \frac{\mathrm{E}[W^2]^2}{\mathrm{Var}[W^2]}$ and $\Omega = E[W^2]$. The parameters m and Ω of Nakagami-m RV can be expressed as

$$\Omega = E[W^2] = (E[X_i^{2/n}])^n = \Gamma\left(1+\frac{1}{n}\right)^n \tag{2.9}$$

and

$$m = \frac{E[W^2]^2}{E[W^4] - E[W^2]^2} = \frac{\Gamma\left(1+\frac{1}{n}\right)^{2n}}{\Gamma\left(1+\frac{2}{n}\right)^n - \Gamma\left(1+\frac{1}{n}\right)^{2n}} \tag{2.10}$$

Applying Taylor series expansion to $\Gamma\left(1+x\right)^{1/x}$, we obtain

$$\Gamma\left(1+x\right)^{1/x} = e^{-\gamma_E} + \frac{\pi^2}{12} e^{-\gamma_E} x + O[x]^2 \tag{2.11}$$

where $\gamma_E = 0.577215...$ is the Euler's constant. With change of variable $\frac{1}{n} = x$ and substituting (2.11) into (2.9), we can approximate Ω as

$$\begin{aligned} \Omega &\approx e^{-\gamma_E} + \frac{e^{-\gamma_E}\pi^2}{12} n^{-1} \\ \Omega &\approx 0.4617 n^{-1} + 0.5614 \end{aligned} \tag{2.12}$$

Similarly, fading parameter of Nakagami-m

$$\begin{aligned} m &\approx \frac{6}{\pi^2} n + (-0.5 - \frac{36\eta^2(1)}{\pi^4}) \\ m &\approx 0.6079 n + 0.3885 \end{aligned} \tag{2.13}$$

where $\eta^2(1) = -2.4041$ and $\eta^s(x)$ is polygamma function with order s which is the $s+1$th derivative of logarithm of Gamma function

$$\eta^s(x) = \frac{d^{s+1}}{dx^{s+1}} \log\Gamma(x) \tag{2.14}$$

In [23], the Nakagami-m parameters are found as $\Omega^{'} = 0.8808n^{-0.9661} + 1.12$ and $m^{'} = 0.6102n + 0.4263$ using MATLAB distribution fitting tool. $\Omega' \approx 2\Omega$ (due to usual normalization for the radio channel). Here, we analytically show that moment matching based approximation can be used to find the same heuristic relationships found in [23]. In the next section, we extend this analysis to Rician channel.

2.3 Moment Based Approximation to Distribution of Product of Rician RVs

Define random variable Z as

$$Z = \left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}} \tag{2.15}$$

which is the geometric mean of Rician random variables Y_i, $i = 1, 2, \ldots, n$ with $\sigma^2 = 1$ and Rician factor K.

$$f_Y(y) = 2(1+K)ye^{-(1+K)y^2-K} I_0(2\sqrt{K(1+K)}y) \tag{2.16}$$

In [19], the distribution of the product of two non-central chi-square random variable was obtained. With change of variables and setting the non-central parameter of chi-square random variable to $K/(K+1)$, we can derive the probability distribution of product of two Rician random variables as

$$f_Y(y) \;=\; e^{-\frac{(K^2}{(K+1)^2}} \sum_{m=0}^{\infty} \frac{K^{2m}}{(K+1)^{2m-1}2^{-2m}} \sum_{j=0}^{m} \frac{G_{0,2}^{2,0}\left(-\frac{y^2}{4}|_{\frac{-m+2j}{2}, \frac{m-2j}{2}}^{-}\right) y}{2(j!(m-j)!)^2} \tag{2.17}$$

The probability distribution of the product of more than two Rician random variables are not available in the literature best to our knowledge. Even for numerical calculation of the distribution for product of two Rician random variables in (2.17), infinite sum needs to be computed which requires high computational complexity.

Thus, we explore the potential of moment matching to approximate the product distribution

of independent random variables. The raw moments of $Y_i^{(1/n)}$ can be written as

$$\begin{aligned}\mu_h = E[Y_i^{h/n}] &= \int_0^\infty x^{h/n} f_{Y_i}(x)dx \\ &= (\frac{1}{1+K})^{h/2n}\Gamma\left(1+\frac{h}{2n}\right) L_{h/2n}[-K] \end{aligned} \tag{2.18}$$

where $L_a[x]$ is the Laguerre's L function defined in terms of hypergeometric function by

$$L_a[x] = {}_1F_1(-a;1;x) \tag{2.19}$$

We match the parameter Ω of Nakagami-m RV as

$$\Omega = (E[Y_i^{2/n}])^n = \frac{1}{1+K}\Gamma\left(1+\frac{1}{n}\right)^n L_{1/n}[-K]^n \tag{2.20}$$

Using the first two terms of the Taylor series expansion of the (2.20), we express Ω in the form of $a_1 n^{-1} + a_0$ where

$$a_0 = \frac{e^{-\gamma_E + L^{1,0}[0,-K]}}{1+K}$$

With the help of (2.21)

$$L^{1,0}[0,-K] = \gamma_E + \log(K) + \Gamma[0,K] \tag{2.21}$$

and (2.22)

$$e^{\Gamma[0,K]} \approx 1 + \frac{e^{-K}e^{-\gamma_E}}{K} \tag{2.22}$$

a_0 and a_1 of Ω become

$$a_0 \approx \frac{K^2 + (1+K)e^{-K}e^{-\Gamma_E}}{K^2+K} \tag{2.23}$$

and

$$a_1 \approx \frac{K - (1+K)e^{-(K+\Gamma_E)}}{K^2+K} \tag{2.24}$$

Finally, we express Ω as

$$\Omega = \frac{K^2 + (1+K)e^{-(K+\Gamma_E)}}{K^2+K} + \frac{K - (1+K)e^{-(K+\Gamma_E)}}{K^2+K} n^{-1} \tag{2.25}$$

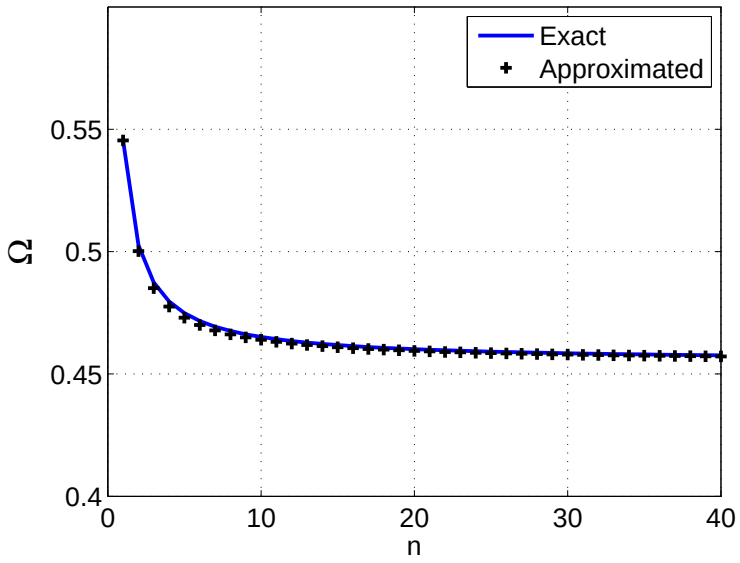

Figure 2.1. Comparison of the approximation and exact Ω values for different diversity order n when Rician factor $K = 10$. Exact (2.20), Approximated (2.25)

Figure 2.1 shows Ω versus n and compares the exact expression in (2.20) with the approximation in (2.25) for $K = 10$. The approximated result indicates that Ω converges to its asymptotic value inversely proportional to both K and n.

Similarly, we can approximate the fading parameter m of Nakagami-m distribution as a function of the original Rician parameter K and the number of fading blocks n.

$$m = \frac{E[Z^2]^2}{E[Z^4] - E[Z^2]^2} \tag{2.26}$$

where m becomes

$$m = \frac{\Gamma\left(1+\frac{1}{n}\right)^{2n} L_{1/n}[-K]^{2n}}{\Gamma\left(1+\frac{2}{n}\right)^{n} L_{2/n}[-K]^{n} - \Gamma\left(1+\frac{1}{n}\right)^{2n} L_{1/n}[-K]^{2n}} \tag{2.27}$$

Again, adopting the Taylor series expansion, we can approximately express (2.27) as

$$m \approx \frac{K}{2}n + \frac{3K+2}{4K+2}. \tag{2.28}$$

The approximated expression above indicates that we can reduce the number transmissions n approximately with the factor of K. Figure 2.2 shows m versus n and compares its

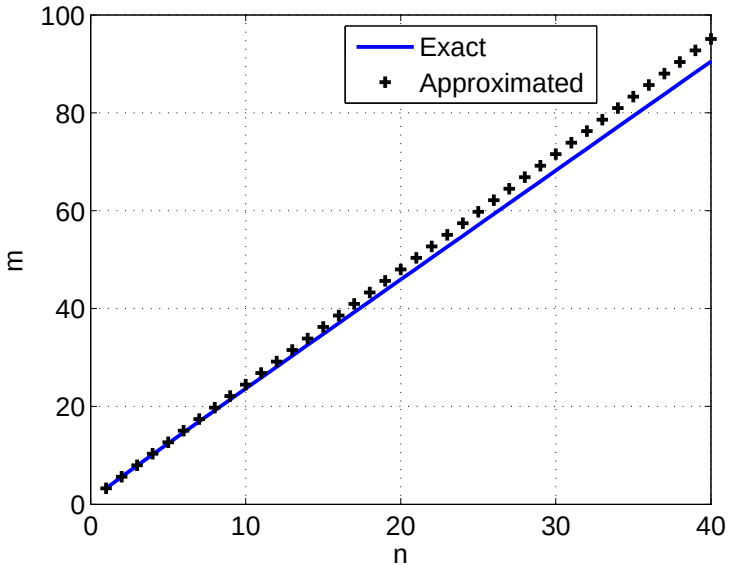

Figure 2.2. Comparison of the approximation and exact m values for varying number of transmissions n when Rician factor $K = 10$. Exact (2.26), Approximated (2.28)

approximation to the exact values at $K = 10$. It can be seen that our approximation of m fits well with the actual values. In the next section, we will use our approximation in calculating the rate offset experienced in ARQ transmissions.

2.4 Rate Offset

Rate offset is defined as the difference between the capacity of AWGN channel and the ϵ-outage capacity. It can be expressed as [22]

$$C^{\epsilon}(\text{SNR}) = \log_2(\text{SNR}) + F_{\epsilon}^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\log_2(|h_n|^2)\right) + o(1) \tag{2.29}$$

where n is the number of blocks and ϵ is the outage constraint. $\log_2(\text{SNR})$ is the AWGN capacity and F_{ϵ}^{-1} is the solution y to the equation $P[X \leq y] = \epsilon$. F_{ϵ}^{-1} gives the highest x value that satisfy the ϵ outage probability. Thus, rate offset N_{∞} is given by

$$N_{\infty} = -F_{\epsilon}^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\log_2(|h_n|^2)\right) \tag{2.30}$$

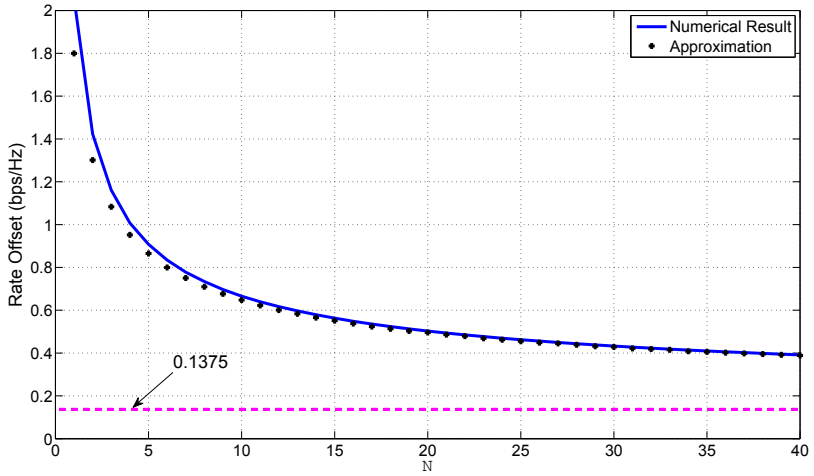

Figure 2.3. Comparison of approximation and exact rate offsets for different diversity order n when Rician factor $K = 10$ and outage $\epsilon = 0.01$. Rate offset converges to $-\log_2(\frac{K}{K+1}e^{\Gamma(0,K)}) = 0.1375$.

We investigate the limiting values of Ω as the diversity order n goes to infinity. As the diversity order increases, the approximated Nakagami-m distribution converges to non-fading AWGN channel with the SNR gain of

$$\lim_{m\to\infty} \Omega = \Omega_0 \tag{2.31}$$

where $\Omega_0 = e^{-\gamma_E}$ for the product of Rayleigh RVs and $\Omega_0 = \frac{K}{K+1}e^{\Gamma(0,K)}$ for the product of Rician RVs. Thus, the rate offset converges to $-\log_2(\Omega_0)$ in both cases. In Figure 2.3, we plot for the rate offset vs. the number of transmission blocks n for Rician RVs. As n goes to infinity, the rate offset converges to $-\log_2(\frac{K}{K+1}e^{\Gamma(0,K)}) = 0.1375$ for $K = 10$. However, the rate offset follows the same trend with diversity order n. For high K values even for one or two rounds of ARQ transmission will be good enough to guarantee the ϵ-outage constraint. In Figure 2.4, we limit the number of the transmissions $n = 8$ and investigate the impact of K on the rate offset. As the K factor increases the rate offset decreases. In Figure 2.5, we adjust the ϵ outage value and investigate the rate offset. Lower outage requires higher rate offset to overcome the effect of the channel fading.

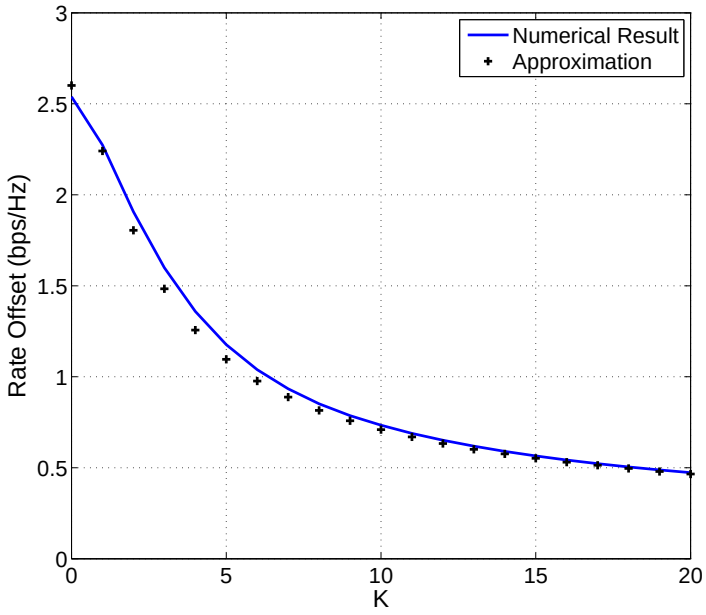

Figure 2.4. Comparison of approximation and exact rate offsets for different Rician factor K when diversity order $n = 8$ and outage $\epsilon = 0.01$

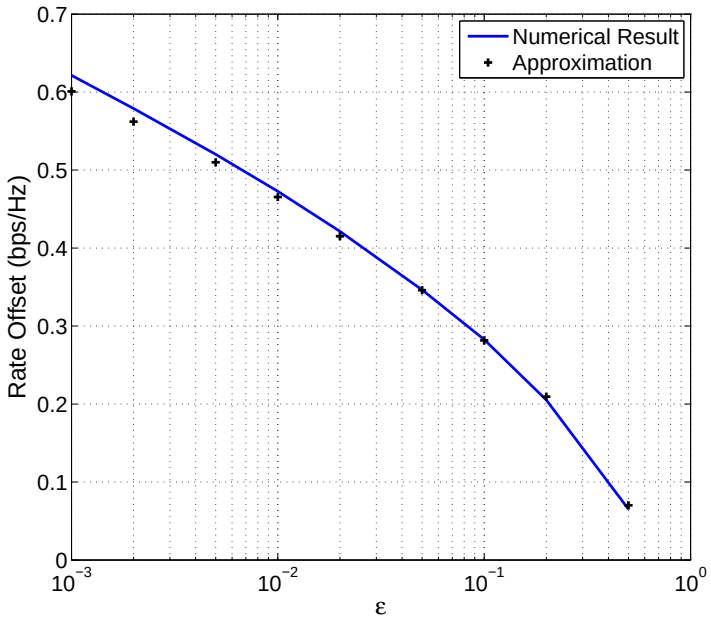

Figure 2.5. Comparison of approximation and exact rate offsets for different ϵ outage probability when diversity order $n = 8$ and Rician factor $K = 20$

2.5 Performance Comparisons

In this section, we compare the performance of Gaussian approximation and proposed Nakagami-m approximation. Define

$$Y_n = \log_2(|h_n|^2) \tag{2.32}$$

where h_n is the channel coefficients. Motivated by Central Limit Theorem (CLT), $\frac{1}{N}\sum_{n=1}^{N} Y_n$ can be approximated by a Gaussian random variable with mean μ and variance $\frac{\sigma^2}{N}$. For Rayleigh channel h_n, μ is expressed as

$$\begin{aligned} \mu &= \int_0^\infty \log_2(y^2) 2y e^{-y^2} dy \\ &= \log_2(e^{-\gamma_E}) = -0.8327 \end{aligned} \tag{2.33}$$

and σ^2 is expressed as

$$\begin{aligned} \sigma^2 &= E[Y_n^2] - \mu^2 = \int_0^\infty \log_2(y^2)^2 2y e^{-y^2} dy - \mu^2 \\ &= \frac{\pi^2}{6\log(2)^2} = 3.42371 \end{aligned} \tag{2.34}$$

In Figures 2.6, 2.7, and 2.8 CDF comparisons of exact, Gaussian and Nakagami-m approximations are shown for Rayleigh channel. As expected by the CLT, as N increases the Gaussian approximation becomes more accurate. Nakagami-m approximation is closer to the exact values compared to Gaussian approximation in the range of N and outage probabilities that we are interested in.

We compare the different approximation's performance on calculating rate offset for ϵ outage over Rayleigh channel in Figures 2.9 and 2.10. Match between Gaussian approximation and exact values is less accurate for very small ϵ values because the tail distribution of Gaussian and exact values do not precisely match. However, Nakagami-m approximation follows the exact values very closely for rate offset even for low ϵ values.

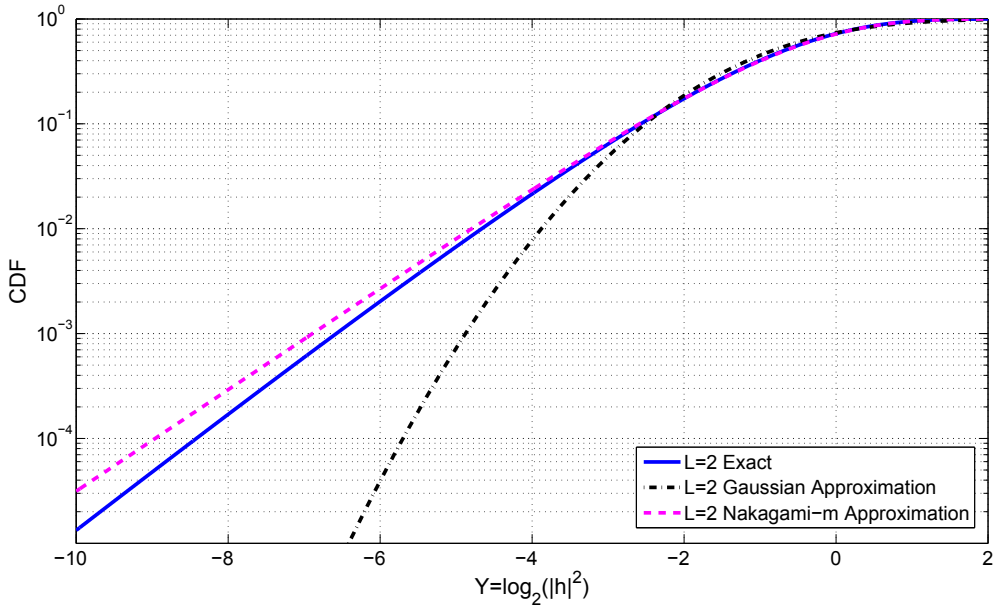

Figure 2.6. CDF comparisons of exact, Gaussian and Nakagami-m approximations of Y for $N = 2$ over Rayleigh channel

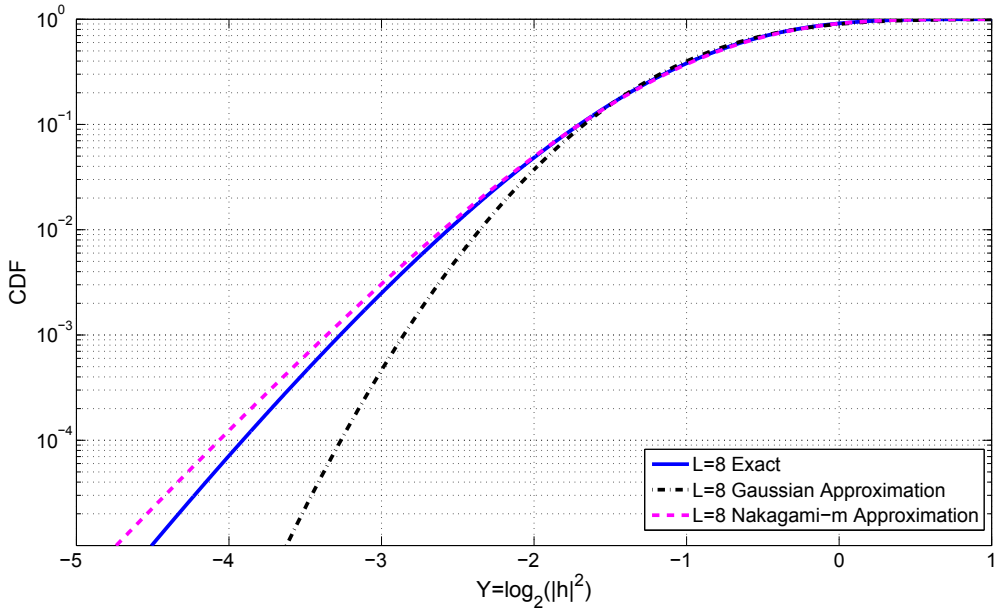

Figure 2.7. CDF comparisons of exact, Gaussian and Nakagami-m approximations of Y for $N = 8$ over Rayleigh channel

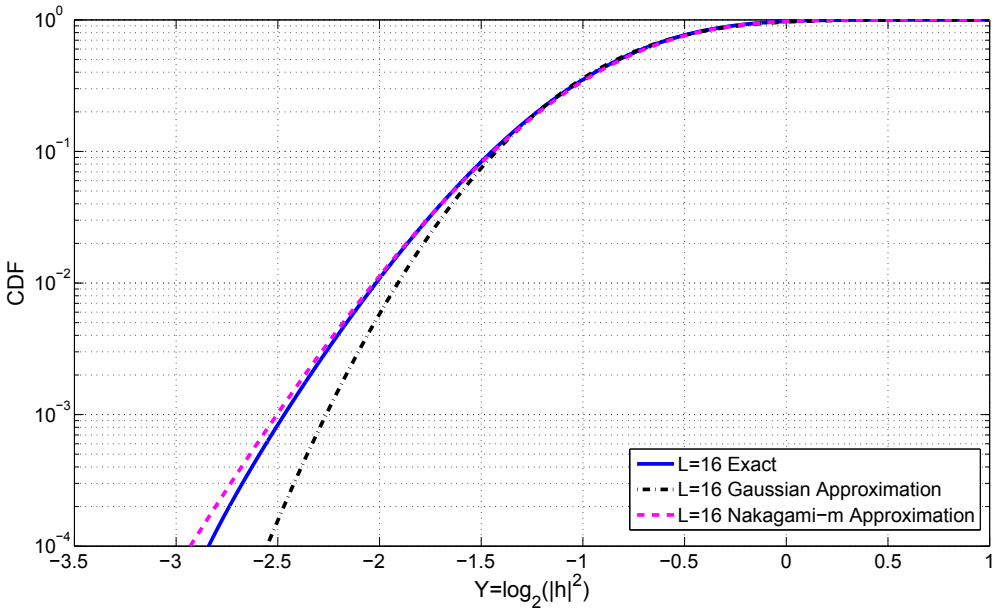

Figure 2.8. CDF comparisons of exact, Gaussian and Nakagami-m approximations of Y for $N = 16$ over Rayleigh channel

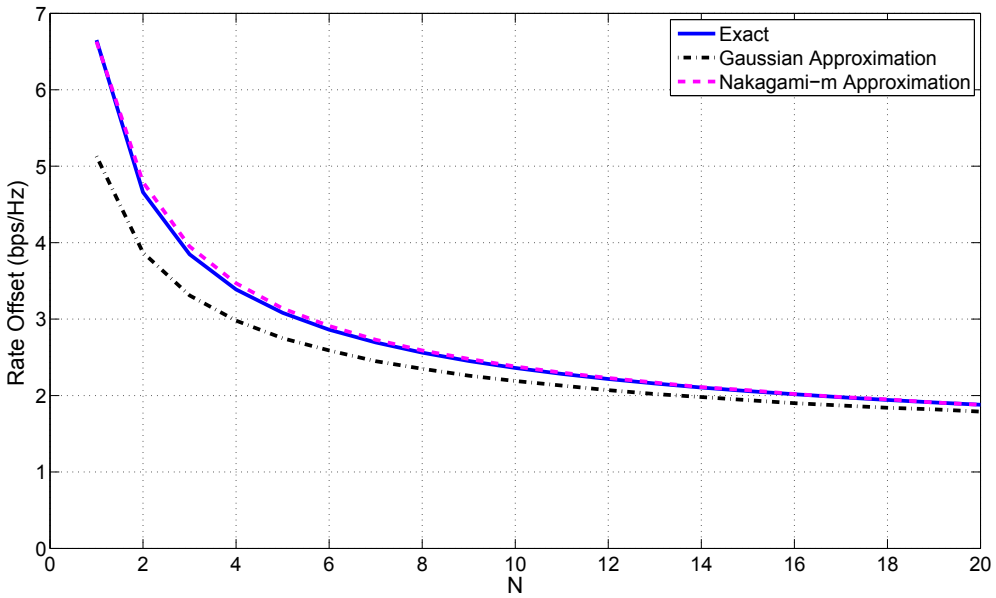

Figure 2.9. Rate offset comparisons of exact, Gaussian and Nakagami-m approximations of Y for different N values with $\epsilon = 0.01$ outage over Rayleigh channel

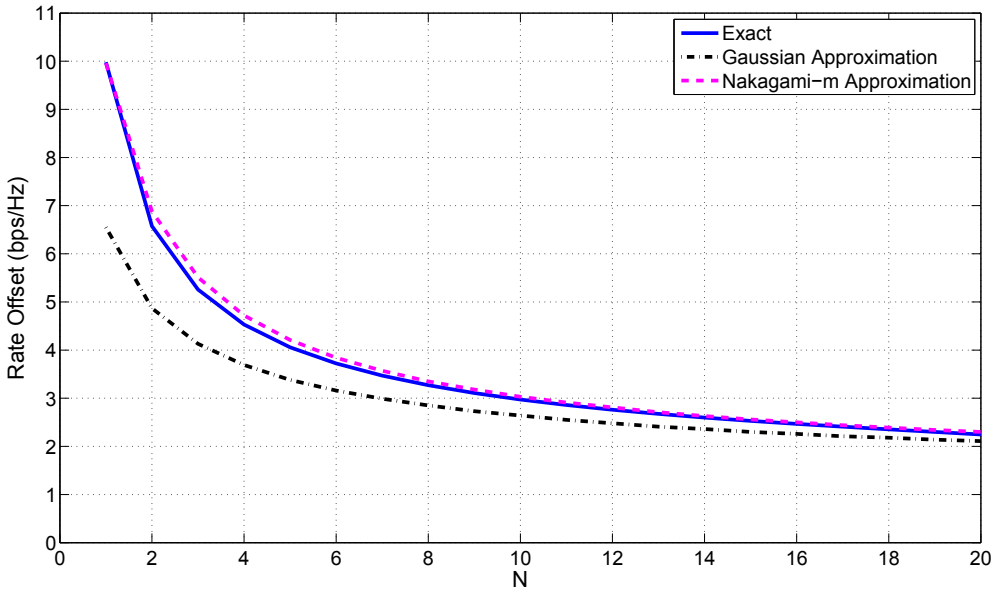

Figure 2.10. Rate offset comparisons of exact, Gaussian and Nakagami-m approximations of Y for different N values with $\epsilon = 0.001$ outage over Rayleigh channel

For Rician channel h_n, $\frac{1}{N}\sum_{n=1}^{N} Y_n$ can be approximated by a Gaussian random variable with mean μ and variance $\frac{\sigma^2}{N}$. For Rician channel, μ is expressed as

$$\begin{aligned} \mu &= \int_0^\infty 2\log_2(y^2) y e^{-(K+1)y^2-K} I_0(2\sqrt{K(1+K)}y)dy \\ &= \log_2(\frac{K}{K+1}) + \frac{\Gamma(0,K)}{\log 2} \end{aligned} \tag{2.35}$$

and σ^2 is expressed as

$$\sigma^2 = E[Y_n^2] - \mu^2 \tag{2.36}$$

where K is the Rician factor. The second order moment of Y_n can be calculated as

$$E[Y_n^2] = \int_0^\infty 2\log_2(y^2)^2 y e^{-(K+1)y^2-K} I_0(2\sqrt{K(1+K)}y)dy \tag{2.37}$$

In (2.37), $I_0(.)$ denotes the modified Bessel function of the first kind of order 0, i.e

$$I_0(2\sqrt{K(1+K)}y) \triangleq \sum_{m=0}^{\infty} \frac{(K(K+1)y^2)^m}{\Gamma(m+1)^2} \tag{2.38}$$

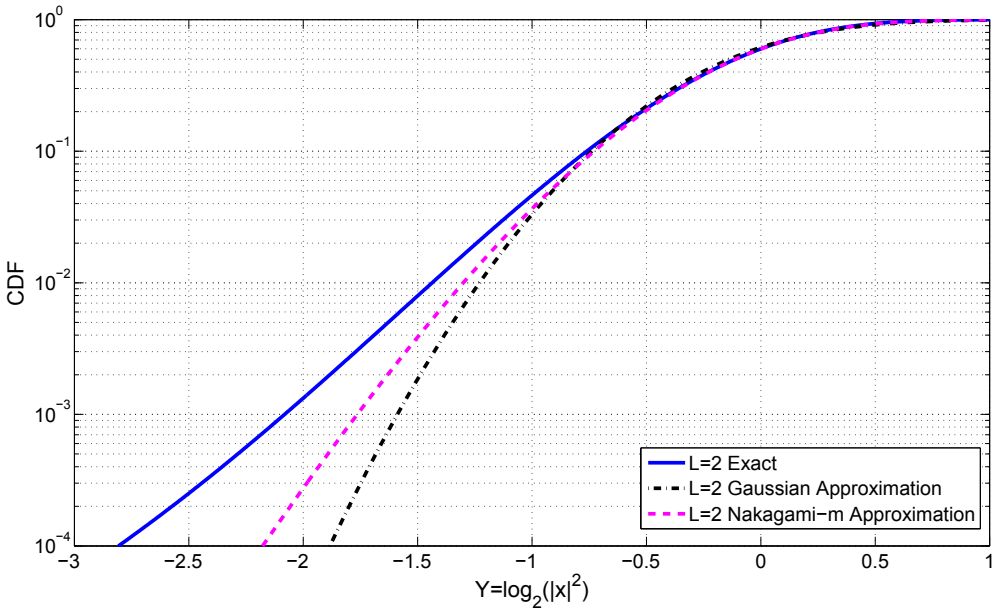

Figure 2.11. CDF comparisons of exact, Gaussian and Nakagami-m approximations of Y for $N = 2$ over Rician channel with $K = 10$

Substituting (2.38) into (2.37)

$$E[Y_n^2] = \sum_{m=0}^{\infty} 2(K+1)e^{-K}\frac{K^m}{\log(2)^2\Gamma(m+1)}\left((\gamma_E - H_m + \log(K+1))^2 + \psi^1(m+1)\right)$$

where $H_m = \sum_{k=1}^{m}\frac{1}{k}$ is the mth harmonic number and $\psi^1(z) = \frac{d^2\log(\Gamma(z))}{dz}$ is the trigamma function.

In Figures 2.11, 2.12 and 2.13 CDF comparisons of exact, Gaussian and Nakagami-m approximations are shown for Rician channel. For Rician channel, Nakagami-m approximation is closer to exact values compared to Gaussian approximation.

We compare the different approximation's performance on calculating rate offset for ϵ outage over Rician channel in Figures 2.14 and 2.15. Also for Rician channel, Nakagami-m approximation follows the exact values very closely for rate offset even for low ϵ values.

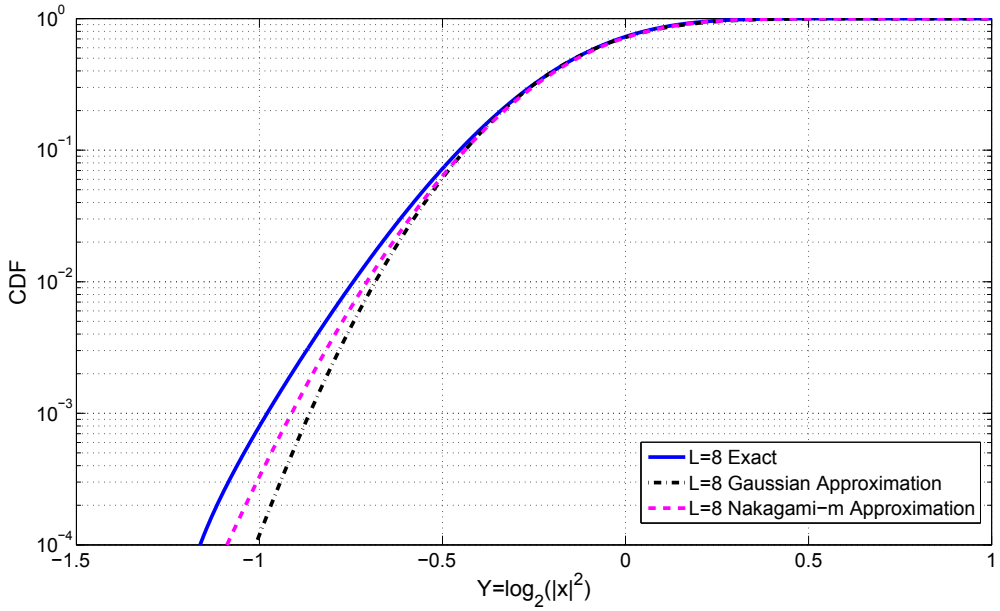

Figure 2.12. CDF comparisons of exact, Gaussian and Nakagami-m approximations of Y for $N = 8$ over Rician channel with $K = 10$

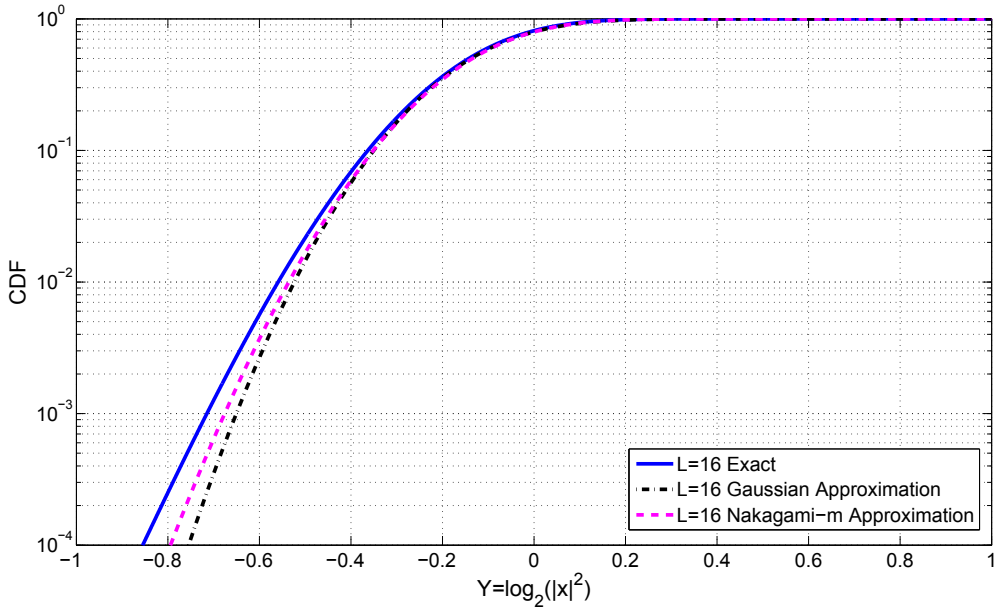

Figure 2.13. CDF comparisons of exact, Gaussian and Nakagami-m approximations of Y for $N = 16$ over Rician channel with $K = 10$

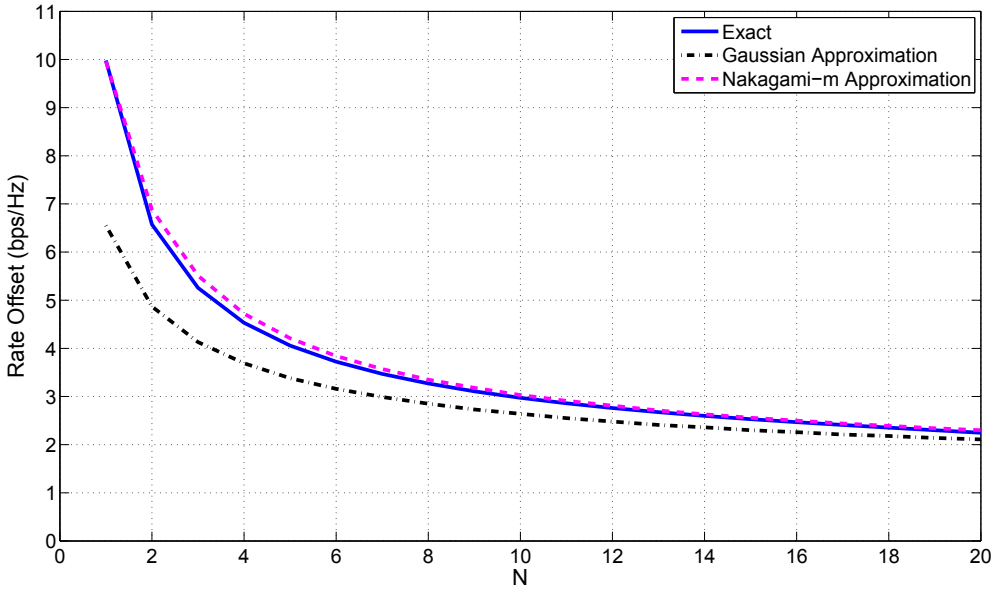

Figure 2.14. Rate offset comparisons of exact, Gaussian and Nakagami-m approximations of Y for different N values with $\epsilon = 0.01$ outage over Rician channel with $K = 10$

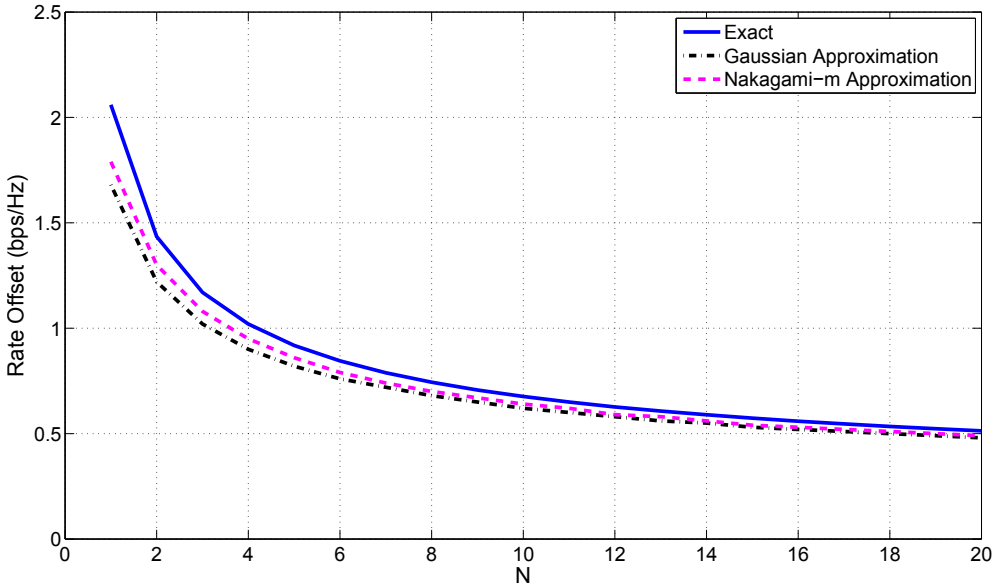

Figure 2.15. Rate offset comparisons of exact, Gaussian and Nakagami-m approximations of Y for different N values with $\epsilon = 0.001$ outage over Rician channel with $K = 10$

2.6 Conclusion

In this chapter, we have proposed to approximate the distribution of products of independent random variables by the widely applicable Nakagami-m distribution by means of moment matching method. Existing methods to evaluate the statistical properties of product distributions are limited to numerical approximations based on distribution fitting tools, Gaussian approximation, or special functions in terms of infinite series. The obtained results show that the moment matching method can be useful even approximating the product distribution of Rician random variables as Nakagami-m distribution. Moment matching also allows us to obtain the insightful results with analytical means. For example, the parameters m and Ω of Nakagami-m random variable are analytically approximated and compared to the results obtained from distribution fitting tools. The analytically tractable Nakagami-m distribution can significantly simplify the performance analysis of ARQ and Hybrid-ARQ transmission systems. We demonstrate the useful of the proposed approximation in rate offset calculations for ARQ systems and also compare the performance of proposed approximation with Gaussian approximation. The limiting behavior of rate offset is also found easily using the new approximation. This work can be extended with the inclusion of power control as in [27] and [28]. In the next chapter, we will focus on the throughput of multi access multi antenna H-ARQ systems.

CHAPTER 3
THRESHOLD BASED HYBRID ARQ SCHEMES FOR MULTI-ACCESS SYSTEMS

In this chapter, interference issues of multi access systems are investigated. Mechanisms to avoid the interference are proposed to efficiently utilize over the air resources in multi access H-ARQ systems.

3.1 System Model

Random multi-access system consists of a base station with multiple receive antennas and multiple K subscriber stations (SSs) as shown in Figure 3.1. We assume that users have continuous downlink transmission, which subscriber stations can use to estimate the channel conditions. On the uplink side, subscriber stations are provided with a common time reference to align their uplink transmissions. Time axis is divided into slots of duration T. Subscriber stations transmit signals aligned with the slots. With these assumptions, a discrete channel model is denoted for the received signal. The received signal y_n, at the base station at slot n, can be expressed as:

$$y_n = \sum_{k \epsilon U(n)} h_{k,n} x_{k,n} + w_n \tag{3.1}$$

where $h_{k,n}, x_{k,n}, w_n$ are the channel conditions of user k in slot n, the transmitted signal of user k in slot n and additive white Gaussian noise (AWGN) in slot n, respectively. The active user set in slot n is denoted with $U(n)$. Each user selects the slots for transmission randomly with a probability of transmission independent of the other users. The optimal distribution of transmission signal $x_{k,n}$ for interference channel systems in (3.1) is not

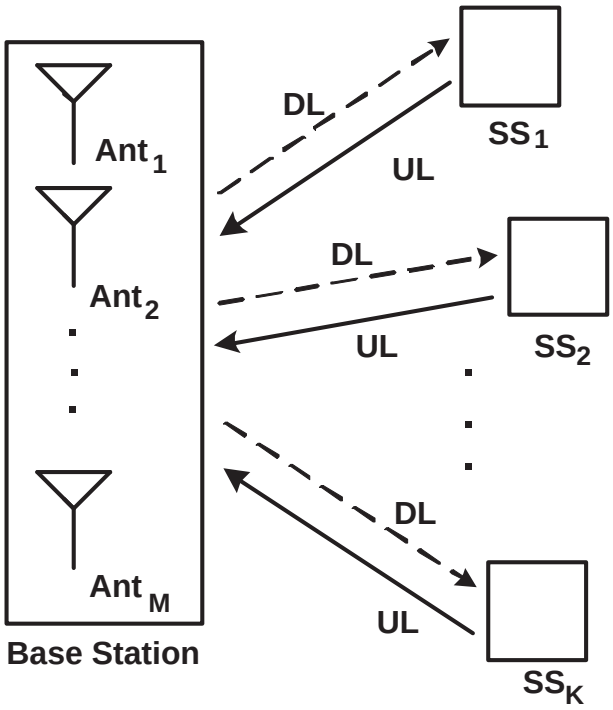

Figure 3.1. System model with $K = u + 1$ users, uplink collision channel (Ant = Antenna, SS = Subscriber Station).

known. For the sake of mathematical tractability of our information theoretic results, we rely on using random Gaussian codebooks for all users. The information theoretic justification of this assumption is provided in [14]. Using Gaussian random codebooks, the probability of decoding failure at slot n, which depends on the distribution of SINR μ and the rate R can be expressed as in [14]

$$p(n) = \begin{cases} \Pr\left(\log_2(1+\mu) \leq R\right)^n & \text{ALOHA} \\ \Pr\left(\log_2(1+\sum_{i=1}^{n}\mu_i) \leq R\right) & \text{CC} \\ \Pr\left(\sum_{i=1}^{n}\log_2(1+\mu_i) \leq R\right) & \text{INR} \end{cases} \tag{3.2}$$

We consider the setting where the base station has perfect channel state information while the transmitter has its own channel information. Each subscriber compares this knowledge with a predetermined threshold to access the medium. This setting can be referred to as open-loop and decentralized system. In order to come up with throughput of H-ARQ schemes we use reward-renewal theorem in [14]. For a user, stopping a transmission is

recognized as a recurrent event with two different rewards X. If the transmission succeeds, we will have reward $X = R$ b/s/Hz and if the transmission fails we have $X = 0$. The consecutive recurrent reward events happen every T seconds. If we apply reward-renewal theorem to throughput of a user $R_1(t)$, the average throughput becomes:

$$\lim_{t\to\infty} \frac{1}{t} R_1(t) = \frac{\mathrm{E}[X]}{\mathrm{E}[T]} \tag{3.3}$$

where the numerator is the mean of the rewards and the denominator is the mean of time between two recurrent events. Thus, the desired throughput expression is

$$\eta = K \frac{\mathrm{E}[X]}{\mathrm{E}[T]} \tag{3.4}$$

where K is the number of users in the system. We consider an optimized system with respect to R and for a given load G, we operate at R values which satisfy $\sup_R \eta$. With these assumptions, the throughput equation becomes

$$\eta = \frac{RG}{\mathrm{E}[T]} = \frac{RG}{\sum_{n=0}^{\infty} p(n)} \tag{3.5}$$

where $p(n)$ is defined in (3.2). We assume a quasi-static Rayleigh fading channel with probability and cumulative density functions with mean SNR $\bar{\gamma}$ as:

$$\begin{aligned} f_{\gamma_{k,n}}(x) &= \frac{1}{\bar{\gamma}} e^{-x/\bar{\gamma}} \\ F_{\gamma_{k,n}}(x) &= 1 - e^{-x/\bar{\gamma}} \end{aligned} \tag{3.6}$$

We can define the received SINR μ with u interfering Rayleigh distributed users as:

$$\mu = \frac{\gamma}{1 + \gamma_{MAI}^{u}} \tag{3.7}$$

We focus on information theoretic capacity and for Aloha transmission, instantaneous capacity for a given SINR value μ can be expressed as

$$R = \log_2(1 + \mu) \tag{3.8}$$

If only the transmitter knows the channel, we can define the outage as

$$C_{\text{out}} = R \Pr(\log_2(1+\mu) < R) \tag{3.9}$$

For one-user access Aloha, outage happens if we have collision where more then one user transmits in a time slot. With outage, the throughput of Aloha for infinite number of users under fading channel becomes

$$\eta_{Aloha}(R) = RGe^{-(2^R-1)/\bar{\gamma}} \tag{3.10}$$

Without outage, if we assume we have infinite delay tolerance, the throughput of Aloha for infinite number of users under fading channel becomes

$$\eta_{Aloha}(R) = RGe^{-(2^R-1)/\bar{\gamma}-(1-2^{-R})G} \tag{3.11}$$

In Figure 3.2, we investigate the affect of the access to the network. We compare the performance of multi-access and one user access. In the case of one user access, the outage happens when there is a collision even if the received $SINR$ is good enough to decode the packet correctly. In multi-user access, outage happens when the information theoretic capacity is not satisfied with the $SINR$. For low target rates R, letting one user to access channel results in performance loss compared to multi-user access. However, for large target rates both schemes behave the same. For high target rates, the information theoretic capacity can be satisfied by the $SINR$, so most of the time when there is a collision the packets are discarded. These results give us the motivation to design a rate dependent threshold based scheme which we will be explained in the next section.

We compute the throughput of the multi-access schemes with some general assumptions such as:

1. Full buffer traffic for all users: An infinite number of information messages are available for each user.

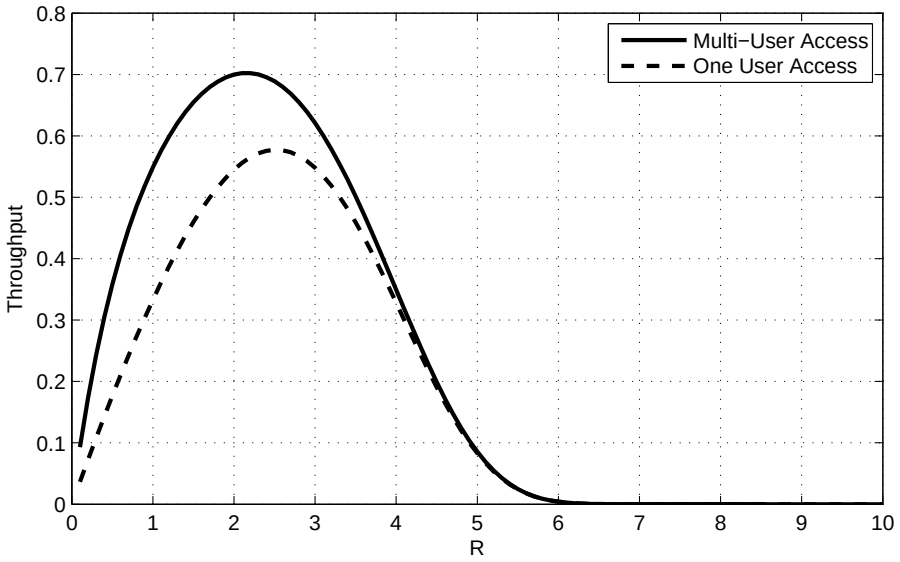

Figure 3.2. Throughput vs. Rate (R) for infinite user system with offered load $G = 1$.

2. The ACK/NACK feedback channel is delay-free and error-free.

3. All users have the same average signal-to-noise ratio (SNR) $\bar{\gamma}$. Pathloss and shadowing are fully compensated by power control.

4. On each slot n, each user selects slot n for transmission with a probability of $p_t = G/K$ where G and K are offered load and number of users, respectively.

5. Base station sends a predetermined threshold γ_T to all of the users.

6. After selecting slot n, user k compares its instantaneous SNR γ_k with the threshold γ_T. In particular, if γ_k equals or exceeds γ_T, that user can transmit during slot n.

7. Channel is constant at each slot. The transmission continues till the aggregated received signal can be decoded without any error.

3.2 Unified Approach

We are proposing a unified approach where the throughput of different H-ARQ schemes can be found, given the characteristic function of the SINR. In Section 3.1, we explain the throughput expression using reward-renewal theorem. In throughput expression (3.3), the denominator is hard to find due to infinite sum of probability of failures $p(n)$ at each slot. Using the characteristic function, we can specify the infinite sum of probability of failures as

$$\sum_{n=0}^{\infty} p(n) = \sum_{n=0}^{\infty} \int_{0}^{2^R-1} \int_{-\infty}^{\infty} \left(\Phi_{\mu}(jv)\right)^n e^{jvx} dv dx \tag{3.12}$$

Given that characteristic function is always smaller than zero, we can move the summation inside the integral and come up with a simple calculation for the denominator of the throughput.

$$\sum_{n=0}^{\infty} p(n) = \int_{0}^{2^R-1} \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \left(\Phi_{\mu}(jv)\right)^n e^{jvx} dv dx = \int_{0}^{2^R-1} \int_{-\infty}^{\infty} \frac{e^{jvx}}{1-\Phi_Z(jv)} dv dx \tag{3.13}$$

As an example to justify the use of unified approach, we want to calculate the throughput of Chase combining where the packets are combined at each transmission until aggregated SNR yields a rate of R. We start with the CDF of instantaneous $SINR$ μ which is given in [14] as

$$f_{\mu}(x) = \sum_{u=0}^{N_u-1} Pr[K=u] \frac{(1+\bar{\gamma}u+x)e^{-x/\bar{\gamma}}}{\bar{\gamma}(1+x)^{u+1}} \tag{3.14}$$

After some derivation using the CDF equation, the expression for $p(n)$ which is defined in (3.2) can be simplified as

$$p(n) = \frac{\beta(n, 2^R-1)}{\Gamma(n)} \tag{3.15}$$

where $\beta(n,x)$ is an incomplete gamma function. We need to take an integral of incomplete gamma function to find the throughput. It is hard to find the throughput using this approach. If we use the proposed unified approach, we can find average time for recurrent

event E[T] as

$$\sum_{n=1}^{\infty} p(n) = \sum_{n=1}^{\infty} \frac{\gamma(n, 2^R - 1/\bar{\gamma})}{\Gamma(n)} = \int_0^{2^R - 1/\bar{\gamma}} \underbrace{\sum_{n=1}^{\infty} \frac{t^{n-1}}{\Gamma(n)}}_{e^t} e^{-t} dt = \frac{2^R - 1}{\bar{\gamma}} \tag{3.16}$$

With the insertion of (3.16) into the throughput equation, throughput of Chase combining for an infinite number of users under fading channel becomes

$$\eta_{CC}(R) = \frac{\mathrm{E}[X]}{\mathrm{E}[T]} = \frac{RG}{1 + (2^R - 1)/\bar{\gamma}} \tag{3.17}$$

3.3 Opportunistic Aloha

In a wireless communication system, a user may have periods of stronger as well as weaker channel conditions than its average channel condition due to time variation of its channel. When many users are present, different users may experience their best channel conditions at different periods of time. By letting users transmit when they have their best channel conditions, we can exploit multi-user diversity. As shown in [29], the sum capacity is maximized by multi-user diversity by allowing only the user with the best channel conditions. However, a centralized scheduler is needed to exploit the multi-user diversity, which is not feasible for random multi-access systems. Random multi-access systems are mostly modelled on Aloha [30]. Exploiting multi-user diversity on Aloha systems has been studied by many researchers. A channel aware Aloha is proposed for the fast fading channel in [31] to exploit multi-user diversity. In [32] and [33], the authors exploit the decentralized channel state information and find the asymptotic stable throughput of random multi-access channels. In [34], the authors investigate the opportunistic Aloha for Orthogonal Frequency-Division Multiple Access (OFDMA) wireless networks. In [35], multi-user diversity is investigated for reservation random access where channel state information is not available to users when they contend for channel access.

We consider exploiting multi-user diversity gains in a distributed way over random multi-access channels with a simple Aloha protocol. We call this new scheme opportunistic Aloha, where the users wait to transmit until their channel conditions are above a predetermined threshold value. We also analyze the performance of the new scheme in terms of throughput. By applying information theoretic capacity and renewal-reward theorem [14], we obtain an expression for throughput based on a predetermined threshold. Moreover, we find the optimum threshold which asymptotically maximizes the throughput. In our scheme, the feedback mechanism is removed by a broadcasted threshold which will be used by users to decide to transmit or not. In [36], the amount of feedback is analyzed to exploit the multi-user diversity. The authors in [36] optimize the throughput by limiting the amount of feedback in the network. In [37], multi-user diversity is exploited by setting multiple levels of thresholds for feedback transmission in order to reduce the feedback load to a minimum level without sacrificing the throughput.

We use the system model specified in Section 3.1. The building block of the subscriber station can be seen in Figure 3.3. A threshold value is broadcasted by the base station to all of the users. The active user set depends on the transmission threshold broadcasted by the base station. The receiver at the base station can estimate the presence of the users who transmit in each slot. In our current analysis, we assume that the threshold is the same for all of the users and denoted as γ_T. At the beginning of each slot, the subscriber station calculates the DL SNR from the pilots or preambles to compare with the predetermined threshold. We assume a TDD system where we can use DL SNR information for our UL transmissions.

For Rayleigh fading channel, the probability and cumulative distribution functions of SNR γ with mean $\bar{\gamma}$ is given in Section 3.6. We define a new SNR distribution $\hat{\gamma}$ where if γ equals or exceeds threshold γ_T, γ equals to the effective SNR $\hat{\gamma}$. The probability distribution of $\hat{\gamma}$

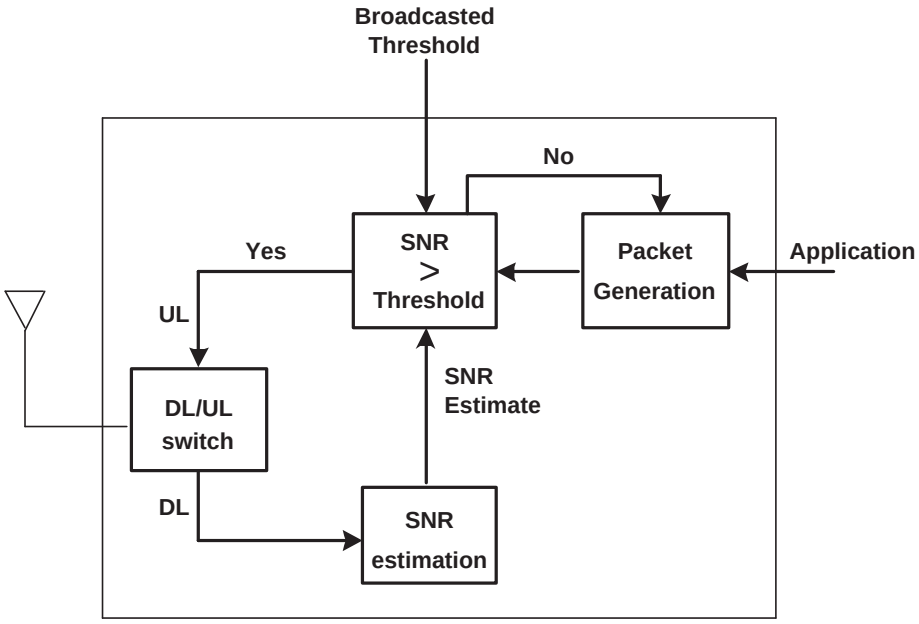

Figure 3.3. Transmission control at the subscriber station.

for user k and slot n can be expressed as:

$$f_{\hat{\gamma}_{k,n}}(\hat{\gamma}) = F_{\gamma_{k,n}}(\gamma_T)\delta[\hat{\gamma}] + f_{\gamma_{k,n}}(\hat{\gamma})U[\hat{\gamma} - \gamma_T] \tag{3.18}$$

where $\delta[\hat{\gamma}]$ and $U[\hat{\gamma}]$ are unit impulse and unit step functions, respectively. We assume all the users have the same threshold γ_T and the same SNR distribution. Instantaneous multi-access interference (MAI) of u users which have the same SNR distribution with threshold γ_T, can be expressed as:

$$\hat{\gamma}^u_{MAI} = \sum_{k=1}^{u} \hat{\gamma}_k \tag{3.19}$$

After some manipulations and substitution of p where $p = F_{\gamma_{k,n}}(\gamma_T)$, the probability density function of u interfering mobile station(s) was found as

$$f_{\hat{\gamma}^u_{MAI}}(\hat{\gamma}) = F_{\gamma_{k,n}}(\gamma_T)^u\delta(\hat{\gamma}) + \sum_{k=0}^{u-1} \left(\begin{array}{c} u \\ k \end{array} \right) \frac{p^k(\hat{\gamma} - \gamma_T(u-k))^{u-k-1}}{\bar{\gamma}^{u-k}(u-k-1)!} e^{-\hat{\gamma}/\bar{\gamma}} U[\hat{\gamma} - (u-k)\gamma_T] \tag{3.20}$$

Our main focus is to find the distribution of instantaneous received SINR μ due to u on-off interferers based on threshold γ_T. The instantaneous SINR μ can be expressed as

$$\mu = \frac{\hat{\gamma}}{1 + \hat{\gamma}^u_{MAI}} \tag{3.21}$$

The derivation details of the cumulative distribution function (CDF) of μ can be found in Appendix A.2. The throughput of the Aloha η is given by [14]

$$\eta = \mathrm{RG}(1 - p_F) \tag{3.22}$$

where R, G and p_F are the rate (bits/s/Hertz), offered load and the probability of decoding failure at a slot, respectively. The probability of decoding failure at a slot is expressed as:

$$p_F = \Pr(\log_2(1 + \mu) < R) \tag{3.23}$$

where μ is the instantaneous SINR of the reference user. After some manipulation, the probability of decoding failure can be expressed as

$$p_F = \sum_{k=0}^{u} \left(\begin{array}{c} u \\ k \end{array} \right) p_t^k (1 - p_t)^{u-k} F_\mu^k (2^R - 1) \tag{3.24}$$

where there are u interfering stations with probability of transmission p_t. The derivation of cumulative distribution function of μ with γ_T can be found in Appendix A.2. In order to find the optimum threshold that maximizes throughput in (3.22), the probability of decoding failure p_F in (3.24) needs to be minimized. The minimization problem can be formulated as

$$\frac{\partial p_F}{\partial \gamma_T} = 0 \tag{3.25}$$

For 2-user case, we find the optimum threshold which satisfies the condition in (3.25) as $\gamma_T = 2^R - 1$. With threshold $\gamma_T = 2^R - 1$, the throughput becomes a function of rate R and offered load G. Maximum throughput is achieved by some finite rate R. Figures 3.5 and 3.6 show the rate values that maximize the throughput. However for more than 2-user

case, the cumulative distribution function in (3.24) becomes so complex that we cannot find a closed form solution to (3.25). We tackle this problem by focusing on the limiting behaviour of the threshold when the offered load G goes to infinity. Asymptotically, the maximum throughput is achieved when only one user transmits in the system [38]. For users with independent, identically distributed (i.i.d) Rayleigh fading channels, the probability of having U active users in a system with offered load G is:

$$Pr(G = U) = \begin{pmatrix} G \\ U \end{pmatrix} (1 - F_{\hat{\gamma}}(\gamma_T))^U (F_{\hat{\gamma}}(\gamma_T))^{G-U} \tag{3.26}$$

where the average number of users becomes

$$\overline{U} = \sum_{l=0}^{G} l Pr(G = l) \tag{3.27}$$

with an i.i.d channel $\overline{U} = G(1 - F_{\hat{\gamma}}(\gamma_T))$. Asymptotically, we expect on average one user at each slot, thus the optimum threshold must satisfy

$$G(1 - F_{\hat{\gamma}}(\gamma_{T_{opt}})) = 1 \tag{3.28}$$

yielding

$$\gamma_{Topt} = \overline{\gamma} \log(G) \tag{3.29}$$

where $\overline{\gamma}$ and G are mean SNR and offered load, respectively. In Figure 3.8, we show that asymptotic threshold $\overline{\gamma} \log(G)$ and the threshold value $2^R - 1$ have similar throughput performance for large G values which verifies our threshold selection. Moreover, we did an exclusive search for the optimum threshold and in Figure 3.4, we plot the optimum threshold which maximizes the throughput in (3.22) for different load G values. In the same figure, we plot the asymptotic optimum threshold value ($\overline{\gamma} \log(G)$) as well. For large G values, the optimum threshold merges with the asymptotic optimum threshold. For opportunistic Aloha with large R, the average number of transmission grows faster than

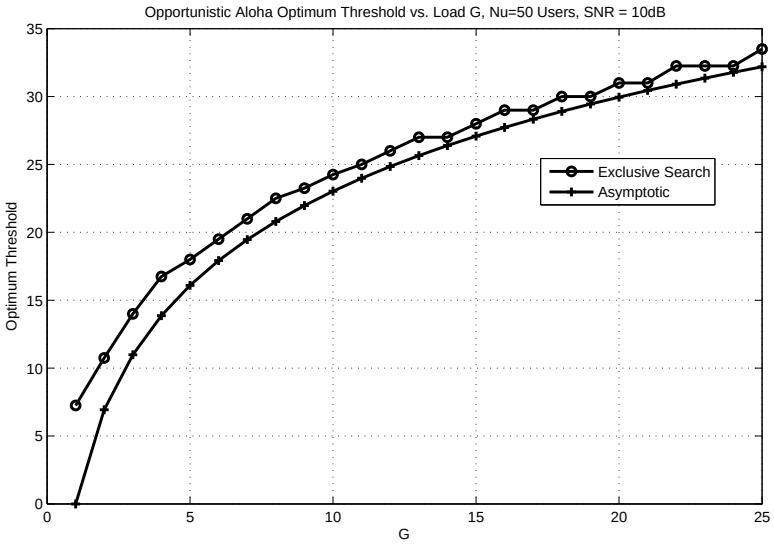

Figure 3.4. Optimum threshold vs. G for $N_u = 50$ user system.

R, thus we expect that throughput will go to 0. For opportunistic Aloha with large R, the probability of decoding failure in (3.24) becomes

$$\lim_{R\to\infty} p_F = \sum_{k=0}^{u} \left(\begin{array}{c} u \\ k \end{array} \right) p_t^k (1-p_t)^{u-k} = 1 \tag{3.30}$$

where $\lim_{R\to\infty} F_\mu^k(2^R - 1) = 1$. Then the throughput in (3.22) becomes

$$\lim_{R\to\infty} \eta = RG(1-p_F) = 0 \tag{3.31}$$

Figure 3.5 and Figure 3.6 show that throughput η goes to 0 for larger R values. For large SNR values, p_F will be always smaller than the probability of failure when there is only one interfering user in the system as expressed in (3.32)

$$\lim_{\gamma\to\infty} p_F < F^1(2^R - 1)(1-p_t) = (1 - e^{-(2^R-1)/\gamma})(1-p_t) \tag{3.32}$$

With (3.32), throughput in (3.22) becomes

$$\lim_{\gamma\to\infty} \eta \geq G\log(1+\gamma)(1-p_t) = \infty \tag{3.33}$$

With the increasing SNR, the throughput of the system also increases without a bound.

We compare analytical throughput results with and without a threshold for 2 users and 50 users systems with the same offered load $G = 1$ over Rayleigh fading channel. For the opportunistic Aloha, the transmission threshold for all of the users is set to $2^R - 1$. Moreover, the effect of increasing load G on throughput is investigated for both 2 and 50 user systems. In Figure 3.5, the average throughput of the 2 user system is compared with the system with and without a threshold. We see performance improvements with proposed threshold based scheme. The reason for the performance improvement is that the users access the random access channel selectively. When their SNR estimates are lower than the threshold, users delay the packet transmissions. By delaying the packet, they also decrease the interference to other random access users. Adding threshold to the system creates multi user diversity gains in the system. However, setting a high threshold will result in a throughput degradation due to under utilization of the channel and unnecessary delays. The results for 50 users with the same load $G = 1$ is shown in Figure 3.6. The results show that increasing number of users from 2 to 50 with the same load will decrease maximum throughput for both of the schemes due to high collision. However, the threshold based opportunistic Aloha scheme has still better performance than that of Aloha. In Figure 3.7, the maximum achievable throughput with respect to different load G is shown for 2-user case. It can be seen that setting an appropriate threshold to the system will increase the throughput for highly loaded system. In highly loaded systems, the collision probability will increase leading to lower overall throughput. The performance gap will increase with increasing load G. In Figure 3.8, we investigate the affect of load G in 50 users system. The opportunistic Aloha schemes performance increases with an increase on load. However, regular Aloha's performance will degrade due to high number of collisions. The proposed opportunistic Aloha scheme offers better throughput compared to the conventional Aloha system, at the expense of delay. Using threshold-based random access will decrease the

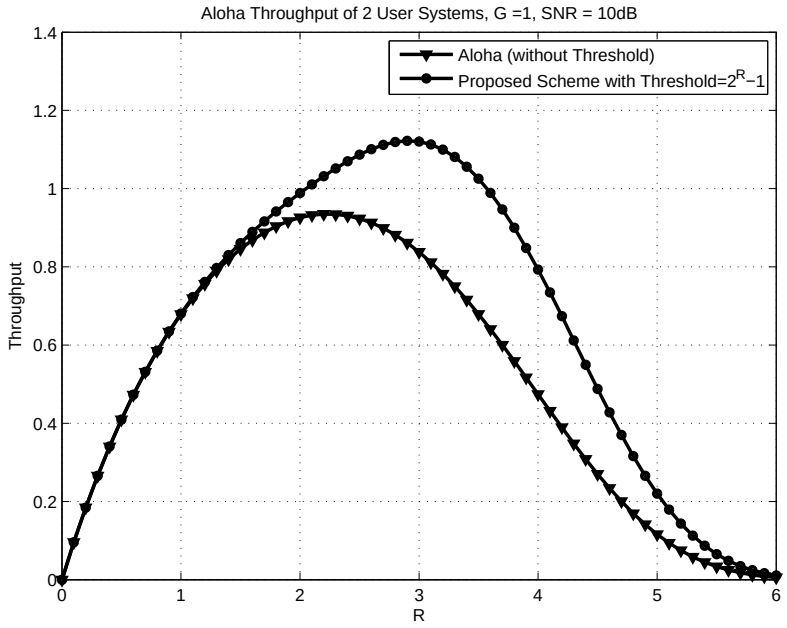

Figure 3.5. Throughput vs. Rate (R) for $N_u = 2$ user system with offered load $G = 1$.

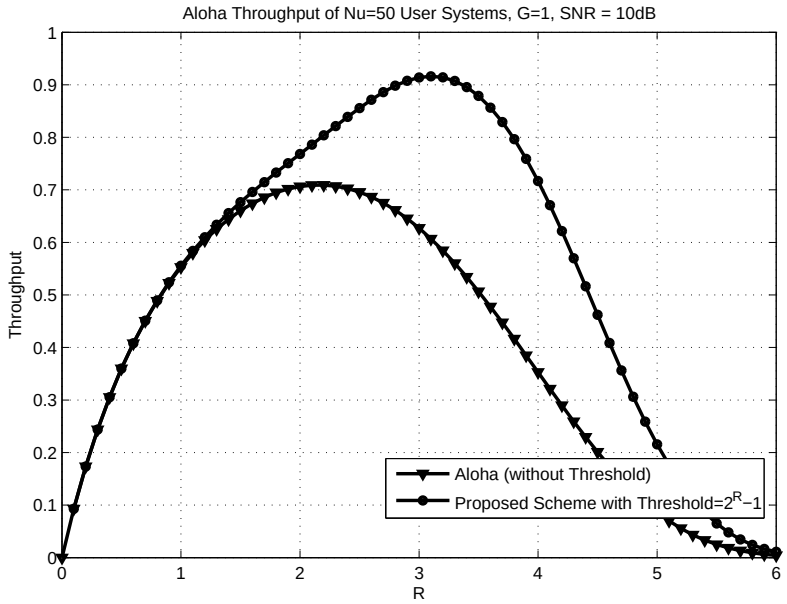

Figure 3.6. Throughput vs. Rate (R) for $N_u = 50$ user system with offered load $G = 1$.

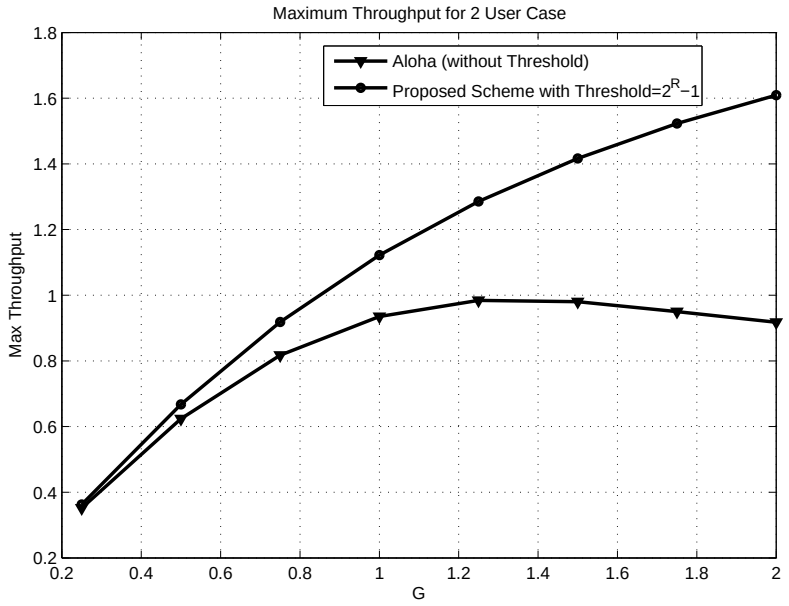

Figure 3.7. Maximum achievable throughput vs. offered load (G) in 2-user system.

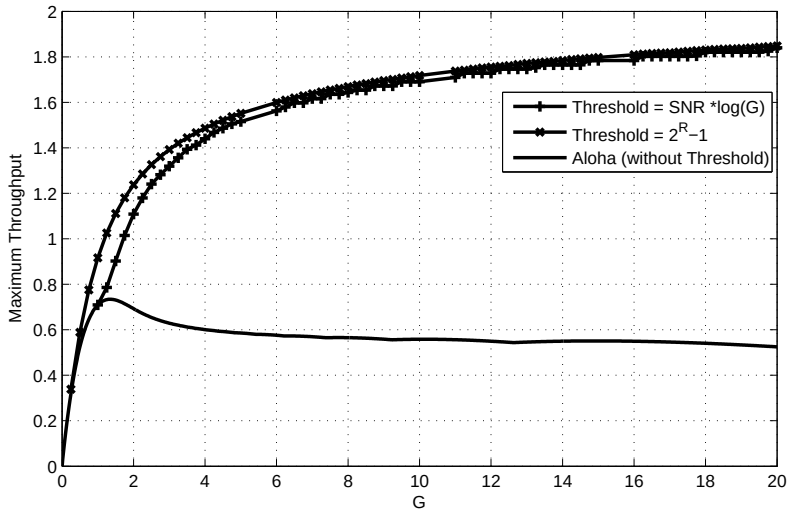

Figure 3.8. Maximum achievable throughput vs. offered load (G) in 50-user system.

collision probability and increase the throughput of the random access channel. Moreover, we found the asymptotic optimum threshold values which maximize the overall throughput of the multi-access system. We extend this work to multi antenna systems in the next section.

3.4 Multiple Antenna Systems

In this section , we consider Hybrid ARQ protocols with multi antennas. We focus on simple Aloha, chase combining (CC) and incremental redundancy (INR) Hybrid-ARQ schemes. Throughput of these Hybrid-ARQ schemes are analyzed for quasi static single input single output (SISO) scheme in [14]. The multi antenna channel is constant during each slot but changes independently from slot to slot. This scenario applies where the time between consecutive active slots is larger than the channel coherence time or for frequency selective fading where each user at each active slot picks different frequency according to a pseudo random hopping scheme. In [39], performance of Hybrid ARQ in block fading multi antenna channels is investigated.

We specify a multiple antenna discrete channel model by index s for the received signal. We express the received signal $\mathbf{y}_s$ at the base station in vector form during slot s with k active users as:

$$\mathbf{y}_s = \mathbf{H}_s \mathbf{x}_s + \mathbf{n}_s \tag{3.34}$$

where we define the transmit signal

$$\mathbf{x}_s = [\mathbf{x}_{1,s}, \mathbf{x}_{2,s}, \cdots, \mathbf{x}_{k,s}]^T \tag{3.35}$$

$$\mathbf{x}_{1,s} = [x_{1,1}, x_{1,2}, \cdots, x_{1,L}]^T$$

and the channel matrix

$$\mathbf{H}_s = [\mathbf{h}_{1,s}, \mathbf{h}_{2,s}, \cdots, \mathbf{h}_{k,s}]$$

where each users channel vector consists of LM element

$$\mathbf{h}_{1,s} = [h_{1,1}, h_{2,1}, \cdots, h_{L,1}, \cdots, h_{1,M}, h_{2,M}, \cdots, h_{L,M}]^T \tag{3.36}$$

where $\mathbf{n}_s$, M, L are AWGN noise vector at slot s, number of receive antennas, and number of bits in each slot, respectively. We assume that the fading of different users are independent of each other with $\frac{1}{LM}\mathrm{E}[\mathbf{H}_s^H \mathbf{H}_s] = \mathbf{I}$. Each user selects the slots for transmission randomly with a probability of transmission independently of the other users. The receiver at the base station can estimate the channels with high accuracy due to slowly varying nature of the channel. We use receive beamforming on the received signal $\mathbf{y}_s$ and the estimate for the kth user $\hat{\mathbf{d}}_{k,s}$ becomes

$$\hat{\mathbf{d}}_{k,s} = \mathbf{w}_{k,s}^H \mathbf{y}_s \tag{3.37}$$

where weight vector $\mathbf{w}_{k,s}$ for each user equals to complex conjugate of its normalized channel response $\mathbf{h}_{k,s}/\|\mathbf{h}_{k,s}\|$ [40]. A similar mechanism to concatenate beamforming and space time coding is presented in [41]. In the next section, we provide throughput analysis of Hybrid-ARQ schemes.

3.5 Throughput of Multiple Antenna Hybrid-ARQ Systems

We start our analysis for Rayleigh fading channel. The instantaneous SINR μ with M receive antennas can be expressed as follows

$$\mu = \frac{X_M}{1+Y} \tag{3.38}$$

where Y is multi-access interference (MAI) and X_M is the sum of $2M$ independent Gaussian random variables and distributed as chi-square with $2M$ degrees of freedom. With simple beamforming, the projection of interference signals over a normalized unit vector has the same distribution as the original signal so beamforming does not change the distribution of

the MAI compared to single antenna MAI. Instantaneous multi-access interference (MAI) of u users with same SNR distribution can be expressed as:

$$Y = \sum_{k=1}^{u} \gamma_k \tag{3.39}$$

The derivation details of the cumulative distribution function (CDF) of μ can be found in Appendix A.1. For the case of $M = 2$ antennas, the CDF of the instantaneous SINR conditioned on u interfering user can be expressed as:

$$F_{\mu,2}(x) \quad = \quad 1 - \frac{e^{-x/\gamma}}{(1+x)^u}\left(1 + \frac{x}{\gamma} + \frac{xu}{1+x}\right) \tag{3.40}$$

The probability of decoding failure at slot n which depends on the SINR distribution and the rate R expressed in Section 3.1 for different H-ARQ schemes. For Aloha, we have a closed form throughput equation using reward-renewal theorem in [14].

$$\eta_{\text{ALOHA}} = \frac{\text{RG}}{\text{E}[T]} = \frac{\text{RG}}{1 + \sum_{m=1}^{\infty} p(1)^m} = \text{RG}(1 - p(1)) \tag{3.41}$$

where $p(1) = \Pr(\log_2(1+\mu) \leq R)$, R and G are first slot decoding failure, rate (bits/s/Hertz) and offered load, respectively. For CC and INR, we can use characteristic function for the first m slots and models the remaining slots as Gaussian distributed random variables.

The average throughput is calculated for Aloha, CC and INR over Rayleigh fading channel. The effect of increasing load G on throughput is also investigated for 50 users systems. The throughput for multiple antenna configurations $M = 1, 2, 3, 4$ are calculated to see the impact of antenna number on the throughput. In Figure 3.9, for $M = 2$ antenna case, the throughput of the 50 users system is plotted with respect to load G for Aloha, CC and INR respectively. For large G, large number of users transmit in every slot. The throughput of Aloha, CC and INR merges to some specific values for large G due to high interference levels. In Figure 3.10, M number of antenna vs. throughput is plotted for highly load $G = 10$ systems. Increasing the number of antennas will increase the throughput of all Hybrid

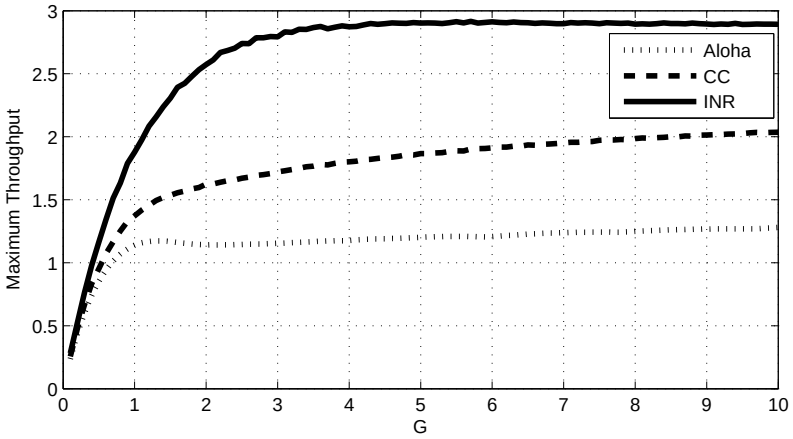

Figure 3.9. Load G vs. throughput η for $M = 2$ antennas, SNR $= 10dB$

ARQ systems on highly loaded systems. For highly loaded systems, throughput improves linearly with the number of antennas. Similar results are found using an experimental testbed in [42].

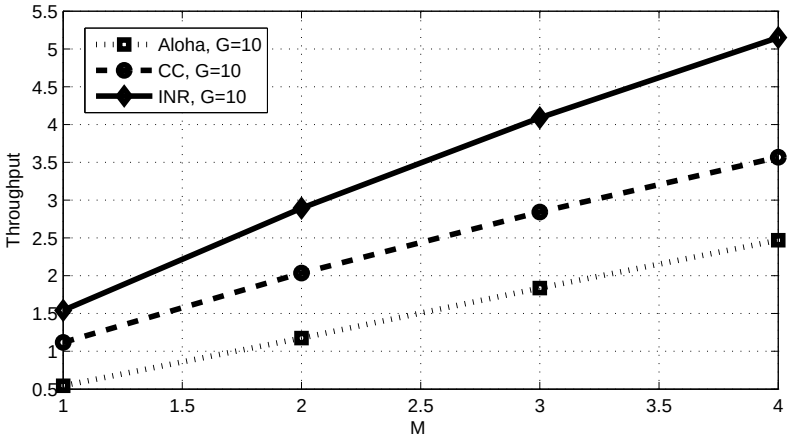

Figure 3.10. M antenna count vs. throughput, $G = 10$, SNR $= 10dB$

3.6 Conclusion

Combining channel coding and retransmission protocols are simple and effective solutions for reliable random multi-access communication. On top of that adding MIMO support to random access schemes improves the reliability of the system by adding more degrees of freedom. In this chapter, we have shown the affect of simple receive beamforming on Aloha, Chase combining and INR Hybrid-ARQ schemes and came up with analytical throughput expressions using reward renewal theorem. In the next chapter, we will focus on the power saving mechanism in the radio resource control.

CHAPTER 4
OPTIMIZATION OF DRX FOR POWER SAVING AND LATENCY

In this chapter, we focus on the power consumption issues of 3GPP Long Term Evolution Advanced (LTE-A) and present analytical models to LTE-A power saving mechanism called Discontinuous reception (DRX). DRX saves battery power of user equipment (UE) usually at the expense of potential increase in latency in the Long Term Evolution (LTE) networks. Therefore, an optimization is needed to find the best tradeoff between latency and power saving. We first develop an analytical model to estimate power saving achieved and latency incurred by DRX mechanism for active and background mobile traffic. A tradeoff scheme is then formulated to maintain a balance between these two performance parameters based on operator's preference for power saving and latency requirement of traffic. The analytical model is validated using system level simulation results obtained from OPNET Modeler. The results show that the proposed tradeoff scheme is efficient in keeping a balance between power saving and latency [43]. The results also indicate that DRX short cycles are very effective in reducing latency for active traffic, while shorter inactivity timer is desirable for background traffic to enhance power saving. We also propose a mechanism to switch DRX configuration based on traffic running at UE, using UE assistance procedure recently adopted by 3GPP in Release 11. DRX configuration switching increases the power saving significantly without any noticeable increase in latency of active traffic.

4.1 Introduction

Long Term Evolution Advanced (LTE-A) is designed to keep up with the today's rapidly growing data traffic, especially with the proliferation of smart phones and mobile applica-

tions. It has become a necessity to enhance LTE system in order to achieve higher data rates with less power consumption. LTE-A is targeted to support 1Gbps peak data rate in downlink (DL) in addition to the existing LTE capabilities. The increased data rate is achieved by adopting higher transmission bandwidth with carrier aggregation, higher order modulation, advanced coding techniques and advanced multiple antenna schemes [3]-[46]. Support for these features and diverse data traffic in LTE-A devices have resulted in battery hungry devices.

With the increased capacity over the air in LTE-A network, diverse data applications (DDAs) such as Facebook, Skype, twitter, email exchange, and messengers have widely become popular. Traffic characteristics of DDA are quite different than those of traditional internet applications [47]. These applications generate short packets in network every few seconds even when they are not being actively used by the users [47]. The short but frequent packets from such applications create crucial challenges as UE keeps changing states between Connected (RRC_Connected) and Idle (RRC_Idle), there by draining significant battery power continuously and causing excessive signalling overhead in the network [47][48]. We have used DDA and background traffic interchangeably throughout the rest of the section. Also, the terms Connected and Idle have been used for LTE device states RRC_Connected and RRC_Idle, respectivelly throughout the rest of the section.

Discontinuous Reception (DRX) has been adopted as an effective power saving mechanism, which largely addresses the power saving and unnecessary state transition issues [49][6]. DRX saves UE battery power by allowing UE to monitor DL control channel less frequently and go to sleep whenever there is no packet activity for the UE [7] [8]. Since power saving in DRX usually comes at the expense of increased end-to-end delay, DRX may create serious latency issues with active traffic where latency may be of prime focus. DRX mechanism in LTE is discussed in more detail in Section 4.2.

DRX configuration adaptation based on power saving and other performance parameters

in LTE is a relatively new research area. Authors in [50] have showed that power saving by DRX mechanism comes at the expense of degraded system utility. The authors proposed an adaptive DRX inactivity timer based on the channel condition, so that UE with bad channel quality will go to sleep for lesser time to increase its transmission opportunity (hence average throughput) at the cost of reduced power saving. Effect of DRX functionality on the traditional LTE scheduling has been presented in [18]. It proposes a DRX-aware scheduling scheme to reduce the packet loss rate due to DRX sleep. Authors in [51] have investigated algorithms to optimize DRX parameters to find a balance between power savings and the experienced throughput for a single user web-browsing session. Power saving achieved by DRX and its impact on QoS of Voice over IP (VoIP) traffic have been studied in [52] for dynamic and semi persistent packet scheduling strategies. The authors have shown that a tradeoff between power saving and spectral efficiency exists. For example, the power savings are larger for semi persistent scheduling than for dynamic packet scheduling at the cost of a lower spectral efficiency. In spite of these works, investigation of a complete framework to enable DRX Configuration adaptation with the help of UE assistance in order to find a tradeoff between power saving and other performance parameters such as latency based on traffic running at UE has not received much attention in the literature.

In this chapter, we first develop an analytical model to estimate the power saving and delay incurred by DRX operation for the cases of active and background traffic. These derived performance parameters are then used to formulate a multi-objective optimization problem through DRX configuration. This problem explores tradeoff between power saving and latency requirements. The tradeoff is customized based on latency requirements of active traffic and operator's preference for power saving. We show that DRX configuration optimization is crucial to leverage the power saving by DRX while maintaining a reasonable latency for active traffic. Since, latency requirements vary widely with applications, a single DRX configuration cannot be optimal for a UE as applications running on the UE change

over time. Therefore, a DRX configuration switching mechanism has been discussed to adapt the DRX setting based on most recent applications running at the UE. The proposed DRX switching mechanism uses UE assistance indication recently adopted in 3GPP Release 11. We have shown that DRX configuration switching, which adapts DRX parameters based on ongoing traffic at UE, can improve the power saving significantly without degrading the delay performance of active traffic.

The rest of the chapter is organized as follows. Section 4.2 describes DRX mechanism in LTE. Active and background traffic models are explained in Section 4.3. An analytical model has been proposed to evaluate performance parameters of DRX in Section 4.4 both for active and background traffic. Power consumption and delay formulations are given in 4.5. A tradeoff between power saving and latency has been explored in Section 4.6 by formulating a multi-objective optimization problem. A DRX configuration switching scheme has been proposed in Section 4.7. Then, in Section 4.8, simulation and analytical results are presented and discussed. The chapter is concluded in Section 4.9.

4.2 DRX Mechanism in LTE

An LTE device can be either in RRC_Connected or RRC_Idle after the power on of the device. As shown in Figure 4.1, the transition from one state to another depends on the traffic activity. Network runs an inactivity timer called RRC Inactivity timer to push the device to Idle state. When there is no packet activity for a long duration (equal to RRC Inactivity Timer), UE is moved to Idle state from Connected state. RRC Inactivity timer is reset with any uplink or downlink data activity.

In LTE, UE may be configured with a DRX mechanism by radio resource control (RRC) in both Connected and Idle states. However, the functionality and configuration parameters of DRX are different in these LTE states. Since data traffic mainly takes place in Connected

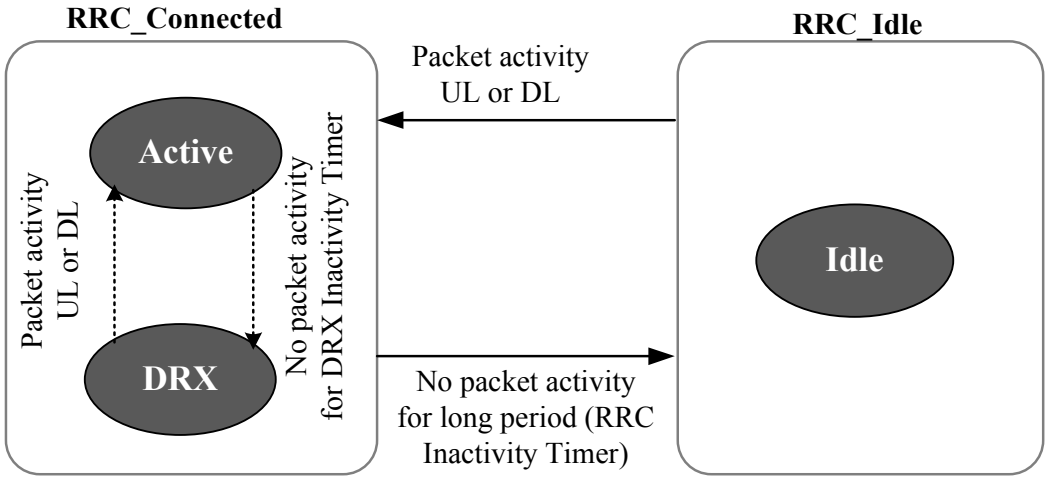

Figure 4.1. LTE and device states

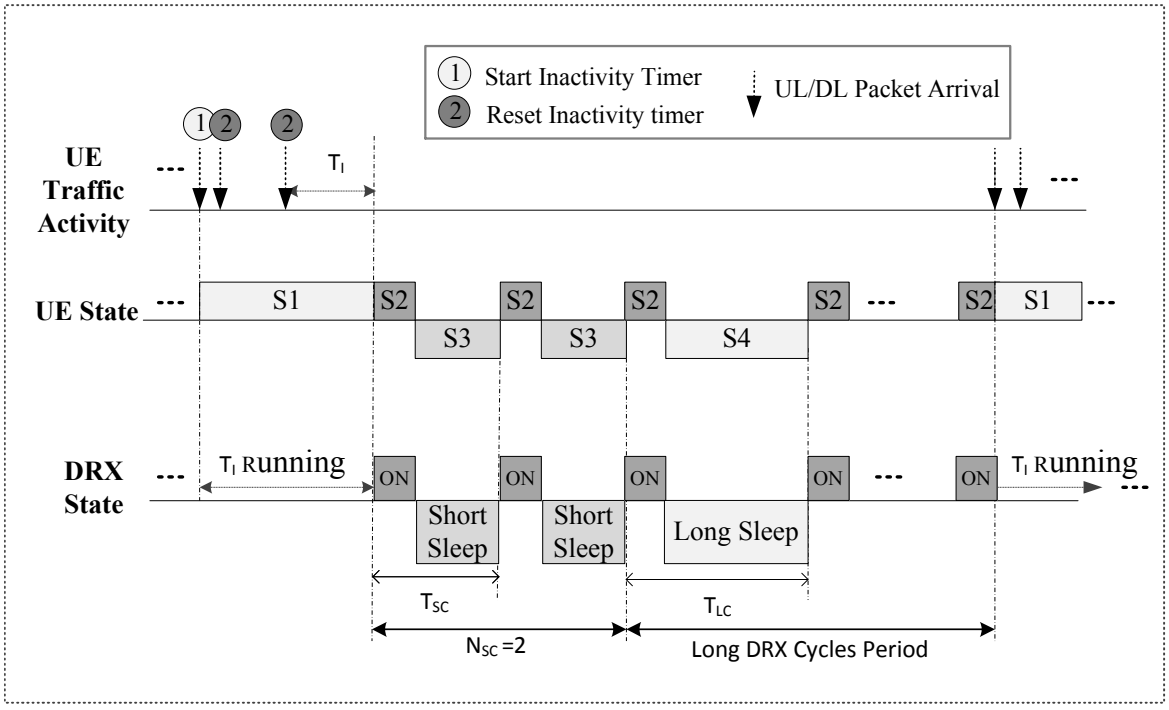

Figure 4.2. A typical scenario of DRX mechanism in LTE-A.

state, we focus on DRX in Connected state in this thesis. DRX can be configured by RRC on per UE basis to control UE's Physical DL Control Channel (PDCCH) monitoring activity in order to save UE's battery power [7][5]. Network runs an inactivity timer called DRX Inactivity timer for the connected UEs to push them to DRX mode. When there is no packet activity for a duration of DRX Inactivity timer, UE moves to DRX mode. DRX inactivity timer is also reset with any uplink or downlink data activity. DRX is configured by parameters such as On Duration Timer (T_{ON}), DRX Inactivity Timer (T_I), Long DRX Cycle (T_{LC}), and optionally the drxShortCycleTimer (N_{SC}) and Short DRX Cycle (T_{SC}) [7]. Figure 4.2 depicts a typical scenario of DRX mechanism considered in this chapter. UE with DRX configuration continuously monitors PDCCH for possible DL packet transmission when T_I is running. If T_I is running, T_I is restarted on any DL packet indication in PDCCH or UL packet transmission. Upon expiry of T_I, UE enters short DRX cycle as described in [7]. After N_{SC} consecutive short cycles, UE enters long DRX cycle. If short DRX cycle is not configured, UE enters long DRX cycle upon the expiry of T_I. Short DRX cycle is desirable for delay sensitive active traffic as it reduces overall latency [53], while it is not configured for background traffic where power saving is of prime focus.

During each DRX cycle, UE remains awake monitoring PDCCH for T_{ON} period and remains asleep for rest of the cycle called sleep period ($T_{SS} = T_{SC} - T_{ON}$ for short cycles and $T_{LS} = T_{LC} - T_{ON}$ for long cycles). DL packet indication in PDCCH or arrival of higher layer data packets for uplink (UL) transmission during T_{ON} terminates the DRX operation immediately and starts T_I. On the other hand, UL packet arrival during DRX sleep at UE may terminate DRX operation immediately or UE may wait for next T_{ON} period to terminate it. In this chapter, we have assumed latter case so that the UL packet arrived during DRX sleep is delayed until next T_{ON} period. For the sake of clarity and simplicity, Idle state is not shown in Figure 4.2. UE states $S_i \in \{1, 2, 3, 4\}$ are discussed in more detail in Section 4.4.

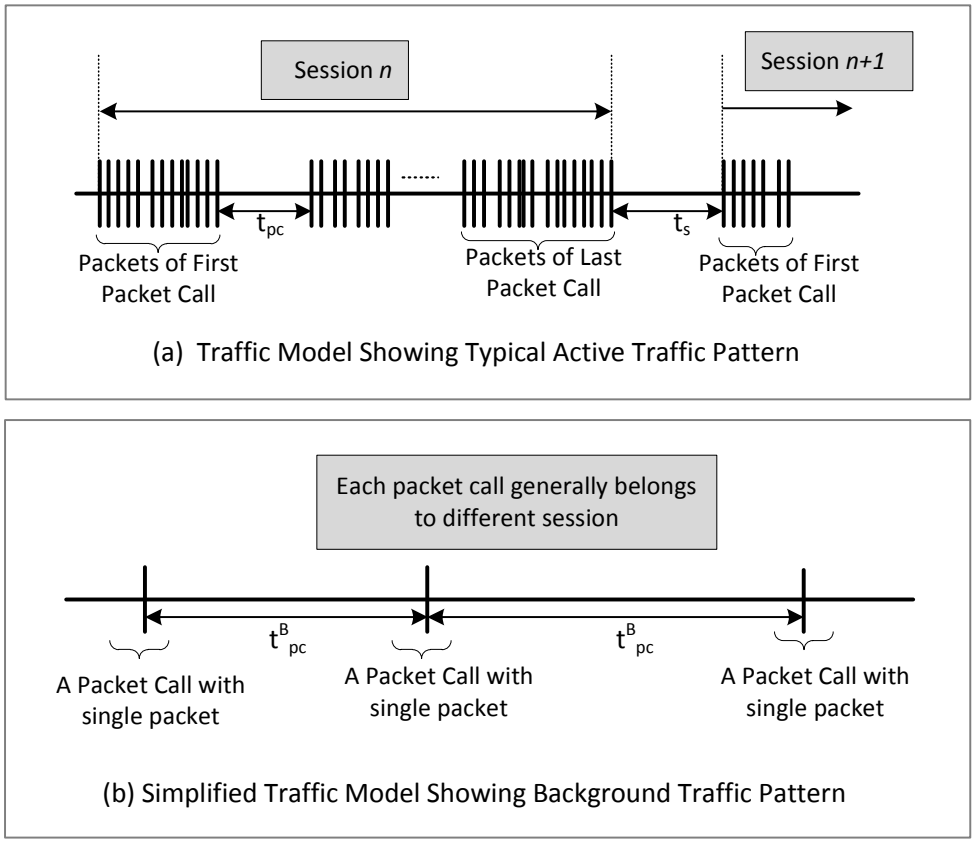

Figure 4.3. Traffic model showing typical active and background data traffic models.

4.3 Traffic Models

In this section, we consider two different traffic types which are active traffic and background traffic. Active traffic is valid when the user is actively using the device. On the other hand, background traffic is valid when the device is unattended but the still the applications are running.

4.3.1 Active Traffic

The active data traffic model usually consists of several sessions with exponentially distributed session inter-arrival time (t_s) as shown in Figure 4.3a. Each session of active traffic consists of multiple packet calls with exponentially distributed packet call inter-

arrival time (t_{pc}). The distribution parameters for active data traffic model considered in this chapter have been summarized in Table 4.1. At any time, the next packet call may either belong to the ongoing session or be the first packet call of a new session with probabilities P_{os} and P_{ns}, respectively. Since η_s has a geometric distribution, we have $P_{ns} = 1/\bar{\eta}_s$ and $P_{os} = 1 - P_{ns}$.

4.3.2 Background Traffic

Due to the increasing popularity of internet access over LTE network from smart devices, several new applications (such as social networking applications, weather update apps, IM apps, online gaming and so on) have recently emerged [54][55]. It has been found that these applications generate short packets frequently even when UE device is in unattended mode [47]. This traffic is known as background traffic. To simplify the traffic model, we assume that these applications generate background traffic which consists of multiple sessions and each session has a single packet call compared to multiple packet calls in active traffic. Based on real traffic traces collected in [47], each packet call usually consists of one or two short packets. Assuming that the inter-arrival time between packets of a packet call being less than T_I, the traffic model for these applications can be simplified as shown in Figure 4.3b. For the further analysis, we can safely assume that each packet call effectively has a single packet. We further assume that inter-arrival time between sessions (which is effectively inter-arrival time between packets in our simplified model) is exponentially distributed with mean inter-arrival time (t_{pc}^B). In the following section, we derive analytical expressions for performance parameters of LTE-A network with DRX mechanism in operation for both active and background traffic.

Table 4.1. Distribution Parameters of Traffic Model Under Consideration

Model Parameters	Distribution	Mean
Session inter-arrival time (t_s)	Exponential	$1/\lambda_s$
Number of Packet Calls per Session (η_s)	Geometric	$\bar{\eta}_s$
Packet call inter-arrival time (t_{pc})	Exponential	$1/\lambda_{pc}$
Number of Packets per Packet Call (η_{pc})	Geometric	$\bar{\eta}_{pc}$
Packet inter-arrival time (t_p)	Exponential	$1/\lambda_p$
Packet call inter-arrival time (t_{pc}^B)	Exponential	$1/\lambda_{pc}^B$

4.4 DRX Mechanism as Semi-Markov Process

We model DRX mechanim as a Semi-Markov process because times between state transitions (holding times) are random variables and also these random variables depend on the two states between the move is made [56][57].

4.4.1 Active Traffic

The Semi-Markov process representing DRX operation is shown in Figure 4.4a. Note that short and long DRX sleep states are shown for active traffic, while only long DRX sleep has been considered in case of background traffic. Based on the DRX operation described in Section 4.2, UE can be in one of the following four states:

- S_1 (Active State with T_I running): This is the highest power consuming state with UE being most active in transmitting/receiving data and/or monitoring PDCCH. Any new data activity restarts T_I. If there is no data activity for a period of T_I, than T_I expires and UE moves to state S_2.

- S_2 (DRX ON State with T_{ON} running): The expiry of T_I moves UE into S_2. Also, the end of DRX sleep moves UE in S_2 if there is no packet activity during the sleep period. Any packet arrival (DL or UL) during S_2 starts T_I and pushes UE back to S_1 immediately. Otherwise, expiry of T_{ON} moves UE to either short DRX sleep (state S_3) or long DRX sleep (state S_4). UE moves to S_4 only if it has been in S_3 for N_{SC}

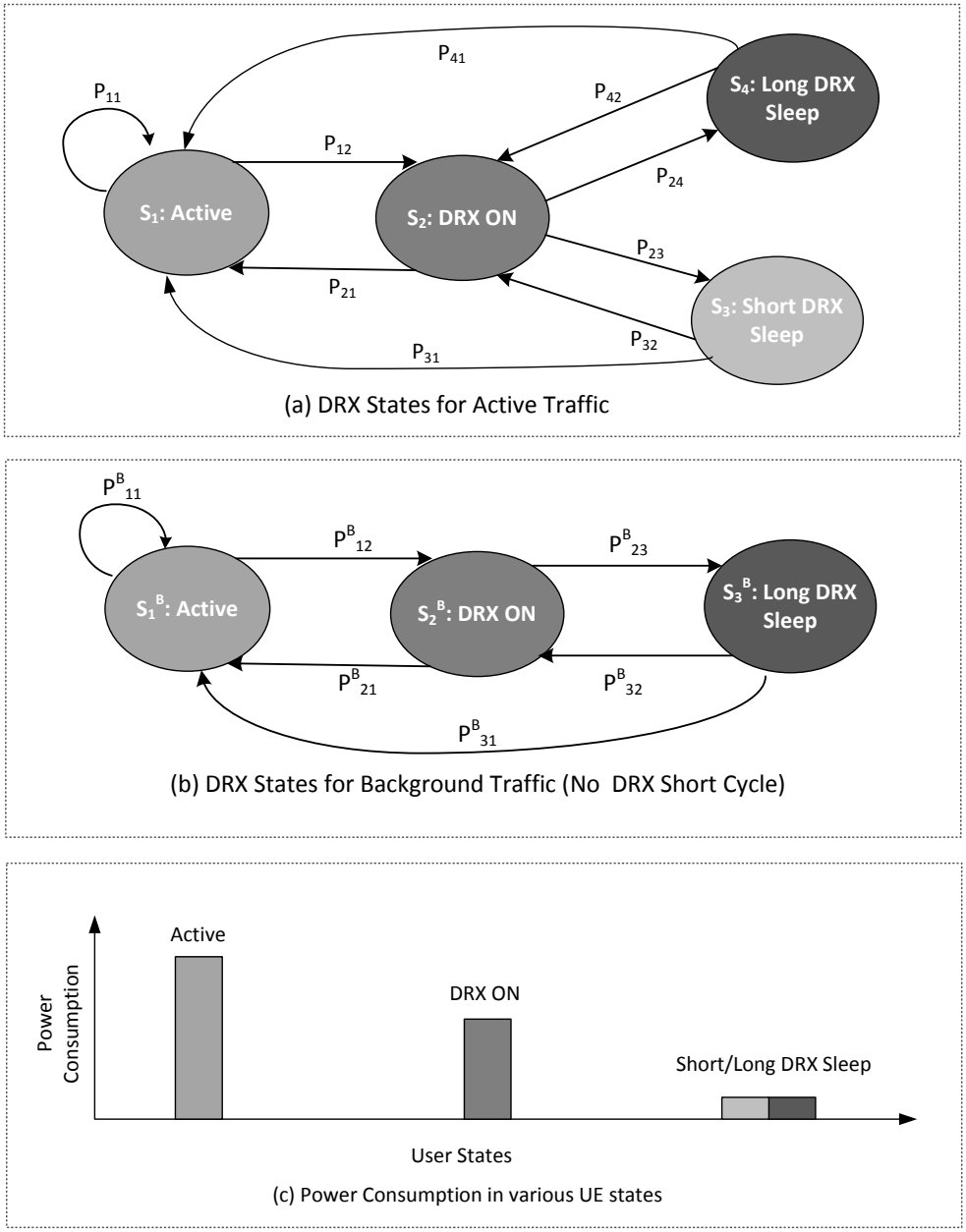

Figure 4.4. Semi-Markov Process Model for the proposed DRX mechanism. Power consumption bars shown in figure are not in exact scale.

times, consecutively. During DRX ON, UE does not receive/transmit data, however, it monitors PDCCH. It is a moderate power consuming state with power consumption being lower than that in S_1 and higher than that in S_3 and S_4.

- S_3 (Short DRX Sleep State): If there is any packet activity during S_3, UE moves to S_1 at the end of short DRX sleep, otherwise moves to S_2 at the end of sleep period. UE neither transmits/receives data nor monitors PDCCH in S_3.

- S_4 (Long DRX Sleep State): If there is any packet activity during S_4, UE moves to S_1 at the end of long DRX sleep, otherwise it moves to the S_2 at the end of the long sleep period. UE neither transmits/receives data nor monitors PDCCH in S_4.

4.4.2 Background Traffic

The Semi-Markov process representing DRX operation is shown in Figure 4.4b. Background traffic applications are usually delay tolerant; therefore, the short DRX cycle is not configured in practice. We assume that for background traffic, device directly enters long DRX after the DRX inactivity timer expiry. With this simplification, the UE running background traffic can be in one of the following three states:

- S_1^B (Active State with T_I running): This is the highest power consuming state with UE being most active in transmitting/receiving data and/or monitoring PDCCH. Any new data activity restarts T_I. If there is no data activity for a period of T_I, the T_I then expires and UE moves to state S_2^B. UE remains in this state for at least one T_I.

- S_2^B (DRX ON State with T_{ON} running): The expiry of T_I moves UE into S_2^B. Also, end of DRX sleep moves UE in S_2^B if there was no packet activity during the sleep period. Any packet arrival (DL or UL) during S_2^B starts T_I and pushes UE back to $S1$ immediately. Otherwise, expiry of T_{ON} moves UE to DRX sleep (state S_3^B).

During DRX ON, UE does not receive/transmit data, however, it monitors PDCCH. It is moderate power consuming state with power consumption being lower than that in S_1^B and higher than that in S_3^B.

- S_3^B (DRX Sleep State): If there is any packet activity during S_3^B, UE moves to S_1^B at the end of DRX sleep, otherwise moves to the S_2^B at the end of sleep period. UE neither transmits/receives data nor monitors PDCCH in S_3^B.

4.5 Power Saving and Delay Analysis

In this section, we analytically compute power saving and delay of DRX mechanism for both active traffic and background traffic.

4.5.1 Active Traffic

We have assumed that once a packet call starts, UE does not move to the DRX until all packets of that packet call are transmitted. Since packet inter-arrival time of packets of a packet call is relatively smaller than T_I, this assumption should not be unreasonable. With this assumption, the transition probabilities for Markov chain is shown in Figure 4.4a where P_{kl} $\forall k, l \in \{1, 2, 3, 4\}$ represent the transition probability from state k to l. When UE is in state S_1, UE restarts T_I if next packet call starts before expiry of T_I, otherwise moves to S_2. Therefore, P_{11} and P_{12} can be computed as

$$\begin{aligned} P_{11} &= \Pr\left[t_{pc} < T_I\right] P_{os} + \Pr\left[t_s < T_I\right] P_{ns} \\ &= P_{os}(1 - e^{-\lambda_{pc} T_I}) + P_{ns}(1 - e^{-\lambda_s T_I}), \end{aligned} \tag{4.1}$$

$$P_{12} = 1 - P_{os}(1 - e^{-\lambda_{pc} T_I}) - P_{ns}(1 - e^{-\lambda_s T_I}), \tag{4.2}$$

where Pr[.] is the probability of occurrence. When UE is in state S_2, UE moves to S_1 if packet arrives before T_{ON} expiry, otherwise it goes to either S_3 or S_4. First N_{SC} expiry of

T_{ON}, moves UE to S_3 and then after any expiry of T_{ON} moves UE to S_4. Therefore, P_{21}, P_{23} and P_{24} can be computed as

$$\begin{aligned} P_{21} &= \Pr\left[t_{pc} < T_{ON}\right] P_{os} + \Pr\left[t_s < T_{ON}\right] P_{ns} \\ &= P_{os}(1 - e^{-\lambda_{pc} T_{ON}}) + P_{ns}(1 - e^{-\lambda_s T_{ON}}), \end{aligned} \tag{4.3}$$

$$\begin{aligned} P_{23} &= P_{os} e^{-\lambda_{pc} T_{ON}} (1 - e^{-\lambda_{pc} N_{SC} T_{SC}}) \\ &\quad + P_{ns} e^{-\lambda_s T_{ON}} (1 - e^{-\lambda_s N_{SC} T_{SC}}), \end{aligned} \tag{4.4}$$

and,

$$\begin{aligned} P_{24} &= P_{os} e^{-\lambda_{pc} T_{ON}} e^{-\lambda_{pc} N_{SC} T_{SC}} \\ &+ P_{ns} e^{-\lambda_s T_{ON}} e^{-\lambda_s N_{SC} T_{SC}}. \end{aligned} \tag{4.5}$$

Similarly, we can compute P_{31}, P_{32}, P_{41} and P_{42} as

$$P_{31} = P_{os}(1 - e^{-\lambda_{pc} T_{SS}}) + P_{ns}(1 - e^{-\lambda_s T_{SS}}), \tag{4.6}$$

$$P_{32} = P_{os}(e^{-\lambda_{pc} T_{SS}}) + P_{ns}(e^{-\lambda_s T_{SS}}), \tag{4.7}$$

$$P_{41} = P_{os}(1 - e^{-\lambda_{pc} T_{LS}}) + P_{ns}(1 - e^{-\lambda_s T_{LS}}), \tag{4.8}$$

$$P_{42} = P_{os}(e^{-\lambda_{pc} T_{LS}}) + P_{ns}(e^{-\lambda_s T_{LS}}). \tag{4.9}$$

Then, we construct the transition probability matrix of the Markov chain as

$$P_T = \begin{bmatrix} P_{11} & P_{12} & 0 & 0 \\ P_{21} & 0 & P_{23} & P_{24} \\ P_{31} & P_{32} & 0 & 0 \\ P_{41} & P_{42} & 0 & 0 \end{bmatrix} \tag{4.10}$$

Let $\varphi_k \ \forall k \in \{1, 2, 3, 4\}$ denote the steady state probability of staying at state S_k of the Markov chain, respectively. Using $\sum_{k=1}^{4} \varphi_k = 1$ and balance equations $\varphi_k = \sum_{l=1}^{4} \varphi_l P_{lk}$,

we can obtain the expressions for steady state probabilities as

$$\varphi_1=\frac{1-P_{23}P_{32}-P_{24}P_{42}}{1-P_{23}P_{32}-P_{24}P_{42}+P_{12}(1+P_{23}+P_{24})}$$
$$\varphi_2=\frac{P_{12}}{1-P_{23}P_{32}-P_{24}P_{42}+P_{12}(1+P_{23}+P_{24})}$$
$$\varphi_3=\frac{P_{12}P_{23}}{1-P_{23}P_{32}-P_{24}P_{42}+P_{12}(1+P_{23}+P_{24})}$$
$$\varphi_4=\frac{P_{12}P_{24}}{1-P_{23}P_{32}-P_{24}P_{42}+P_{12}(1+P_{23}+P_{24})}$$

In the following, we obtain the expected value $E[.]$ of the holding time of different states of the DRX Semi-Markov process. Let us assume Θ_i for $i \in \{1,2,3,4\}$ denotes the holding time of the state S_i.

Holding time of state S_1: UE enters state S_1 after getting a packet call. We have assumed that once UE is active it transmits all packets of the ongoing packet call, before entering further DRX operation. After serving packets of ongoing packet call, UE leaves this state if there is no new packet call start within a duration of T_I. Otherwise, any new packet call arrival before T_I expiry restarts T_I and moves UE back to S_1. $E[\Theta_1]$ can be given as

$$E[\Theta_1] = t_{service} + \bar{t}_{T_I} \tag{4.11}$$

where $t_{service}$ is the average service time for transmitting all packets of a packet call and $\bar{t}_{T_I}$ is the average time duration for T_I restarts or expiry after serving an ongoing packet call. If t_{RTT} is the service time of a packet which consists of duration from transmission of a packet from UE (or eNB) to getting ACK from eNB (or UE), $t_{service}$ can be given as

$$t_{service} = \frac{E[\eta_{pc}-1]t_{RTT}}{\rho} = \frac{\bar{\eta}_{pc}-1}{\lambda_p}, \tag{4.12}$$

where $\rho = t_{RTT}\lambda_p$ is the packet-level offered load.

Also, $\bar{t}_{T_I}$ can be calculated as

$$\begin{aligned}\bar{t}_{T_I} &= P_{os}\left(\int_0^{T_I} t\Pr[t_{pc}=t]dt + \int_{T_I}^{\infty} T_I\Pr[t_{pc}=t]dt\right) \\ &+ P_{ns}\left(\int_0^{T_I} t\Pr[t_s=t]dt + \int_{T_I}^{\infty} T_I\Pr[t_s=t]dt\right) \\ &= P_{os}\left(\frac{1-e^{\lambda_{pc}T_I}}{\lambda_{pc}}\right) + P_{ns}\left(\frac{1-e^{\lambda_s T_I}}{\lambda_s}\right). \end{aligned} \tag{4.13}$$

Holding time of state S_2: When UE is in state S_2, there can be two cases: (I) UE moves to state S_1 due to a packet arrival before expiry of T_{ON}. (II) No packet arrives during T_{ON} and UE moves to states S_3 or S_4. Therefore, the holding time of state S_2 can be calculated as

$$\begin{aligned}E[\Theta_2] &= P_{os}\left(\int_0^{T_{ON}} tPr[t_{pc}=t]dt + \int_{T_{ON}}^{\infty} T_{ON}Pr[t_{pc}=t]\right) \\ &+ P_{ns}\left(\int_0^{T_{ON}} tPr[t_s=t]dt + \int_{T_{ON}}^{\infty} T_{ON}Pr[t_s=t]\right) \\ &= P_{os}\left(\frac{1-e^{\lambda_{pc}T_{ON}}}{\lambda_{pc}}\right) + P_{ns}\left(\frac{1-e^{\lambda_s T_{ON}}}{\lambda_s}\right). \end{aligned}$$

Holding times of states S_3 and S_4. Since any packet (UL or DL) arriving during DRX sleep period will be buffered until the next T_{ON}, $E[\Theta_3]$ and $E[\Theta_4]$ can be computed as

$$E[\Theta_3] = T_{SS} = T_{SC} - T_{ON}. \tag{4.14}$$

$$E[\Theta_4] = T_{LS} = T_{LC} - T_{ON}. \tag{4.15}$$

After knowing $E[\Theta_m]$ and φ_m $\forall m \in \{1,2,3,4\}$, we can now derive the expression for performance parameters such as the percentage of time spent by UE in sleep time (Γ_{SLEEP}) and buffering delay incurred by DRX. Since DRX sleep is the minimum power consuming state of LTE-A in Connected state, Γ_{SLEEP} can be a good metric to predict the power

saving by DRX. Γ_{SLEEP} can be obtained as

$$\Gamma_{\text{SLEEP}} = \frac{\varphi_3 E[\Theta_3] + \varphi_4 E[\Theta_4]}{\varphi_1 E[\Theta_1] + \varphi_2 E[\Theta_2] + \varphi_3 E[\Theta_3] + +\varphi_4 E[\Theta_4]}. \tag{4.16}$$

Overall delay expression can be given by

$$\delta_{\text{DRX}} = \sum_{l=1}^{4} P_{S_l} \delta_l \tag{4.17}$$

where P_{S_l} is the probability of a packet call arriving during state S_l and δ_l is the average delay during state S_l. Since we have assumed that any UL/DL packets arrived at UE/eNB for a sleeping UE are buffered until UE comes out of DRX sleep, there should be a delay only for the state S_3 and S_4. For simplicity, we assume that wireless systems is lightly loaded so that during active and ON duration (i.e. states S_1 and S_2) delays are $\delta_1 = \delta_2 \approx 0$. When a packet or a packet call arrives during a short sleep time, it suffers from a delay which can vary from 0 to T_{SS}. Similarly, a buffering delay of 0 to T_{LS} is imposed to a packet or a packet call arriving during the long sleep. We can compute δ_3 as

$$\begin{aligned} \delta_3 &= P_{os} \int_0^{T_{SS}} (T_{SS} - t) \Pr[t_{pc} = t] dt \\ &+ P_{ns} \int_0^{T_{SS}} (T_{SS} - t) \Pr[t_s = t] dt \\ &= T_{SS} - \frac{P_{os}}{\lambda_{pc}} \left(1 - e^{-\lambda_{pc} T_{SS}}\right) - \frac{P_{ns}}{\lambda_s} \left(1 - e^{-\lambda_s T_{SS}}\right) \end{aligned} \tag{4.18}$$

and similarly δ_4 as

$$\delta_4 = T_{LS} - \frac{P_{os}}{\lambda_{pc}} \left(1 - e^{-\lambda_{pc} T_{LS}}\right) - \frac{P_{ns}}{\lambda_s} \left(1 - e^{-\lambda_s T_{LS}}\right)$$

where $T_{SS} = T_{SC} - T_{ON}$ and $T_{LS} = T_{LC} - T_{ON}$. Now, the average buffering delay due to DRX mechanism can be estimated as

$$\delta_{\text{DRX}} = P_{S_3} \delta_3 + P_{S_4} \delta_4. \tag{4.19}$$

where P_{S_3} can be computed as

$$\begin{aligned} P_{S_3} = & \quad \alpha_{pc} \sum_{i=1}^{N_{SC}} e^{-i\lambda_{pc}T_{SC}} + \alpha_s \sum_{i=1}^{N_{SC}} e^{-i\lambda_s T_{SC}} \\ = & \quad \alpha_{pc} e^{-\lambda_{pc}T_{SC}} \left(\frac{1-e^{-N_{SC}\lambda_{pc}T_{SC}}}{1-e^{-\lambda_{pc}T_{SC}}} \right) \\ + & \quad \alpha_s e^{-\lambda_s T_{SC}} \left(\frac{1-e^{-N_{SC}\lambda_s T_{SC}}}{1-e^{-\lambda_s T_{SC}}} \right). \end{aligned} \tag{4.20}$$

where $\alpha_{pc} = P_{os}(1 - e^{-\lambda_{pc}T_{SS}})e^{-\lambda_{pc}(T_I - T_{SC} + T_{ON})}$, $\alpha_s = P_{ns}(1 - e^{-\lambda_s T_{SS}})e^{-\lambda_s(T_I - T_{SC} + T_{ON})}$.

And similarly P_{S_4} can be computed as

$$P_{S_4} = \beta_{pc} \sum_{k=1}^{\infty} e^{-k\lambda_{pc}T_{LC}} + \beta_s \sum_{k=1}^{\infty} e^{-k\lambda_s T_{LC}}, \tag{4.21}$$

where $\beta_{pc} = P_{os}(1 - e^{-\lambda_{pc}T_{LS}})e^{-\lambda_{pc}(T_I + N_{SC}T_{SC} - T_{LC} + T_{ON})}$ and
$\beta_s = P_{ns}(1 - e^{-\lambda_s T_{LS}})e^{-\lambda_s(T_I + N_{SC}T_{SC} - T_{LC} + T_{ON})}$. Note that both summations in (4.21) are geometric series with common ratios $e^{-\lambda_{pc}T_{LC}}$ and $e^{-\lambda_s T_{LC}}$. Since, $\lambda_{pc} \geq 0$, $\lambda_s \geq 0$ and $T_{LC} \geq 0$, we have $|e^{-\lambda_{pc}T_{LC}}| \leq 1$ and $|e^{-\lambda_s T_{LC}}| \leq 1$. Therefore, both summations in (4.21) converge and P_{S_4} can be simplified as

$$P_{S_4} = \beta_{pc} \left(\frac{e^{-\lambda_{pc}T_{LC}}}{1 - e^{-\lambda_{pc}T_{LC}}} \right) + \beta_s \left(\frac{e^{-\lambda_s T_{LC}}}{1 - e^{-\lambda_s T_{LC}}} \right). \tag{4.22}$$

4.5.2 Background Traffic

In this section, we focus on background traffic and simplified analytical model proposed in Section 4.5.1. Generally, short DRX cycle is not configured in case of background traffic as it is delay tolerant traffic [51]. Therefore, the DRX mechanism in this case is simplified having only three states as shown in Figure 4.4b.

Let $P_{kl}^B \; \forall k, l \in \{1, 2, 3\}$ represent the transition probability from state k to l of the Markov chain for background traffic case as seen in Figure 4.4b.

Then the transition probability matrix of the Markov chain for background case can be given as

$$P_T^B = \begin{bmatrix} P_{11}^B & P_{12}^B & P_{13}^B \\ P_{21}^B & P_{22}^B & P_{23}^B \\ P_{31}^B & P_{32}^B & P_{33}^B \end{bmatrix} = \begin{bmatrix} 1 - e^{-\lambda_{pc} T_I} & e^{-\lambda_{pc} T_I} & 0 \\ 1 - e^{-\lambda_{pc} T_{ON}} & 0 & e^{-\lambda_{pc} T_{ON}} \\ 1 - e^{-\lambda_{pc} T_{LS}} & e^{-\lambda_{pc} T_{LS}} & 0 \end{bmatrix} \tag{4.23}$$

Let $\varphi_m^B,\ \forall m \in \{1,2,3\}$ denote the steady state probability of staying at states S_m^B of the embedded Markov chain for background traffic, respectively. Then using $\sum_{m=1}^{3} \varphi_m^B = 1$ and balance equations $\varphi_m^B = \sum_{n=1}^{3} \varphi_n^B P_{nm}^B$, we can obtain the expressions for stationary distribution for background traffic case as

$$\boldsymbol{\varphi^B} = \begin{cases} \varphi_1^B &= \dfrac{1 - P_{23}^B P_{32}^B}{1 + P_{12}^B + P_{12}^B P_{23}^B - P_{23}^B P_{32}^B} \\ \varphi_2^B &= \dfrac{P_{12}^B}{1 + P_{12}^B + P_{12}^B P_{23}^B - P_{23}^B P_{32}^B} \\ \varphi_3^B &= \dfrac{P_{12}^B P_{23}^B}{1 + P_{12}^B + P_{12}^B P_{23}^B - P_{23}^B P_{32}^B} \end{cases} \tag{4.24}$$

$\Theta_m^B,\ \forall m \in \{1,2,3\}$ denote the holding times of states S_m^B for background traffic case, then $\Theta_m^B,\ \forall m \in \{1,2,3\}$ can be calculated as

$$E[\Theta_1^B] = \frac{1 - e^{\lambda_{pc} T_I}}{\lambda_{pc}}, \tag{4.25}$$

$$E[\Theta_2^B] = \frac{1 - e^{\lambda_{pc} T_{ON}}}{\lambda_{pc}}, \tag{4.26}$$

$$E[\Theta_3^B] = T_{LS} = T_{LC} - T_{ON}. \tag{4.27}$$

The percentage of time spent by UE in sleep time in case of background traffic (Γ_{SLEEP}^B) can be given as

$$\Gamma_{\text{SLEEP}}^B = \frac{\varphi_3^B E[\Theta_3^B]}{\varphi_1^B E[\Theta_1^B] + \varphi_2^B E[\Theta_2^B] + \varphi_3^B E[\Theta_3^B]} \tag{4.28}$$

Proceeding similar to the case of active traffic in Section 4.5.1, the buffering delay in case of background traffic can be calculated as

$$\begin{aligned}\delta^B_{\text{DRX}} &= \sum_{i=1}^{\infty}\left[e^{-\lambda_{pc}(T_I+T_{ON}+(i-1)T_{LC})}(1-e^{-\lambda_{pc}T_{LS}})\right]\delta^B_3 \\ &= \beta_B\left(\frac{e^{-\lambda_{pc}T_{LC}}}{1-e^{-\lambda_{pc}T_{LC}}}\right)\delta^B_3, \end{aligned} \tag{4.29}$$

where $\beta_B = (1-e^{-\lambda_{pc}T_{LS}})e^{-\lambda_{pc}(T_I+T_{ON}-T_{LC})}$ and δ^B_3 is the average delay during long sleep in case of background traffic which is given by

$$\begin{aligned}\delta^B_3 &= \int_0^{T_{LS}}(T_{LS}-t)\Pr[t_{pc}=t]dt \\ &= T_{LS}-\frac{1}{\lambda_{pc}}\left(1-e^{-\lambda_{pc}T_{LS}}\right). \end{aligned} \tag{4.30}$$

4.6 Optimization Formulation for Power Saving and Latency

Usually power saving in DRX comes at the expense of degraded delay performance [52]. A trade-off between delay and power saving may be required in a practical network. In this section, we propose a scheme to optimize DRX parameters in order to minimize DRX buffering delay and maximize power saving jointly. This can be achieved by using a concept of multi-objective optimization in the problem formulation. In multi-objective optimization, we may introduce preferences of objectives by introducing *apriori* weights to the objective functions [58][59]. These weights can be correlated with the LTE-A operator's preference for power saving over latency.

Since maximizing DRX sleep (which reflects power saving) is equivalent to minimizing active time γ_{ACT}, we can select γ_{ACT} as one of the objectives. γ_{ACT} is $(1-\gamma_{\text{SLEEP}})$ and $(1-\gamma^B_{\text{SLEEP}})$ for the cases of active and background traffic, respectively. The other objective can be taken as minimizing DRX buffering delay, say δ_{NORM}. Note that buffering delay is the main component contributed by DRX to increase the end-to-end latency. To get a single objective function by combining these two objectives, we first need to normalize

the individual objectives to transfer them to dimensionless normalized objectives. γ_{ACT} is dimensionless and already normalized as $0 \leq \gamma_{\text{ACT}} \leq 1$. Second normalized objective can be taken as $\delta_{\text{NORM}} = P_{S_3}\delta_3/T_{SS} + P_{S_4}\delta_4/T_{LS}$ for active traffic or $\delta_{\text{NORM}} = \delta^B_{\text{DRX}}/T_{LS}$ for background traffic. Then, the multi-objective optimization problem can be formulated as **P1:**

$$\min_{\mathcal{T}_{\text{SC}},\mathcal{N}_{\text{SC}},\mathcal{T}_{\text{LC}}} \quad w\Gamma_{\text{ACT}} + (1-w)\delta_{\text{NORM}} \tag{4.31}$$

$$\text{subject to} \quad T_{SC} \in \mathcal{T}_{\text{SC}}, N_{SC} \in \mathcal{N}_{\text{SC}}, T_{LC} \in \mathcal{T}_{\text{LC}} \tag{4.32}$$

$$0 \leq w \leq 1,$$

where w defines the operator's preference for power saving over DRX induced latency. $w = 1$ represents power minimized DRX configuration while $w = 0$ represents latency minimized DRX configuration [60]. $\mathcal{T}_{\text{SC}}$, $\mathcal{N}_{\text{SC}}$ and $\mathcal{T}_{\text{LC}}$ are sets of possible values of T_{SC}, N_{SC} and T_{LC}, respectively as provided in 3GPP standard [8]. In case of background traffic, $\mathcal{T}_{\text{SC}}$ and $\mathcal{N}_{\text{SC}}$ are null sets. T_I usually has less impact on γ_{ACT} and δ_{NORM} compared to DRX parameters considered for optimization in **P1**. Therefore, we may select a suitable value of T_I and then optimize the problem to simplify the calculation. Since DRX parameters are sets of small number of known values, even exhaustive search is feasible to get the solution of the optimization problem **P1** in real time.

The weight w may depend on ongoing applications at UE and therefore can vary over time. In the following, we present a DRX switching mechanism to adapt DRX to the ongoing traffic.

4.7 DRX Configuration Switching with UE Assistance Mechanism

Recently, 3GPP Release 11 has adopted a UE assistance mechanism which enables UE to send its power preference to the LTE-A network [8]. For this purpose, a single bit UE

assistance information called power preference indication (PPI) can be sent by UE to the network. PPI = 1 indicates UE's preference to go to the low power consumption state while PPI = 0 represents UE's intention to be in a normal state which is preferable for delay sensitive traffic. PPI triggering has been left to UE vendors, while how to configure low power or low delay state has been left to network vendor. In this section, we present a use case of DRX configuration switching using PPI. We assume that UE has some level of intelligence to identify the active versus background traffic and trigger sending PPI (PPI=0 for active traffic and PPI = 1 background). eNB optimized the DRX configuration based on traffic information (i.e. PPI) from UE.

For active traffic, eNB can solve **P1** either with $w = 0$ or the preference defined by operator for active traffic such as lower w [61]. Alternately, the tradeoff between power saving and delay can also be optimized based on QoS requirement of the active traffic running at UE. For example, active traffic may have a delay constraint which must be satisfied by the network. For such a traffic, **P1** can be formulated as problem **P2** to maximize the power saving as follows.

$$\min_{\mathcal{T}_{\mathrm{SC}},\mathcal{N}_{\mathrm{SC}},\mathcal{T}_{\mathrm{LC}}} \quad \Gamma_{\mathrm{ACT}} \tag{4.33}$$

$$\text{subject to} \quad T_{SC} \in \mathcal{T}_{\mathrm{SC}}, N_{SC} \in \mathcal{N}_{\mathrm{SC}}, T_{LC} \in \mathcal{T}_{\mathrm{LC}},$$

$$\delta_{\mathrm{DRX}} \leq \delta_{\mathrm{MAX}}, \tag{4.34}$$

where δ_{MAX} is the upper bound on buffering delay defined by operator for active traffic. On the other hand, for background traffic which practically does not have delay requirement, eNB can find the DRX parameters by solving **P1** with $w = 1$ or higher $w \leq 1$. Alternately, the operator can set a target for power saving in case of background traffic as given in problem **P3**.

$$\min_{\mathfrak{T}_{\mathrm{LC}}} \quad \delta^{B}_{\mathrm{DRX}} \tag{4.35}$$

$$\text{subject to} \quad T_{LC} \in \mathfrak{T}_{\mathrm{LC}},$$

$$\Gamma_{\mathrm{ACT}} \leq \Gamma_{\mathrm{MAX}}, \tag{4.36}$$

where Γ_{MAX} ensures minimum power saving set by operator. DRX switching mechanism has been summarized in Algorithm 1.

Algorithm 1 DRX Configuration Switching

```
ppi ← PPI from UE
if ppi == 0 then
    Solve P1 with w = 0 or Solve P2
    OptimalDRXConfiguration ← {T_LC,optimal, T_SC,optimal, N_SC,optimal}
else
    Solve P1 with w = 1 or Solve P3
    OptimalDRXConfiguration ← {T_LC,optimal}
end if
return OptimalDRXConfiguration
```

4.8 Performance Evaluation

We have used OPNET Modeler 17.1 [62] to obtain the system level simulation results while MATLAB has been used to obtain the analytical results based on the analytical model proposed in Section 4.4. We used OPNET build in application and TCP/IP stacks, however we adjusted traffic characteristics in OPNET application layer to align with the traffic parameters used in analytical results. In Figure 4.5, protocol stack of both UE and eNB are shown. LTE protocol stack consists of LTE Non-access stratum (NAS) and LTE Access stratum (AS). Non-access stratum (NAS) is a functional layer between a core network and user equipment and supports signalling and traffic between those two elements. On the other hand, access stratum (AS) is a functional layer between radio network and

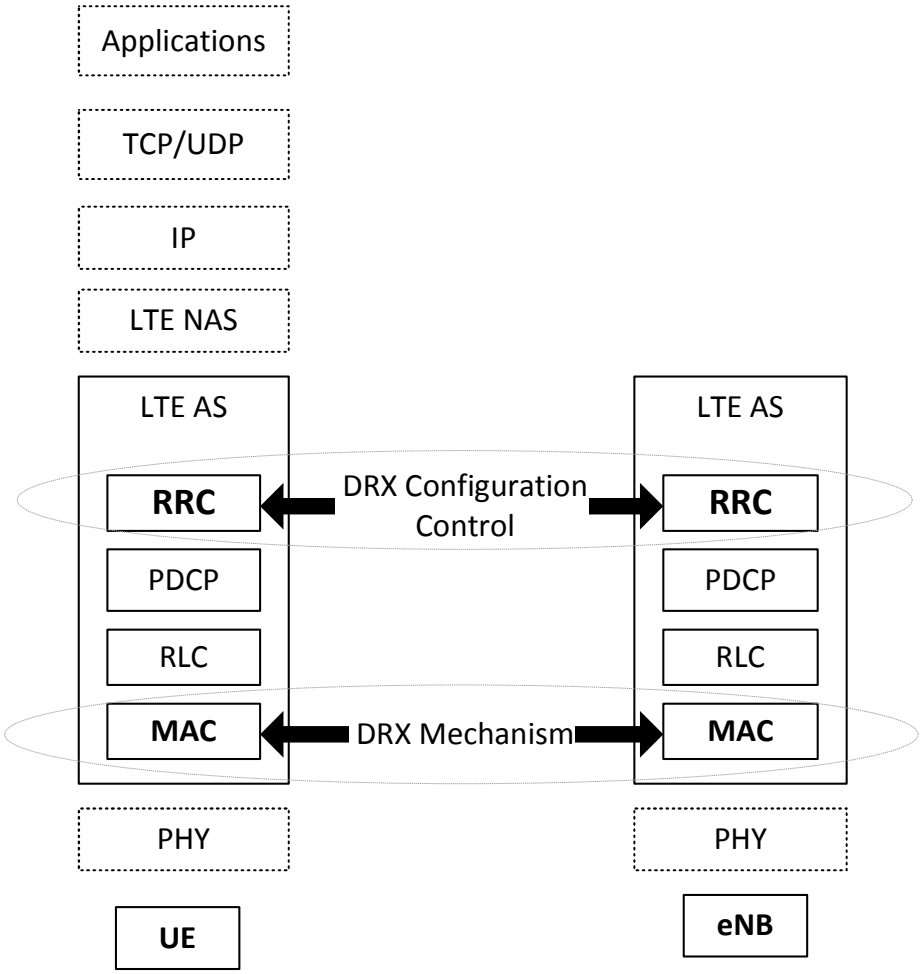

Figure 4.5. LTE Protocol stacks in OPNET implementation and the protocol layers impacted by DRX mechanism

user equipment. In this chapter, we focus on access stratum and modify some of the functionalities of the access stratum to enable our proposed algorithms.

Radio Resource Control (RRC) handles the control plane signalling which includes functions for connection establishment and release, broadcast of system information, radio bearer establishment/reconfiguration and release, RRC connection mobility procedures, DRX configuration, and paging notification. Packet Data Convergence Protocol (PDCP) layer is responsible for IP header compression and decompression and also for the transfer of user data to RLC. Radio Link Control (RLC) layer functionality consists of error correction

through ARQ, concatenation, segmentation, reordering, and duplicate detection. Medium Access Control (MAC) layer is responsible of creating the frames that will be send over physical layer. Medium access layer is also responsible of link adaptation, downlink/uplink scheduling, random access, and user data rate control.

In our simulations, we modify the RRC layer to dynamically change the DRX configuration and also we changed the MAC layer to support both short and long DRX cycles. Details of simulation parameters are provided in Table 4.2.

Table 4.2. OPNET Simulation Parameters for DRX Optimization

Parameter	Value
LTE Bandwidth/Duplexing	10MHz/FDD
Sub-carrier spacing	15 KHz
Sub-frame length (TTI)	1 ms
Number of symbols per TTI	14
Number of data/control symbols per TTI	11/3
DRX ON Duration (ms)	4
DRX Inactiviy Timer (ms)	30
Short DRX Cycle (ms)	20, 40, 80, 160, 320, 640
Long DRX Cycle	4 times Short DRX Cycle
$1/\lambda_s$ (s)	60
$1/\lambda_{pc}$ (s)	0.5
$1/\lambda_p$ (s)	0.01
$\bar{\eta}_s$	6
$\bar{\eta}_{pc}$	20
λ_{pc}^B (s)	10

We have considered UE's battery power saving and added latency due to DRX operation as two major performance metrics in order to evaluate the performance of DRX operation. Since DRX sleep is the minimum power consuming state of UE in Connected state, the percentage of time spent in DRX sleep has been chosen to estimate the UE's battery power saving.

Figure 4.6 and Figure 4.7 show the variation of sleep time and delay with various long DRX cycles, respectively. Both Figure 4.6 and Figure 4.7 depict results obtained from the

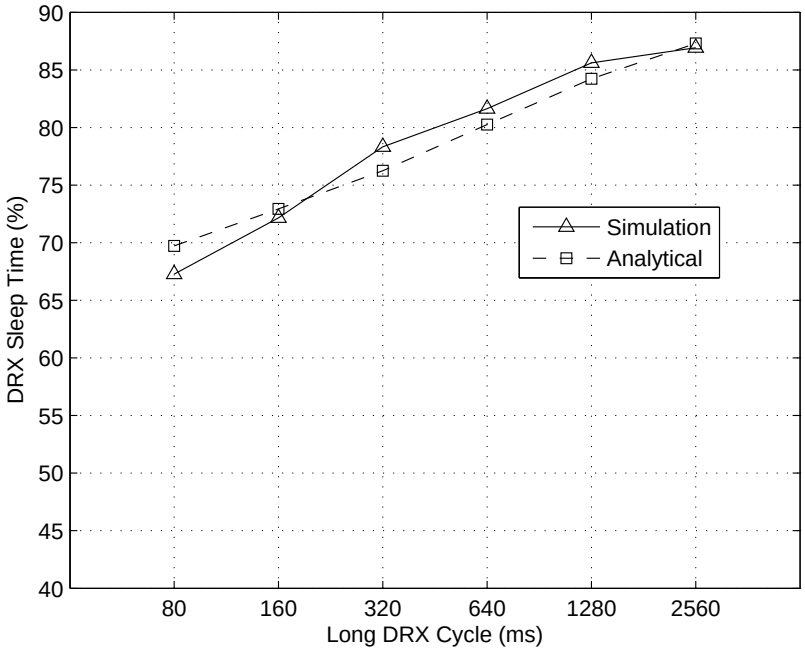

Figure 4.6. Time spent in DRX sleep versus DRX cycles for active traffic. $T_{ON} = 4ms,\ T_I = 30ms,\ T_{LC} = 4 * T_{SC}$ and $N_{SC} = 2$.

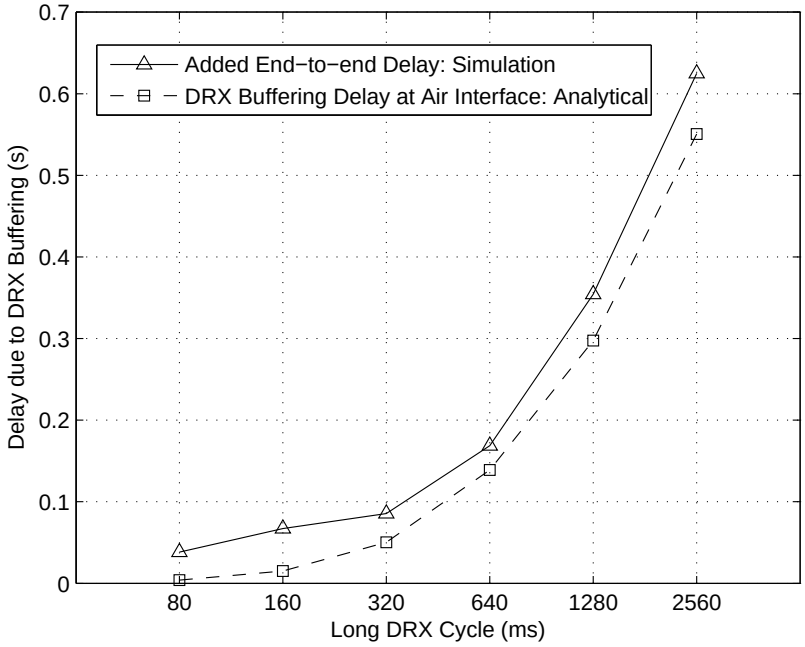

Figure 4.7. Buffering delay due to DRX sleep and its effect on end-to-end delay versus DRX cycles for active traffic. $T_{ON} = 4ms,\ T_I = 30ms,\ T_{LC} = 4 * T_{SC}$ and $N_{SC} = 2$.

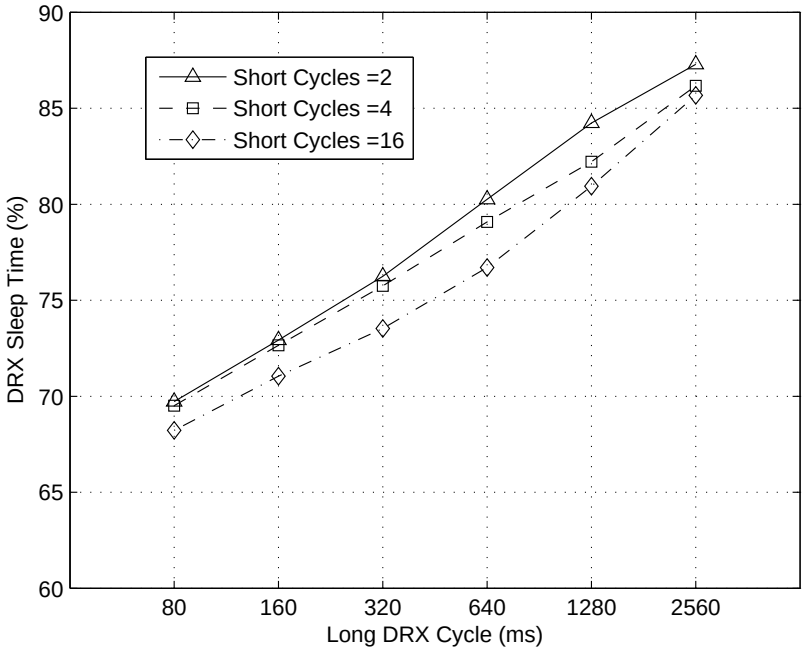

Figure 4.8. Time spent in DRX sleep versus DRX cycles for various N_{SC} for active traffic. $T_{ON} = 4ms,\ T_{LC} = 4 * T_{SC}$ and $T_I = 30ms$.

analytical model developed in this chapter and those obtained from system level simulation using OPNET simulator. Since analytical and simulation results of sleep time are closely matched, these results verify the accuracy of analytical model to represent DRX operation. Also, as seen from Figure 4.7, the variation of added end-to-end packet delay (obtained from simulation results) due to DRX has similar characteristics as that of DRX buffering delay obtained from the analytical model. It is obvious from Figure 4.6 that longer DRX cycles are preferable to maximize power saving. The reason is that for given T_{ON} and N_{SC}, longer DRX cycles increase the sleep time in each cycle. However, latency also increases with longer DRX cycles due to longer buffering in each cycle as depicted in Figure 4.7. Shorter DRX cycles are preferable for delay sensitive active traffic at the cost of lowered sleep time. From Figure 4.7, added end-to-end delay is always higher than analytical buffering delay. As analytical model establishes a lower bound for the end-to-end delay.

Figures 4.6 and 4.7 clarify the need for optimizing DRX configuration for a tradeoff between

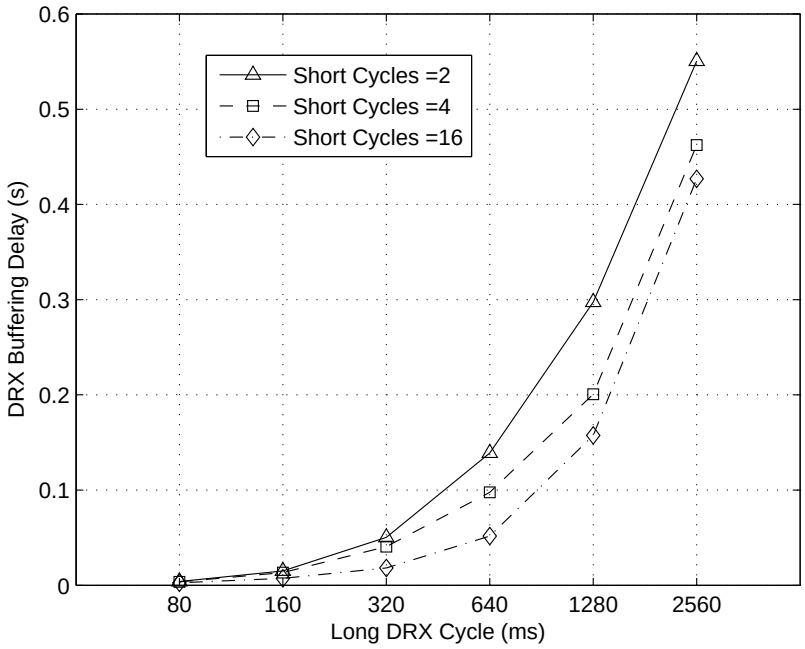

Figure 4.9. Buffering delay due to DRX sleep versus DRX cycles for various N_{SC} for active traffic. $T_{ON} = 4ms,\ T_{LC} = 4 * T_{SC}$ and $T_I = 30ms$

power saving and latency for active traffic. The tradeoff between power saving and latency is also quite dependent on the number of short cycles (N_{SC}) as shown in Figures 4.8 and 4.9. Higher N_{SC} decreases the probability of entering the long DRX cycle and hence decreases the buffering delay. Therefore, higher N_{SC} decreases the latency as seen in Figure 4.9 and should be preferable for delay-sensitive applications particularly when long DRX cycle is very large. For example for $T_{LC} = 2560ms$, a buffering delay improvement of about 29% can be achieved by changing N_{SC} from 2 to 16 as seen in Figure 4.9, while the power saving reduces by less than 2% for the same configuration change as seen in Figure 4.8. The use of short cycle is therefore recommended for active traffic where latency is crucial.

Figure 4.10 shows the variation of sleep time with various long DRX cycles for background traffic. Figure 4.10 depicts results obtained from the analytical model and system level simulation using OPNET simulator. Analytical and simulation results of sleep time are closely matched for background traffic as well. Background traffic activity is very light

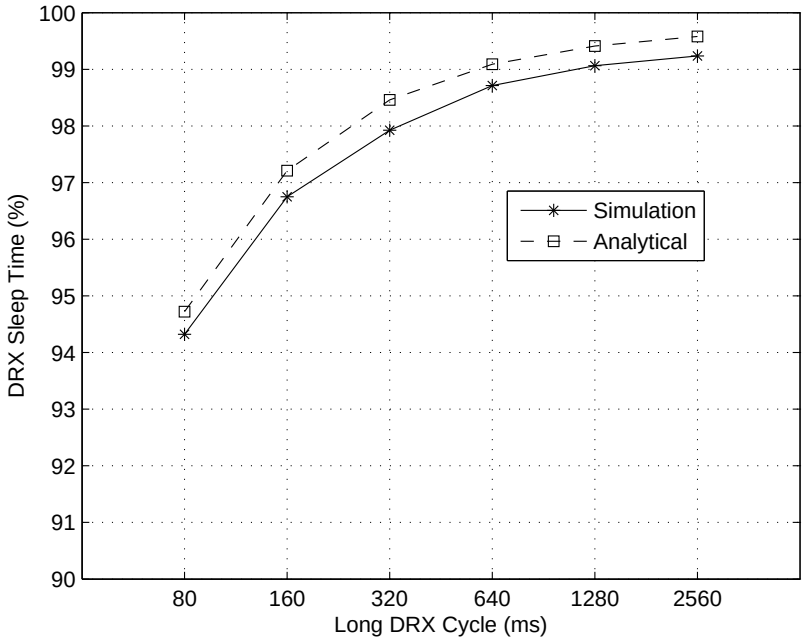

Figure 4.10. Time spent in DRX sleep versus DRX cycles for background traffic. $T_{ON} = 4ms,\ T_I = 30ms$ and $N_{SC} = 0$.

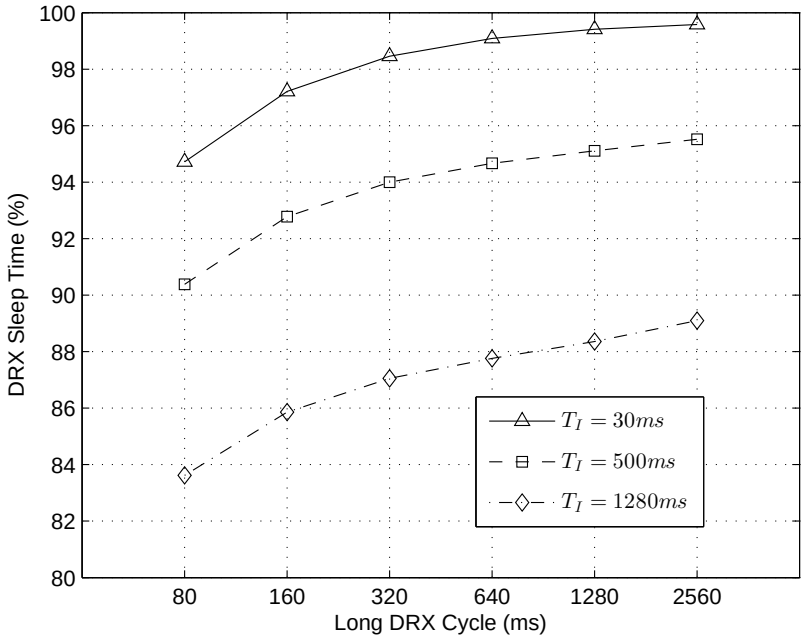

Figure 4.11. Time spent in DRX sleep versus DRX cycles for various T_I for background traffic. $T_{ON} = 4ms,\ N_{SC} = 0$.

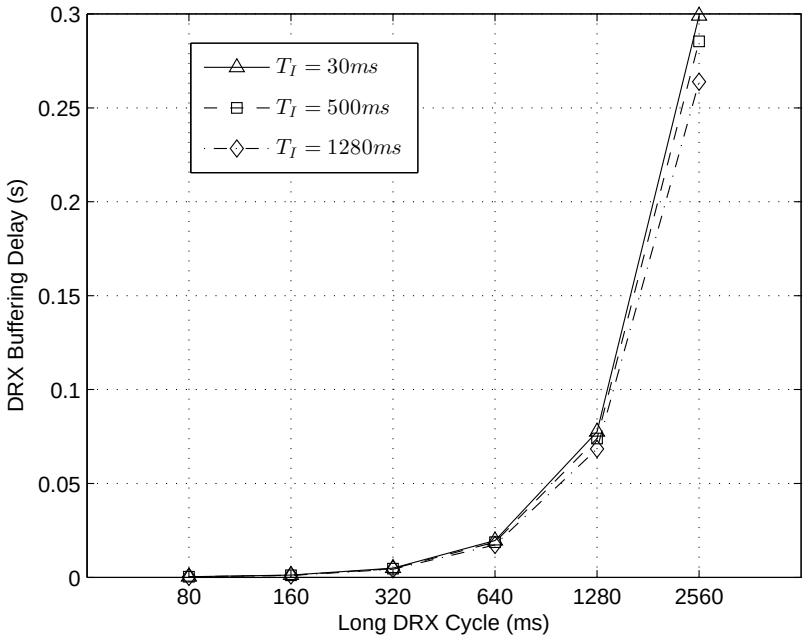

Figure 4.12. Buffering delay due to DRX sleep versus DRX cycles for various T_I for background traffic. $T_{ON} = 4ms$ and $N_{SC} = 0$.

compared to active traffic, so even for smaller DRX long cycles we can have very high sleep percentages compared to active traffic.

For background traffic, DRX inactivity timer (T_I) becomes an important DRX parameter to optimize power saving as shown in Figure 4.11. Larger values of inactivity timer keeps the UE in active longer there by adversely affecting the potential power saving. On the other hand, DRX Inactivity timer has minimal impact on end to end delay for background traffic as seen in Figure 4.12. For background traffic, longer DRX cycles with short DRX inactivity timer will be optimum. If the device figures out that it is running background traffic, it would be beneficial to send a PPI and ask for lower DRX inactivity timer settings.

Table 4.3 depicts an example of DRX configuration optimization for active traffic for various latency requirement (i.e., δ_{MAX}) of the traffic by solving **P2**. It is evident that stringent latency constraint (i.e., lower δ_{MAX}) requires DRX configuration with smaller T_{SC} and higher N_{SC} so that UE does not enter long DRX frequently. T_{LC} is also smaller so that

Table 4.3. Optimal DRX Configuration for Various Maximum Delay Bounds Set by Operator for Active Traffic. ($T_I = 30ms,\ T_{ON} = 4ms,\ 1/\lambda_p = 10ms,\ 1/\lambda_{pc} = 500ms,\ 1/\lambda_s = 60s$.)

$\delta_{\text{MAX}}(ms)$	25	50	100	200	400
$T_{SC}(ms)$	80	128	160	256	512
N_{SC}	10	8	4	3	3
$T_{LC}(ms)$	320	512	640	1024	2048
$\Gamma_{\text{SLEEP}}(\%)$	74.3	76.6	79.1	82.0	85.2
$\delta_{\text{DRX}}(ms)$	24.4	49.2	97.6	189.2	374.7

Table 4.4. Optimal DRX Configuration for Various Minimum Power Saving Requirements Set by Operator for Background Traffic. ($T_{ON} = 4ms,\ 1/\lambda_{pc} = 10s$.)

$\Gamma_{\text{MAX}}(\%)$	0.5	1	2	4	8	16
$T_{LC}(ms)$	1024	512	256	128	64	32
$T_I(ms)$	10	20	40	80	100	300
$\Gamma_{\text{SLEEP}}(\%)$	99.5	99.0	98.0	96.1	92.8	84.9
$\delta_{\text{DRX}}(ms)$	50.03	12.56	3.08	0.73	0.16	0.03

even the UE enters long DRX, the long sleep does not effect the latency. For the same reasons, power saving decreases with the stringent latency requirement.

Optimization of DRX parameters in case of background traffic for various operator's minimum power saving targets (i.e., γ_{MAX}) are presented in Table 4.4 based on **P3**. When higher power saving is required, UE must use longer T_{LC} with smaller T_I so that UE enters the long DRX as soon as possible.

Table 4.5 illustrates the benefits of DRX switching based on traffic characteristics in enhancing the power savings while satisfying the delay constraint set by operator for active traffic. Table 4.5 also presents power saving improvements during active traffic sessions using DRX configuration optimized for active traffic (obtained from **P2**) compared to that in case of using DRX configuration optimized for background traffic (obtained from **P3**). Increase in power saving by 10.7% to 48.8% can be achieved depending on the delay requirements as can be seen from Table 4.5.

Table 4.5. Example Showing Benefits of DRX Switching during Active Traffic Session.

With DRX Parameters Optimized for Background Traffic			
$\Gamma_{\text{MAX}}(\%)$	4	8	16
$\Gamma_{\text{SLEEP}}(\%)$	66.8	61.4	41.2
$\delta_{\text{DRX}}(ms)$	10.0	2.3	0.3
$\Gamma_{\text{SLEEP}}(\%)$ Increment With DRX Parameters Optimized for Active Traffic			
$\delta_{\text{MAX}} = 25ms$	10.7	16.1	36.3
$\delta_{\text{MAX}} = 50ms$	13.2	18.6	38.8
$\delta_{\text{MAX}} = 100ms$	16.5	21.9	42.1
$\delta_{\text{MAX}} = 200ms$	20.3	25.7	45.8
$\delta_{\text{MAX}} = 400ms$	23.2	28.6	48.8

4.9 Conclusion

In this chapter, we have demonstrated the necessity of optimizing DRX configuration in order to maintain a tradeoff between power saving and latency. We have developed an analytical model for the DRX operation and derived expressions for performance parameters such as power saving and buffering delay due to DRX in context of both active and background traffic. An optimization problem was then formulated to find tradeoff between power saving and latency based on operator's preference for power saving and latency requirement of traffic. A DRX switching mechanism was explored to adapt DRX configuration based on most recent traffic running at UE. We have validated the analytical model through system level simulations. Results showed that the proposed tradeoff scheme is highly efficient in tuning DRX configuration to achieve required power saving while ensuring reasonably low latency. It is also observed that higher number of short DRX cycles is preferable for active traffic and shorter values of inactivity timer is preferable for background traffic. Results have revealed the necessity of DRX switching based on the traffic running at UE to maximize the benefits of DRX functionality. In the next chapter, we will focus modeling and simulation of 3G/4G cellular networks.

CHAPTER 5

ADVANCED WIRELESS COMMUNICATION CASE STUDIES

We present a common network simulation environment which models realistic communication link and data traffic for advanced wireless systems [63]. This chapter of dissertation validates the proposed models and algorithms in the presence of realistic physical layer and data traffic. Most of simulation studies to date make excessive simplifications of the parameters at the different layers. Typically either physical layer parameters or data traffic are strongly simplified to study the network layer. Although this provides useful insights to the small-scale interactions between layers, it may not provide an accurate view of the large-scale effects and performance of the network. In order to accurately simulate data traffic, the legacy traffic such as HTTP, VoIP and FTP are analyzed then new traffic models are investigated.

Figure 5.1 shows a HTTP traffic pattern for a typical web browsing session. Characteristics of HTTP traffic mainly depends on structure of web page as well as human interaction. Usually, each session of a HTTP traffic consists of ON and OFF periods because of requirement of the human interaction. The ON period also called packet call represents user's request for information and the sequence of packets to download a web page from HTTP server. The OFF period also called reading time when the user reads the information before moving to another web page. The reading time depends on amount of web page content and the user behaviour. A web page consists of a main object followed by zero or more embedded objects. Each packet call also has ON and OFF durations similar to those of a session. However, the ON and OFF periods of packet call are depends on machine interaction between user and server. The ON time represents periods of downloading main object

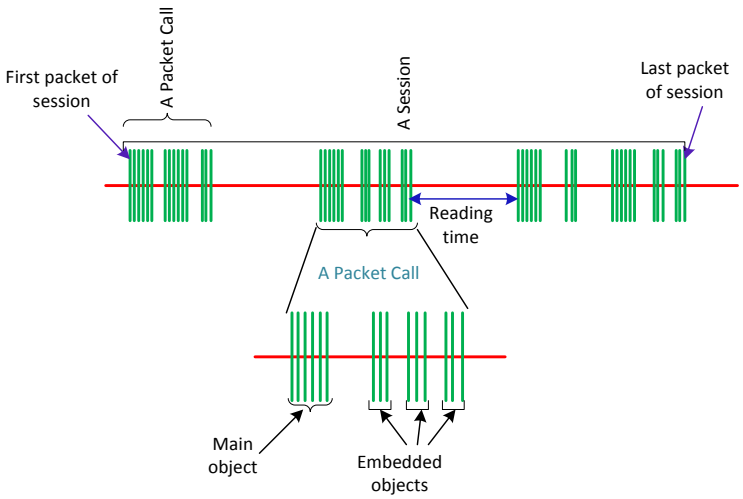

Figure 5.1. Typical HTTP traffic session

or embedded object, while OFF time represents parsing time of user request. Reading time and parsing times are usually modelled as exponentially distributed variable with mean values of 30s and 0.13s, respectively.

A session of an FTP application consists of multiple file transfers as shown in Figure 5.2a. The time between end of previous file transfer and user request for next file transfer is called reading time and is exponentially distributed with a mean value of 180s. The duration of packet call in an FTP session depends on the throughput of the user and FTP file size.

A typical VoIP session is depicted in Figure 5.2b. A VoIP session lasts for the entire user call duration which consists of multiple active and silence (inactive) periods. Packets are generated every frame duration + network packet delay jitter during the active state. Frame duration varies from 10ms to 30ms depending on the voice coder and value of network packet delay jitter varies from 0 to 120ms. Note that silence period should not be considered as UE being inactive because UE is listening to other side called/calling party.

Overall, all of these legacy traffic has ON/OFF nature. If the activity starts, it continues for a specific time and when the activity stops there will be no activity for a specific time. However, current social networking and emerging traffic characteristics are not suitable

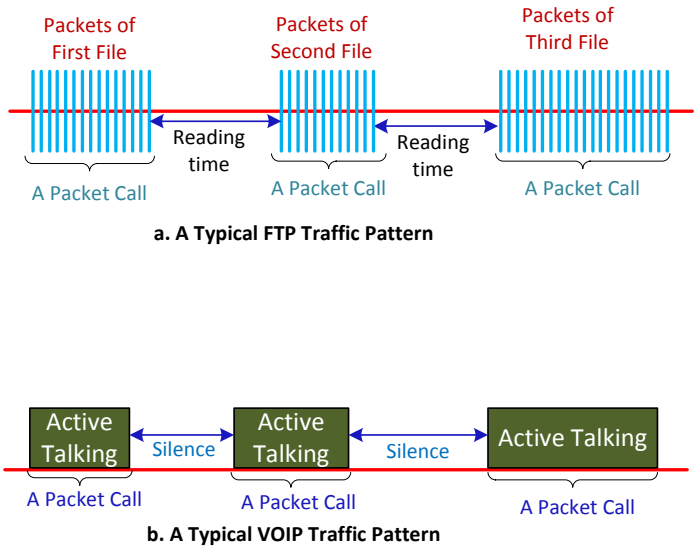

Figure 5.2. Typical FTP and VOIP traffic sessions

for this kind of ON/OFF modeling. In the next section, we focus on the emerging traffic characteristics.

5.1 Explosive Growth of Mobile Internet Traffic

It is shown in [64] that eightfold increase in global IP traffic has been observed over the period of past 5 years, and threefold increase is predicted over the next 5 years. IP traffic is expected to grow at annual growth rate of 29% from 2011 to 2016 which will result in expected annual IP Traffic surpassing 1.3 Zettabytes per year (109.5 exabytes per month) by 2016. Global internet data trend is summarised based on [64] in Table 5.1.

In terms of number of connected devices, it will increase from 1 connected device per capita in 2011 to 3 connected devices per capita in 2016 [64]. Furthermore, non-PC IP traffic is increasing significantly due to recent explosion in the popularity of use of smart phones and tablets. The consumer IP traffic originated from these devices will grow at a rate higher than 119%, while that originated from PC will grow at a rate of 26%. As a result, IP traffic from wireless devices will account for majority (about 61%) of IP traffic by 2016 [64].

Table 5.1. Global Internet Data Trend from 2011 to 2016 (AGR = Annual Growth Rate, PB = Petabytes).

	2011	2012	2013	2014	2015	2016	AGR (2011-2016)
Consumer IP Traffic, 2011-2016 (PB/Month)							
Fixed Internet	20121	29155	40132	53319	64024	73463	30%
Managed IP	5,269	7,270	9,482	11,202	12,953	14592	23%
Mobile data	402	879	1717	3116	5213	8313	83%
Total	25792	37304	51332	67637	82189	96367	30%
Consumer Mobile IP Traffic , 2011-2016 (PB/Month)							
File Sharing	77	115	155	205	261	362	36%
Internet Video	308	737	1546	2918	4882	7615	90%
Web, email, data	159	297	482	732	1125	1775	62%
Online Gaming	7	14	24	41	78	118	76%
VoIP	8	10	12	15	23	36	36%
Total	402	879	1717	3116	5213	8313	83%

Mobile data traffic is expected to grow 3 times faster than that of fixed IP traffic during the period 2011 to 2016 at a predicted annual growth rate of 78%. This will result in a mobile IP traffic of 10.8 exabytes per month by 2016 which will be 10% of total IP traffic in 2016. Note that mobile data traffic was only 2% of total IP traffic in 2011 [64].

Along with increase in mobile traffic share of total IP traffic, the nature of emerging mobile traffic is also diversified in the sense that their characteristics are significantly different than those of traditional applications. Some of the examples of applications which are gaining popularity over wireless networks are weather update, news update, VoIP (such as Skype and Google talk), social networking services (such as Facebook, twitter, FindMe), Geo services (such as Google places, location targeted advertisement), messaging (SMS and instant messaging) and online games. These applications create frequent idle to connected transitions, due to frequent live updates, periodic keep-alive, frequent start or stop of services, frequent push to talk services. Note that the state transition also includes series of actions such as wake-up from idle, connection reestablishment, network attachment for bearer establishment, send and/or receive background traffic (such as update/live

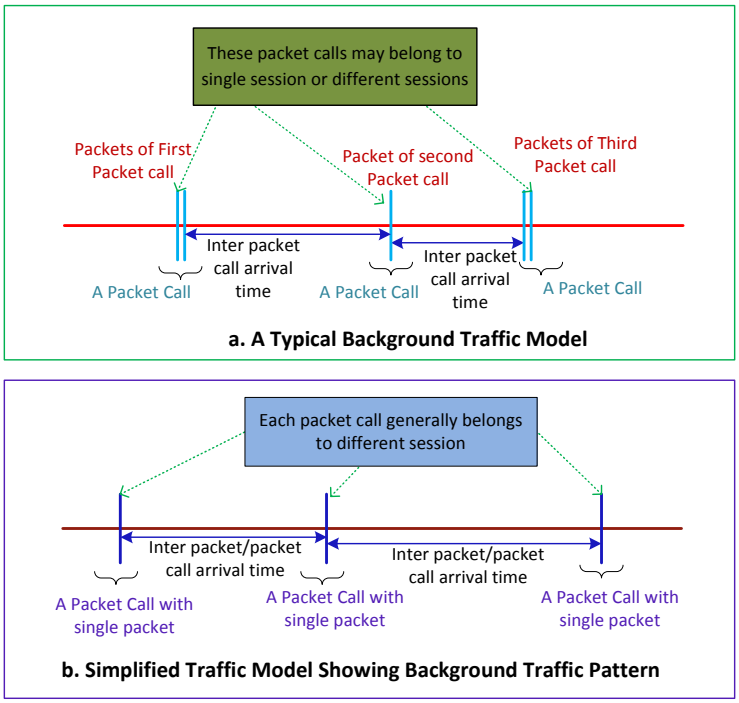

Figure 5.3. Typical background traffic characteristics

messages), go back to idle and so on [47][48]. Note that these applications continue the above mentioned activities even when the user device is unattended. The packet generated by these activities are termed as background traffic in [47]. Background traffic does not have any ON/OFF nature. Traffic generated during background mode is usually very small packets around every few seconds [47]. Based on real traces collected in [47], each background traffic usually consists of one or two small packets. Background traffic model is depicted in Figure 5.3.

In the next section, we will investigate different wireless technologies and show how a common network simulation environment can simulate different radio resource control mechanisms. The wireless technologies are explained in a chronological order starting with CDMA 2000 networks.

5.2 CDMA2000 Case Study

The CDMA2000 standard provides the opportunity for IP applications over wireless links. Serious difficulties stand in the way because the Transfer Control Protocol (TCP) which was designed for wireline links does not perform well over wireless links. TCP assumes that packet loss is mostly due to queue overflows and network congestion, but packet losses over wireless channels have a very different characteristic, leading to severe degradation of TCP performance. In this section, we investigate and model the interactions of TCP with the Radio Link Protocol (RLP) in CDMA 2000. Radio Link Protocol (RLP) is an automatic repeat request (ARQ) fragmentation protocol used over CDMA 2000 air interface. An OPNET module was developed to investigate the impact of reverse link data rate selection on 1xEV DO forward link data rates. Similar analysis of performance of CDMA systems can be found in [65] over a satellite communication network. A new receiver is also proposed in [66] for CDMA devices.

5.2.1 Optimization of TCP/IP Parameters

The OPNET simulation model consists of multiple point-to-point links. The mobile nodes are connected to the base station by a static wire that represents the physical medium of transmission. Although the current simulation environment is set up without Radio Link Protocol (RLP) and assumes BER= 0, both forward and reverse link scheduler are built to segment TCP/IP packets according the payload sizes of 1xEV DO.

Figure 5.4 shows the OPNET simulation scenario used in this section. The OPNET simulator developed for the project can simulate up to 16 high-rate data users served with a single base station. Each mobile is defined as a workstation node, which is connected to a server node through the base station, as shown in Figure 5.4. The simulation model includes workstation nodes as mobile users and a base station node which is connected to server. Workstations can communicate with the base station via two point-to-point links

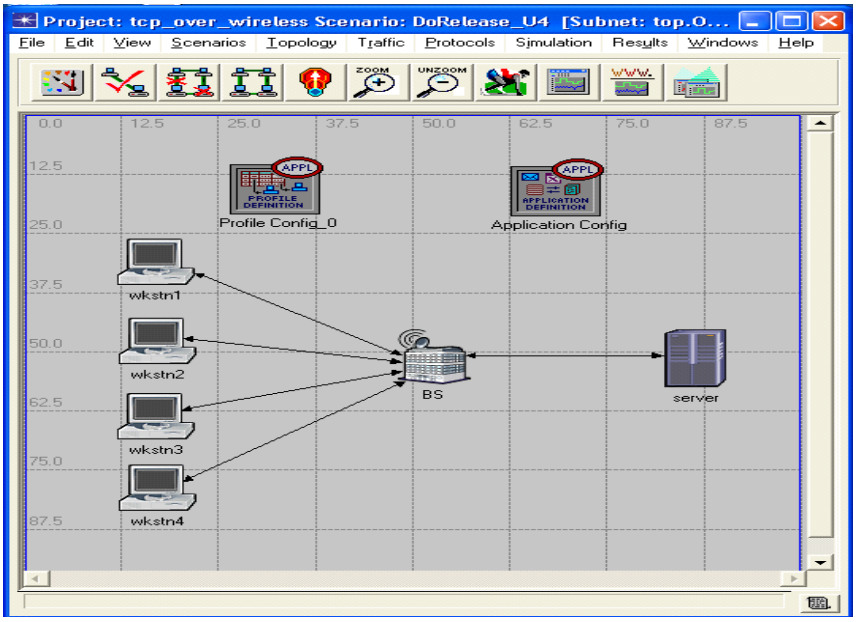

Figure 5.4. Simulation setup in OPNET

(forward and reverse link). Data rates and multiple access are emulated according to 1xEV DO forward and reverse link data rates. These models are integrated with the built-in OPNET models (applications, TCP/IP, etc). Figure 5.5 depicts the protocol stack used in the simulation. Similar protocol stack implementation is used in [67].

In this work, we analyze the required reverse link rates to achieve the forward link data rates depending on the number of users in the system. As expected, in order to achieve higher throughput in the forward link, devices need to be allocated higher speed data links in the reverse link to send ACK/NACKs on time. In Table 5.2.1, the results are shown for 1 user and 2 users. If the number of users increases, the users need to share the available reverse link rates where the minimum reverse link rate is 9600kbps (1x).

5.3 WIMAX Case Study

Based on IEEE 802.16 [68] air interface standard, the WiMAX is deployed in many countries as broadband wireless metropolitan network and millions of people benefit from it. The

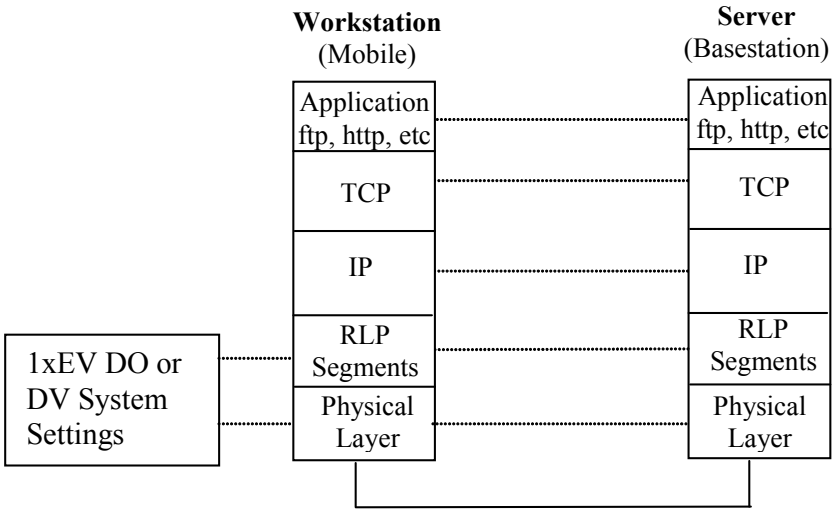

Figure 5.5. Protocol stack model for TCP/IP/RLP/1xEV simulation

Peak Forward Rate	1 User	2 Users
614400	1x	1x
1288000	2x	1x
1843200	2x	2x
2457600	4x	2x

Table 5.2. Best Reverse Link Rates for 1xEV DO Peak Forward Link Rates

WiMAX technology has advantages in high data rates, quality of service, security and scalability, all which attract tremendous investment from all around the world. With using high data rates of WiMAX, operators can support multiple Voice over IP (VoIP) mobile stations. VoIP is a technology that allows you to make real-time voice calls using a broadband Internet connection instead of regular (or analog) phone lines [69]. In order to differentiate VoIP and other services, WIMAX defines five different Quality of Service (QoS) classes. These classes are Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), extended real-time Polling Service (ertPS), non-real-time Polling Service (nrtPS) and Best Effort (BE) Service [68]. With these QoS classes, WIMAX became more flexible and robust for different kind of applications. However most of the current WIMAX deployments in the market only provides BE services for all kinds of Internet applications. In best effort QoS,

there is no minimum bandwidth grant for any connection and no differentiation between real-time traffic and non-real-time traffic. Therefore real-time voice traffic may suffer packet losses and long delays when there is a heavy load on non-real time applications such as ftp uploading.

In this section, we first analyze voice capacity over WIMAX Best effort QoS class and provide methods to improve voice performance using OPNET WIMAX module. We evaluate the VoIP capacity over both error free channel and additive white gaussian noise (AWGN) channel. The performance of well known header compression methods to solve the overhead problem of VoIP packets is also investigated [70].

In our simulations, we are using 5ms Time Division Duplex (TDD) OFDMA over 10MHz bandwidth. WIMAX supports smaller bandwidths such as 5 MHz; however 10 MHz deployment is the most common deployment among operators. The other simulation parameters can be found in Table 5.3.

In Figure 5.6, we depict a sample WIMAX deployment where a BS is located in the center. All the mobile users are paired. Each mobile station belongs to either calling party or called party. In each pair, one mobile station can only call the other mobile station. This pairing is used to have same uplink and downlink load in the WIMAX simulations.

In our simulations, our aim is to mimic a realistic modulation and coding scheme (MCS) distribution as shown in Table 5.4. Usually in a network, mobile stations will use different modulation and coding schemes based on their Signal-Noise Ratio (SNR). In the simulation platform, we define exit and entry SNR thresholds for each MCS, as shown in Table 5.4. For example, if the current SNR is 15dB, then device will use 16QAM1/2; if SNR changes to 9dB, 9dB is less than exit threshold of 16QAM1/2 and in region of exit and entry threshold of QPSK3/4, so MCS will be changed to QPSK3/4. We did exclusive search to find the distance requirements for each device where they can support different MCSs like 64QAM, 16QAM or QPSK. Finally, we came up with a set of circles. Each region with different

Table 5.3. Parameters of OPNET Simulation for WiMAX Network

Parameters	Value
—Network Parameters—	
PHY Profile	OFDMA/10MHz
Frame/Symbol Duration	5ms/102.86μs
Number of subcarriers	1024
Frame Preamble(symbols)	1
TTG/RTG	106 μs
UL Subframe size(symbols)	21
DL/UL-MAP Repetition Count	No
—Base Station Parameters—	
Maximum number of SS nodes	400
Maximum Transmission Power	0.5w
Transmit/Receive Antennas	1
Altitude (m)	160
—Mobile Station Parameters—	
Modulation and Coding	Adaptive
Average SDU Size	1500bytes
Buffer Size	64Kbytes
ARQ	Enable
Window Size(ARQ blocks)	512
Block Size(bytes)	8
Retransmission Time Interval	0.1s
Block Lifetime Interval	3s
ACK Type	Selective
Feedback Time Interval	0.01s
HARQ Parameters	Enable
Maximum Retransmissions	4
Explicit ACK Delay(frames)	1
Number of Channels(UL/DL)	8/16
Buffer Size Constant(K)	20
Transmit/Receive Antennas	1
Maximum Transmission Power	0.5w
—Voice Application Parameters—	
Frame Size	20 ms
Coding Rate	6.3Kbps
Silence Length(in/out)	exp(1.25s)
Talk Spurt Length (in/out)	exp(1.25s)
Type of Service	Best Effort(0)

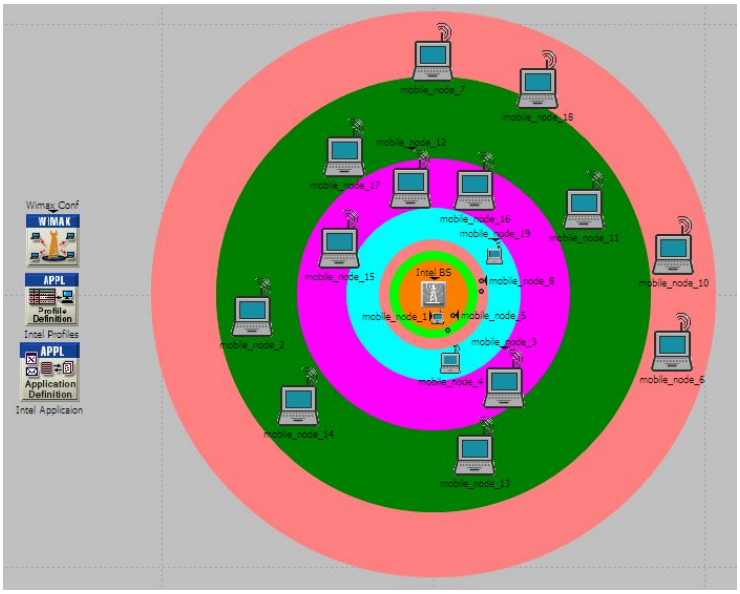

Figure 5.6. Example of a WiMAX network deployment

Table 5.4. MCS Distribution and Threshold Definition

MCS	% distribution	Exit-Entry Threshold(dB)
QPSK1/2	39%	$0.1 - 4$
QPSK3/4	14%	$5.0 - 10$
16QAM1/2	22%	$12 - 18$
16QAM3/4	16%	$19 - 22$
64QAM2/3	5%	$23 - 24$
64QAM3/4	3%	$25 - 26$
64QAM5/6	1%	$29 - 30$

color stands for a certain type of MCS. Figure 5.6 shows an example network layout with one base station and multiple mobile stations.

5.3.1 Optimization of Control Overhead for VoIP

We investigate voice capacity in both error free channel and AWGN channel with multipath using ITU Pedestrian B model and Vehicle A model. We use Mean Opinion Score (MOS) as a metric to evaluate quality of voice communication. Instead of using bad or good, MOS defines a numerical indication of the perceived quality of the voice received after being

Table 5.5. MOS Value for VoIP

MOS Value	Quality	Description
5	Excellent	Perfect, like face to face talk
4	Good	Fair, sound very clear but little annoying can be detected
3	Fair	Annoying, but still good enough to communicate
2	Poor	Very annoying, nearly impossible to communicate
1	Bad	Very bad, impossible to communicate

Table 5.6. Simulation Results for Error Free and AWGN Channels.

Scenarios	number of users	DL Usage	UL Usage	Control Overhead	Average MOS
Error free channel	150	58.7%	84.9%	24%	3.7
	160	62.7%	85.2%	17%	2.1
AWGN Pedestrian B	130	60.1%	82.1%	25.2%	3.2
	140	69.4%	85.2%	20.2%	2.5
AWGN Vehicle A	100	47.3%	69.8%	23.1%	3.1
	110	58.2%	81.1%	23.3%	2.8

compressed, transmitted and decompressed. MOS is expressed in number from 1 to 5 as expressed in Table 5.5. Our constraint is to satisfy the average MOS score of all of the users 3 and higher. We are using OPNET's default MOS calculation method which depends on the end to end delay. In order to find the capacity, we increase the number of users one by one till the average MOS score requirement can not be satisfied for a user due to scheduling delay and latency.

Table 5.6 shows results for error free and AWGN with Pedestrian B and Vehicle A channel models. It can be seen from the results that we can support about 150-160 VoIP user over error free channel, on a Pedestrian B channel capacity is about 130-140 users and on a Vehicle A channel capacity is about 100-110 users. As expected, due to errors and retransmission, the number of users that can be supported are less in AWGN channels compared to error free channel. Our results show that the uplink is the bottleneck to support more VoIP users. Also from these results, we found that VoIP control overhead is very high and around 20-25%.

In traditional IP networks, application data is encapsulated with headers before it is transmitted to link layer. The header overhead is not a big problem when data payload is large. But in VoIP, voice data payload is usually very small so that header part of the data becomes proportionally big. Generally in voice applications, voice data is encapsulated with Real-time Transport Protocol (RTP) packet, which is then encapsulated into UDP packets. In a voice packet, IP header size is 160 bits; UDP header size is 64 bits; and RTP header size is 96 bits. Total header size becomes around 320 bits for a single VoIP packet. OPNET voice encoder creates VoIP packets which have payload size of 128 bits. So in order to send 128 bits, 320 bits of overhead needs to be added. Header compression is a technique to reduce this overhead. In our simulation, we apply header compression at MAC layer of OPNET WIMAX module. Because voice data has been wrapped with RTP and UDP header, we have to identify first whether the packet is a voice packet or not. When an IP packet comes from higher layers, we abstract its data field and check its packet format. After identifying the VoIP packets, we compress the overall 320 bits header into 32 bits. Performance of header compression is investigated in multiple scenarios where we adjust the active number of users in the network. These scenarios consist of one base station and $50 - 100 - 150 - 200$ mobile stations. Table 5.7 shows the impact of header compression on both uplink and downlink burst usage as well as the MOS score. Table 5.7 shows that with header compression, downlink and uplink burst usage can be decreased by 50%. Moreover, the header compression becomes crucial when we hit the capacity of the network. For 50, 100, 150 users, we may not need header compression to satisfy MOS requirement. However, for 200 users without header compression, we can not satisfy MOS requirement.

Table 5.7. Simulation Results for Different Number of Users.

number of users	Compression	DL Burst Usage	UL Burst Usage	MAP Usage	Average MOS
50	Off	13.3%	30.5%	10.1%	3.9
	On	6.5%	14.8%	10.2%	3.9
100	Off	28.1%	54.9%	16.6%	3.9
	On	14.9%	26.5%	16.7%	3.9
150	Off	40.1%	80.7%	16.7%	3.8
	On	22.2%	39.3%	18.5%	3.9
200	Off	48.8%	85.2%	16.3%	1.1
	On	31.3%	52.1%	24.7%	3.9

5.4 LTE Case Study

3GPP Long Term Evolution Advanced (LTE-A) has been established as a promising next generation 4G wireless communication system which is capable of offering higher data rate over the air. The higher data rate is achieved by the use of higher order modulation, advanced coding techniques, higher transmission bandwidth, and multiple advanced antenna schemes [3][5]. As a result, user equipment (UE) needs computationally complex circuitry draining battery more rapidly than ever to support the 4G features. The increased capacity over the air in LTE-A network and development of smart phones, tablet and e-readers have changed the internet usage paradigm drastically attracting more users to access internet over the LTE-A network frequently [54]. These smart devices have also given birth to several new mobile internet applications called diverse data application (DDA). The characteristics of traffic generated by of DDA are quite different than those of traditional internet applications [47]. For example, access of applications like facebook, twitter, skype, gtalk, IM, and gaming over LTE-A network is increasing rapidly. These applications send small packets to network every few seconds even when these applications are not being used by the user actively [47]. In general, DDA generates a short burst of one or more small sized packets frequently even when UE is in unattended state. As a result, UE keeps changing states between connected and idle draining a lot of battery power and causing sig-

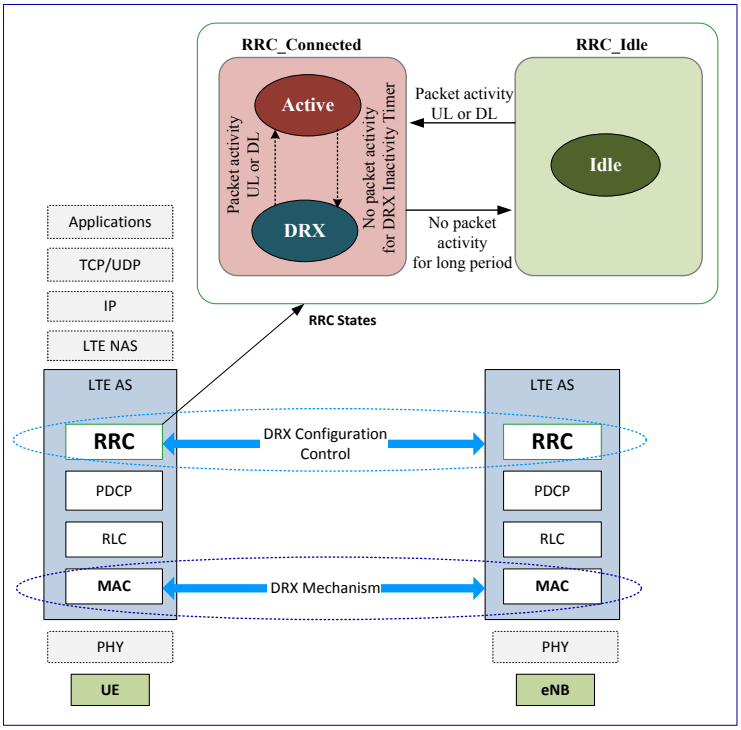

Figure 5.7. Various UE states in LTE-A

nificant signalling overhead. Discontinuous Reception (DRX) can be configured in LTE-A network as an effective power saving mechanism, which largely addresses the power saving and unnecessary state transition issues of DDA [71].

In LTE-A, UE can stay in RRC Connected and RRC Idle states as shown in Figure 5.7. UE is registered to network in RRC Connected state and has established radio bearers (signaling and data radio bearers) with a serving eNB in order to transmit/receive uplink/downlink data and control messages. UE in RRC Connected state is known to the network at cell level. This is the state where UE is most active and all data packet transmission related to the UE takes place. When there is no data activity for a long duration, UE releases radio bearers and enters RRC Idle state. In Idle state UE is known to the network at tracking area level and can be reached by the network through paging message [72].

LTE-A network support always connected feature during the RRC CONNECTED state in

which UE continuously monitors the downlink (DL) control channel called Physical Downlink Control Channel (PDCCH) in order to check for possible DL data. DRX saves UE power by allowing UE to monitor PDCCH less frequently and go for sleep whenever there is no traffic activity [7] [8]. If DRX is configured, UE sleeps whenever there is no data activity and wakes up only for a short duration called DRX ON duration. The durations of UE sleep state and UE active state vary based on the DRX parameters setting. Longer and frequent DRX sleep improves the power saving at the cost of probable increased end-to-end latency. Obviously shorter and less frequent DRX sleep should improve the latency performance of the applications running at UE. In real life a tradeoff between power saving and latency is required based on the application's latency requirement. Therefore, application specific DRX configuration should improve the benefits of DRX in LTE-A network. A single set of DRX configuration may not be optimal for all applications with various latency requirements. In this section, we analyze and investigate the benefits of DRX switching which enables adjusting DRX configuration based on the characteristics of current applications running at UE. We defined different sets of DRX configurations some of which are low power consumption DRX configurations while some are low latency DRX configurations. We consider latency as a crucial user experience parameter and therefore use performance-optimized DRX configuration and low-latency DRX configuration interchangeably throughout the section. We considered two kinds of traffic for our analysis; a background traffic without latency requirement and an active traffic with stringent latency requirement. When background traffic is running with non-continuous active traffic sessions, we showed that DRX switching can improve the latency performance of active traffic while maintaining similar level of power saving during background traffic only sessions. We also discussed the mechanism to enable DRX switching based on the traffic characteristics.

DRX mechanism can be configured in both RRC Idle and RRC Connected states. However, the functionality and configuration parameters are different in each state. In RRC

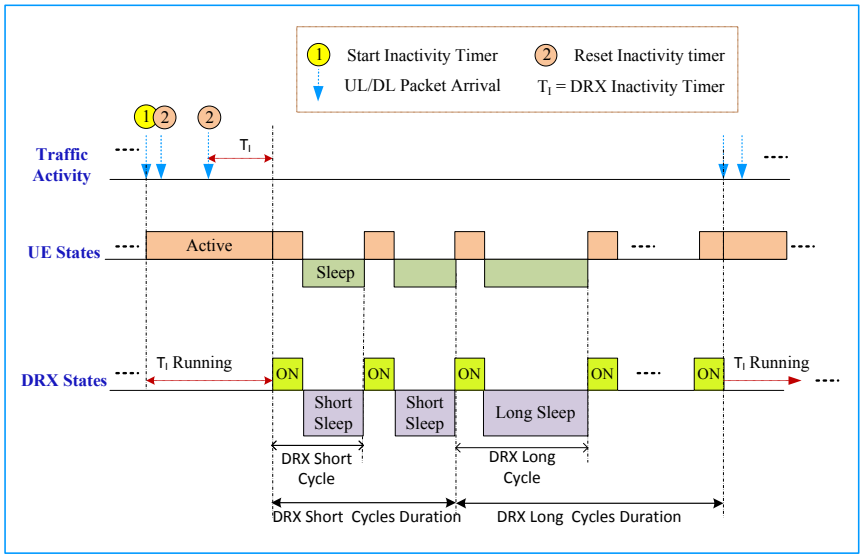

Figure 5.8. UE states based on DRX cycle and traffic activity

Connected state, DRX mechanism may be configured by the radio resource control (RRC) on per UE basis. DRX controls PDCCH monitoring activity of a UE in the LTE-A system [7]. In RRC Idle mode, the UE monitors a paging channel periodically as configured by DRX cycle, for incoming traffic. In this section, we focus on the RRC Connected state DRX mechanism, which is configured by the RRC using the timers/parameters: *on Duration Timer* (T_{ON}), *DRX Inactivity Timer* (T_I), *Long DRX Cycle* (T_{LDC}), $drxStartOffset$, and optionally the $drxShortCycleTimer$ and *short DRX Cycle* (T_{SDC}) [7]. Figure 5.8 shows the DRX functionality we have considered in this section. RRC Idle state is not shown in the Figure 5.8 for simplicity. Generally when there is no packet activity during the RRC Connected state for a time duration called idle inactivity timer, UE moves to RRC Idle.

If DRX is configured, UE continuously monitors PDCCH for possible DL packet transmission when T_I is running. Upon DL packet indication in PDCCH or UL packet arrival at UE, the T_I is reset. When T_I expires, UE enters DRX cycles by starting the timer T_{LDC} as we have not considered short DRX cycles. After expiration of T_I, DRX cycle may not

start immediately. Next subframe where DRX cycle shall start after the expiration of T_I is determined by $drxStartOffset$ as described in [7]. During each DRX cycle, UE monitors PDCCH for T_{ON} period and remains asleep for rest of the cycle. DL packet indication in PDCCH during T_{ON} terminates the DRX operation and starts T_I. Also, arrival of higher layer data packets for uplink transmission during T_{ON} terminates the DRX cycle and starts T_I. The UL packet arrival during DRX sleep at UE may terminate DRX operation immediately or waits for next T_{ON} period to abort the DRX operation. We have assumed latter case so that UL packet arrived during DRX sleep is delayed until the next T_{ON} period.

5.4.1 Optimization of DRX Parameters

DRX in RRC Connected state is defined to help reduce UE power consumption. In TR 36.822 [47], UE power consumption and latency (a user visible metric) are classified as some of the output metrics of interest. Latency requirements may vary widely among applications. For example, applications such as social networking, background traffic and IM traffic usually do not have low latency requirements and hence may need DRX parameters setting to maximize the UE power saving, while real time applications like video streaming and gaming may require DRX parameters setting to meet stringent latency constraints. Since a UE can run various kinds of applications over time, use of a single set of DRX parameters all the time does not seem to be reasonable to achieve maximum benefit from the DRX functionality. Therefore, we should define multiple DRX sets of different categories such as low power DRX category and low latency DRX category. In case of multiple DRX parameters sets, most favourable DRX parameters set at any time can be selected based on the current information such as characteristic of applications running at UE, network conditions and UE battery power remaining.

In this section, we focus on investigating the importance of DRX switching in context of UEs running different applications over the time. We assume that the user always remains

in RRC Connected mode (no RRC Idle mode) and will directly enter long DRX after the DRX inactivity timer expiry as shown in Figure 5.8. We refer to DRX opportunity as mentioned in [7] as DRX sleep [72]. We assume that different sets of DRX parameters are available at the E-UTRAN for every UE. In addition, we assume that there exists an efficient mechanism by which the UE and eNB negotiate and select one of the available DRX parameter sets at the E-UTRAN. The negotiation mechanism may allow UE to request a DRX configuration which may not be in the predefined DRX parameter sets at E-UTRAN.

We have used OPNET MODELER 17.1 [62] to implement multiple DRX configuration sets and required DRX switching mechanism to enable UE and eNB to select most suitable DRX configuration based on the characteristics of current traffic. Multiple sets of DRX settings can be configured by user as UE attribute 'DRX Parameters'. Note that DRX switching must select one of the DRX settings from this user-configured DRX list. Similarly, multiple DRX settings can be defined as 'DRX Parameters' attribute at eNB. These DRX setting are exchanged between UE and eNB during the network attach procedure [73]. In real network, these DRX sets can be exchanged by including these sets as an information element (IE) in the RRC messages.

The DRX switching mechanism is depicted in Figure 5.10. Initially, UE and eNB select one of the DRX parameter sets during the UE attach process. For this purpose, one of the DRX sets can be declared as default set. Alternately, UE may request for a specific DRX setting based on expected traffic characteristics. However, final decision of DRX selection is always made by eNB. After initial DRX parameters selection, UE continues DRX operation with the selected DRX setting unless there is a need for DRX parameters change. DRX parameters change may be triggered by new applications with different traffic characteristics than that of current application(s). In our implementation, we focus on delay-tolerant background traffic sessions and delay sensitive active traffic sessions for DRX switching. Furthermore, for now DRX Switching is implemented in such a way that

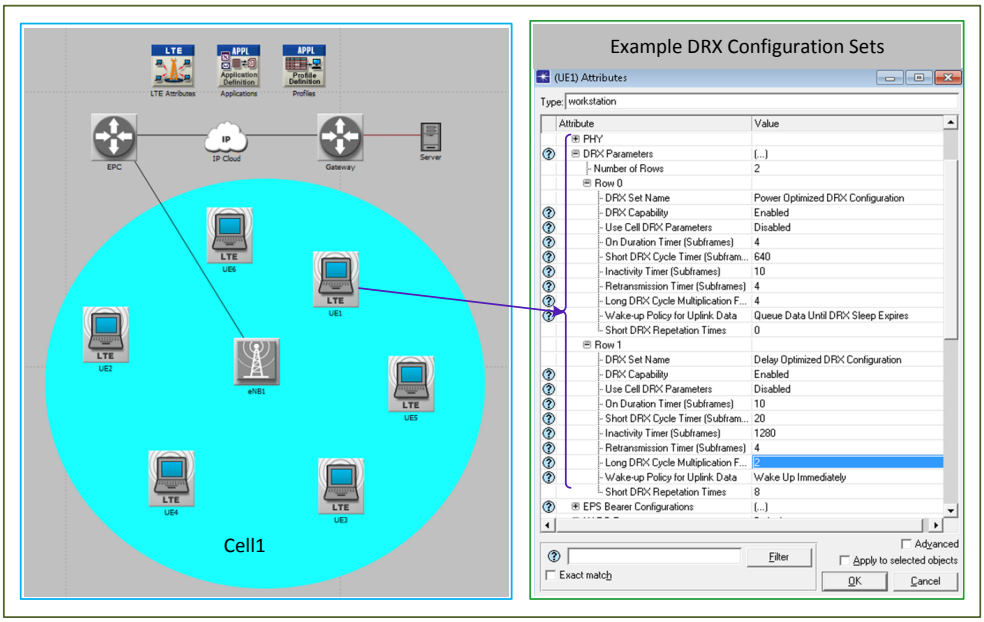

Figure 5.9. A typical scenario showing implementation of multiple sets of DRX configurations in OPNET Modeler.

it is aligned with start/end of different applications at UE. DRX switching triggered by change in traffic characteristic in real-time is a topic for our future work. Whenever UE finds change in traffic (active versus background and vice versa in our implementation), it requests eNB for its preferred new DRX set. eNB may or may not grant switching to new DRX set. If granted, UE adopts new DRX parameters. Otherwise, it continues with current DRX setting [74].

In all of our analysis, we used background traffic based on the background traces included in TR 36.822 [47]. Background traffic is running at UE throughout the simulation duration. Active traffic session represents data with frequent packet arrival and larger size of packets compared to that of background traffic. The simulation parameters and assumptions are provided in the Appendix. For simulation evaluation, we have considered the percentage of time spent in DRX sleep duration which can be correlated to the power savings achieved in UE power consumption and end to end user latency as the output metrics. It is better to choose higher DRX cycle value for maximal power savings for a given inactivity timer. The

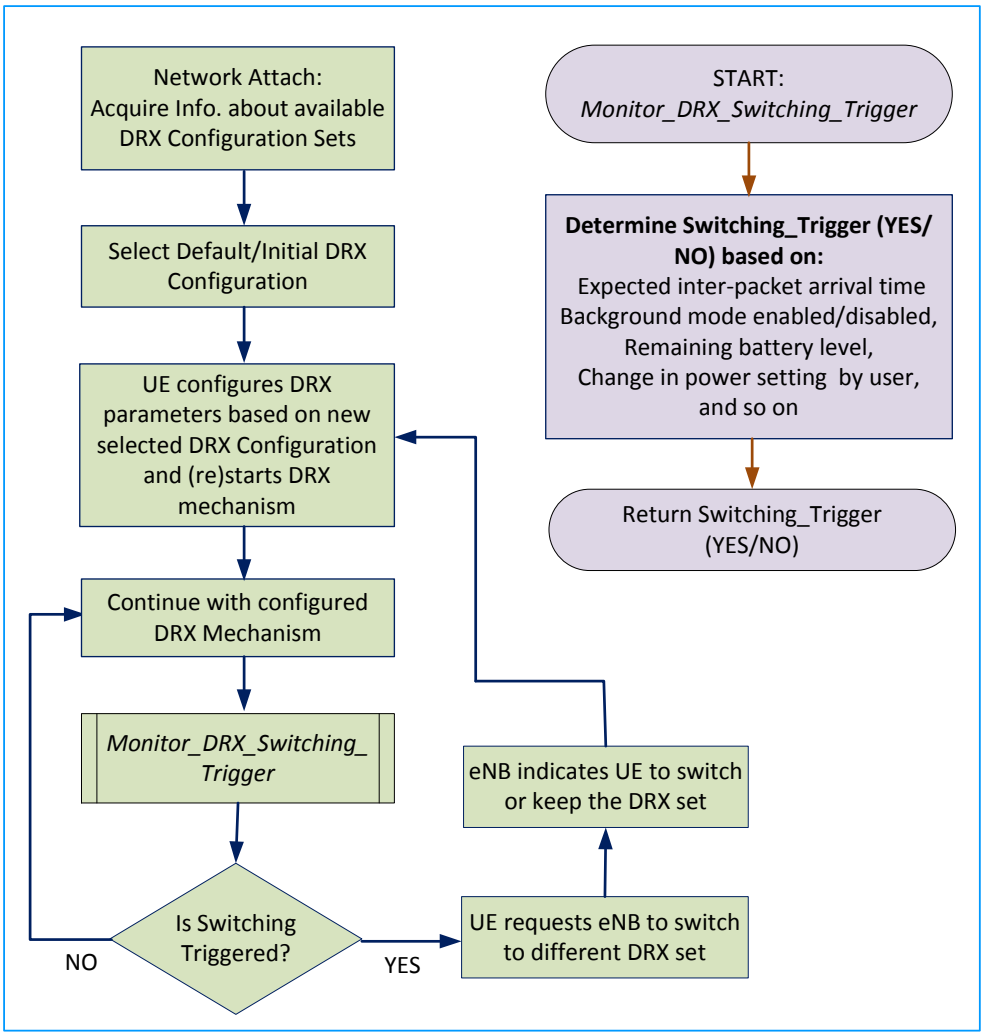

Figure 5.10. Trigger mechanism for DRX switching

Table 5.8. Active Time Percentage

Background Traffic					
DRX Cyle	Inactivity Timer				
	10ms	30ms	300ms	500ms	1280ms
40ms	25.1	25.2	26.8	27.8	29.2
80ms	12.6	12.8	14.7	15.8	17.7
160ms	6.3	6.5	8.6	9.9	11.4
320ms	3.2	3.5	5.6	7.5	10.2
640ms	1.7	1.8	3.9	5.4	8.1
1024ms	1.1	1.2	2.6	3.5	5.7
1280ms	0.9	1.0	2.4	3.3	6.5
Active Traffic					
DRX Cyle	Inactivity Timer				
	10ms	30ms	300ms	500ms	1280ms
40ms	28.4	28.6	33.1	35.8	41.0
80ms	16.3	16.8	22.0	25.3	31.9
160ms	10.4	11.1	17.0	20.2	28.2
320ms	7.7	8.2	14.9	18.6	28.1
640ms	5.6	6.0	12.8	15.5	26.2
1024ms	4.5	4.8	8.6	11.4	22.0
1280ms	4.3	4.6	8.7	11.7	25.5

reason is that longer DRX cycle for a given DRX ON duration keeps UE in sleep for longer duration in each cycle. However, this may not be suitable in terms of delay performance for active traffic as longer sleep causes higher latency.

Tables 5.8 provides the active time percentage (defined as (100− DRX Sleep Percentage)) for the selected DRX cycles and inactivity timers for UE with background and active traffic respectively. For example, in Table 5.8 (UE with background traffic), for an inactivity timer of 500ms, when the DRX cycle is changed from 160ms to 640 ms, there is 4.5% improvement (in terms of power saving). Similarly, there is 11% improvement in sleep percentage time between DRX cycle of 80ms to that of 640ms for an inactivity timer of 30ms as shown marked in the figure (12.8%, 1.8% respectively). Even as the inactivity timer is changed from 30ms to 500ms, for a DRX cycle of 160ms, the active time % increases by 3.4% to

Table 5.9. Delay for Background and Active Traffic (s)

Background Traffic					
DRX Cyle	Inactivity Timer				
	10ms	30ms	300ms	500ms	1280ms
40ms	0.02	0.01	0.02	0.02	0.01
80ms	0.03	0.03	0.04	0.02	0.02
160ms	0.07	0.07	0.08	0.06	0.03
320ms	0.15	0.15	0.16	0.14	0.04
640ms	0.31	0.31	0.29	0.26	0.12
1024ms	0.51	0.49	0.48	0.49	0.36
1280ms	0.63	0.62	0.60	0.46	0.32
Active Traffic					
DRX Cyle	Inactivity Timer				
	10ms	30ms	300ms	500ms	1280ms
40ms	0.18	0.18	0.18	0.18	0.18
80ms	0.20	0.20	0.20	0.20	0.19
160ms	0.24	0.24	0.24	0.23	0.22
320ms	0.32	0.32	0.31	0.30	0.28
640ms	0.48	0.47	0.45	0.44	0.40
1024ms	0.66	0.67	0.63	0.61	0.54
1280ms	0.79	0.80	0.75	0.74	0.63

show that there is an effect of change in DRX inactivity timer as well. As is expected, the active time % for active traffic decreases with increasing DRX cycle length. For example, for an inactivity timer of 30ms, there is a 10.8% difference between DRX cycle length of 80ms versus that of 640ms.

While power consumption constraints are more important for UE when running only background traffic; when active traffic is introduced, depending on the type of application, different delay performance can be achieved as shown in Table 5.9. As is expected the delay difference generally increases as the DRX cycle is increased. In order to meet tight latency requirements for active traffic, for example, to ensure that it is less than 200ms, the UE needs to have DRX cycle length of 80ms or lower (i.e. 40ms). Referring to Table 5.9, depending on the type of active application, the need in terms of latency varies widely

based on the long DRX cycle chosen. If a value of 640ms is chosen to be nominal for power savings, it will introduce an average of 450ms delay that may be unacceptable for some active applications.

For background traffic, we assume that the delay is not a constraint and that power constraint is more important condition to be met. As for active traffic, for the same DRX inactivity timer and DRX cycle length of 30 ms and 80 ms respectively, the delay is within 200ms requirement, but there is a trade-off in the power savings in that it is almost 11% less power efficient than that of 640ms of DRX cycle length for the same DRX inactivity timer. Thus, DRX switching can be beneficial to optimize between power and performance for a UE running different kinds of applications at different times.

UE assistance may help to choose the appropriate DRX parameters in order to improve the quality of user experience. For delay tolerant application users, power optimized DRX parameters will provide high power saving with longer battery life. On the other hand, for active traffic users delay optimized DRX parameters will satisfy their delay requirement at the expense of more power consumption.

5.5 Conclusion

In this chapter, we have presented a common platform for wireless communication systems. First, we have investigated mechanisms to optimize TCP/IP throughput over CDMA2000 network. Secondly, we have optimized the VoIP capacity of WIMAX systems with header compression. Finally, we have shown the trade off between power saving and latency of DRX mechanism in LTE-Advanced.

CHAPTER 6
FUTURE DIRECTIONS AND CONCLUSIONS

With the increase of the usage of mobile devices such as smart phones, tablets etc. that offer internet applications like social networking, email and web browsing, user expectations rise in terms of higher data rates and minimum power consumption. In this dissertation, we have proposed radio resource control mechanisms to tackle these two problems.

ARQ is used to improve the data rate and reliability over the cellular networks. For the optimization of the ARQ parameters, the distribution of the product of independent random variables is needed. In this dissertation, we have proposed an approximation to the distribution of products of independent random variables by the widely applicable Nakagami-m distribution by means of moment matching method. Existing methods to evaluate the statistical properties of product distributions are limited to numerical approximations based on distribution fitting tools, Gaussian approximation, or special functions in terms of infinite series. The obtained results show that the moment matching method can be useful even approximating the product distribution of Rician random variables as Nakagami-m distribution. Moment matching also allows us to obtain the insightful results with analytical means. The analytically tractable Nakagami-m distribution can significantly simplify the performance analysis of ARQ and Hybrid-ARQ transmission systems. We have demonstrated the usefulness of the proposed approximation in rate offset calculations for ARQ systems and compared the performance of proposed approximation with Gaussian approximation. This work can be extended to analyse performance of ARQ systems as a function of average SNR and outage probability.

Combining channel coding and retransmission protocols are simple and effective solution

for high data rate and reliable random multi-access communication. On top of that adding MIMO support to random access schemes improve the reliability of the system by adding more degrees of freedom. We have proposed a threshold based multi-access system then we have extended analysis to multi antenna systems. We have showed the affect of simple receive beamforming on Aloha, Chase combining and INR Hybrid-ARQ schemes and came up with analytical throughput expressions using reward renewal theorem. In our analysis, we have assumed that the power of each user is fixed. However, in real networks power control is another way to combat the channel error. This work can be extended with the inclusion of power control.

For power consumption issues, we have demonstrated the necessity of optimizing DRX configuration. With DRX configuration optimization, networks can maintain a tradeoff between power saving and latency. We have developed an analytical model for the DRX operation in context of both active traffic and background traffic. We have derived expressions for performance parameters such as power saving and buffering delay. We then formulated an optimization problem to find tradeoff between power saving and delay based on operator's preference for power saving and latency. We have validated the analytical model through system level simulations. Results showed that proposed tradeoff scheme is highly efficient in tuning DRX configuration to achieve required power saving while ensuring reasonably low latency. This work can be extended to find the optimum DRX parameters for VoIP application.

Broadband wireless access has gained traction during the past few years by deployment of OFDMA based LTE and WIMAX systems around the world. Network modeling and simulation of these cellular technologies are crucial to improve the performance and to evaluate new features. Our unified simulation environment has provided unique ways to simulate and enhance these cellular technologies.

APPENDIX

A.1 Instantaneous SINR Conditioned on the Number of Interfering Users for Single Antenna Systems

The CDF of instantaneous SINR μ (3.21) with u on-off interferers based on threshold γ_T

$$\begin{aligned} F_{\mu}^{u}(z) &= \Pr\{\mu \le z\} = \Pr\{\hat{\gamma} \le (1+\hat{\gamma}_{MAI}^{u})z\} \\ &= \int_0^{\infty} dy \int_0^{z(1+y)} dx f_{\hat{\gamma}}(x) f_{\hat{\gamma}_{MAI}^{u}}(y) \\ &= \int_0^{\infty} F_{\hat{\gamma}}(z(1+y)) f_{\hat{\gamma}_{MAI}}(y) dy \end{aligned}$$

After some manipulations the CDF becomes

$$\begin{aligned} F_{\mu}^{u}(z) &= pU[z] + p^{u}\left(e^{-\gamma_T/\gamma} - e^{-z/\gamma}\right) U[z-\gamma_T] \\ &+ \sum_{K=0}^{u-1} \binom{u}{K} p^{K} \left\{ \left(e^{-\gamma_T/\gamma} F_{u-K}\left[\frac{\gamma_T/z-1}{\gamma}\right] - e^{-z/\gamma} G(\frac{\gamma_T/z-1}{\gamma}, u, K) \right) \right. \\ &\times (U[z] - U[z-\gamma_T/((u-K)\gamma_T+1)]) \\ &+ \left. \left(e^{-\gamma_T(u-K+1)/\gamma} - e^{-z/\gamma} \frac{e^{-(u-K)\gamma_T(z+1)/\gamma}}{(z+1)^{u-K}} \right) U[z-\gamma_T/((u-K)\gamma_T+1)] \right\} \end{aligned}$$

where

$$\begin{aligned} p &= 1 - e^{-\gamma_T/\gamma} \\ G(v_1, u, K) &= \int_{v_1-\gamma_T(u-K)}^{\infty} \frac{e^{-\gamma_T(u-K)(z+1)/\gamma} t^{u-K-1}}{\gamma^{u-K}(u-K-1)!} e^{-t(z+1)/\gamma} dt \qquad v = v_1/\gamma \\ F_N(v) &= e^{-N\gamma_T/\gamma} \sum_{k=0}^{N-1} \frac{v^k}{k!} e^{-v} \end{aligned}$$

The CDF of the instantaneous SINR for $K = 2$ users ($u = 1$ interfering user) can be expressed as:

$$\begin{aligned} F_{\mu}^{1}(z) &= (1 - e^{-\gamma_T/\gamma})U[z] + \frac{ze^{-(z\gamma_T+\gamma_T-z)/(z\gamma)}}{z+1}(U[z] - U[z - \frac{\gamma_T}{\gamma_T+1}]) \\ &+ (e^{-2\gamma_T/\gamma} - \frac{e^{-(\gamma_T+(\gamma_T+1)z)/\gamma}}{z+1})U[z - \frac{\gamma_T}{\gamma_T+1}] \\ &+ (1 - e^{-\gamma_T/\gamma})(e^{-\gamma_T/\gamma} - e^{-z/\gamma})U[z - \gamma_T] \end{aligned}$$

A.2 Instantaneous SINR Conditioned on the Number of Interfering Users for Multiple Antenna Systems

As it is shown in [14], the cumulative distribution function of instantaneous SINR μ with u on-off interferers with M receive antennas

$$\begin{aligned} F_{\mu,M}(x) &= \Pr\{\mu \leq x\} = \Pr\{X_M \leq (1+Y)x\} \\ &= \int_0^{\infty} dy \int_0^{x(1+y)} dx f_{X_M}(x) f_Y(y) \\ &= \int_0^{\infty} F_{X_M}(x(y+1)) f_Y(y) dy \end{aligned}$$

After some manipulations the CDF becomes

$$\begin{aligned} F_{\mu,M}(x) &= 1 - \frac{e^{-x/\gamma}}{(1+x)^u} \\ &\left[\sum_{k=0}^{M-1} \sum_{l=0}^{k} \frac{(x/\gamma)^{k-l}}{(k-l)!} (\frac{x}{x+1})^l \begin{pmatrix} u-1+l \\ u-1 \end{pmatrix}\right] \end{aligned}$$

where u is the number of interfering users.

REFERENCES

[1] "ITU. global cellular subscriptions." `http://www.itu.int/ITU-D/ict/statistics/index.html`.

[2] L. G. Roberts, *Aloha Packet System with and without slots and capture.* Stanford Research Institute, first ed., 1972.

[3] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 10 –22, Jun. 2010.

[4] R. Kim and S. Mohanty, "Advanced power management techniques in next-generation wireless networks," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 94 –102, May 2010.

[5] C. Bontu and E. Illidge, "DRX mechanism for power saving in LTE," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 48 –55, Jun. 2009.

[6] S. Gao, H. Tian, J. Zhu, and L. Chen, "A more power-efficient adaptive discontinuous reception mechanism in LTE," in *Proc. IEEE VTC'11-Fall*, pp. 1 –5, Sept. 2011.

[7] "Evolved universal terrestrial radio access (E-UTRA); medium access control (MAC); protocol specification," *3GPP TS 36.321, Rel. 10, V.10.2.0*, Jun. 2011.

[8] "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," *3GPP TS 36.331, Rel. 10, V.10.3.0*, Sept. 2011.

[9] J. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WIMAX.* Prentice Hall, first ed., 2007.

[10] H. Holma and A. Toskala, *HSDPA/HSUPA for UMTS: High Speed Radio Access for Mobile Communications.* Wiley, first ed., 2006.

[11] S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications.* Prentice-Hall, first ed., 1983.

[12] R. Link, S. Kallel, and S. Bakhtiyari, "Throughput performance of memory ARQ scheme," *IEEE Trans. on Veh. Technol.*, vol. 48, no. 1, pp. 891–899, May 1999.

[13] R. Cam and C. Leung, "Throughput analysis of some ARQ protocols in the presence of feedback errors," *IEEE Trans. on Commun.*, vol. 45, no. 1, pp. 35–44, Jan. 1997.

[14] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. on Information Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001.

[15] P. Wu and N. Jindal, "Analysis of fixed outage transmission schemes: A finer look at the full multiplexing point," in *Proc. IEEE Int. Conf. on Commun.*, May 2008.

[16] H. E. Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel:diversity-multiplexing-delay tradeoff," *IEEE Trans. on Information Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.

[17] L. Zhou, H. Xu, H. Tian, Y. Gao, L. Du, and L. Chen, "Performance analysis of power saving mechanism with adjustable DRX cycles in 3GPP LTE," in *Proc. IEEE Veh. Technol. Conf.*, pp. 1 –5, Sept. 2008.

[18] H. Bo, T. Hui, C. Lan, and Z. Jianchi, "DRX-aware scheduling method for delay-sensitive traffic," *IEEE Commun. Lett.*, vol. 14, pp. 1113 –1115, Dec. 2010.

[19] W. T. Wells, R. L. Anderson, and J. W. Cell, "The distribution of the product of two central or non-central chi-square variates," *The Annals of Mathematical Statistics*, pp. 1016–1020, Jan. 1962.

[20] J. Salo, H. M. El-Sallabi, and P. Vainikainen, "Statistical analysis of the multiple scattering radio channel," *IEEE Trans. on Antennas and Propagation*, vol. 54, pp. 3114–3124, Nov. 2006.

[21] J. Salo, H. M. El-Sallabi, and P. Vainikainen, "The distribution of the product of independent rayleigh random variables," *IEEE Trans. on Antennas and Propagation*, vol. 54, no.2, pp. 639–643, Feb. 2006.

[22] P. Wu and N. Jindal, "Performance of hybrid-arq in block-fading channels: A fixed outage probability analysis," *IEEE Trans. on Commun.*, vol. 58, no. 4, pp. 1129–1141, Apr. 2010.

[23] H. Lu, Y. Chen, and N. Cao, "Accurate approximation to the pdf of the product of independent rayleigh random variables," *IEEE Antenna and Wireless Propagation Letters*, vol. 10, pp. 1019–1022, Jun. 2011.

[24] Y. Chen, G. K. Karagiannidis, H. Lu, and N. Cao, "Novel approximations to the statistics of products of independent random variables and their applications in wireless communications," *IEEE Trans. on Vehicular Technology*, vol. 61, pp. 443–453, Feb. 2012.

[25] A. T. Koc and M. Torlak, "Analytical approximation to the distribution of the product of independent random variables," *submitted, IEEE Antennas and Propagation Letters, Jan. 2013*.

[26] M. K. Simon and M.-S. Alouni, *Digital Communications over Fading Channels*. Wiley Series, 1th ed., 2000.

[27] C. Suh, A. T. Koc, and S. Talwar, "Interference aware power control for cellular systems," in *Proc. IEEE Globecom*, (Honolulu, Hawaii), Nov. 2009.

[28] G. Miao, N. Himayat, G. Y. Li, A. T. Koc, and S. Talwar, "Interference aware energy efficient power optimization," in *Proc. IEEE Int. Conf. on Commun.*, Jun. 2009.

[29] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," pp. 331–335, Jun. 1995.

[30] N. Abramson, "The aloha system-another alternative for computer communications," in *Proc. Fall Joint Comput. Conf.*, (Houston, TX), pp. vol. 44, 281–285, Apr. 1970.

[31] X. Qin and R. Berry, "Exploiting multiuser diversity for medium access control in wireless networks," in *Proc. IEEE Infocom Conf.*, (San Francisco, CA), pp. 1084–1094, 2003.

[32] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *IEEE Trans. on Information Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.

[33] G. Mergen and L. Tong, "On the asymptotic stable throughput of opportunistic random access," in *Conf. on Information Sciences and Systems, Princeton University*, (Princeton), pp. 1–6, Mar. 2004.

[34] K. Bai and J. Zhang, "Opportunistic multichannel aloha: Distributed multiaccess control scheme for OFDMA wireless networks," pp. 848–855, vol. 55, no. 3, May 2006.

[35] N. Zhang, B. Vojcic, M. Souryal, and E. G. Larsson, "Exploiting multiuser diversity in reservation random access," *IEEE Trans. on Wireless Commun.*, vol. 5, no. 9, pp. 2548–2554, Sept. 2006.

[36] D. Gesbert and M. Alouini, "How much feedback is multi-user diversity really worth?," in *Proc. IEEE Int. Conf. on Commun., Paris, France*, pp. 234–238, Jun. 2004.

[37] V. Hassel, M. Alouini, D. Gesbert, and G. E. Qien, "Exploiting multiuser diversity using multiple feedback thresholds," in *Proc. IEEE Veh. Technol. Conf., Stockholm, Sweden*, pp. 1302–1306, Jun. 2005.

[38] M. K. Simon and M. Alouini, "Performance analysis of generalized selection combining with threshold test per branch(T-GSC)," *IEEE Trans. on Veh. Technol.*, vol. 51, no. 5, pp. 1018–1029, Sept. 2002.

[39] A. T. Koc and M. Torlak, "Performance of hybrid arq in block fading multiantenna channels," *Proc. IEEE Veh. Technol. Conf., Taipei, Taiwan*, pp. 1–5, May 2010.

[40] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge University Press, first ed., 2005.

[41] A. T. Koc and M. Torlak, "Concatenated beamspace-time coding for directional wireless channels," in *Proc. IEEE Asilomar Conf. on Signals,Systems and Computers*, (Pacific Grove,CA), Nov. 2002.

[42] A. T. Koc, B. Kecicioglu, D. Kim, and M. Torlak, "Experimental investigation of four transmit antenna space-time block codes," in *Proc. IEEE Wireless and Microwave Technology Conf., Orlando, Florida*, Dec. 2006.

[43] A. T. Koc, S. C. Jha, R. Vannithamby, and M. Torlak, "Adaptive drx configuration to optimize device power saving and latency of mobile applications over lte advanced network,," *submitted to Proc. IEEE ICC 2013*, Jun. 2013.

[44] F.-S. Chu, K.-C. Chen, and G. Fettweis, "Green resource allocation to minimize receiving energy in OFDMA cellular systems," *IEEE Commun. Lett.*, vol. 16, pp. 372 –374, Mar. 2012.

[45] G. Berardinelli, T. Sorensen, P. Mogensen, and K. Pajukoski, "Transmission over multiple component carriers in LTE-A uplink," *IEEE Trans. on Wireless Communication*, vol. 18, no. 4, pp. 67 –73, Aug. 2011.

[46] "Technological progression of LTE-advanced in release 11," *IEE Commun. Mag.*, vol. 50, pp. 24 –26, Nov. 2012.

[47] "LTE RAN enhancements for diverse data applications," *3GPP TR 36.822 V0.2.0*, Nov. 2011.

[48] "Study on non-mtc mobile data applications impacts (release 12)," *3GPP TR 22.801, Rel. 12, V.12.0.0*, Dec. 2011.

[49] J. Puttonen, E. Virtej, I. Keskitalo, and E. Malkamaki, "On LTE performance tradeoff between connected and idle states with always-on type applications," in *IEEE Int. Symp. PIMRC'12*, pp. 981 –985, Sept. 2012.

[50] S. Gao, H. Tian, J. Zhu, and L. Chen, "A more power-efficient adaptive discontinuous reception mechanism in LTE," in *Proc. IEEE VTC'11-Fall, San Francisco, United States*, pp. 1 –5, Sept. 2011.

[51] T. Kolding, J. Wigard, and L. Dalsgaard, "Balancing power saving and single user experience with discontinuous reception in LTE," in *IEEE Int. Symp. ISWCS'08*, pp. 713 –717, Oct. 2008.

[52] M. Polignano, D. Vinella, D. Laselva, J. Wigard, and T. Sorensens, "Power savings and QoS impact for VoIP application with DRXDTX feature in LTE," in *IEEE Proc. VTC-Spring'11, Budapest, Hungary*, (Budapest, Hungary), pp. 1 –5, May 2011.

[53] K.-C. Ting, H.-C. Wang, C.-C. Tseng, and F.-C. Kuo, "Energy-efficient DRX scheduling for QoS traffic in LTE networks," in *IEEE Int. Symp. ISPA'11*, pp. 213 –218, May 2011.

[54] "LTE RAN enhancements for diverse data applications," *3GPP RAN Plenary Contribution, RP-110410*, Mar. 2011.

[55] S. Baghel, K. Keshav, and V. Manepalli, "An investigation into traffic analysis for diverse data applications on smartphones," in *Nat. Commun. Conf. NCC'12*, pp. 1 –5, Feb. 2012.

[56] J. Si, A. Barto, W. Powell, and D. Wunsch, "Hierarchical decision making," in *Handbook of Learning and Approximate Dynamic Programming*, pp. 203 –232, Wiley-IEEE Press, 1 ed., 2004.

[57] J. Janssen and R. Manca, *Applied Semi-Markov Processes*. Springer, Oct. 2010.

[58] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. and Multidisciplinary Optim.*, vol. 26, pp. 369–395, 2004.

[59] T. Murata, H. Ishibuchi, and H. Tanaka, "Multi-objective genetic algorithm and its applications to flowshop scheduling," *Comput. and Ind. Eng.*, vol. 30, no. 4, pp. 957–968, 1996.

[60] S. Jha, A. T. Koc, R. Vannithamby, and M. Torlak, "Optimizing drx configuration to improve battery power saving and latency of active mobile applications over lte-a network," *Proc. IEEE WCNC 2013*, Apr. 2013.

[61] A. T. Koc, S. C. Jha, R. Vannithamby, and M. Torlak, "Device power saving and latency optimization for emerging mobile applications in lte-a networks through drx configuration," *submitted, IEEE Trans. on Mobile Computing, Dec. 2012*.

[62] "OPNET: Applications and network performance," *OPNET Technologies, Inc., OPNET Modeler Version: 17.1.A*.

[63] A. T. Koc, S. C. Jha, and M. Torlak, "Network modeling and simulation of 3g/4g cellular networks," *submitted, IEEE Trans. on Wireless Communication Magazine, Jan. 2013*.

[64] "Cisco visual networking index: Forecast and methodology, 2011-2016.," *White Paper, Cisco Systems, Inc., Visual Networking Index Forecast*, pp. 1 –16, May 2012.

[65] A. T. Koc, O. Ozdemir, M. Torlak, and W. P. Osborne, "Opnet simulation of random access cdma networks," *Proc. OPNETWORK 2002*, Aug. 2002.

[66] A. T. Koc, O. Ozdemir, and M. Torlak, "Reduced rank noncohenerent interference suppression for cdma communications," in *Proc. IEEE Asilomar Conf. on Signals,Systems and Computers, Pacific Grove, CA*, Nov. 2004.

[67] A. T. Koc, O. Ozdemir, and M. Torlak, "Simulation based study of random access cdma networks," *Journal of Naval Science and Engineering*, vol. 2, pp. 49–76, Jul. 2004.

[68] "Ieee 802.16-2009: Part 16:air interface for broadband wireless access systems," Jun. 2009.

[69] E. Haghani and N. Ansari, "Voip traffic scheduling in wimax networks," in *Proc. IEEE Globecom*, (New Orlean, LA), Nov. 2008.

[70] A. T. Koc and L. C. Zhu, "Voip capacity enhancements using opnet wimax module," *Proc. OPNETWORK 2010*, Aug. 2010.

[71] S. C. Jha, A. T. Koc, and R. Vannithamby, "Optimization of discontinuous reception (DRX) for mobile internet applications over LTE," in *Proc. IEEE VTC'12-Fall*, pp. 1 –5, Sept. 2012.

[72] "E-utra and e-utran overall description, stage 2," *3GPP TS 36.300 V11.1.0*, Mar. 2012.

[73] A. T. Koc, S. C. Jha, M. Tarradell, and S. Bangaloe, "Drx configuration switching to optimize user power saving and application performance requirements fro lte advanced networks," *Proc. OPNETWORK 2012*, Aug. 2012.

[74] S. Jha, A. Koc, and R. Vannithamby, "Optimization of discontinuous reception (drx) for mobile internet applications over lte," *Proc. IEEE Veh. Tech. Conference*, Apr. 2012.

VITA

Ali Taha Koç was born in Ankara, Turkey, on June 5, 1980, the son of Şadiye and Atilla Koç. He received the Bachelor of Science degree in Electrical & Electronics Engineering from Bilkent University in June, 2001 in Ankara, Turkey. He then joined the graduate school at The University of Texas at Dallas in August, 2001. He received the Master of Science degree in Electrical Engineering from The University of Texas at Dallas in December, 2005. He has been working at Intel Corporation since January 2006 as a network software engineer. He has worked on power control, interference mitigation and power saving mechanisms for IEEE 802.16e and IEEE 802.16m networks in the past. His current research interests include device power saving and user experience enhancement mechanisms in 3GPP LTE-A networks.

Permanent address: 2356 NW Edinburg Drive
Hillsboro, Oregon 97124

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.