Performance Models for LTE-Advanced Random Access

by

Revak Tyagi

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved April 2014 by the
Graduate Supervisory Committee:

Martin Reisslein, Chair
Cihan Tepedelenlioglu
Michael McGarry
Yanchao Zhang

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

LTE-Advanced networks employ random access based on preambles transmitted according to multi-channel slotted Aloha principles. The random access is controlled through a limit W on the number of transmission attempts and a timeout period for uniform backoff after a collision. The LTE-Advanced random access system is modeled by formulating the equilibrium condition for the ratio of the number of requests successful within the permitted number of transmission attempts to those successful in one attempt. It is analytically proved that for $W \leq 8$ there is only one equilibrium operating point, and for $W \geq 9$ there are three operating points if the request load ρ is between load boundaries ρ_1 and ρ_2. These load boundaries as well as the corresponding system operating points are analytically identified. Analytical expressions for the throughput and delay of successful requests at the operating points are found and validated through simulations. Further, the results are generalized using a steady-state equilibrium based approach and models for single-channel and multi-channel systems are developed, incorporating the barring probability P^B. Ultimately, the de-correlating effect of parameters O, P^B, and $T_o^{\max}$ is identified and the Poissonization effect due to the backlogged requests in a slot is introduced. The impact of Poissonization on different traffic is studied. Further research directions are discussed at the conclusion.

DEDICATION

I dedicate this thesis,

To my grandparents, who are not alive to see this thesis come to completion, for their love and stories that are integral to myself.

To my parents Dr. Lokeshwar Sharma and Dr. Vijay Tyagi, for their help, support, and motivation.

To my brother Maj. Rachak Tyagi, for successfully prodding me to pursue graduate studies.

To my sister Kadambari, for taking care of everyone.

To my wife Isha, for her love, care, and affection which helped me remain focused and motivated while working on this thesis.

To our son Shiven, who is quick to crawl at the time of writing this thesis.

ACKNOWLEDGEMENTS

I am grateful to Prof. Martin Reisslein, without whose constant guidance and support this research could not have materialized. I would also like to express my gratitude to Prof. Frank Aurzada for his help with mathematical analysis and for insightful discussions. I would like to thank Dr. Ki-Dong Lee for help with LTE standard. I would also like to thank Prof. Cihan Tepedelenlioglu and Prof. Michael McGarry for their help during the problem exploration, and Prof. Yanchao Zhang whose class on network security was also helpful during problem exploration.

I gratefully acknowledge the Arizona State University, National Science Foundation, SenSIP center, and LG Electronics Mobile Research, San Diego for financial support.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Long Term Evolution (LTE) Advanced (LTE-Advanced) Larmo *et al.* (2009) is a popular Radio Access Network (RAN) protocol standard for 4G cellular networks, which has been chosen by many service providers worldwide 4G Americas (2012); Boudriga *et al.* (2008). Given the prominence of LTE-Advanced in 4G networks it is important to thoroughly analyze its protocol features. In this study, we focus on the slotted Aloha based random access procedure, which distributed user equipment (UE) nodes must complete to establish a connection to the central enhanced Node B (eNB). For applications with frequent small data transmissions, such as periodic monitoring of patient vitals in ubiquitous health care systems Lee and Vasilakos (2011), the random access procedure must be completed for each data transmission, thus efficient and low-delay completion of the random access is highly important.

In brief, the LTE-Advanced random access protocol consists of an access barring check Lien *et al.* (2012), which may bar (block) a UE from attempting to connect to the eNB for a prescribed time period, followed by preamble contention. The preamble contention follows essentially the principles of a multi-channel slotted Aloha system Rom and Sidi (1989) with a limit on the number of retransmissions. Specifically, in a given time slot, the UEs with connection requests randomly select a preamble from among a set of O orthogonal preambles. If two or more UEs select the same preamble, a collision occurs. A UE with a collided preamble retransmits in a later slot, provided it has not exhausted the W permitted transmission attempts. In this study, we focus on the preamble contention and leave the access barring for future work.

We model the preamble contention for an infinite UE population generating requests according to a Poisson processes through a system equilibrium condition. We define a function $h(t)$ to represent the ratio of the expected number of requests successfully completed within the W attempts to the number of attempts completed in one attempt for a given expected number of transmitting UEs t. We analyze the equilibrium condition by examining the intersections of the function $h(t)$ with the line t/ρ whose slope is inversely proportional to the load ρ. From this analysis, we show that for $W \leq 8$ there is only a single equilibrium operating point. For $W \geq 9$ we analytically specify load boundaries ρ_1 and ρ_2 that depend only on W, such that for loads ρ in the range $(\rho_1,\ \rho_2)$, three equilibrium operating points exist, which we analytically specify. For loads ρ outside the $[\rho_1,\ \rho_2]$ range, only one equilibrium exists for $W \geq 9$. While slotted Aloha systems have been analyzed extensively in the literature (see Section 1.1), to the best of our knowledge, our study is the first to analytically specify the boundaries of the load range giving rise to multiple equilibria for $W \geq 9$ as well as to analytically specify the throughput and delay at these operating points.

The widely expected increase in the number of nodes combined with new services, such as ubiquitous healthcare applications, machine-type and smartphone communication (MTSC) Lien *et al.* (2011a), and other small data applications, will frequently generate small data sets. Low delay is often a key requirement for these frequent small data sets, which result in a high random access load. One possible strategy for ensuring low delays is to limit the traffic load. For instance, for a system with transmission attempt limit $W \leq 8$, which has only a single operating point, our delay analysis can be used to limit the load ρ so as to keep the mean delay below a tolerable delay target. For systems with $W \geq 9$, the load can be limited to below the lower load boundary ρ_1 to avoid the performance degradations due to multiple operating points.

1.1 Related Works

The throughput-delay performance of slotted Aloha type random access without a limit on the number of transmission attempts, which corresponds to $W \to \infty$, has been examined in a number of seminal studies for single-channel systems Carleial and Hellman (1975); Dai (2012); Ferguson (1975); Jenq (1980a); Kamal and Mamoud (1979); Kleinrock and Lam (1975a); Murali and Hughes (1998a); Naware *et al.* (2005a) and multi-channel systems Pountourakis and Sykas (1992); Shen and Li (2002); Szpankowski (1983); Yue (1991); Zhang and Liu (1993). For the infinite node model, several of these seminal studies, e.g., Carleial and Hellman (1975); Jenq (1980a); Kamal and Mamoud (1979); Kleinrock and Lam (1975a); Murali and Hughes (1998a); Szpankowski (1983) found that for (per channel) loads $\rho < 1/e$, slotted Aloha has three equilibrium operating points, namely one low-delay high-throughput operating point, an intermediate point corresponding to moderate delay and throughput, and a saturation point corresponding to high delay and low throughput; for loads $\rho > 1/e$ (whereby $1/e$ corresponds to our load boundary ρ_2 for large W), only a single saturation point exists. For a multi-channel slotted Aloha system with fast retry (i.e., immediate retransmission in the next slot) or uniform backoff (i.e., retransmission after a uniformly distributed backoff time), we extend these results as follows: We show that for a finite limit on the number of transmission attempts W, $W \geq 9$, there is a load boundary ρ_1 below which only a single operating point exists; for large W, the ρ_1 asymptotically behaves as $(\log W)/W$.

Slotted Aloha based random access with a transmission attempt limit W has been simulated in Lüders and Haferbeck (1994), while the impact of retransmissions on a general packet (cell) queueing system has been analyzed in Grishechkin *et al.* (2003) and a limit of $W = 3$ has been shown to minimize delays in lightly loaded

slotted Aloha in Simon and Votta (1985). Kwak et al. Kwak *et al.* (2005) analyzed the effects of limiting the number of transmission attempts on general exponential backoff. Kim Kim (1992) formulated an equilibrium condition and observed from exhaustive numerical exploration the existence of either one or three equilibrium operating points for different limits on the number of transmission attempts. Similarly, Liu Liu (2002) formulated an equilibrium condition for slotted Aloha with transmission attempt limit and explored this equilibrium condition numerically. In contrast, we formally analyze our equilibrium condition to identify the multiple operating points as well as the corresponding throughputs and delays.

Sakakibara *et al.* Sakakibara *et al.* (2000, 2003) and Noguchi *et al.* Onozato *et al.* (1986); Wang *et al.* (1991) modeled slotted Aloha systems with the formalisms of catastrophe theory Onozato and Noguchi (1985b,a). Through a cusp theory approximation within the catastrophe theory formalism, Sakakibara et al. proved that for $W \leq 8$ transmissions there exists only a single equilibrium operating point, whereas for $W \geq 9$ there are load boundaries within which multiple operating points exist. In contrast, we model the underlying slotted Aloha random access dynamics directly through an elementary equilibrium equation. We not only prove the existence of single and multiple operating points, but also analytically identify the load boundaries that give rise to multiple operating points for $W \geq 9$ as well as analytically identify all operating points. We also analytically derive the throughput-delay performance corresponding to the operating points.

Sarker *et al.* Sarker and Halme (2000a); Sarker (2006); Sarker and Mouftah (2008, 2013) investigated the impact of limiting the number W of transmission attempts on the throughput. Their work is complementary to ours in that their main focus is on controlling the number of transmission attempts so as to maximize throughput by operating the system near the classical stability limit of $1/e$ of the channel bandwidth

(which approximately corresponds to our upper load boundary ρ_2). In contrast to the work of Sarker et al., we include the delay in our evaluations and identify the impact of the single or multiple operating points on throughput and delay.

Recently, the various aspects of LTE random access have attracted significant research interest. Yilmaz *et al.* Yilmaz *et al.* (2011) identified optimization problems for LTE random access. Vukovic and Filipovich Vukovic and Filipovich (2011) examined the impact of different physical random access configurations, such as possible non-uniform distribution of random access opportunities over the slot in the LTE time structure. Kwan and Leung Kwan and Leung (2011) examined the inter-cell interference caused by neighboring eNBs. Wei *et al.* Wei *et al.* (2012) examined group paging in an LTE network, where each UE in the group has only one requests. Yun Yun (2012) comprehensively described the physical (PHY) layer and medium access control (MAC) layer of 3GPP Universal Terrestrial Radio Access, which is related to LTE-Advanced, with a Markov chain model. Our study is complementary to Yun (2012) in that we focus on the medium access control and analyze in detail its implications for the existence of a single or multiple operating points, which are not explicitly considered in Yun (2012), as well as the resulting throughput and delay. Seo and Leung Seo and Leung (2011a) studied the uniform backoff in LTE relative to the exponential backoff in IEEE 802.16 WiMAX. In the context of this backoff study, Seo and Leung briefly analyzed the implications of these backoff mechanisms on system operating points in the limit $W \to \infty$ for saturated (high traffic) conditions. Similarly, in Seo and Leung (2012), Seo and Leung analyzed the initial random access with infinite retransmission limit for LTE semi-persistent scheduling. Our study focused on the impact of the finite transmission attempt limit W and covers the full range of load conditions.

Slotted Aloha based contention with limited number of trials arises also when a mobile reader scans RFID tags Sarangan *et al.* (2008). Multiple equilibrium operating points for the mobile RFID reader have been observed by Alcaraz *et al.* Alcaraz *et al.* (2011) and considered in the setting of RFID system parameters. Our analysis complements the Alcaraz *et al.* model in that we analyze the specific underlying conditions that give rise to the multiple operating points and identify these points.

Chapter 2

MODEL OF LTE-ADVANCED RANDOM ACCESS SYSTEM

2.1 Random Access Protocol

We consider a single cell in a cellular system, whereby the cell is comprised of a central node, referred to as evolved Node B (eNB), and of multiple User Equipment (UE) nodes. Considering a single cell is not a limitation of our model since the RA procedure in LTE is an interaction between a prescribed UE and its most relevant cell, which is chosen while staying in RRC_IDLE mode or while staying in RRC_CONNECTED mode with time synchronization lost (e.g., when the timing alignment timer expires).

The user equipment (UE) nodes try to establish a communication flow with the eNB, which can be thought of as a circuit-switched connection for the purposes of the present model. (Generally, LTE operates in the packet-switched mode while providing a circuit-switched mode as a fallback; however these details are not relevant for the present model.)

As illustrated in Figs. 2.1 and 2.2, the overall Random Access (RA) procedure in LTE-Advanced (as well as in the LTE standard preceding LTE-Advanced) consists of an access barring check followed by random access preamble transmission and response. The access barring check allows a UE with probability (access barring factor) P_b to immediately transmit a preamble; while the UE has to wait with probability $1 - P_b$ for an ac-BarringTime. In the present study we focus on the analysis of the preamble transmission by setting $P_b = 1$ and leave the incorporation of the access barring check into our model for future work.

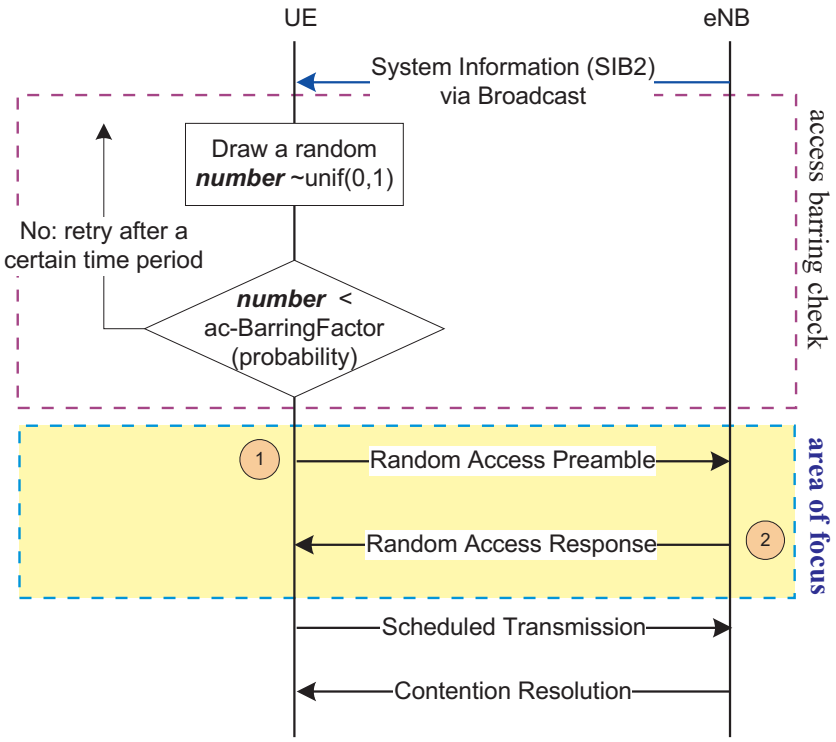

Figure 2.1: Illustration of LTE-Advanced Random Access (RA) procedure and the model for the preamble contention: (a) The UEs that have passed through the access barring check start transmitting a preamble, and the contention-based transmissions from multiple UEs generates a request load in the random access system.

UEs use one of the Random Access Channel (RACH) opportunities configured by the physical (PHY) layer. The RACH is a set of logical resources defined in the 3-dimensional domain of time-frequency-preamble, whereby the UE randomly chooses a preamble from among O, $O > 1$, allowable preambles.

The eNodeB receives RA requests from UEs during a time slot of duration T_s. If multiple UEs transmitted their requests using the same preamble in the same slot, then those RA requests are considered to have collided. We note that physical layer considerations, such as different levels of transmission power among the UEs, can

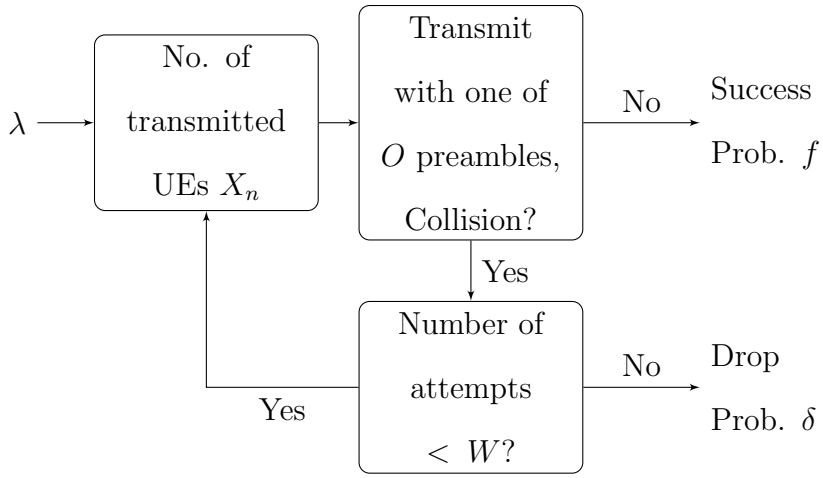

Figure 2.2: Illustration of LTE-Advanced Random Access (RA) procedure and the model for the preamble contention: (b) Requests for new preamble transmissions arriving at rate λ [requests/slot] and collided transmissions that have not yet exhausted their W permitted attempts contribute to the total number X_n of UEs transmitting a preamble in a slot n. Each transmitting UE randomly selects one of the O preambles.

influence the success or collision of RA requests. The focus of this present study is on capturing the MAC layer behavior and thus detailed physical layer considerations are beyond the scope of this study. When a collision occurs, contention (for RA) is not considered to be resolved, i.e., contention resolution failed. UEs are able to identify the contention resolution result at the fourth step of the RA of LTE 36.321 (2011). If contention is resolved, the UE enters the RRC_CONNECTED mode.

When contention is not resolved, the UE may repeat the preamble transmission. Specifically, if the UE has had less than W transmission attempts so far, it re-transmits. On the other hand, if the Wth preamble transmission has failed, then the UE drops the request. Before its re-transmission, the UE waits according to a prescribed backoff interval $T_o^{\max}$ ranging from 0 ms to 960 ms, which is signaled by the eNB. For simplicity, we initially set the backoff interval to $T_o^{\max} = 0$ ms, that is,

UEs whose preamble transmissions collide in a given slot may re-transmit in the next slot (non-zero backoff intervals are examined in Appendix D). The setting $T_o^{\max} = 0$ corresponds to fast retry in Choi *et al.* (2006). Each re-transmitting UE uniformly randomly selects a new preamble from among O preambles, independently of the preceding preamble selection.

2.2 Performance Metrics

The two key performance metrics directly related to the random access procedure are the mean (steady-state) delay D and the mean (steady-state) throughput T of the random access system in equilibrium. We define the delay D of random access as the time period from the instant a UE generates a preamble to the instant the UE is notified about the accepted connection; whereby only requests that are successful within the W transmission attempts are considered in the delay evaluation. We define the throughput T as the long-run average rate at which connection acceptances are granted.

Chapter 3

SYSTEM ANALYSIS

3.1 Definition of System Characteristics

Table 3.1: Summary of Stability Analysis Model Notations

Slotted Aloha based preamble contention	
O	Number of preambles (equivalent to number of channels in multi-channel slotted Aloha)
T_s	Slot duration for slotted Aloha contention [in seconds]
$T_o^{\max}$	Maximum backoff time [in slots] of uniform backoff
W	Maximum number of transmission attempts
Request traffic model	
λ	Request generation rate [in requests/slot]
$\rho = \frac{\lambda}{O}$	Request load [in requests/slot] per preamble
Random access system model	
X_n	Total number of UEs transmitting a preamble in slot n
ξ_n	Number of UEs transmitting a preamble for a newly generated request in slot n
Z_n	Total number of unsuccessful UEs (with collided preambles) in slot n
f	Probability of successful preamble transmission (without collision) by an UE in a given slot
Continued on the next page	

Table 3.1: *Continued*

$\delta = (1-f)^W$	Probability that a UE request collides in all W attempts
x	Expected number of UEs transmitting a preamble (from both new and previously collided requests) in a given slot in steady state
$t = \frac{x}{O}$	Expected number of transmitting UEs per preamble
$y = e^{-t}$	Substitution to simplify notation in system balance equation
$h(t) = \frac{1-\delta}{f} = \frac{1-(1-y)^W}{y}$	Ratio of probability of success within W attempts to probability of success in one attempt
Load boundaries	
$\rho_1,\ \rho_2$	For $W \geq 9$, there is one operating point for loads ρ outside $[\rho_1,\ \rho_2]$; there are three equilibrium operating points if $\rho_1 < \rho < \rho_2$; $\rho_1,\ \rho_2$ depend only on W as per Eqns. (4.4) and (4.5)
Performance metrics	
D	Mean delay from request generation to successful contention completion [in slots]
T	Mean throughput of successful contention completions [in request/slot per preamble]

We model the initial request generation with a Poisson process with a prescribed rate λ [requests/slot]. This model corresponds to an infinite population of "virtual" UEs in the cell, whereby each UE can request a circuit with the eNB.

We define X_n to be a random variable denoting the number of UEs that are sending a preamble in a given slot n. We note that both newly generated requests

and the re-transmissions of old requests contribute to X_n as analyzed in detail in Section 3.2.

We let ξ_n be a random variable denoting the number of UEs that transmit a preamble for a newly generated request in slot n. For the considered Poisson request arrival process with rate λ, the number of newly generated requests per slot has expected value $E[\xi_n] = \lambda$.

We define f to denote the (steady-state) probability that a UE successfully sends a preamble, i.e., sends a preamble without collision, in a given slot, i.e., any slot in which the UE participates in preamble contention. Note that we model f to be indifferent to the UE's age in retransmission. The probability f is derived in Section 3.3 and simulations verifying the model accuracy are presented in Section 5.2.

We define δ to denote the (steady-state) probability that a UE request is unsuccessful in all its W transmission attempts, and as a result drops its request. A UE's attempt in a given slot is unsuccessful with probability $1 - f$, thus the probability that the UE is unsuccessful in all its W attempts can be modeled as

$$\delta = (1 - f)^W. \tag{3.1}$$

Note that $1 - \delta$ is the probability that the UE is successful in one of its (at most W) transmission attempts. The model notation is summarized in Table 3.1.

3.2 System Balance (Equilibrium) Formulation

We develop a recursion for X_n by noting that the UEs sending a preamble in slot n are either (A) UEs that have generated a new request during the preceding slot and are now sending their preamble for the first time in slot n, or (B) UEs that experienced a preamble collision in one (or several) preceding slot(s) and have not yet exhausted the maximum number of preamble transmissions W.

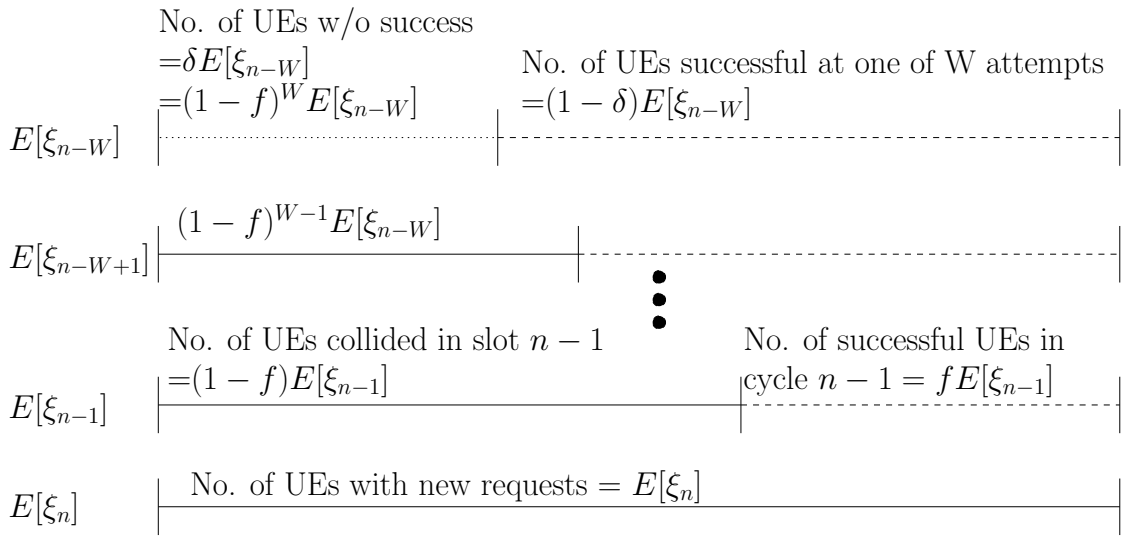

Figure 3.1: Illustration of dynamics leading to recursion (3.3) for the expected number of UEs $E[X_n]$ transmitting a preamble in slot n. UEs that have been unsuccessful in slots $n-1$ through $n-W+1$ (represented by the left solid-line portions) retransmit a preamble in slot n. Additionally considering that $E[\xi_n]$ UEs with newly generated requests transmit a preamble in slot n and that $\delta E[\xi_{n-W}]$ UEs drop after having had no success in W attempts leads to the recursion (3.2).

Note that X_{n-1} UEs sent a preamble in slot $n-1$; whereby, these UEs either had generated a new request for preamble transmission (during slot $n-2$) and this new request is transmitted for the first time in slot $n-1$, or had a preamble collision in one (or several) preceding slot(s). In steady state, an expected number of $fE[X_{n-1}]$ UEs successfully transmitted a preamble in slot $n-1$. The remaining $(1-f)E[X_{n-1}]$ UEs had a preamble collision and will re-try in slot n, provided they have not exhausted the maximum number of preamble transmissions W. In particular, those UEs that transmitted a preamble for the first time in slot $n-W$ and experienced collisions in all slots $n-W,\ n-W+1,\ \ldots, n-2$, and $n-1$ have exhausted their maximum number of preamble transmissions and drop out. Noting that an expected number of

$E[\xi_{n-W}]$ UEs had transmitted a preamble for a first time in slot $n-W$, $\delta E[\xi_{n-W}]$ UEs drop out after the preamble contention in slot $n-1$. Thus, there are an expected number of

$$E[X_n] = E[\xi_n] + (1-f)E[X_{n-1}] - \delta E[\xi_{n-W}] \tag{3.2}$$

UEs transmitting a preamble in slot n. In the illustration in Fig. 3.1, $E[X_n]$ corresponds to the sum of the left (solid line) portions of $E[\xi_{n-1}], \ldots, E[\xi_{n-W+1}]$, plus all of $E[\xi_n]$. Note that these left portions correspond to $(1-f)E[\xi_{n-1}], \ldots, (1-f)^{W-1}E[\xi_{n-W+1}]$ UEs. Thus, alternatively, we obtain the expected number of UEs transmitting a preamble in slot n as

$$E[X_n] = \sum_{t=0}^{W-1} (1-f)^t E[\xi_{n-t}]. \tag{3.3}$$

Proceeding from (3.2), we define x to denote the long-run (steady-state) expected value of X_n, noting that in steady state $x = E[X_n] = E[X_{n-1}]$. Thus,

$$x = \lambda + (1-f)x - \delta\lambda. \tag{3.4}$$

Recalling from (3.1) that $\delta = (1-f)^W$ and rearranging terms gives the steady-state system balance equation

$$\frac{x}{\lambda} = \frac{1-\delta}{f} \tag{3.5}$$

$$= \frac{1-(1-f)^W}{f}. \tag{3.6}$$

Intuitively, Eqn. (3.5) expresses that the ratio of the expected total number x of transmitting UEs to the expected number λ of UEs transmitting a newly generated request equals the ratio of probability $1-\delta$ of eventual success after at most W attempts to the probability f of success in one attempt. For very low loads, this ratio is one since transmissions are successful in the first attempt ($f \to 1$) and thus all

transmission are new requests ($x \to \lambda$). As the load increases, some transmissions fail on the first attempt and lead to an increase in the proportion of retransmissions relative to new transmissions and a commensurate increase in the probability of success after W attempts relative to the success probability in the first attempt. For very high loads, the success probability in a given slot becomes small ($f \to 0$) and the probability of success after W attempts approaches fW [as $(1-f)^W \approx 1 - fW$ in (3.6) for small f]. Correspondingly, the expected number of transmitting UEs x approaches the expected number of requests generated in W slots, i.e., λW. Thus, both sides of (3.5) approach the number of transmission attempts W.

In the following section, we evaluate the probability f of a successful transmission in a slot for the specific preamble transmission procedure in LTE-Advanced. Then, we examine the resulting system balance equation and its implications for system stability.

3.3 Probability of Successful Preamble Transmission f

Let Z_n be a random variable denoting the total number of unsuccessful UEs in slot n. Note that

$$E[Z_n] = (1-f)x. \tag{3.7}$$

We denote $\alpha_i,\ i = 1, \ldots, X_n$, for the preamble randomly selected by UE i. Note that the preambles α_i are independent random variables that are uniformly distributed in $\{1, 2, \ldots, O\}$. A collision occurs if two distinct UEs i and $j,\ j \neq i$, select the same preamble, i.e.,

$$Z_n = \sum_{i=1}^{X_n} 1_{\{\exists j \in \{1,\ldots,X_n\}, j \neq i:\ \alpha_i = \alpha_j\}}. \tag{3.8}$$

We evaluate the expectation of the number of unsuccessful UEs Z_n in Appendix A as

$$E[Z_n] \approx x\left[1 - \mathrm{e}^{-x/O}\right]. \tag{3.9}$$

Thus, from (3.7) and (3.9),

$$f = \mathrm{e}^{-x/O}. \tag{3.10}$$

3.4 Summary

We proceed by inserting (3.10) in (3.5). For improved readability we define the preamble load (request generation rate normalized by number of preambles) $\rho := \lambda/O,\ \rho \geq 0$, and the normalized expected number of UEs transmitting in a slot as $t := x/O,\ t \geq 0$. The resulting form of the balance equation is

$$\frac{t}{\rho} = \frac{1-(1-\mathrm{e}^{-t})^W}{\mathrm{e}^{-t}}. \tag{3.11}$$

While this non-linear equation has no closed-form analytical solutions, it can be solved with standard numerical methods. We show in Section 4 that depending on the values of ρ and W, (3.11) has one, two, or three solutions for t. From the numerically obtained solutions for t, we obtain the expected numbers of UEs transmitting in a slot as $x = tO$ and the probabilities of successful UE transmission through (3.10). In order to facilitate the analysis of the balance equation (3.11), we define for its right-hand side

$$h(t) := g(\mathrm{e}^{-t}) := \frac{1-(1-\mathrm{e}^{-t})^W}{\mathrm{e}^{-t}}. \tag{3.12}$$

3.5 Numerical Results

In Figs. 4.1 and 4.2 we compare the ratio of the probability of contention success after at most W attempts to the probability of success in one attempt as given by the function $h(t)$ in Eqn. (3.12), denoted by `Eq`, with simulations, denoted by `Sim`. (The lines related to t_0, t/ρ_1, and t/ρ_2 in Figs. 4.1 and 4.2 are examined in Section 4 and should be ignored for now.) The simulation model was implemented using OMNeT++ Varga (2001) libraries in C++. Statistics collection and execution

management was done using Akaroa2 Erwing *et al.* (1999). In these simulations, we held the number of transmitting UEs t at a prescribed value and observed the mean and 90 % confidence interval of the ratio h. We observe from Figs. 4.1 and 4.2 that the analytical model for the ratio h given by Eqn. (3.12) closely matches the simulation results and thus accurately models the preamble contention.

Chapter 4

ANALYSIS OF EQUILIBRIUM OPERATING POINTS

4.1 Preliminaries

The left-hand side of (3.11) is a line through the origin with slope $1/\rho$. Intersections of the line t/ρ and the function $h(t)$ defined in (3.12) specify the operating points (equilibrium points) of the system where the balance equation (3.11) is satisfied.

As shown in Appendix B, $h(t)$ is a strictly increasing function starting at $h(0) = 1$ and ending at $h(\infty) = W$. Furthermore, $h(t)$ has one inflexion point at t_0, whereby the function has one convex piece in the domain $[0, t_0]$ and one concave piece in the domain $[t_0, \infty)$.

4.2 Single Equilibrium Point for $W \leq 8$

In Appendix B, we show that $h(t)$ has precisely one convex piece (on $[0, t_0]$) and one concave piece (on $[t_0, \infty)$), which implies that the intersection of $h(t)$ and a linear function ($t \mapsto t/\rho$) can have at most three solutions. On the other hand, since $h(0) = 1$ and $h(\infty) = W$, there must be at least one solution.

We now examine the tangent of $h(t)$ at the inflexion point t_0. Note that (3.11) has three solutions for some ρ if and only if this tangent crosses the y-axis below zero, see Figs. 4.1 and 4.2, that is, if and only if

$$h(t_0) - t_0 h'(t_0) < 0. \tag{4.1}$$

This equation can readily be checked numerically for any value of W following the equations in Appendix B: calculate the unique solution $z_0 \in (0, 1)$ of $p(z) = 0$ in (B.7)

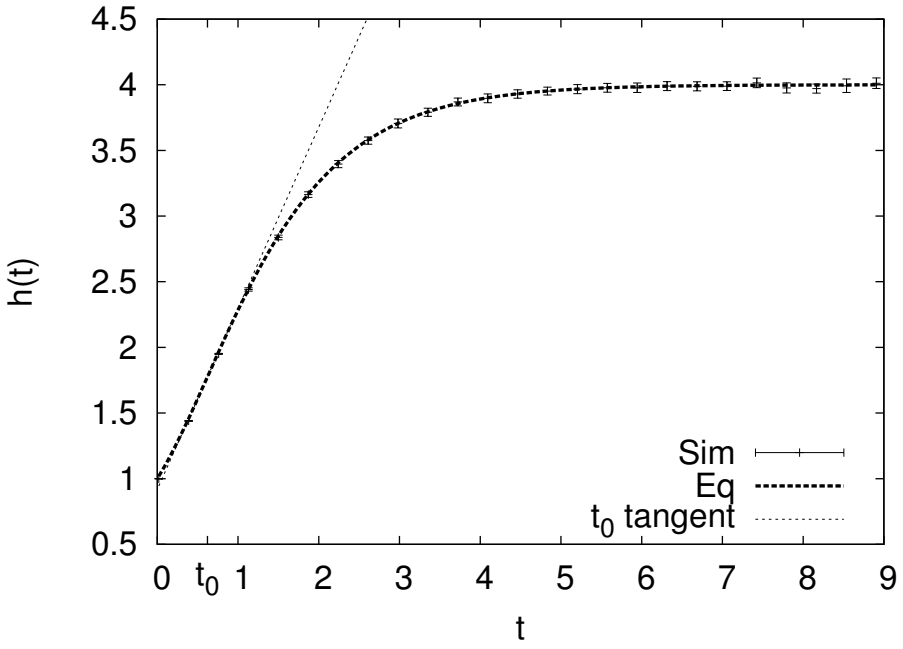

Figure 4.1: Comparison of function $h(t)$ from Eqn. (3.12) denoted by `Eq` with corresponding simulations (`Sim`) for $W = 4$. The figure also illustrates the tangent to function $h(t)$ at inflexion point t_0 for $W = 4$. Generally, for $W \leq 8$, this tangent crosses the y-axis above the origin and the line t/ρ has a single intersection with $h(t)$ for any load ρ. Thus, only a single equilibrium operating point exists.

and then t_0 via (B.8) and check condition (4.1). It turns out that condition (4.1) is violated for all $W \leq 8$ and satisfied for all $W \geq 9$. Thus, for $W \leq 8$ transmission attempts, the balance equation (3.11) has a single unique solution, i.e., the system has a single equilibrium operating point.

4.3 Multiple Equilibrium Points for $W \geq 9$.

As noted in Section 4.2, for all $W \geq 9$, Eqn. (4.1) is satisfied, i.e., the tangent on $h(t)$ at the inflexion point t_0 crosses the y-axis below zero, as illustrated in Fig 4.2 where the t_0 tangent crosses the x-axis near $t = 1$. Thus, by the piecewise convex

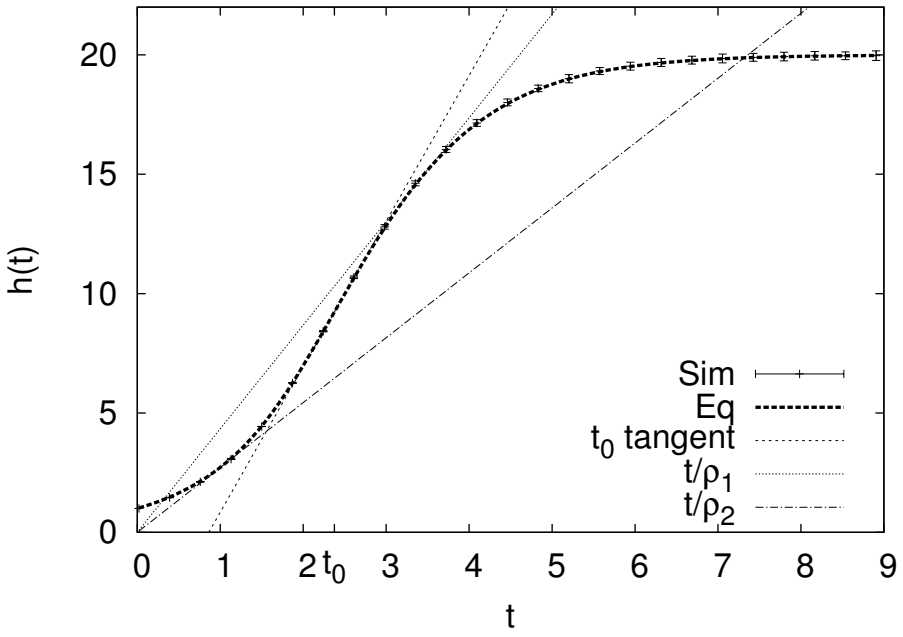

Figure 4.2: Comparison of function $h(t)$ from Eqn. (3.12) (`Eq`) with simulations (`Sim`) for $W = 20$. The figure also illustrates the tangent to function $h(t)$ at inflexion point t_0 for $W = 20$. Generally, for $W \geq 9$, this tangent crosses the y-axis below the origin and the line t/ρ has a single intersection with $h(t)$ for load $\rho < \rho_1$, three intersections for $\rho_1 < \rho < \rho_2$, and one intersection for $\rho > \rho_2$. Thus, a single, or up to three equilibrium operating points exist and are specified by the intersection(s) of t/ρ with $h(t)$.

and concave property of $h(t)$ shown in Appendix B, there are two tangents on $h(t)$ crossing the origin, illustrated by t/ρ_1 and t/ρ_2 in Figure 4.2. These two tangents satisfy

$$h'(t) = \frac{1}{\rho} \qquad \text{and} \qquad h(t) = \frac{1}{\rho}t, \tag{4.2}$$

for *some* ρ and t. Substituting $y = \mathrm{e}^{-t}$ these two equations become

$$g'(y)(-y) = \frac{1}{\rho} = \frac{g(y)}{t}, \tag{4.3}$$

which is

$$t\left[1-(1-y)^W - yW(1-y)^{W-1}\right] = 1-(1-y)^W. \tag{4.4}$$

Solving (4.4) for t gives the solutions t_1, t_2, which are those t values where the tangents touch the function $h(t)$. The corresponding slopes $\rho_i,\ i=1,2$, of the tangents are obtained from (4.2) as $\rho_i = t_i/h(t_i)$:

$$\rho_i = \frac{t_i \mathrm{e}^{-t_i}}{1-(1-\mathrm{e}^{-t_i})^W}. \tag{4.5}$$

Note that ρ_1 and ρ_2 specify the boundaries of the domain of loads ρ where multiple equilibrium operating points exist. In summary, through the analysis in Section 4.2 and this section, based on the properties of the function $h(t)$ shown in Appendix B, we have proven the following theorem.

Theorem 1. *For $W \geq 9$ transmission attempts, there are load boundaries $\rho_1, \rho_2,\ 0 < \rho_1 < \rho_2 < \infty$, that only depend on W according to Eqns. (4.4) and (4.5) such that*

- *for $\rho < \rho_1$ the random access system has a single unique equilibrium point;*
- *for $\rho = \rho_1$ the random access system has exactly two equilibrium points;*
- *for $\rho_1 < \rho < \rho_2$ the random access system has exactly three equilibrium points;*
- *for $\rho = \rho_2$ the random access system has exactly two equilibrium points;*
- *for $\rho > \rho_2$ the random access system has a single unique equilibrium point.*

The one, two, or three equilibrium operating points for a prescribed load ρ are given by the solutions for t of the balance equation (3.11). If $\rho < \rho_1$ or $\rho > \rho_2$, then, by Theorem 1, the balance equation (3.11) gives one solution for t; whereas for other ρ values, Theorem 1 states that there are two or three solutions for t. For a given solution t of the balance equation, the corresponding equilibrium operating point in

terms of the total expected number x of UEs transmitting a preamble in a slot is given as $x = tO$.

4.3.1 Asymptotics for Large Number of Transmission Attempts W

We proceed to examine the asymptotics for the load boundaries ρ_1 and ρ_2 as the transmission attempt limit W becomes large. For large W one solution of (4.4) is $t_2 \sim 1$, giving

$$\rho_2 \sim \mathrm{e}^{-1}, \tag{4.6}$$

which corresponds to the case $Wy \to \infty$. For the case $Wy \to 0$, we show in Appendix C that

$$\rho_1 \sim \frac{\log \frac{W-1}{2\mathrm{e}^{-1}} + \log\log \frac{W-1}{2\mathrm{e}^{-1}} - 1}{W} \tag{4.7}$$

$$\sim \frac{\log W}{W}. \tag{4.8}$$

4.3.2 Numerical Results

In Fig. 4.3, we plot the load boundaries ρ_1 and ρ_2 as a function of the transmission attempt limit W. We observe that the exact upper boundary ρ_2 from Eqns. (4.4) and (4.5) closely approaches the asymptotic boundary $\rho_2 \sim 1/e$ from (4.6) even for relatively small W values; for $W \geq 15$, the exact ρ_2 essentially coincides with the asymptotic boundary $1/e$. On the other hand, the exact lower load boundary ρ_1 from Eqns. (4.4) and (4.5) approaches the asymptotics given by (4.7) and (4.8) relatively slowly, with the detailed asymptotic (4.7) giving a somewhat better approximation for moderately large W values than the simplified asymptotic (4.8). Thus, for LTE system evaluations, the upper load boundary ρ_2 can be readily approximated by the asymptotic $1/e$ for moderately large W. For the lower boundary ρ_1, the exact analytical characterization through Eqns. (4.4) and (4.5) should be used since the

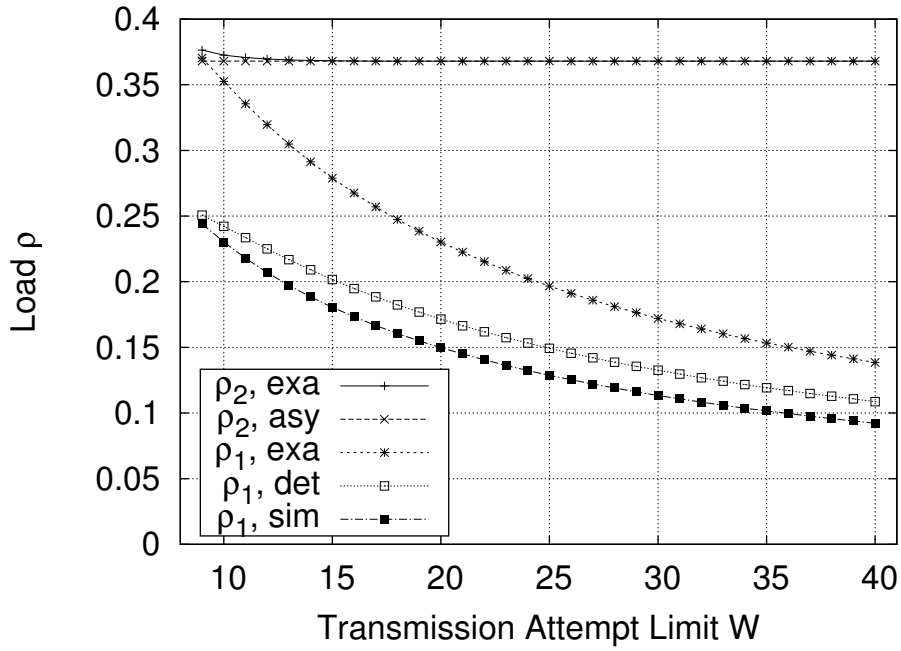

Figure 4.3: Lower boundary ρ_1 and upper boundary ρ_2 of load range with multiple equilibrium operating points as a function of number W of permitted transmission attempts. Exact results are obtained with Eqns. (4.4) and (4.5), while detailed (`det`) and simplified (`sim`) asymptotics for ρ_1 are from Eqns. (4.7) and (4.8), respectively, and the asymptotic for ρ_2 is from (4.6).

asymptotics overestimate the load range with multiple equilibrium points, especially for small or moderate W values.

We also observe from Fig. 4.3 that for $W \geq 9$, the width $\rho_2 - \rho_1$ of the load range with multiple equilibrium operating points widens considerably with increasing W, e.g., from $\rho_2 - \rho_1 = 0.1$ for $W = 15$ to 0.2 for $W = 30$. For $W = 200$, the maximum transmission attempt limit in LTE-Advanced, which is not included in Fig. 4.3 to allow a detailed view of the small W values, ρ_1 drops to 0.0376. That is, multiple operating points exist for loads between 0.0376 and $1/e$ for $W = 200$.

Chapter 5

THROUGHPUT-DELAY ANALYSIS

In this section we examine the throughput and delay metrics defined in Section 2.2. New requests are generated by the UEs with rate λ [requests/slot], which normalized by the number of preambles O is $\rho = \lambda/O$, and a given UE is successful within the permitted W transmission attempts with probability $1 - \delta$. Thus, the mean throughput of successful requests [requests/slot per preamble] is

$$T = \rho(1 - \delta). \tag{5.1}$$

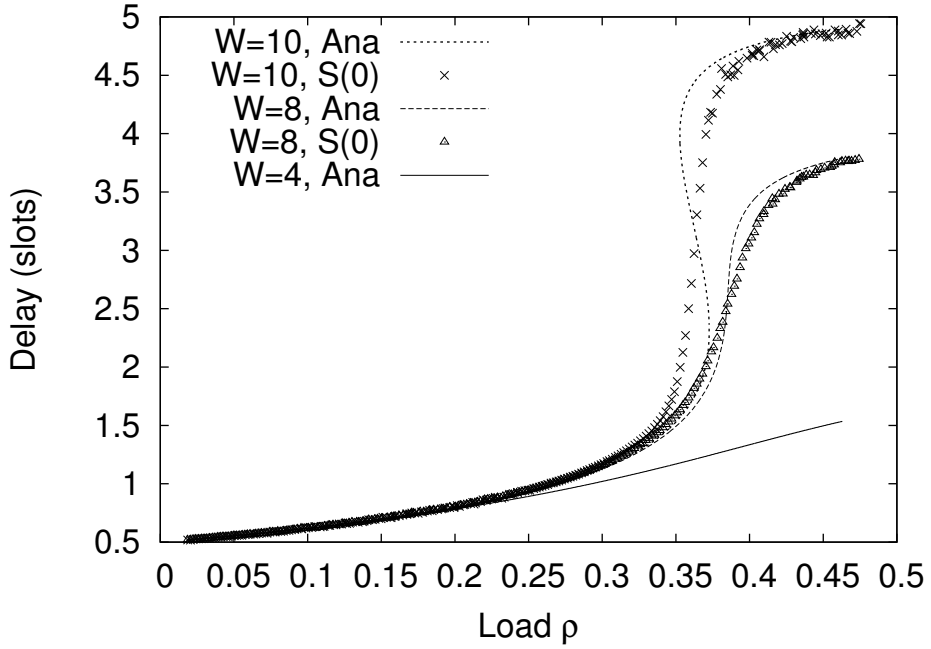

Figure 5.1: Delay Plot for $W = 4, 8, 10$, $T_o^{\max} = 0$

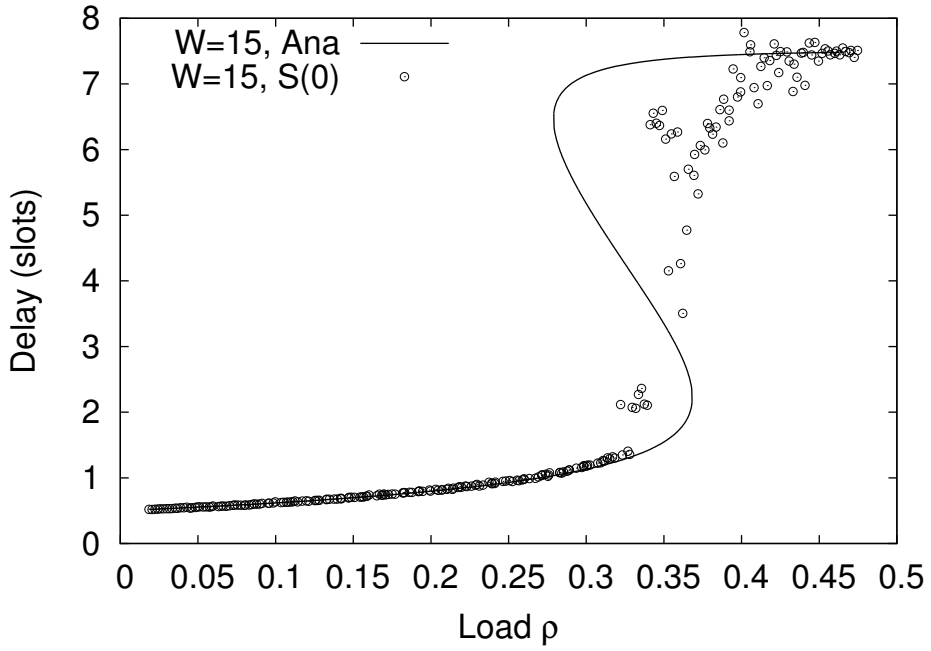

Figure 5.2: Delay Plot for $W = 15$, $T_o^{\max} = 0$

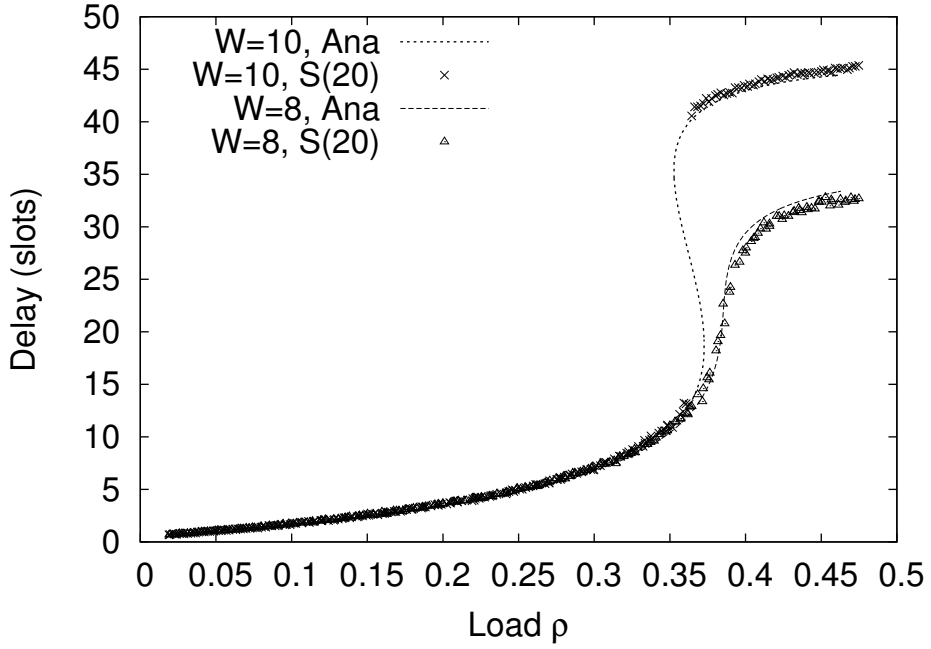

Figure 5.3: Delay Plot for $W = 4, 8, 10$, $T_o^{\max} = 20$

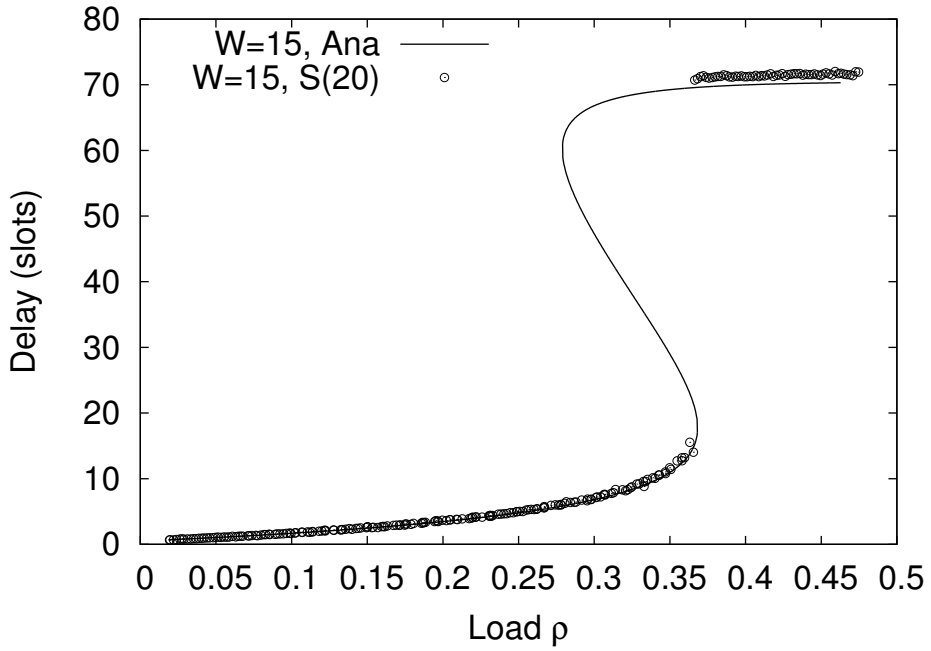

Figure 5.4: Delay Plot for $W = 15$, $T_o^{\max} = 20$

5.1 Delay Analysis

Recall that a preamble transmission is successful with probability f. Thus, the probability of exactly $n,\ n = 0, 1, \ldots, W-1$, collisions before a success, can be modeled as $(1-f)^n f$. Hence, the probability of a UE to experience n collisions, given that it sends (i.e., experiences any number $k,\ k = 0, 1\ \ldots, W-1$ collisions) is

$$\frac{(1-f)^n f}{\sum_{k=0}^{W-1}(1-f)^k f}. \tag{5.2}$$

Each collision increases the delay by one slot of duration T_s. Thus, the expected delay due to collisions is

$$D_c = T_s \cdot \sum_{n=0}^{W-1} n \cdot \frac{(1-f)^n f}{\sum_{k=0}^{W-1}(1-f)^k f}. \tag{5.3}$$

We model the delay from the instant of request generation to the next time slot boundary (backward recurrence time) Heyman and Sobel (2003) with the additive

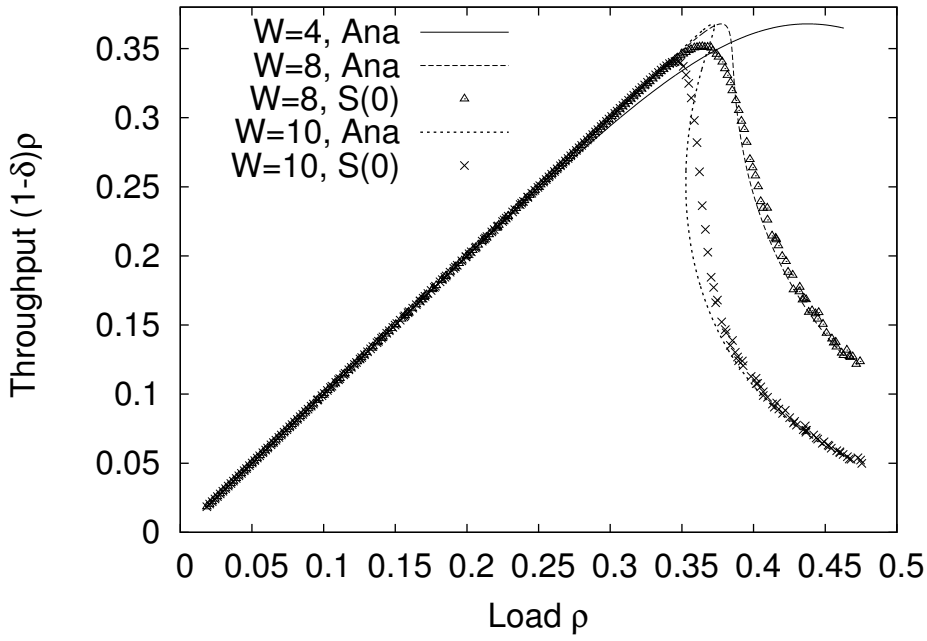

Figure 5.5: Throughput as a Function of Load for $W = 4, 8, 10$, $T_o^{\max} = 0$

constant $T_s/2$. We further employ the summation identity for $0 < y < 1$,

$$\sum_{k=0}^{W-1} k \cdot y^k = y \cdot \frac{1 + (W-1)y^W - Wy^{W-1}}{(1-y)^2}. \tag{5.4}$$

Hence, for the preamble transmission success probability f obtained through Eqns. (3.11) and (3.10),

$$\begin{aligned} D &= T_s \left(\frac{1}{f} - 1\right) \frac{1 + (W-1)(1-f)^W - W(1-f)^{W-1}}{1 - (1-f)^W} \\ &\quad + \frac{T_s}{2}. \end{aligned} \tag{5.5}$$

With uniform backoff with $T_o^{\max}$, as outlined in Appendix D, the average delay caused by a collision increases from T_s to $\left(1 + \frac{T_o^{\max}}{2}\right) T_s$, i.e., T_s has to be replaced by $\left(1 + \frac{T_o^{\max}}{2}\right) T_s$ in the first summand of Eqn. (5.5).

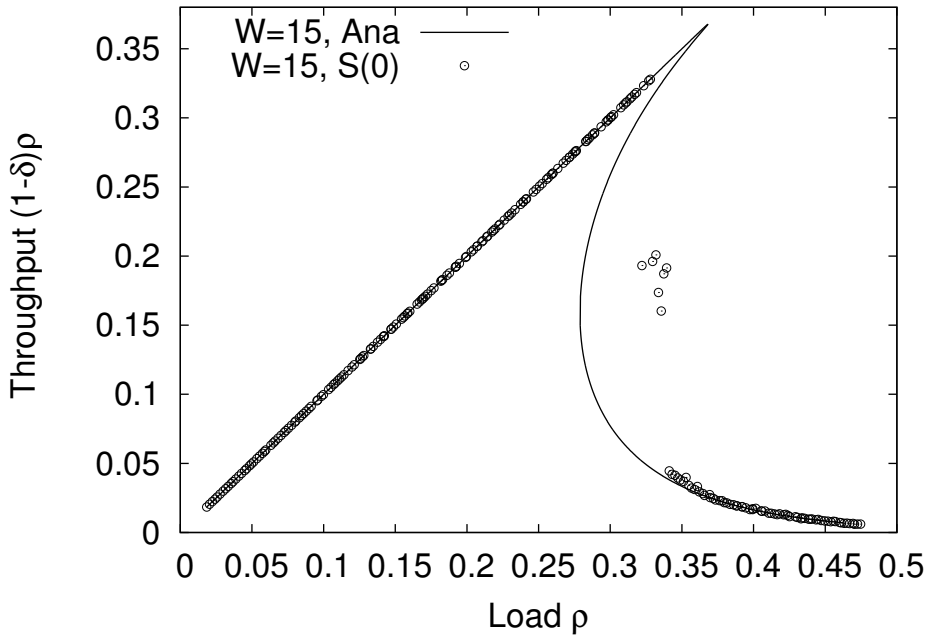

Figure 5.6: Throughput as a Function of Load for $W = 15$, $T_o^{\max} = 0$

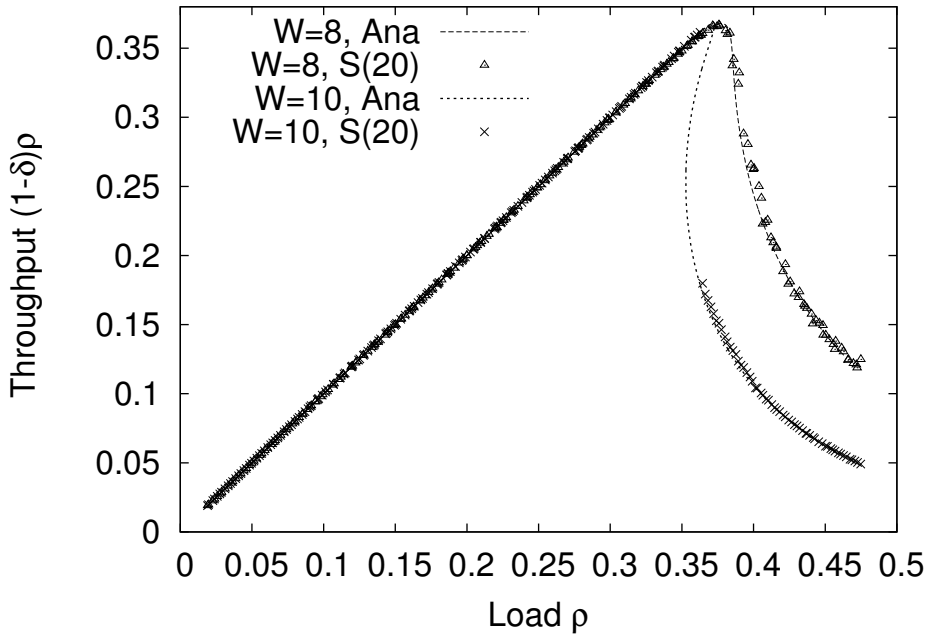

Figure 5.7: Throughput as a Function of Load for $W = 4, 8, 10$, $T_o^{\max} = 20$

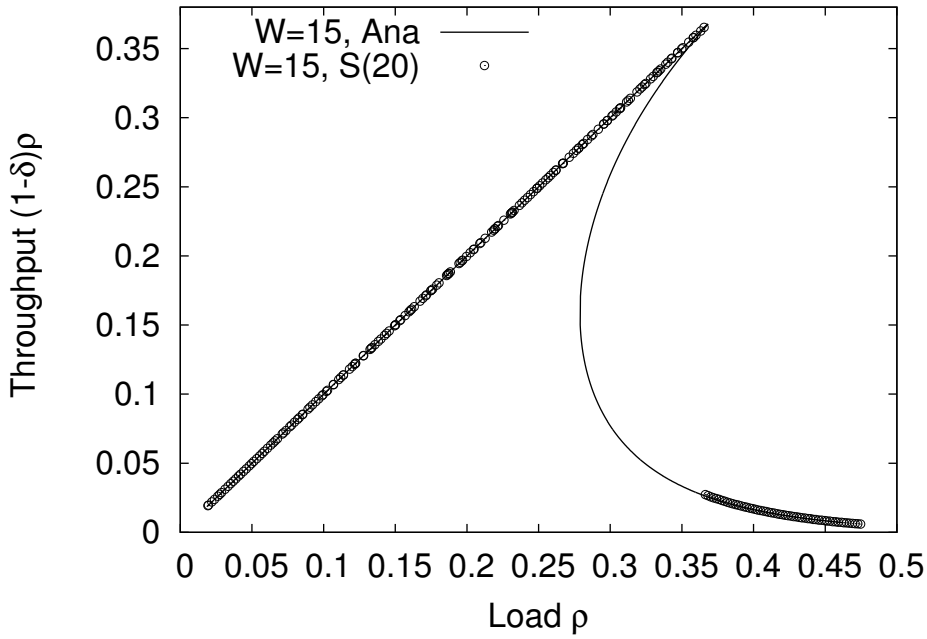

Figure 5.8: Throughput as a Function of Load for $W = 15$, $T_o^{\max} = 20$

5.2 Evaluation Results

In Figs. 5.1- 5.4, Mean delay D [in slots] of successful requests, as a function of load ρ [in new requests per slot per preamble] for a range W of transmission attempts. The load boundaries for multiple operating points are $\rho_1 = 0.353$, $\rho_2 = 0.373$ for $W = 10$; while for $W = 15$ they are $\rho_1 = 0.279$, $\rho_2 = 0.368$. For $W \geq 9$ for $\rho_1 < \rho < \rho_2$, we plot the delay values corresponding to the three equilibrium points. S($T_o^{\max}$) denotes simulation results for a given $T_o^{\max}$ value. In Figs. 5.5- 5.8, throughput $T = \rho(1 - \delta)$ (in successful requests per slot per preamble) as a function of load ρ [in new requests per slot per preamble] for a range W of transmission attempts. The load boundaries for multiple operating points for $W = 10$ are $\rho_1 = 0.353$, $\rho_2 = 0.373$, while the boundaries for $W = 15$ are $\rho_1 = 0.279$, $\rho_2 = 0.368$. Simulation results for $T_o^{\max} = 0$ are denoted by S(0), while S(20) denotes simulation results for $T_o^{\max} = 20$ slots.

In Figs. 5.1- 5.4 and 5.5- 5.8, we plot the mean delay [in slots] experienced by a successful request and the throughput $T = \rho(1 - \delta)$ of successful requests per slot per preamble as a function of the load ρ. For a relatively small transmission attempt limit $W = 4$ without backoff, i.e., $T_o^{\max} = 0$, we observe that the successful requests experience a low delay of less than $D = 1.17$ slots at load $\rho = 0.35$, see Fig. 5.1, while Fig. 5.5 indicates that the throughput is $T = 0.334$. For $W = 8$, the highest transmission attempt limit that guarantees a single operating point for all loads ρ, without backoff, the delays are moderately higher with $D = 1.49$ slots while the throughput is very slightly higher $T = 0.348$ at load $\rho = 0.35$ compared to $W = 4$. At this moderately high load level, having the unsuccessful transmissions attempt more re-transmissions is beneficial in that it slightly increases the throughput, while only moderately increasing the delay.

As the load increases beyond 0.37, we observe from Figs. 5.1- 5.4 and 5.5- 5.8 that the contention with $W = 8$ leads to rapidly increasing delays while the throughput drops sharply. In contrast, for $W = 4$, the system degrades more gracefully, with the mean delay starting to level out around $D = 1.8$ slots for load levels of $\rho = 0.6$ (i.e., outside the plotted range) while the throughput drops to $T = 0.1$ for a load of $\rho = 0.925$. For high loads, the success probability f becomes small ($f \to 0$), and the mean delay due to collisions D_c given by (5.3) correspondingly approaches $\lim_{f\to 0} D_c = (W-1)T_s/2$; adding the mean waiting time $T_s/2$ from request generation to next slot boundary gives a maximum expected delay of $WT_s/2$. Intuitively, for small success probability f, almost all requests undergo W transmission attempts (and nearly all requests drop after their Wth attempt). The few requests that are successful, experience their success after a number of attempts that is approximately uniformly distributed over $[1,\ W]$ (corresponding to $[0, W-1]$ experienced collisions).

The sharp throughput drop at very high loads, e.g., load $\rho = 0.45$, for $W = 8$ compared to $W = 4$ observed in Fig. 5.5 is mainly due to the exacerbation of the overload conditions by the higher number of transmission attempts with $W = 8$. Specifically, for load $\rho = 0.45$, we found from the numerical evaluation of the analysis in Section 3 that for $W = 4$, there are on average $x = 57.5$ preambles transmitted per slot with a steady-state success probability $f = 0.345$ resulting in a throughput of $T = 0.367$. In contrast, with $W = 8$ transmission attempts, there are $x = 165.4$ transmitted preambles with success probability $f = 0.047$ and throughput $T = 0.143$. At this particular high load level, the doubled number of transmission attempts with $W = 8$ roughly triples the number of contenting transmissions in each slot, which reduces the success probability to roughly one seventh of the success probability for $W = 4$. The higher number of contending requests with $W = 8$ somewhat compensates for this dramatically lower success probability, but the throughput with $W = 8$ is still less than half compared to $W = 4$.

For $T_o^{\max} = 20$ slots for $W = 4$ and 8, we observe a close to ten-fold increase of the mean delays in Figs. 5.3, and 5.4 compared to Figs. 5.1, and 5.2, while the throughput remains essentially unchanged in Figs. 5.5- 5.8. The $T_o^{\max} = 20$ slot timeout increases the delay introduced by a collision from one slot to on average $(1+T_o^{\max}/2) = 11$ slots [see discussion immediately following Eqn. (5.5)]. On the other hand, as outlined in Appendix D, a static timeout $T_o^{\max}$ does not affect the steady-state drop probability δ and thus preserves the steady-state throughput.

In the simulations for this section, the Poisson generation rate λ of new requests was incremented with a step size of 0.1, corresponding to a step size for $\rho = \lambda/O$ of 0.00185 for $O = 54$ preambles, which is a typical operational O value for LTE-Advanced networks. Each point represents the mean of a simulation run long enough such hat the 90 % confidence intervals of both performance metrics are less than 10 %

Table 5.1: Comparison of finite UE simulation model (N UEs, idle UE generates new request with probability p in a slot) with simulation and analysis infinite UE model ($N \to \infty$, Poisson request generation rate $\lambda = Np$ req./slot) for $W = 4$

	$\rho = 0.30$		$\rho = 0.35$		$\rho = 0.40$		$\rho = 0.45$	
N	D	T	D	T	D	T	D	T
100 (sim)	0.96	0.30	1.05	0.34	1.15	0.38	1.22	0.43
1000 (sim)	1.02	0.29	1.15	0.33	1.30	0.36	1.44	0.38
10000 (sim)	1.02	0.29	1.17	0.33	1.33	0.36	1.47	0.37
∞ (sim)	1.03	0.29	1.16	0.33	1.33	0.36	1.48	0.36
∞ (ana)	1.02	0.29	1.17	0.33	1.33	0.36	1.50	0.37

of their corresponding sample means. We observe from Figs. 5.1- 5.4, and 5.5- 5.8 that the analytical model closely approximates the simulation results for $W = 8$. The simulation results for $W = 4$, which match very closely to the plotted analysis results, were omitted to avoid clutter.

In Table 5.1, we compare the Poisson request generation model with rate λ, which represents an infinite UE population, with a corresponding simulation model for a finite number of N UEs, whereby each of the N UEs generates a new request only when it is idle with probability p in a slot. We observe that compared to the Poisson model, the finite UE model gives smaller delays D and higher throughputs T as the number N of UEs decreases and the load ρ increases. For smaller N and higher ρ relatively more of the UEs are backlogged with a collided request that is being retransmitted, thus reducing the effective request generation rate. The Poisson model represents a worst-case request generation model in that the generation rate of new requests stays constant, irrespective of the number of backlogged requests.

Turning to the results for $W = 10$ and 15, we observe that the analytical model and simulations closely match for loads outside the $(\rho_1,\ \rho_2)$ range. For loads inside the ρ_1 to ρ_2 range, a plotted simulation point for a given load gives the mean of the respective performance metric (delay or throughput) experienced in a very long simulation run. That is, the simulation point reflects the multiple operating points and the delays and throughputs experienced at these operating points weighted by the sojourn times at these operating points.

We turn to the effect of backoff for $W = 15$. We observe from the simulation results that the uniform backoff with $T_o^{\max} = 20$ slots helps to achieve essentially zero drop probability and consequently throughput equal to the traffic load for loads up to approximately 0.361 in Fig. 5.8, whereas without backoff ($T_o^{\max} = 0$), drop probabilities of close to zero occur only for loads up to around 0.320 in Fig. 5.6. (In Figs. 5.2 and 5.4, these load values correspond to the loads where the delays "jump up" from the respective lower segments of the S-shaped delay curves.) The backoff uniformly redistributes the collided UEs from a given slot that have not exhausted their W attempts over the following $T_o^{\max} + 1$ slots. This redistribution effectively "smoothes" the number of UEs rejoining the contention and lowers the probability of the system entering the operating points with higher drop probabilities and delays. Note that this smoothing effect comes at the expense of greatly increased mean delay. For instance, for a load of $\rho = 0.31$, the mean delay is 7.77 slots with $T_o^{\max} = 20$ slots compared to a mean delay of 1.29 slots with $T_o^{\max} = 0$. We note that the results in Figs. 5.5- 5.8 indicating relatively small benefits of uniform backoff for contention with a typical number of $O = 54$ preambles are complementary to the results displayed in (Seo and Leung, 2011a, Fig. 2), which considers the special case of $O = 1$ preamble. The results in (Seo and Leung, 2011a, Fig. 2) indicate significant throughput increases albeit at the expense of substantially increased delay

due to uniform backoff. With uniform backoff, the collided UEs from a given slot are effectively randomly redistributed to the $O \cdot (T_o^{\max} + 1)$ preambles occurring over the next $T_o^{\max} + 1$ slots. Thus, for a very small number O of preambles, the uniform backoff can help in reducing future collisions, thus increasing throughput. For the typical, moderately large numbers on the order of tens of preambles, the effect of backoff diminishes, as observed in Figs. 5.1, 5.2, 5.3, and 5.4.

Turning to the comparison of the performance for $W = 10$ and 15 with the $W = 4$ and 8 values without multiple operating points, we observe from Fig. 5.1 that the mean delay for $W = 10$ at its $\rho_1 = 0.353$ load is approximately 1.88 slots compared to 1.61 slots for $W = 8$ at the 0.353 load. The mean delays for $W = 15$ and $W = 8$ for $T_o^{\max} = 0$ at $\rho_1 = 0.279$ (for $W = 15$) are essentially the same 1.03 slots. Notice also that for $W = 10$ and 15, the throughput is close to the request arrival rate for loads $\rho < \rho_1$. We furthermore observe that for $W = 10$ and 15 with $T_o^{\max} = 0$, the performance can degrade quite considerably for loads $\rho > \rho_1$, especially toward the middle and upper end of the ρ_1 to ρ_2 range.

Chapter 6

EQUILIBRIUM APPROACH, LTE-A PARAMETERS, AND POISSONIZATION

6.1 Introduction

Machine-to-Machine (M2M) communications or Machine Type Communications (MTC) is a research area of immense interest Lawton (2004); Wu *et al.* (2011); Fan *et al.* (2014); Kim *et al.* (2014); Marsch *et al.* (2012). As an enabler for the futuristic Internet of Things (IoT), the performance and efficiency of MTC is of high importance. LTE/LTE-A mobile standards by 3GPP consortium is a viable choice for MTC infrastructure deployment due to wide deployment of mobile communication technologies and ongoing roll outs of LTE across the world Lien *et al.* (2011b); Taleb and Kunz (2012).

MTC differs from conventional human-to-human (H2H) communication in several ways. A few important differences being the relatively low amount of data payload and high frequency of calls. Essentially, a client machine will repeatedly access the server to inform of the current status or to query the server for updated status. These status messages are of small sizes. In the context of LTE/LTE-A, the client machines are known as User Equipments (UEs). UEs gain access to the network through evolved-NodeB (eNodeB). Due to relatively long idle times between successive transmissions by devices, it is prudent that they detach themselves from the server and wait until the data is needed to be sent again. This allows for significant statistical multiplexing, implying support for a large number of UEs. Various features of MTC, in context of 3GPP, are described in Taleb and Kunz (2012).

Since most UEs have to access the network for a very short duration. The success in gaining admission to the channel becomes the bottleneck. In case of LTE/LTE-A, this bottleneck exists at accessing the eNodeB through random access channel (RACH). The medium access procedure for LTE/LTE-A RACH access is described in 36.321 (2011).

Congestion and overload control are some of the challenges posed during random access by MTC based on LTE/LTE-A networks Taleb and Kunz (2012); Laya *et al.* (2014). Some of the specific challenges arise when a large number of UEs try to transmit in a short duration e.g. after a power outage or when a large set of sensors responds to a common event Laya *et al.* (2014). In these cases, number of UEs that will try to communicate is usually unknown and can cause brief periods of contention. If the number of UEs supported by the eNodeB is large, which is desirable to keep infrastructure costs low, then the brief outages can result in longer overloaded states Tyagi *et al.* (2013, 2012, 2014) during stable operation, causing the network to be inaccessible to the UEs for relatively large durations.

In this chapter, keeping the importance of MTC, and availability and suitability of LTE/LTE-A for the MTC in view, we consider the use of LTE-A as standard used for providing wireless access to UEs. LTE-A uses a slotted-ALOHA like random access mechanism for UEs to gain access to eNodeB. Essentially, a UE goes through a self barring check before attempting a connection request and goes through contention using a random preamble. The eNodeB listens for RA requests during a slot. If multiple UEs use the same preamble in the same slot then due to collision, the UEs have to re-attempt. A UE can re-attempt after a random wait. A UE is allowed to make a limited number of attempts, before it should stop trying and drop the request. The eNodeB routinely broadcasts the parameters related to RA via System

Information Block Type 2 (SIB2) or via Media Access Control (MAC) Protocol Data Unit (PDU). The RA procedure is describe in detail in Section 6.3.

Back-off procedure used by LTE/LTE-A, seen in the context of MTC, is the main focus of this paper. Each collision results in a back-off cycle, and introduces a delay. LTE/LTE-A uses uniform back-off after each collision. We look at the impact of back-off in generic terms and ascertain how exactly it impacts the RA procedure. Specifically, we explore how analysis of back-off can provide guidance on number of UEs that can be supported, under various traffic models. We establish how the performance of these models converge through particular usage of back-off interval. A detailed understanding of back-off will help design MTC systems where, a trade-off of various back-off parameters impacting delay and success probabilities determines the number of UEs supported by eNodeB.

The RA procedure in LTE/LTE-A can be understood in terms of three stochastic processes, the arrival process, the departure process and the backlog process. The backlog process arises as a result of collisions from the arrival process and is thus positively correlated to the arrival process. We approach the system analysis with a focus on the backlog. Backlog is an important consideration due to the following reasons:

- Backlog effectively acts as the buffer of the system in a queueing model and hence, determines the capacity of the system, in terms of RA traffic supported by eNodeB.
- The eNodeB can manipulate the Backlog through the periodic broadcast via SIB2 or MAC PDU.
- Backlog directly impacts the success probabilities of incoming requests.
- Backlog positively correlates with the expected back-off duration.

We present a model which is based on basic steady state features of a generic process. In particular, we use a Poisson arrival process for mathematical tractability, and because, as we shall prove, other models will converge to a Poisson arrival process model under suitable conditions. This convergence of different arrival models to one of Poisson arrival process is very helpful in gaining a deterministic understanding of expected success and drop probabilities for a particular system.

We will discuss the related work in Section 6.2. The RA model for LTE/LTE-A and the backlog process is discussed in Section 6.3. In Section 6.4, we will use steady state analysis to establish different system models, and discuss the impact of back-off. Section 6.5 will detail the numerical work and simulation data to support the analytical conclusions. We shall conclude the paper in Section 6.6.

6.2 Related Work

M2M service technologies and features of M2M traffic are presented in Kim *et al.* (2014); Lien *et al.* (2011b); Hasan *et al.* (2013). Various classifications of M2M platforms and the associated requirements and functionalities are discussed. Specific details on M2M, in context of healthcare and smart grid are presented in Fan *et al.* (2014). Impact of retransmission Limits on RA in LTE/LTE-A in context of healthcare is discussed in Tyagi *et al.* (2013).

ALOHA was introduced by Abramson (1970). It has been widely studied since then. Several works have examined the capacity and delay performance of these systems without retransmission limits Carleial and Hellman (1975); Ferguson (1975); Ghez *et al.* (1988); Jenq (1980b); Kamal and Mahmoud (1979); Kleinrock and Lam (1975b); Murali and Hughes (1998b, 1997); Naware *et al.* (2005b) and infinite number of users. The throughput for slotted ALOHA is found to be e^{-1}. Typically Markovian models have been used to analyze the systems in these works. These models do not

focus on the backlog. Carleial and Hellman (1975) indicated that average based steady state measures are poor indicators of system performance. Since our model proves that for large backlogs Poisson arrival process is valid, our model allows to estimate second order statistical measures, such as variance (which is equal to mean for Poisson arrival model) and thus hopes to provide better indication of system performance and greater insight into system dynamics.

Analytical models for slotted ALOHA system, with a limit on the number of attempts, have been developed in Lüders and Haferbeck (1994); Sarker and Halme (1997, 1998, 2000b). However, these models examine only the steady state expectations and do not consider the backlog and its impact. In addition to using a more basic equilibrium based approach, which derives from the existence of steady state itself, we prove that backlog fundamentally affects the arrival process and dominates it as supported user base increases. Additionally, several works have examined the bistability in slotted ALOHA systems predicted by analytical models Carleial and Hellman (1975); Ferguson (1975). Recent works have also estimated limits for the bistability Sakakibara *et al.* (2000, 2003). Our model is applicable to low rates of arrival and is not impacted by the bistability problem. Specifically, for numerical work we restrict ourselves to a maximum of 7 retransmissions and thus do not bistable zones Tyagi *et al.* (2013).

Some recent studies, such as Seo and Leung (2012, 2011c,b, 2010); Rivero-Angeles *et al.* (2006), have examined the second moments of delay in slotted ALOHA system in LTE, impact of retransmission limits, and compared various back-off strategies, using Markov models. This is different from our analysis, as we focus on the backlog as a fundamental influence. Understanding of backlog provides greater insight into the impact of back-off. Also, the equilibrium based approach depends only on existence of steady state and gives fundamental insights into the system. Retransmission back-

off policies are analyzed in Joseph and Raychaudhuri (1988), for slotted ALOHA channels.

Optimization of back-off interval for random access is studied in Haas and Deng (2003). However, the proposed scheme is based on channel state information. Our models do not assume that channel state information is available and hence are applicable to wider range. A class dependent back-off schemes for LTE/LTE-A MTC has been presented in Jian *et al.* (2013). We do not consider multiple classes for our analysis, and consider that all traffic has equal priority. However, this is not a limitation, since the equilibrium based approach used by us is easily extensible to multiple traffic categories with summations for different categories replacing individual cases. Detailed delay analysis for OFDMA-ALOHA is presented in Mutairi *et al.* (2013).

Impact of different parameters such as access class barring (ac-BarringTime), separation of resources, back-off only, etc. are presented in Kouzayha *et al.* (2013); Yang *et al.* (2012); Amirijoo *et al.* (2009). These are very preliminary studies and do not go into details as in current paper. Automatic configuration of RACH parameters based on a desired delay performance in LTE is considered in Choi *et al.* (2011).

Throughput analysis for M2M RA in LTE is done in Lee *et al.* (2011). RACH collision probability for MTC has been discussed in Cheng *et al.* (2012). Two different interpretations of collision probabilities are presented. In this chapter, we consider collision probability from the perspective of attempts in a slot. In Gerasimenko *et al.* (2012), energy and delay analysis is done for MTC under LTE-A. Overload conditions are generally considered. Prioritized RA in LTE for MTC is dicussed in Lin *et al.* (2014). The focus is on dynamic ac-BarringTime to provide Quality of Service (QoS). Adaptive Traffic Load Slotted ALOHA (ATL-S-ALOHA) based RA control is proposed in Li *et al.* (2013). An information theoretic analysis of RA for multiple user case is presented in Minero *et al.* (2012).

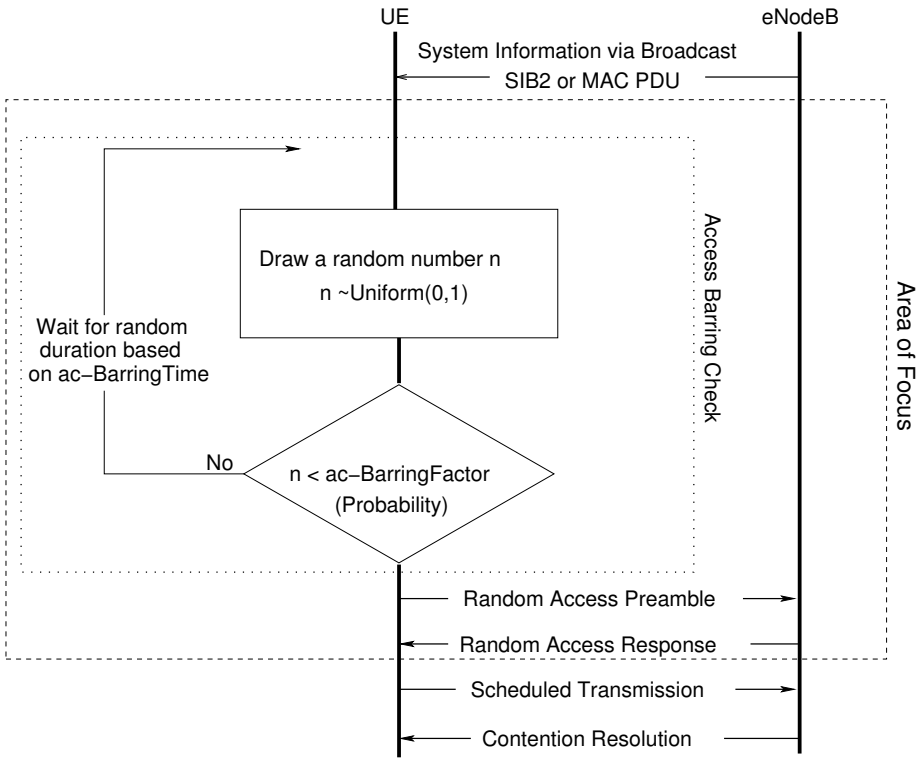

Figure 6.1: Illustration of Random Access (RA) procedure in LTE/LTE-A.

Limits on variance of unimodal distributions have been well established for years. A proof on maximum variance for unimodal distributions is presented in Muilwijk (1966). Variance limits on unimodal distributions is also discussed in Jr. *et al.* (1985). Based on these limits, we will consider a process in which arrivals in a slot are Bernoulli distributed. A variable rate Poisson arrival process based simulation methodology is proposed for MTC for wireless networks in Paiva *et al.* (2011). We use a 2-state Markov Modulated Poisson Process in our study.

6.3 RA Procedure

In this section, we will discuss the RA procedure in LTE/LTE-A systems. A pictorial representation of the RA procedure is given in Figure 6.1 Lee *et al.* (2011); Tyagi *et al.* (2013). Each UE, when it has data to send, starts the RA procedure by

undergoing a local barring check. The parameter for this check ac-BarringFactor is broadcast by eNodeB via SIB2 or MAC PDU. The UE generates a uniform random number and clears local barring if probability of generated number is greater than the probability specified by ac-BarringFactor. If the generated number is less than that specified by ac-BarringFactor, then the UE has to wait for a long duration dependent on ac-BarringTime. ac-BarringTime is also periodically broadcast by eNodeB.

After clearing local barring check, the UE uniformly selects a random preamble from a set of O preambles available for RA. It then transmits the RA request using the chosen preamble.

An eNodeB periodically transmits the RA parameters using SIB2 or MAC PDU. The minimum duration for transmitting updated parameters is 10 ms. The minimum period for which an eNodeB listens for UE requests before sending a response is 0.5 ms, which is equivalent to a sub-slot, a slot being of 1 ms duration. For the rest of this paper, we will refer to the sub-slot as a slot, considering this as the standard listening period for the eNodeB.

During a slot, if more than one UE transmits using the same preamble, then a collision is considered to have occurred on that preamble. Although, it may be possible to retrieve some request information in this case due to Capture Effect, we model a worst case scenario and consider that all requests which used the particular preamble are irrecoverable. Requests which used a different preamble are not affected due to collisions on another preamble.

In case of a collision, a UE, on failing to receive a response from eNodeB, initiates a back-off. The UE uniformly generates a random back-off interval, with the maximum duration of back-off interval being $T_o^{\max}$. When in this stage, a UE is considered to be backlogged.

At the end of back-off, the UE will again undergo local barring and on clearing local barring will re-attempt RA request using another randomly chosen preamble. A UE may fail up to a maximum of W attempts, after which it must drop the request. The eNodeB periodically informs UEs of W through SIB2 or MAC PDU along with other RA parameters.

6.4 System Models

In this section, we will discuss the basic backlog buildup and introduce further notation. We first model the random access contention in a given time slot $n+1$ in section 6.4.1. We will use steady state equilibrium conditions to establish steady state analytical solutions for the Poisson arrival process. We will consider the single preamble case in section 6.4.2, and will then extend to multiple preamble case in section 6.4.3. In these sections, we will relate three RA control parameters, ac-BarringTime (Barring Probability), O (Number of preambles to choose from), and W (Maximum attempts before dropping the request) to the steady state solution for the system. In section 6.4.4, we will discuss the impact of $T_o^{\max}$, on the backlog and the system dynamics. Table 6.1 summarizes the notation used in this and later sections.

Table 6.1: Summary of main model notations

Given system parameters (constants) for preamble contention	
O	Number of available preambles
$T_o^{\max}$	Maximum backoff time [in slots]: Requests that collide in a slot n are re-transmitted in a (uniformly distributed) slot
	Continued on the next page

Table 6.1: *Continued*

	$n+1,\ n+2,\ \ldots, n+T_o^{\max}+1$
W	Maximum number of transmission attempts
λ	Poisson generation rate of new requests (in requests/slot)
Numbers of newly generated, backlogged, and dropped requests	
a_{n+1}	Number (random var.) of newly generated (arrived) requests for transmission in slot $n+1$
$\lambda = \mathbb{E}a_{n+1}$	Expected value of number of newly generated requests for a slot
$o_n^{(i)}$	Number (random var.) of backlogged (old) requests at end of slot n for retransmission in slot $n+i,\ i=1,2,\ldots,T_o^{\max}+1$
$\hat{X}_n$	Number of backlogged (old) requests for retransmission in nth slot
$\hat{X}_{n,i}$	Number of backlogged (old) requests for retransmission in ith slot, scheduled in nth slot
d_n	Number (random var.) of dropped requests (that have failed in W transmission attempts) at end of slot n
Total number of transmitted requests	
$t_{n+1} =: \theta$	Total number (random variable) of requests transmitted in slot $n+1$
$\vartheta = \mathbb{E}\theta$	Expected (steady-state) value of total number of transmitted requests in a slot

Continued on the next page

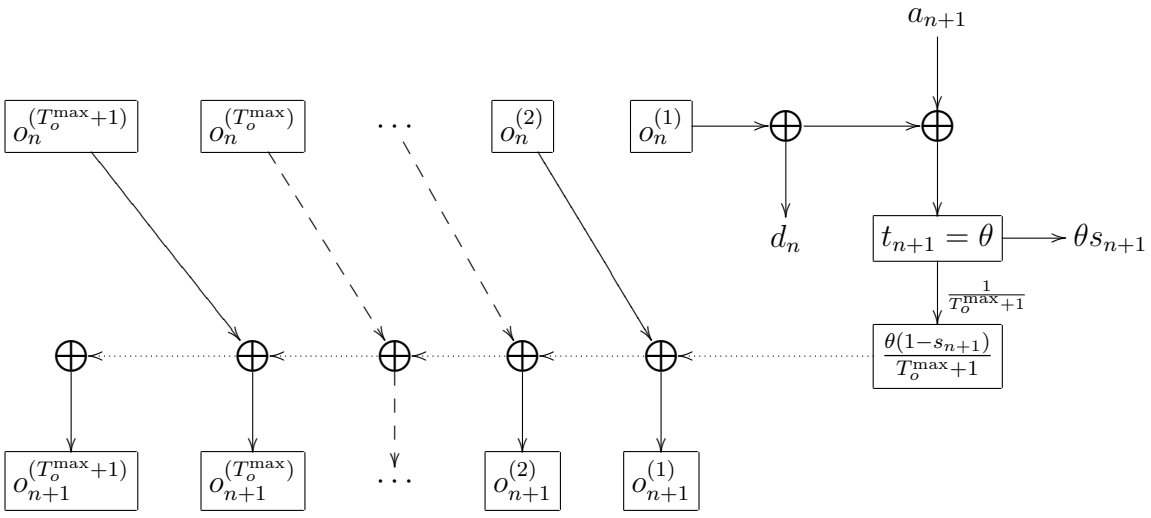

Figure 6.2: Illustration of model of random access system with backoff. In slot $n+1$, $o_n^{(1)}$ backlogged requests minus d_n requests that have exhausted their W attempts plus a_{n+1} new requests give a total of $t_{n+1} =: \theta$ transmitted preambles. With success probability s_{n+1}, there are θs_{n+1} successful requests and $\theta(1 - s_{n+1})$ failed requests that are uniformly distributed over the next $T_o^{\max} + 1$ slots for retransmission. These failed requests join the requests that had been previously scheduled for retransmission in one of the next $T_o^{\max}$ slots.

Table 6.1: *Continued*

Contention success/failure model	
s_{n+1}	Probability of successful preamble transmission (without collision) in slot $n+1$
$\varsigma = \mathbb{E}[s_{n+1}]$	Steady-state success probability
δ_{n+1}	Probability of a request failing in W transmission attempts based on success probability s_{n+1}
$\delta = \mathbb{E}[\delta_{n+1}]$	Steady-state drop probability

6.4.1 System Dynamics in a Given Slot $n+1$

Figure 6.2 shows the basic back-off process during RA. For a given slot $n+1$, we let a_{n+1} be a random variable denoting the number of UEs that transmit a preamble for a newly generated request in slot $n+1$. We let $o_n^{(1)}$ be a random variable denoting the number of UEs with previously collided (old) preambles that backed off and scheduled their next preamble transmission attempt for slot $n+1$.

We let d_n be a random variable denoting the number of UEs that drop their requests at the end of slot n due to having exhausted the number of permissible transmission attempts W. Thus, as illustrated in Fig. 6.2, $o_n^{(1)} - d_n$ UE requests are actually re-transmitted in slot $n+1$. Hence, there are a total of

$$t_{n+1} = a_{n+1} + o_n^{(1)} - d_n := \theta \tag{6.1}$$

preamble transmissions contending in slot $n+1$; we denote θ for this random variable to reduce clutter in the subsequent analysis.

We denote s_{n+1} for the probability that a UE successfully transmits a preamble, i.e., that the preamble transmission does not collide, in slot $n+1$. The success probability s_{n+1} depends on number of contending preamble transmissions t_{n+1} and the number of available preambles O. Thus, θs_{n+1} preamble transmissions are successful in slot $n+1$. On the other hand, $\theta(1-s_{n+1})$ preamble transmissions are unsuccessful (collide) and back off according to the backlog model in the following section.

If a UE request was unsuccessful in its Wth transmission attempt, we keep it for now in the backlog model and drop the request from consideration just before the contention in slot $n+1$. We denote δ_{n+1} for the probability that a given UE request collides W times and model this probability as

$$\delta_{n+1} = (1 - s_{n+1})^W. \tag{6.2}$$

6.4.2 Single Preamble Steady State Model

In this section, we will develop a steady state model for the preamble contention using equilibrium criteria assuming that only one preamble is in use. We will discuss and justify the applicability of the model and the Poisson arrivals in Section 6.4.4.

Let us consider that the system is in steady state. For a system in steady state the equilibrium condition must hold. Since the system is neither a source nor a sink for UE requests, the arrivals and departures of the request must balance. Thus,

$$\mathbb{E}[\text{No. of Arrivals per slot}] = \mathbb{E}[\text{No. of Departures per slot}], \tag{6.3}$$

$$\mathbb{E}[\text{No. of Arrivals per slot}] = \mathbb{E}[\text{No. of Successes per slot}] + \mathbb{E}[\text{No. of Drops per slot}]. \tag{6.4}$$

Let us consider the expected number of request arrivals per slot to eNodeB be denoted by $\mathbb{E}[a_{n+1}] = \lambda$. Let us further model these arriving requests by a Poisson arrival process. Since we have a Poisson arrival process, the distribution of number of arrivals per slot will be Poisson distributed. λ is the mean of this distribution of counts of arrivals per slot.

Let us denote the number of UEs that can transmit in a particular slot by X, $\mathbb{E}[X] = x$, and the barring probability by P^B. Then, t_{n+1} denote the number of UEs which clear barring and actually transmit in a slot. Then, the expected number of UEs actually attempting in a slot is $\mathbb{E}[t_{n+1}] = (1 - P^B)\mathbb{E}[X] = \vartheta$. For analytical tractability, we model X as a Poisson arrival process, x being the mean of resulting Poisson distributed counts. Hence, expected number of UEs transmitting in a slot is $(1 - P^B)x = \vartheta$. Then, the steady state expectation of successful UEs correspond to

the case when attempting population is exactly one i.e.,

$$\mathbb{E}[\text{Number of Successes per slot}] = k \cdot \mathbb{P}[k]|_{k=1} \tag{6.5}$$

$$= k \cdot \left. \frac{\mathrm{e}^{(-\vartheta x)} \left(\vartheta\right)^k}{k!} \right|_{k=1} \tag{6.6}$$

$$= k \cdot \left. \frac{\mathrm{e}^{\left(-(1-P^B)x\right)} \left((1-P^B)x\right)^k}{k!} \right|_{k=1} \tag{6.7}$$

$$= (1-P^B)x \cdot \mathrm{e}^{-(1-P^B)x}. \tag{6.8}$$

Using Eq. (6.8), we can say that the steady state probability of success for a UE which attempts to transmit after clearing barring is,

$$\varsigma = \frac{\mathbb{E}[\text{Number of Successes per slot}]}{\mathbb{E}[\text{Number of Attempting UEs per slot}]} \tag{6.9}$$

$$= \frac{\vartheta \cdot \mathrm{e}^{-\vartheta}}{\vartheta} \tag{6.10}$$

$$= \mathrm{e}^{-\vartheta} = \mathrm{e}^{-(1-P^B)x}. \tag{6.11}$$

We further consider that the subsequent attempts have little or no correlation and hence, can be considered independent. We will discuss the correlation among successive attempts in greater detail in section 6.4.4. Taking the considered independence in account, we can then say that the steady state drop probability is probability of failure in W attempts i.e. $\delta = (1-\varsigma)^W$, where W is the number of attempts after which a UE must drop the request. Thus,

$$\delta = (1-\varsigma)^W \tag{6.12}$$

$$= \left(1 - \left(1 - \mathrm{e}^{-\vartheta}\right)^W\right) = \left(1 - \left(1 - \mathrm{e}^{-\left(1-P^B\right)x}\right)^W\right). \tag{6.13}$$

Observing that the expected number of drops per slot $\mathbb{E}[d_n]$ is a fraction of expected arrivals, we have,

$$\mathbb{E}[d_n] = \lambda \cdot \delta \tag{6.14}$$

$$= \lambda \cdot \left(1 - \left(1 - \mathrm{e}^{-\vartheta}\right)^W\right) = \lambda \cdot \left(1 - \left(1 - \mathrm{e}^{-(1-P^B)x}\right)^W\right). \tag{6.15}$$

While the relation of successes to the total population is understandable, to better clarify the dependence of drops on arrivals, let us consider the case where after each failure the UE attempts in next immediate slot. In this case if there are a_{n+1} arrivals in a slot then after W slots only a_{n+1} requests can be in the system and exhaust exactly W attempts. Thus, following the example, the average drops should be considered as a fraction of expected arrivals per slot.

Substituting the results in Eq. (6.4), we have,

$$\lambda = \vartheta \cdot \varsigma + \lambda \cdot \delta = \left(1 - P^B\right) x \cdot \varsigma + \lambda \cdot \delta \tag{6.16}$$

Rearranging,

$$\lambda(1 - \delta) = \left(1 - P^B\right) x \cdot \varsigma \tag{6.17}$$

$$\text{or,} \qquad \frac{\left(1 - P^B\right) x}{\lambda} = \frac{1 - \delta}{\varsigma} \tag{6.18}$$

$$\text{or,} \qquad \frac{\left(1 - P^B\right) x}{\lambda} = \frac{1 - \left(1 - \mathrm{e}^{-\left(1-P^B\right)x}\right)^W}{\mathrm{e}^{-(1-P^B)x}}. \tag{6.19}$$

Eq. (6.19) can be solved numerically for x as,

$$\left(1 - P^B\right) x \cdot \mathrm{e}^{-\left(1-P^B\right)x} - \lambda \left(1 - \left(1 - \mathrm{e}^{-\left(1-P^B\right)x}\right)^W\right) = 0 \tag{6.20}$$

To simplify, let us look at the Eq. (6.19) in terms of ϑ , we then have,

$$\frac{\vartheta}{\lambda} = \frac{1 - \left(1 - \mathrm{e}^{-\vartheta}\right)^W}{\mathrm{e}^{-\vartheta}} \tag{6.21}$$

6.4.3 Multi-Preamble Steady State Model

In this section, we will extend the model for single preamble contention to the multi-preamble case.

In addition to the notation used in Section 6.4.2, let us denote the number of preambles by O. For the case of multi-preamble systems, the expected number of UEs actually attempting in a slot per preamble will be,

$$\mathbb{E}[t_{n+1}] = \frac{(1-P^B)\mathbb{E}[X]}{O} \tag{6.22}$$

Using a Poisson arrival process model for X as in Section 6.4.2, expectation of successes per preamble is,

$$\begin{aligned}
\mathbb{E}[\text{Successes}] &= k \cdot \mathbb{P}[k]|_{k=1} \\
&= k \cdot \left. \frac{\mathrm{e}^{\left(-(1-P^B)x/O\right)} \left(\left(1-P^B\right)x/O\right)^k}{k!} \right|_{k=1} && (6.23) \\
&= \frac{\left(1-P^B\right)x}{O} \cdot \mathrm{e}^{\left(-(1-P^B)x/O\right)}. && (6.24)
\end{aligned}$$

The success probability per preamble is then,

$$\begin{aligned}
\varsigma &= \frac{\left(1-P^B\right)x/O \cdot \mathrm{e}^{-\left(1-P^B\right)x/O}}{\left(1-P^B\right)x/O} && (6.25) \\
&= \mathrm{e}^{-\left(1-P^B\right)x/O}. && (6.26)
\end{aligned}$$

And, the drop probability per preamble is,

$$\delta = (1-\varsigma)^W = (1-(1-\mathrm{e}^{-(1-P^B)x/O})^W). \tag{6.27}$$

Using the equilibrium condition in Eq. (6.4) and considering that the expected arrivals per preamble is now λ/O, we have,

$$\frac{\lambda}{O} = \frac{\left(1-P^B\right)x}{O} \cdot \varsigma + \frac{\lambda}{O} \cdot \delta$$

Rearranging and substituting,

$$\frac{\left(1-P^B\right)x/O}{\lambda/O} = \frac{1-\left(1-\mathrm{e}^{-(1-P^B)x/O}\right)^W}{\mathrm{e}^{-(1-P^B)x/O}} \tag{6.28}$$

$$\frac{\left(1-P^B\right)x}{\lambda} = \frac{1-\left(1-\mathrm{e}^{-(1-P^B)x/O}\right)^W}{\mathrm{e}^{-(1-P^B)x/O}} \tag{6.29}$$

In terms of $\vartheta = (1-P^B)x/O$,

$$\frac{\vartheta}{\lambda/O} = \frac{1-\left(1-\mathrm{e}^{-\vartheta}\right)^W}{\mathrm{e}^{-\vartheta}}. \tag{6.30}$$

6.4.4 Impact of $T_o^{\max}$ and the Poisson Arrival Process Assumption for X

In this section, we will introduce the parameter $T_o^{\max}$ and associate it to the Poisson arrival process model for X. We will prove that an increasing $T_o^{\max}$ causes to modify the underlying process of X to a Poisson process, irrespective of the underlying process, if $\lambda \ll X$. We will consider the impact of other factors such as P^B, and O in section 6.5 during the discussion of simulations.

Let us introduce the maximum back-off delay parameter $T_o^{\max}$ from the LTE standard. After suffering a collision, a UE waits for a duration uniformly distributed between 0 and $T_o^{\max}$. We can divide this distribution in slots and consider $T_o^{\max}$ as an integer which represents the number of slots over which failed transmissions in a slot are uniformly rescheduled. Henceforth, we will assume $T_o^{\max}$ to mean this integer number, unless otherwise stated. A non-zero $T_o^{\max}$ acts to reduce the correlation among the number of re-transmitting UEs in subsequent slots.

To mathematically model the impact of $T_o^{\max}$, let us consider the underlying process for X, if we have a_n arrivals in nth slot i.e. new transmissions, and $\hat{X}_n = o_n^{(1)} - d_n$ retransmissions then,

$$X_n = a_n + \hat{X}_n \tag{6.31}$$

If $\mathbb{E}[a_n] \ll \mathbb{E}[\hat{X}]$, then the process of $\hat{X}$ can be considered to dominate the arrival process. By dominating the arrival process, we imply that $\hat{X}$ can be considered a reasonable approximation for X and a_n is insignificant in comparison, and that this holds in general, over time.

$\hat{X}$ constitutes retransmissions accumulated during previous $T_o^{\max}$ slots. Let us consider that after contention in nth slot, $\hat{X}_{n,n+i} : i \in [0, T_o^{\max}]$ retransmissions are scheduled for transmission in the $(n+i)$th slot. Since, i is uniformly distributed between 0 and $T_o^{\max}$, the probability p of choosing any slot is,

$$p = \frac{1}{1 + T_o^{\max}} \tag{6.32}$$

Thus, we have $\hat{X}_n$ requests, and the probability of each request choosing $(n+i)$th slot is p. We model $T_o^{\max}$ so that it is not changing and thus p is constant as the slot from which request is being rescheduled n, increases. Total number of requests which can be scheduled for transmission in slot i are,

$$\hat{X}_i = \hat{X}_{i-(1+T_o^{\max}),i} + \hat{X}_{i-T_o^{\max},i} + \ldots + \hat{X}_{i-1,i} \tag{6.33}$$

$$\approx (1 + T_o^{\max}) \cdot \mathbb{E}[\hat{X}] \tag{6.34}$$

Since, $a_i \ll \hat{X}$, $\mathbb{E}[\hat{X}] \approx \mathbb{E}[X]$,

$$\hat{X}_i \approx (1 + T_o^{\max}) \cdot \mathbb{E}[X] = x(1 + T_o^{\max}) \tag{6.35}$$

Each of the $x(1 + T_o^{\max})$ requests essentially undergoes a Bernoulli trial to select ith slot with probability p of choosing the slot. Total number of successes in $xT_o^{\max}$ Bernoulli trials is Binomial distributed with parameters $x(1 + T_o^{\max})$ and p. Thus,

$$\mathbb{E}[\hat{X}] = x(1 + T_o^{\max}) \cdot p = x(1 + T_o^{\max}) \cdot \frac{1}{1 + T_o^{\max}} = x. \tag{6.36}$$

As $T_o^{\max}$ and x increases, the Binomial distribution tends towards Poisson distribution, implying that counts in a slot for retransmissions become Poisson distributed. This

indicates that the underlying process for X tends to become a Poisson arrival process, as $T_o^{\max}$ and x increases. The variance for $\hat{X}$ thus tends to x.

6.5 Numerical Analysis

In this section we will describe the simulation setup and discuss various results. We use simulation models developed in C++ using C++ standard library, OMNeT++ Varga (2001) based libraries, and simulation tools Akaroa2 Erwing *et al.* (1999) and GNU parallel Tange (2011). Akaroa2 is used for Multiple Replications in Parallel to achieve 95% confidence intervals with 5% relative error for Poisson distributed and Bernoulli distributed arrival counts. GNU parallel is used for managing multiple simulations on multi-core computing systems. While we use the back-off period limits acceptable to the current LTE/LTE-A standard, for the ac-BarringTime duration we use the same permissible delays as those for back-off. For the understanding of the system dynamics, the smaller values are sufficient.

In the simulation results, we use success probabilities (ς) and drop probabilities (δ) as representative of accuracy of the models. Subscripts 'ana' and 'sim' are used to denote analytical and simulated results, respectively. ρ, on horizontal axis denotes the normalized count of arrivals per slot i.e. λ/O.

6.5.1 Poisson Arrivals

Let us consider the Poisson arrivals. Figs. 6.3 and 6.4 show a comparison of solution of Eq. 6.20 and simulations for the case when only one preamble is available for RA. As it can be seen, the model is close but not very good for these parameters.

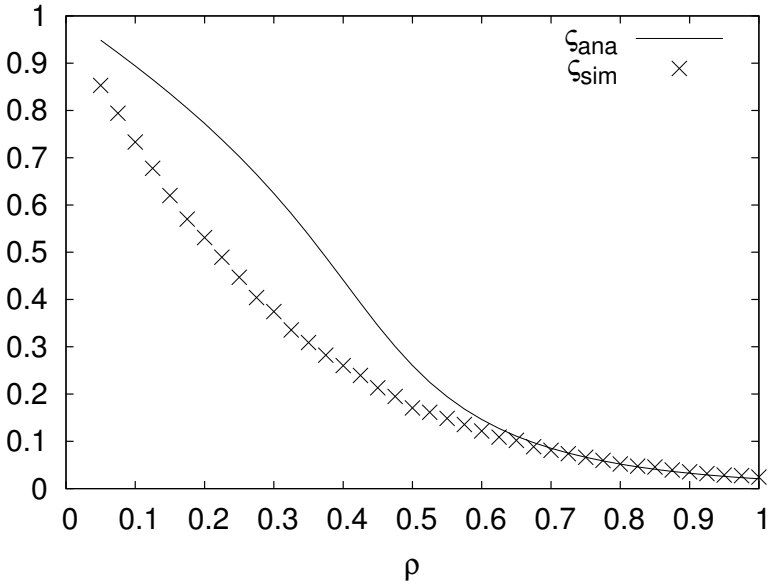

Figure 6.3: Steady Success State Probabilities for Poisson Arrivals. $O = 1$, $W = 4$, $P^B = 0.1$, and $T_o^{\max} = 0$.

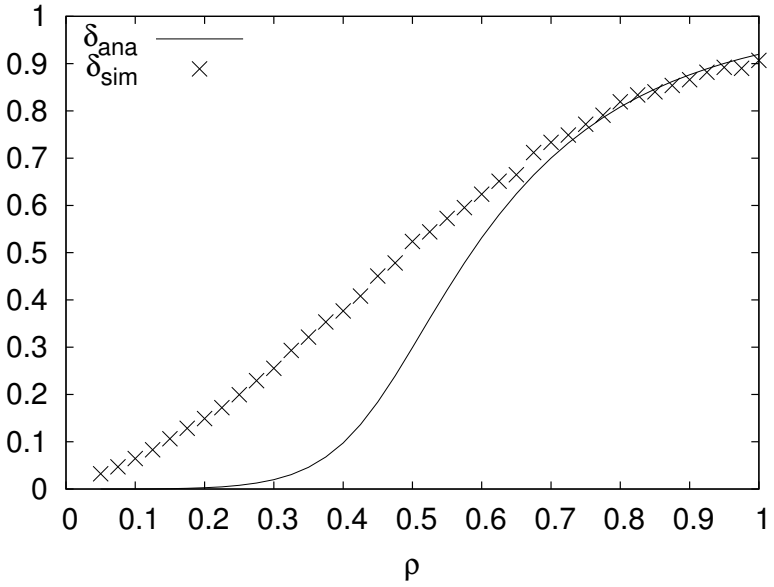

Figure 6.4: Steady Drop State Probabilities for Poisson Arrivals. $O = 1$, $W = 4$, $P^B = 0.1$, and $T_o^{\max} = 0$.

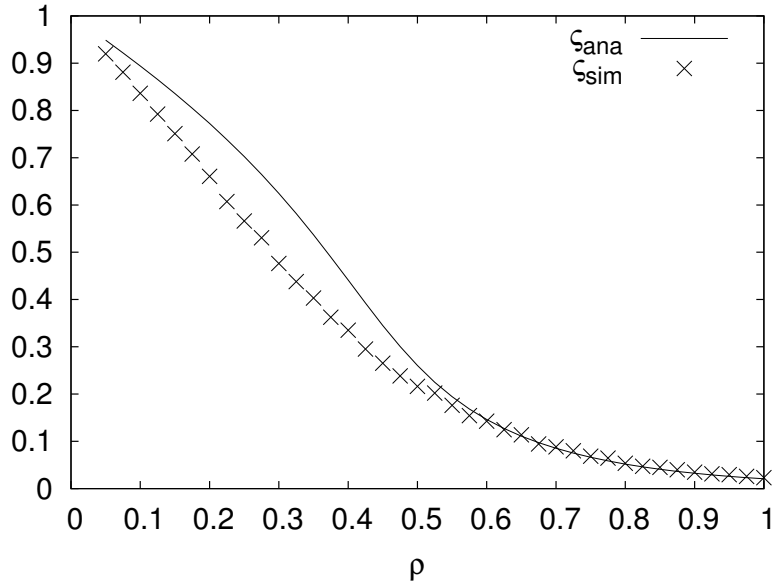

Figure 6.5: Steady Success State Probabilities for Poisson Arrivals. $O = 1$, $W = 4$, $P^B = 0.5$, and $T_o^{\max} = 0$.

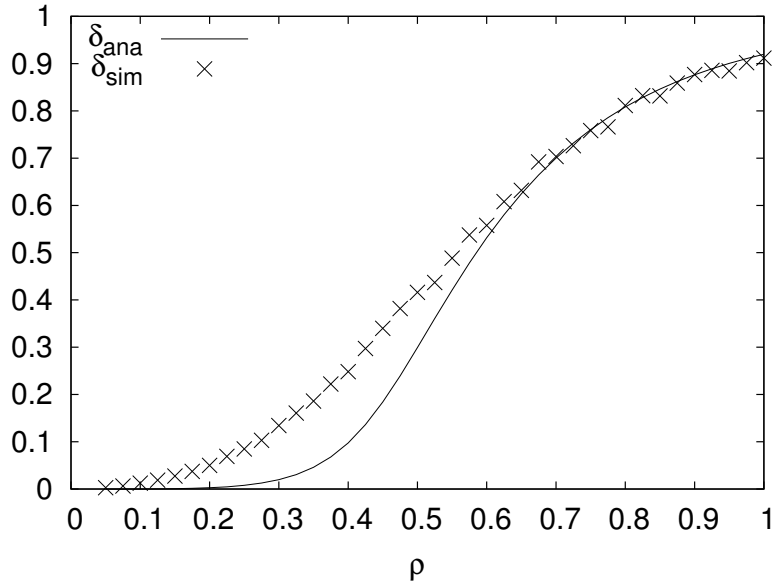

Figure 6.6: Steady Drop State Probabilities for Poisson Arrivals. $O = 1$, $W = 4$, $P^B = 0.5$, and $T_o^{\max} = 0$.

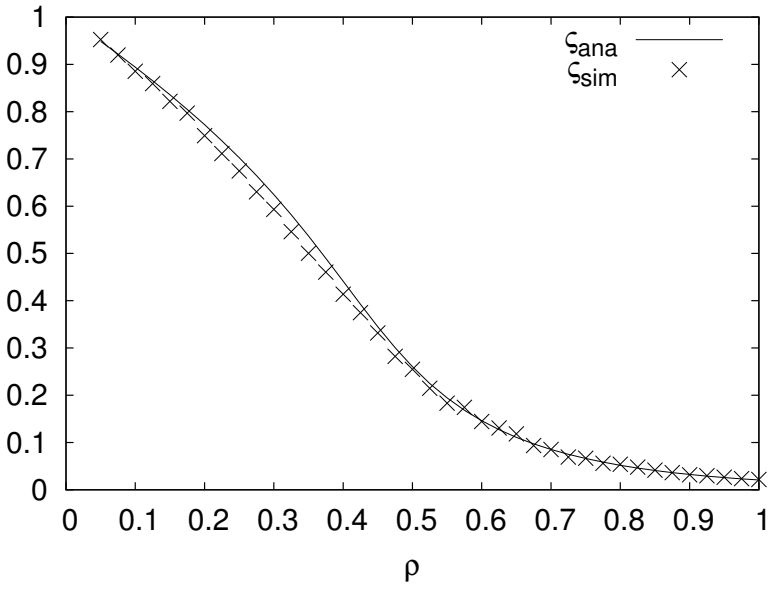

Figure 6.7: Steady State Success Probabilities for Poisson Arrivals. $O = 1$, $W = 4$, $P^B = 0.9$, and $T_o^{\max} = 0$.

Impact of P^B

However, increasing P^B to 0.5, we see that the results have greater agreement in Figs. 6.5 and 6.6. Further, in Figs. 6.7 and 6.8 the results match very well.

In essence, P^B has a de-correlating effect on the subsequent contention periods.

Impact of O

Figs. 6.9 and 6.10 shows the comparison for $O = 10$. Considering that in Eq. 6.28 $\vartheta = (1 - P^B)x/O$, we can say that O has a similar de-correlating effect as P^B and a value of $O = 10$ should reduce the impact of a low $P^B = 0.1$ from $0.9x$ to $0.09x \approx 0.1x$, which is similar to $P^B = 0.9, \vartheta = 0.1x$. Thus, the small values of both P^B and O can be combined together for greater de-correlation. This is reflected well in Figs. 6.9 and 6.10 which matches closely with simulation results, similar to the case for $O = 1$, and $P^B = 0.9$.

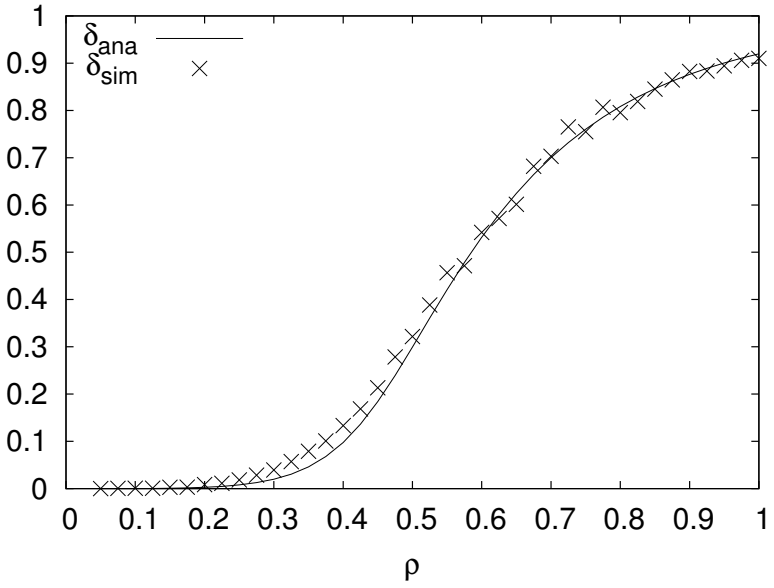

Figure 6.8: Steady State Drop Probabilities for Poisson Arrivals. $O = 1$, $W = 4$, $P^B = 0.9$, and $T_o^{\max} = 0$.

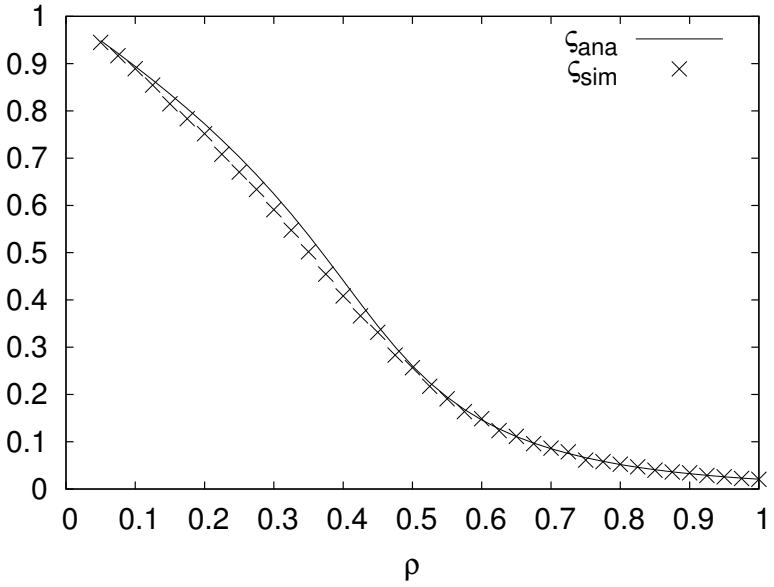

Figure 6.9: Steady State Success Probabilities for Poisson Arrivals. $O = 10$, $W = 4$, $P^B = 0.1$, and $T_o^{\max} = 0$.

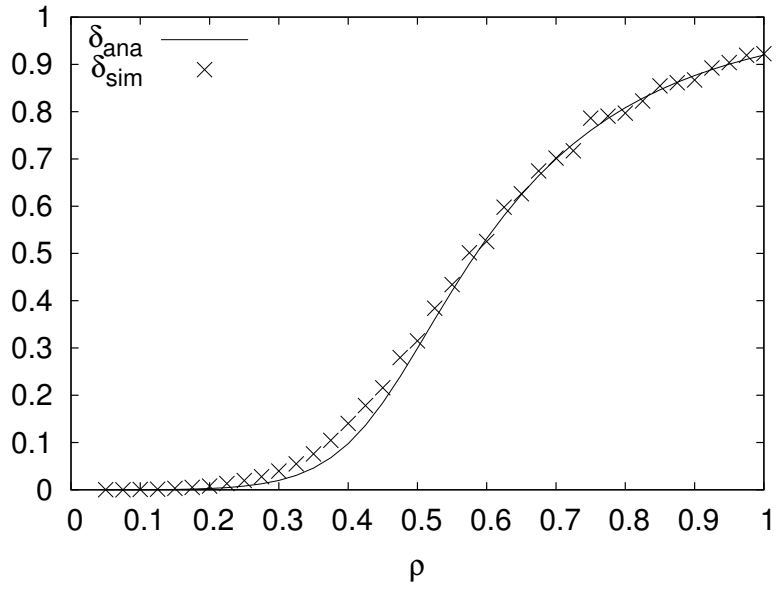

Figure 6.10: Steady State Drop Probabilities for Poisson Arrivals. $O = 10$, $W = 4$, $P^B = 0.1$, and $T_o^{\max} = 0$.

Impact of W

In Figs. 6.11 and 6.12, it can be seen that the disparity between the model probabilities $(\varsigma_{ana}, \delta_{ana})$ and the simulated probabilities $(\varsigma_{sim}, \delta_{sim})$ increases with increase in W. We attribute this increase in disparity to increased impact of correlation on higher values of W. However, the impact of de-correlation remains strong and for $P^B = 0.9$, the model matches well with the simulations in Fig. 6.13 and 6.14.

6.5.2 Bernoulli Distributed Arrival Counts

We simulate using Bernoulli counts as well. Note that for $\hat{X}$ minimum value is 0 and that this is a discrete parameter, let us consider the maximum variance the distribution of $\hat{X}$ may have. It is well established that the maximum variance for a discrete distribution with minimum value a and maximum value b is $(b-a)^2/4$ Jr. *et al.* (1985); Muilwijk (1966). This is the case for a Bernoulli variable with equally

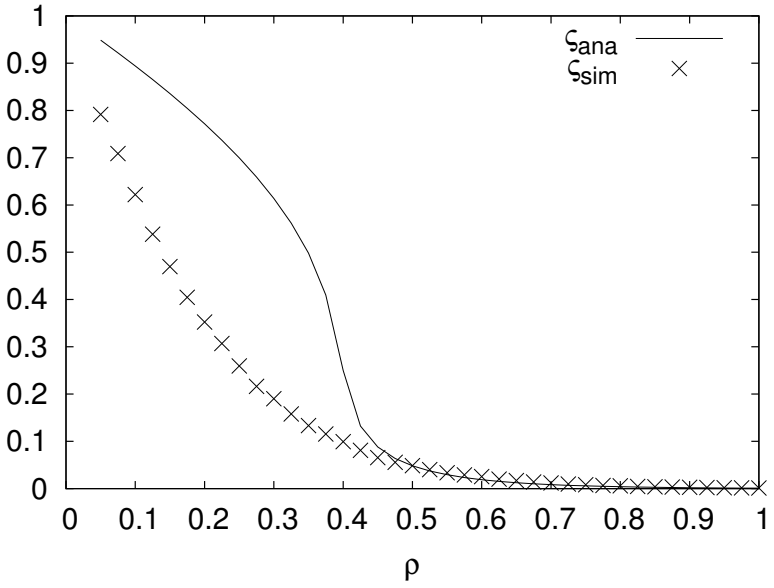

Figure 6.11: Steady State Probabilities for Poisson Arrivals. $O = 1$, $W = 7$, $P^B = 0.1$, and $T_o^{\max} = 0$.

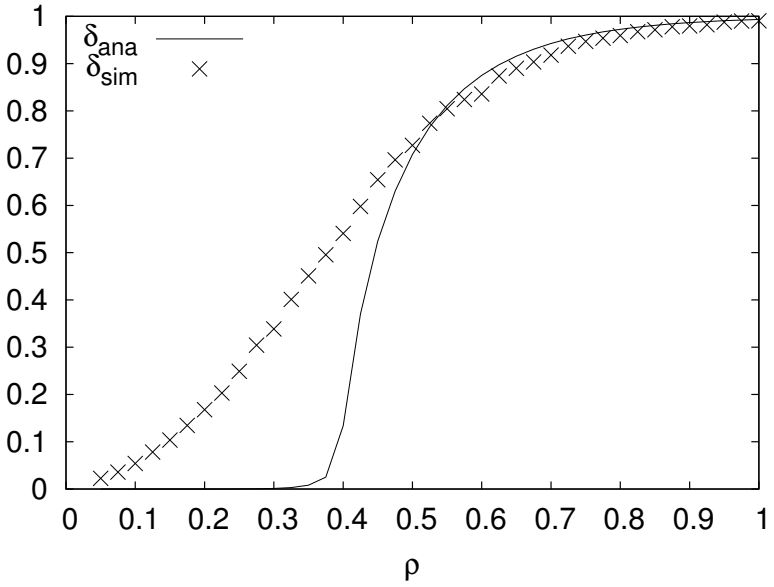

Figure 6.12: Steady State Probabilities for Poisson Arrivals. $O = 1$, $W = 7$, $P^B = 0.1$, and $T_o^{\max} = 0$.

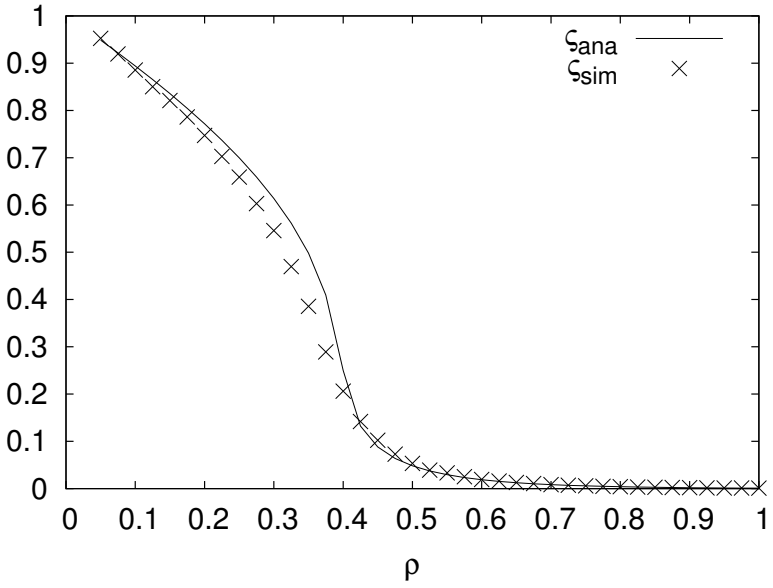

Figure 6.13: Steady State Probabilities for Poisson Arrivals. $O = 1$, $W = 7$, $P^B = 0.9$, and $T_o^{\max} = 0$.

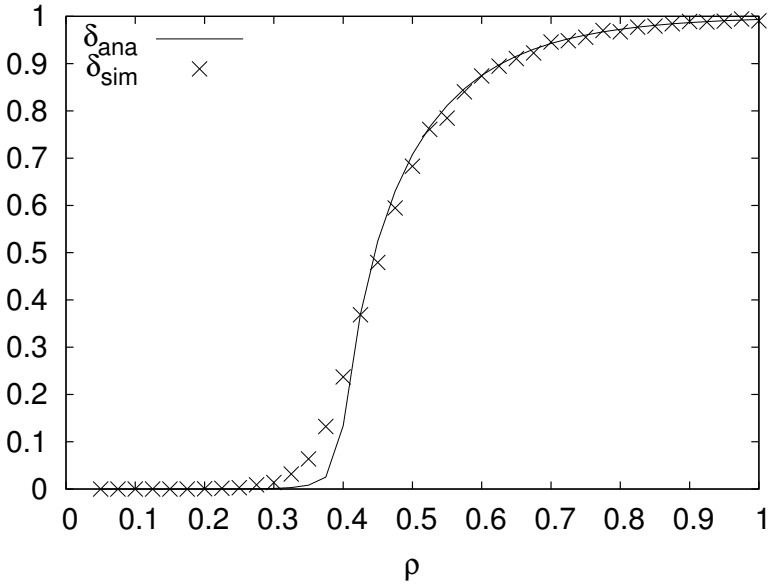

Figure 6.14: Steady State Probabilities for Poisson Arrivals. $O = 1$, $W = 7$, $P^B = 0.9$, and $T_o^{\max} = 0$.

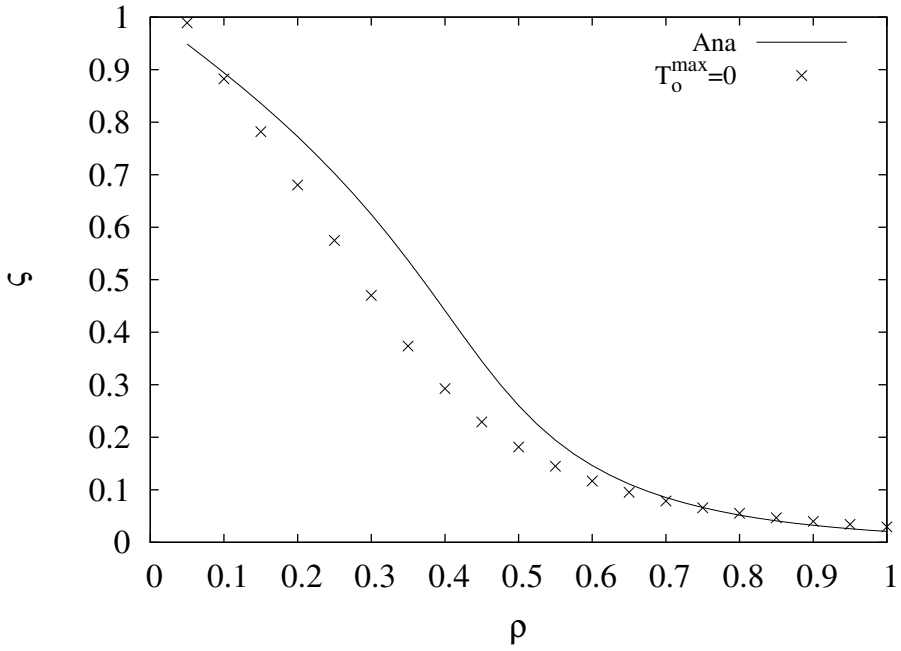

Figure 6.15: Steady State Success Probabilities for Bernoulli Distributed Arrival Counts. $O = 10$, $W = 4$, $P^B = 0.1$, and $T_o^{\max} = 0$.

likely b and a. For $\hat{X}$, $a = 0$ and the mean $x = (a + b)/2 = b/2$ i.e. $b = 2x$. The maximum variance for the original $\hat{X}$ with mean x can be $(2x - 0)^2/4 = x^2$. We thus consider an arrival process based on Bernoulli counts in this study as one of the worst case situations.

For simulations we consider a maximum variance case based on mean arrivals λ, with the two possible count values being either 0 or 2λ. In simulations, we do not find good match between the model Eq. 6.28 and simulated values for small values of O. In this case, the impact of O is greater on de-correlating values compared to P^B. We find that the results in reasonable agreement for $O = 10$, as shown in Figs. 6.15 and 6.16. However, for generally used value of $O = 54$ and a high barring probability $P^B = 0.9$, the results match closely. Figs. 6.17 and 6.18 shows the match for these values.

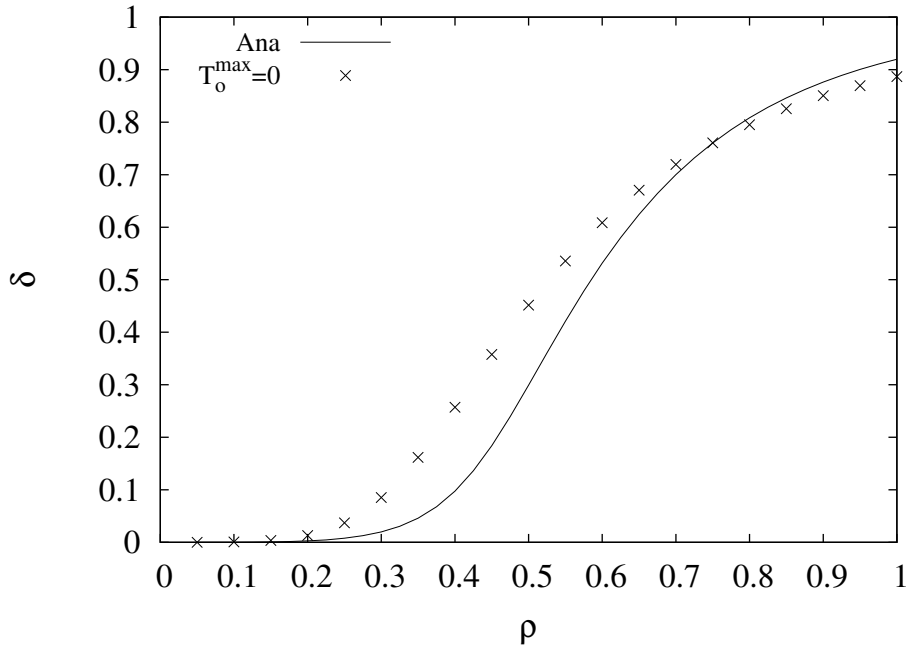

Figure 6.16: Steady State Drop Probabilities for Bernoulli Distributed Arrival Counts. $O = 10$, $W = 4$, $P^B = 0.1$, and $T_o^{\max} = 0$.

6.5.3 MMPP Process Based Arrivals and Impact of $T_o^{\max}$

As another extreme case, we simulate a 2 state Markov Modulated Poisson Process (MMPP). The MMPP we use has, for a mean arrival rate of λ, a Poisson arrival process with mean $\lambda/5$ as low rate process and another Poisson arrival process with mean 5λ as high rate process. The transition probability from high rate process to low rate process is 0.05 and the transition probability from low rate process to high rate process is 0.01.

A 2-state MMPP model is not unimodal and likely to be more difficult to de-correlate. We now use the $T_o^{\max}$ parameter as well. As can be seen from Figs. 6.19 and 6.20, a large value of $T_o^{\max} = 500$ has sufficiently de-correlated the 2-state MMPP arrival process to result in great accuracy with Poisson arrivals based model.

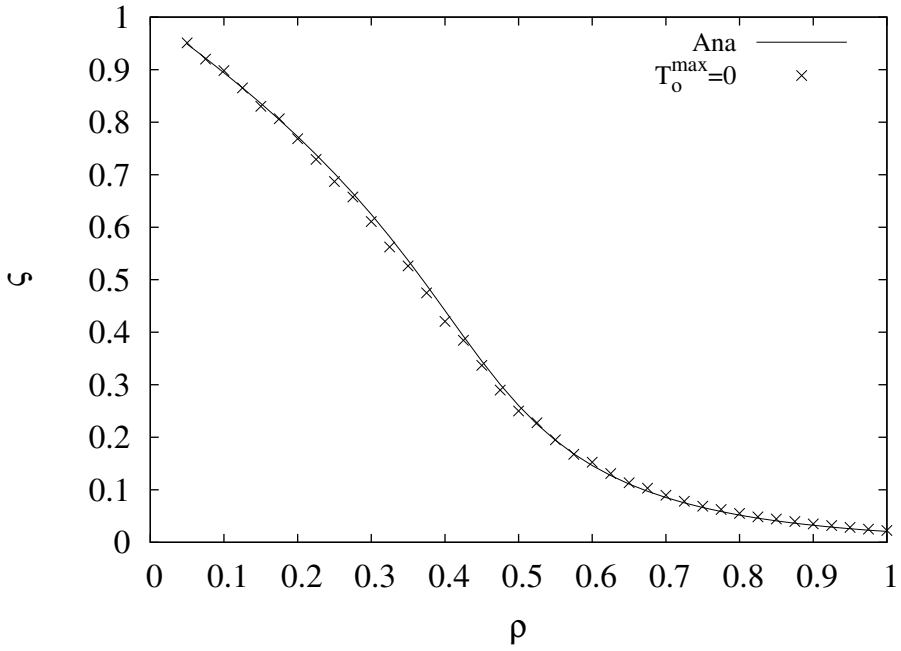

Figure 6.17: Steady State Success Probabilities for Bernoulli Distributed Arrival Counts. $O = 54$, $W = 4$, $P^B = 0.9$, and $T_o^{\max} = 0$.

The impact of an increasing $T_o^{\max}$ is thus, to reduce the variance, maximum x^2 for unimodal distributions, of the original $\hat{X}$ process forcing it to result in a Poisson-ized $\hat{X}$ process with mean and variance x. The value $x^2 - x = x(x-1)$ can thus be considered a measure of maximum reduction needed to ensure that $\hat{X}$ acts as a Poisson arrival process, for unimodal arrival processes.

Particular value of $T_o^{\max}$ which achieves the de-correlation, is the maximum value of $T_o^{\max}$ which impacts the process. Any further increase in $T_o^{\max}$ does not affect the system other than increasing delays. For the case, when $\lambda \ll x$, $X \approx \hat{X}$ and can be considered as a Poisson arrival process.

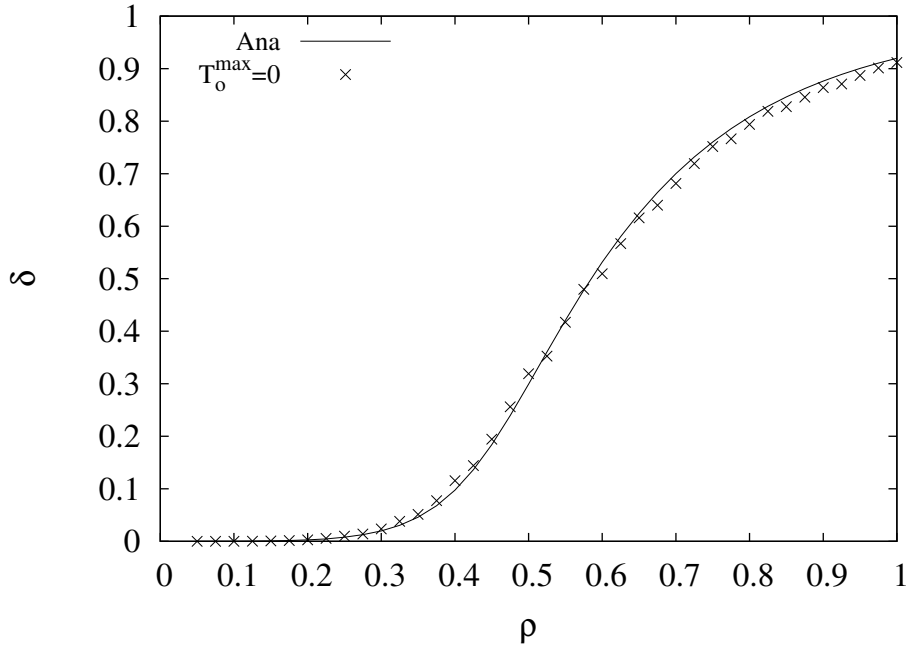

Figure 6.18: Steady State Drop Probabilities for Bernoulli Distributed Arrival Counts. $O = 54$, $W = 4$, $P^B = 0.9$, and $T_o^{\max} = 0$.

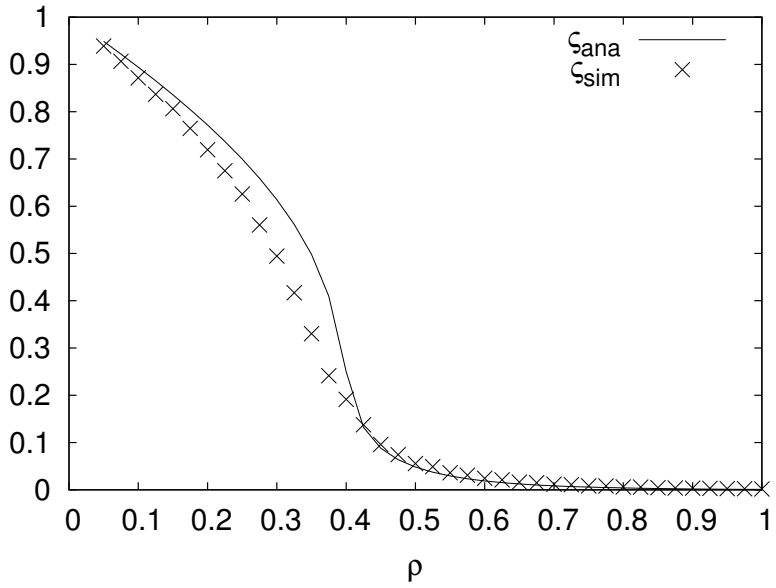

Figure 6.19: Steady State Success Probabilities for 2 State MMPP Distributed Arrival Counts. $O = 54$, $W = 7$, $P^B = 0.9$, and $T_o^{\max} = 500$.

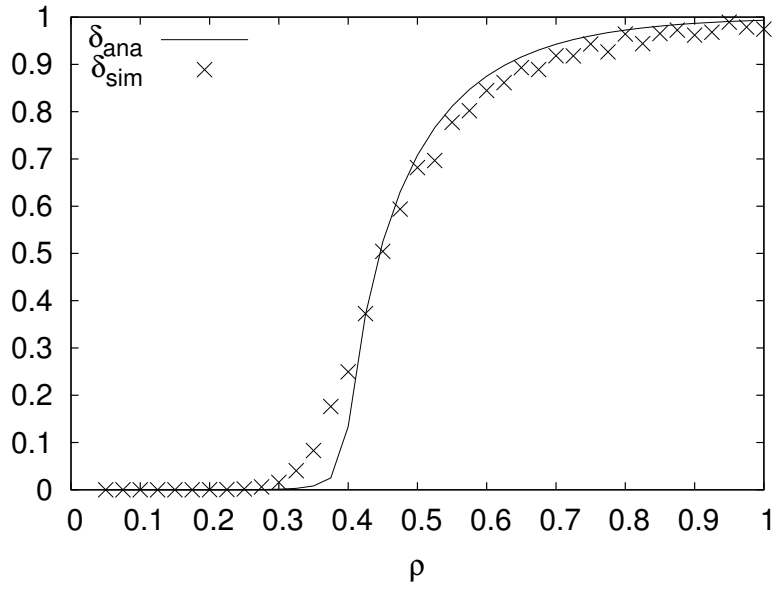

Figure 6.20: Steady State Drop Probabilities for 2 State MMPP Distributed Arrival Counts. $O = 54$, $W = 7$, $P^B = 0.9$, and $T_o^{\max} = 500$.

6.6 Conclusion

In this chapter, we have derived steady state models for LTE/LTE-A RA procedure. The performance of RA procedure is of critical importance in upcoming M2M technologies. We have used an equilibrium based approach to derive the steady state, which is more fundamental and dependent only on existence of a steady state. Our analysis is applicable to both single and multiple preamble case.

Further, we have explored the impact of various parameters associated with the RA procedure. Through simulations and analytical understanding, we have found that suitable sets of these parameters can be found to ensure that a steady state exists. We have also established that such a resulting steady state will closely resemble a Poisson arrivals based system.

With the understanding of various RA parameters, further algorithms can be explored for quick descent to steady state or for various objectives. We have not explored these optimizations in current study and are for future work.

Chapter 7

CONCLUSION

We have analyzed the impact of the number W of transmission attempts on the throughput and delay of the slotted Aloha based preamble contention in the LTE-Advanced random access system. Our study provides analytical characterizations for the combinations of transmission attempt limit W and request load ρ that results in one, two, or three equilibrium operating points. Specifically, for $W \geq 9$ transmission attempts, which are a necessary condition for multiple operating points, we analyze the load region $(\rho_1,\ \rho_2)$ that results in three operating points. We analytically characterize the throughputs and delays at these operating points.

The numerical investigations with our analysis results and verifying simulations indicate that for the examined scenario with $O = 54$ preambles, a small to moderately large transmission attempt limit W around ten without backoff gives good throughput-delay performance. Uniform backoff achieves only relatively small throughput improvements at the expense of substantially increased delays. For reliable low-delay service, a network with $W \geq 9$ should be operated with a load below the boundary ρ_1, which ensures that the network does not experience high-delay operating points. For $W \leq 8$, our delay analysis can be used to identify load limits for low-delay service.

We have extended these results to single channel case. A variance based model has been developed to estimate the variance, which can be used to design backoff schemes based on statistical characteristics of user observed contention.

We have also established a Poissonization effect, due to which UE requests distribute according to a Poisson distribution whose mean equals λ, when uniform backoff

is employed. When the introduced latency is not a problem, this technique can be used to shape the distribution of incoming requests.

There are many important directions for future research. One example direction is to examine service differentiation Cheng *et al.* (2011); Hu *et al.* (2012) whereby different service classes employ different transmission attempt limits W. Another direction is to study the internetworking of LTE-Advanced networks with local networks, such as body area networks, attached to the UE and backhaul networks, such as Ethernet Passive Optical Networks (EPONs) Aurzada *et al.* (2014); Coimbra *et al.* (2013); Lim *et al.* (2013); Maier *et al.* (2009); Milosavljevic *et al.* (2012); Aurzada *et al.* (2011). An integration with LTE-Advanced networks on one end of the EPON and a sensor network on the other end Hossen and Hanawa (2011); Seema and Reisslein (2011), is an interesting area of exploration as well. Use of long reach EPONs will make the area more challenging Kantarci and Mouftah (2012); Mercian *et al.* (2013). Another exciting research direction is the integration with metropolitan area optical networks Bianco *et al.* (2013); Maier and Reisslein (2004); Maier *et al.* (2003); Scheutzow *et al.* (2003); Yang *et al.* (2003), attached to the eNB.

REFERENCES

36.321, G. T., *Evolved Universal Terrestrial Radio Access (E-UTRA): Medium Access Control (MAC) Protocol Specification, v.10.4.0* (2011).

4G Americas, "4G Mobile Broadband Evolution: 3GPP Release 10 and Beyond, HSPA+, SAE/LET and LTE-Advanced", (2012).

Abramson, N., "THE ALOHA SYSTEM: Another Alternative For Computer Communications", in "Proc. of The Joint Computer Conference", AFIPS '70 (Fall), pp. 281–285 (ACM, New York, NY, USA, 1970).

Alcaraz, J. J., E. Egea-Lopez, J. Vales-Alonso and J. Garcia-Haro, "Dynamic system model for optimal configuration of mobile RFID systems", Computer Networks **55**, 1, 74–83 (2011).

Amirijoo, M., P. Frenger, F. Gunnarsson, J. Moe and K. Zetterberg, "On self-optimization of the random access procedure in 3G Long Term Evolution", in "Proc. of IFIP/IEEE International Symposium on Integrated Network Management-Workshops", pp. 177–184 (2009).

Aurzada, F., M. Levesque, M. Maier and M. Reisslein, "FiWi access networks based on next-generation PON and gigabit-class WLAN technologies: A capacity and delay analysis", IEEE/ACM Trans. Netw., in print (2014).

Aurzada, F., M. Scheutzow, M. Reisslein, N. Ghazisaidi and M. Maier, "Capacity and Delay Analysis of Next-Generation Passive Optical Networks (NG-PONs)", IEEE Trans. Comm. **59**, 5, 1378–1388 (2011).

Bianco, A., T. Bonald, D. Cuda and R.-M. Indre, "Cost, power consumption and performance evaluation of metro networks", IEEE/OSA J. Opt. Comm. Netw. **5**, 1, 81–91 (2013).

Boudriga, N., M. Obaidat and F. Zarai, "Intelligent network functionalities in wireless 4G networks: Integration scheme and simulation analysis", Comp. Commun. **31**, 16, 3752–3759 (2008).

Carleial, A. and M. Hellman, "Bistable Behavior of ALOHA-Type Systems", IEEE T. Commun. **23**, 4, 401–410 (1975).

Cheng, J.-P., C.-H. Lee and T.-M. Lin, "Prioritized random access with dynamic access barring for RAN overload in 3GPP LTE-A networks", in "Proc. of IEEE Globecom Workshops", pp. 368–372 (2011).

Cheng, R.-G., C.-H. Wei, S.-L. Tsao and F.-C. Ren, "RACH Collision Probability for Machine-Type Communications", in "Proc. of IEEE Vehicular Technology Conference", pp. 1–5 (2012).

Choi, S., W. Lee, D. Kim, K.-J. Park, S. Choi and K.-Y. Han, "Automatic Configuration of Random Access Channel Parameters in LTE Systems", in "Proc. of IFIP Wireless Days (WD)", pp. 1–6 (2011).

Choi, Y.-J., S. Park and S. Bahk, "Multichannel random access in OFDMA wireless networks", IEEE Journal on Selected Areas in Communications **24**, 603–613 (2006).

Coimbra, J., G. Schultz and N. Correia, "A game-based algorithm for fair bandwidth allocation in Fibre-Wireless access networks", Optical Switching and Networking **10**, 2, 149 – 162 (2013).

Corless, R., G. Gonnet, D. Hare, D. Jeffrey and D. Knuth, "On the Lambert W function", Adv. Comp. Math. **5**, 329–359 (1996).

Dai, L., "Stability and delay analysis of buffered Aloha networks", IEEE Trans. Wireless Commun. **11**, 8, 2707–2719 (2012).

Erwing, G., K. Pawlikowski and D. McNickle, "Akaroa2: Exploiting Network Computing by Distributing Stochastic Simulation", in "Proc. of European Simulation Multiconference (ESM)", pp. 175–181 (1999).

Fan, Z., R. Haines and P. Kulkarni, "M2M Communications for E-Health and Smart Grid: An Industry and Standard Perspective", IEEE Wirel. Commun. **21**, 1, 62–69 (2014).

Ferguson, M., "On the Control, Stability, and Waiting Time in a Slotted ALOHA Random-Access System", IEEE T. Commun. **23**, 11, 1306–1311 (1975).

Gerasimenko, M., V. Petrov, O. Galinina, S. Andreev and Y. Koucheryavy, "Energy and Delay Analysis of LTE-Advanced RACH Performance under MTC Overload", in "Proc. of IEEE GLOBECOM Workshops (GC Wkshps)", pp. 1632–1637 (2012).

Ghez, S., S. Verdu and S. C. Schwartz, "Stability Properties of Slotted Aloha with Multipacket Reception Capability", IEEE T. Automat. Contr. **33**, 7, 640–649 (1988).

Grishechkin, S., M. Devetsikiotis, I. Lambadaris and C. Hobbs, "Multistability in Queues with Retransmission and Its Relationship with Large Deviations in Branching Processes", Theor. Probab. Appl+ **47**, 1, 139–150, URL `http://epubs.siam.org/doi/abs/10.1137/S0040585X97979585` (2003).

Haas, Z. and J. Deng, "On Optimizing the Backoff Interval for Random Access Schemes", IEEE T. Commun. **51**, 12, 2081–2090 (2003).

Hasan, M., E. Hossain and D. Niyato, "Random Access for Machine-to-Machine Communication in LTE-Advanced Networks: Issues and Approaches", IEEE Commun. Mag. **51**, 6, 86–93 (2013).

Heyman, D. P. and M. J. Sobel, *Stochastic Models in Operations Research: Volume 1: Stochastic Processes and Operating Characteristics* (Courier Dover, 2003).

Hossen, M. and M. Hanawa, "Network architecture and performance analysis of MULTI-OLT PON for FTTH and wireless sensor networks", Int. J. Wireless & Mobile Networks **3**, 6, 1–15 (2011).

Hu, N., X.-L. Li and Q.-N. Ren, "Random access preamble assignment algorithm of TD-LTE", Advances in Computer, Communication, Control and Automation, Lec. Notes in Electr. Eng. **121**, 701–708 (2012).

Jenq, Y.-C., "On the stability of slotted ALOHA systems", IEEE Trans. on Commun. **COM-28**, 11, 1936–1939 (1980a).

Jenq, Y.-C., "On the Stability of Slotted ALOHA Systems", IEEE T. Commun. **28**, 11, 1936–1939 (1980b).

Jian, X., Y. Jia, X. Zeng and J. Yang, "A Novel Class-Dependent Back-off Scheme for Machine Type Communication in LTE Systems", in "Proc. of Wireless and Optical Communication Conference (WOCC)", pp. 135–140 (2013).

Joseph, K. and D. Raychaudhuri, "Analysis of Generalized Retransmission Backoff Policies for Slotted-ALOHA Multiaccess Channels", in "Proc. of IEEE International Conference on Communications (ICC)", vol. 1, pp. 430–436 (1988).

Jr., J. W. S., P. S. Odell and D. M. Young, "Maximum Variance Unimodal Distributions ", Stat. Probabil. Lett. **3**, 5, 255–260 (1985).

Kamal, S. and S. Mahmoud, "A Study of Users' Buffer Variations in Random Access Satellite Channels", IEEE T. Commun. **27**, 6, 857–868 (1979).

Kamal, S. and S. Mamoud, "A study of users' buffer variations in random access satellite channels", IEEE Transactions on Communications **27**, 6, 857–868 (1979).

Kantarci, B. and H. Mouftah, "Bandwidth distribution solutions for performance enhancement in long-reach passive optical networks", IEEE Commun. Surv. Tut. **14**, 3, 714–733 (2012).

Kim, J., J. Lee, J. Kim and J. Yun, "M2M Service Platforms: Survey, Issues, and Enabling Technologies", IEEE Commun. Surveys Tuts. **16**, 1, 61–76 (2014).

Kim, S. W., "Frequency-hopped spread-spectrum random access with retransmission cutoff and code rate adjustment", IEEE J. on Selected Areas in Commun. **10**, 2, 344–349 (1992).

Kleinrock, L. and S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation", IEEE Transactions on Communications **23**, 4, 410–423 (1975a).

Kleinrock, L. and S. Lam, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation", IEEE T. Commun. **23**, 4, 410–423 (1975b).

Kouzayha, N., N. C. Taher and Y. Ghamri-Doudane, "Towards a better support of Machine Type Communication in LTE-networks: Analysis of random access mechanisms", in "Proc. of International Conference on Advances in Biomedical Engineering (ICABME)", pp. 57–60 (2013).

Kwak, B.-J., N.-O. Song and L. Miller, "Performance analysis of exponential backoff", IEEE/ACM Transactions on Networking **13**, 2, 343–355 (2005).

Kwan, R. and C. Leung, "On Collision Probabilities in Frequency-Domain Scheduling for LTE Cellular Networks", IEEE Commun. Lett. **15**, 9, 965–967 (2011).

Larmo, A., M. Lindstrom, M. Meyer, G. Pelletier, J. Torsner and H. Wiemann, "The LTE link-layer design", IEEE Communications Magazine **47**, 4, 52–59 (2009).

Lawton, G., "Machine-to-machine technology gears up for growth", Computer **37**, 9, 12–15 (2004).

Laya, A., L. Alonso and J. Alonso-Zarate, "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives", IEEE Commun. Surveys Tuts. **16**, 1, 4–16 (2014).

Lee, K.-D., S. Kim and B. Yi, "Throughput Comparison of Random Access Methods for M2M Service over LTE Networks", in "Proc. of IEEE GLOBECOM Workshops (GC Wkshps)", pp. 373–377 (2011).

Lee, K.-D. and A. Vasilakos, "Access stratum resource management for reliable u-healthcare service in LTE networks", Wireless Networks **17**, 7, 1667–1678 (2011).

Li, J., H. Tian, L. Xu and Y. Huang, "An Optimized Random Access Algorithm for MTC Users over Wireless Networks", in "Proc. of IEEE Vehicular Technology Conference", pp. 1–5 (2013).

Lien, S., K. Chen and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications", IEEE Communications Magazine **49**, 4, 66–74 (2011a).

Lien, S.-Y., K.-C. Chen and Y. Lin, "Toward Ubiquitous Massive Accesses in 3GPP Machine-to-Machine Communications", IEEE Commun. Mag. **49**, 4, 66–74 (2011b).

Lien, S.-Y., T.-H. Liau, C.-Y. Kao and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications", IEEE Trans. on Wireless Commun. **11**, 1, 27–32 (2012).

Lim, W., K. Kanonakis, P. Kourtessis, M. Milosavljevic, I. Tomkos and J. M. Senior, "Flexible QoS differentiation in converged OFDMA-PON and LTE networks", in "Proc. OFC", (2013).

Lin, T.-M., C.-H. Lee, J.-P. Cheng and W.-T. Chen, "PRADA: Prioritized Random Access with Dynamic Access Barring for MTC in 3GPP LTE-A Networks", IEEE T. Veh. Technol. **PP**, 99, 1–1 (2014).

Liu, Y.-S., "Performance analysis of frequency-hop packet radio networks with generalized retransmission backoff", IEEE Transactions on Wireless Communications **1**, 4, 703–711 (2002).

Lüders, C. and R. Haferbeck, "The Performance of the GSM Random Access Procedure", in "Proc. of IEEE Vehicular Technology Conference", vol. 2, pp. 1165–1169 (1994).

Maier, M., N. Ghazisaidi and M. Reisslein, "The audacity of fiber-wireless (FiWi) networks", in "Proc. of AccessNets", vol. 6 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 16–35 (Springer, 2009).

Maier, M. and M. Reisslein, "AWG-based metro WDM networking", IEEE Commun. Mag. **42**, 11, S19–S26 (2004).

Maier, M., M. Reisslein and A. Wolisz, "A hybrid MAC protocol for a metro WDM network using multiple free spectral ranges of an arrayed-waveguide grating", Computer Networks **41**, 4, 407–433 (2003).

Marsch, P., B. Raaf, A. Szufarska, P. Mogensen, H. Guan, M. Farber, S. Redana, K. Pedersen and T. Kolding, "Future Mobile Communication Networks: Challenges in the Design and Operation", IEEE Veh. Technol. Mag. **7**, 1, 16–23 (2012).

Mercian, A., M. McGarry and M. Reisslein, "Offline and online multi-thread polling in long-reach PONs: A critical evaluation", IEEE/OSA J. Lightwave Techn. **31**, 12, 2018–2228 (2013).

Milosavljevic, M., M. Thakur, P. Kourtessis, J. Mitchell and J. Senior, "Demonstration of Wireless Backhauling Over Long-Reach PONs", IEEE/OSA J. Lightwave Techn. **30**, 5, 811–817 (2012).

Minero, P., M. Franceschetti and D. N. C. Tse, "Random Access: An Information-Theoretic Perspective", IEEE T. Inform. Theory **58**, 2, 909–930 (2012).

Muilwijk, J., "Note on a Theorem of M. N. Murthy and V. K. Sethi", Sankhyā Ser. B **28**, 1/2, 183–183 (1966).

Murali, R. and B. Hughes, "Random Access with Large Propagation Delay", IEEE ACM T. Network. **5**, 6, 924–935 (1997).

Murali, R. and B. Hughes, "Coding and stability in frequency-hop packet radio networks", IEEE Trans. Comm. **46**, 2, 191–199 (1998a).

Murali, R. and B. Hughes, "Coding and Stability in Frequency-Hop Packet Radio Networks", IEEE T. Commun. **46**, 2, 191–199 (1998b).

Mutairi, A., S. Roy and G. Hwang, "Delay Analysis of OFDMA-Aloha", IEEE T. Wirel. Commun. **12**, 1, 89–99 (2013).

Naware, V., G. Mergen and L. Tong, "Stability and delay of finite-user slotted ALOHA with multipacket reception", IEEE Transactions on Information Theory **51**, 7, 2636–2656 (2005a).

Naware, V., G. Mergen and L. Tong, "Stability and Delay of Finite-User Slotted ALOHA with Multipacket Reception", IEEE T. Inform. Theory **51**, 7, 2636–2656 (2005b).

Onozato, Y., J. Liu, S. Shimamoto and S. Noguchi, "Effect of propogation delays on ALOHA systems", Comp. Netw. **12**, 329–337 (1986).

Onozato, Y. and S. Noguchi, "On the thrashing cusp in slotted Aloha systems", IEEE Trans. on Commun. **COM-33**, 11, 1171–1182 (1985a).

Onozato, Y. and S. Noguchi, "A unified analysis of steady state behavior in random access schemes", Computer Networks and ISDN Systems **10**, 111–122 (1985b).

Paiva, R., R. Vieira and M. Saily, "Random Access Capacity Evaluation with Synchronized MTC Users over Wireless Networks", in "Proc. of IEEE Vehicular Technology Conference", pp. 1–5 (2011).

Pountourakis, I. and E. Sykas, "Analysis, stability and optimization of Aloha-type protocols for multichannel networks", Computer Communications **15**, 10, 619–629 (1992).

Rivero-Angeles, M., D. Lara-Rodriguez and F. Cruz-Perez, "Gaussian Approximations for the Probability Mass Function of the Access Delay for Different Backoff Policies in S-ALOHA", IEEE Commun. Lett. **10**, 10, 731–733 (2006).

Rom, R. and M. Sidi, *Multiple Access Protocols: Performance and Analysis* (Springer Verlag, 1989).

Sakakibara, K., H. Muta and Y. Yuba, "The Effect of Limiting the Number of Retransmission Trials on the Stability of Slotted ALOHA Systems", IEEE T. Veh. Technol. **49**, 4, 1449–1453 (2000).

Sakakibara, K., T. Seto, D. Yoshimura and J. Yamakita, "Effect of Exponential Backoff Scheme and Retransmission Cutoff on the Stability of Frequency-Hopping Slotted ALOHA Systems", IEEE T. Wirel. Commun. **2**, 4, 714–722 (2003).

Sarangan, V., M. Devarapalli and S. Radhakrishnan, "A framework for fast RFID tag reading in static and mobile environments", Computer Networks **52**, 5, 1058–1073 (2008).

Sarker, J. and H. Mouftah, "Self-stability of slotted ALOHA by limiting the number of retransmission trials in infrastructure-less wireless networks", Telecommunication Systems **52**, 2, 435–444 (2013).

Sarker, J. H., "Stable and Unstable Operating Regions of Slotted ALOHA with Number of Retransmission Attempts and Number of Power Levels", IEE Proceedings - Commun. **153**, 3, 355–364 (2006).

Sarker, J. H. and S. J. Halme, "The Prudence Transmission Method I (PTM I): A Retransmission Cut-Off Method for Contention based Multiple-Access Communication Systems", in "Proc. of IEEE Vehicular Technology Conference", vol. 1, pp. 397–401 (1997).

Sarker, J. H. and S. J. Halme, "The Prudence Transmission Method II (PTM II): On the Maximum and Minimum Values of Retransmission Cut-Off Method for Slotted ALOHA", in "Proc. of IEEE Vehicular Technology Conference", vol. 2, pp. 1315–1320 (1998).

Sarker, J. H. and S. J. Halme, "An Optimum Retransmission Cut-Off Scheme for Slotted ALOHA", Wirel. Pers. Comm. **13**, 185–202 (2000a).

Sarker, J. H. and S. J. Halme, "An Optimum Retransmission Cut-Off Scheme for Slotted ALOHA", Wireless Pers. Commun. **13**, 185–202, URL `http://dx.doi.org/10.1023/A:1008930002745`, 10.1023/A:1008930002745 (2000b).

Sarker, J. H. and H. T. Mouftah, "Stability of multiple receiving nodes slotted ALOHA for wireless ad hoc networks", in "Proc. of IEEE Globecom", pp. 1–5 (2008).

Scheutzow, M., M. Maier, M. Reisslein and A. Wolisz, "Wavelength reuse for efficient packet-switched transport in an AWG-based metro WDM network", IEEE/OSA J. Lightwave Techn. **21**, 6, 1435–1455 (2003).

Seema, A. and M. Reisslein, "Towards efficient wireless video sensor networks: A survey of existing node architectures and proposal for a Flexi-WVSNP design", IEEE Comm. Surv. & Tut. **13**, 3, 462–486 (2011).

Seo, J. and V. Leung, "Design and analysis of backoff algorithms for random access channel in UMTS-LTE and IEEE 802.16 system", IEEE Trans. on Vehicular Techn. **60**, 8, 3975–3989 (2011a).

Seo, J.-B. and V. C. M. Leung, "The Effect of Retransmission Cutoff in S-ALOHA Systems with Binary Exponential Backoff", in "Proc. of International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)", pp. 1452–1456 (2010).

Seo, J.-B. and V. C. M. Leung, "Design and Analysis of Backoff Algorithms for Random Access Channels in UMTS-LTE and IEEE 802.16 Systems", IEEE T. Veh. Technol. **60**, 8, 3975–3989 (2011b).

Seo, J.-B. and V. C. M. Leung, "Design and Analysis of Cross-Layer Contention Resolution Algorithms for Multi-Packet Reception Slotted ALOHA Systems", IEEE T. Wirel. Commun. **10**, 3, 825–833 (2011c).

Seo, J.-B. and V. C. M. Leung, "Performance Modeling and Stability of Semi-Persistent Scheduling with Initial Random Access in LTE", IEEE T. Wirel. Commun. **11**, 12, 4446–4456 (2012).

Shen, D. and V. Li, "Stabilized multi-channel ALOHA for wireless OFDM networks", in "Proc. of IEEE Globecom", pp. 701–705 (2002).

Simon, B. and L. Votta, "The optimal retry distribution for lightly loaded slotted Aloha systems", IEEE Trans. Comm. **33**, 7, 724–725 (1985).

Szpankowski, W., "Packet switching in multiple radio channels: Analysis and stability of a random access system", Computer Networks **7**, 1, 17–26 (1983).

Taleb, T. and A. Kunz, "Machine Type Communications in 3GPP Networks: Potential, Challenges, and Solutions", IEEE Commun. Mag. **50**, 3, 178–184 (2012).

Tange, O., "GNU Parallel - The Command-Line Power Tool", The USENIX Mag. **36**, 1, 42–47 (2011).

Tyagi, R., F. Aurzada, K.-D. Lee, S. Kim and M. Reisslein, "Impact of Retransmission Limit on Preamble Contention in LTE-Advanced Network", IEEE Syst. J. **PP**, 99, 1–14 (2013).

Tyagi, R., F. Aurzada, K.-D. Lee and M. Reisslein, "Equilibrium analysis for connection establishment in LTE/LTE-A Networks: Parameters and their impact (extended version)", Tech. rep., Arizona State University, http://trace.eas.asu.edu/CEext.pdf (2014).

Tyagi, R. R., K.-D. Lee, F. Aurzada, S. Kim and M. Reisslein, "Efficient Delivery of Frequent Small Data for U-Healthcare Applications over LTE-Advanced Networks", in "Proc. of ACM International Workshop on Pervasive Wireless Healthcare", MobileHealth '12, pp. 27–32 (ACM, New York, NY, USA, 2012), URL `http://doi.acm.org/10.1145/2248341.2248354`.

Varga, A., "The OMNeT++ Discrete Event Simulation System", in "Proc. of European Simulation Multiconference (ESM)", pp. 319–324 (2001).

Vukovic, I. and I. Filipovich, "Throughput analysis of TDD LTE Random Access Channel", in "Proc. of International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)", pp. 1652–1656 (2011).

Wang, X., J. Kaniyil, Y. Onozato and S. Noguchi, "Heterogeneous ALOHA networks: A sufficient condition for all equilibrium states to be stable", Computer Netw. ISDN Systems **22**, 213–224 (1991).

Wei, C.-H., R.-G. Cheng and S.-L. Tsao, "Modeling and estimation of one-shot random access for finite-user multichannel slotted ALOHA systems", IEEE Commun. Let. **16**, 8, 1196–1199 (2012).

Wu, G., S. Talwar, K. Johnsson, N. Himayat and K. Johnson, "M2M: From Mobile to Embedded Internet", IEEE Commun. Mag. **49**, 4, 36–43 (2011).

Yang, H.-S., M. Maier, M. Reisslein and W. Carlyle, "A genetic algorithm-based methodology for optimizing multiservice convergence in a metro WDM network", IEEE/OSA J. Lightwave Techn. **21**, 5, 1114–1133 (2003).

Yang, X., A. Fapojuwo and E. Egbogah, "Performance Analysis and Parameter Optimization of Random Access Backoff Algorithm in LTE", in "Proc. of IEEE Vehicular Technology Conference", pp. 1–5 (2012).

Yilmaz, O., J. Hamalainen and S. Hamalainen, "Self-optimization of random access channel in 3rd Generation Partnership Project Long Term Evolution", Wirel. Comm. Mob. Comp. **11**, 12, 1507–1517 (2011).

Yue, W., "The effect of capture on performance of multichannel slotted ALOHA systems", IEEE Trans. Comm. **39**, 6, 818–822 (1991).

Yun, J., "Cross-layer analysis of the random access mechanism in Universal Terrestrial Radio Access", Computer Networks **56**, 1, 315–328 (2012).

Zhang, Z. and Y.-J. Liu, "Comments on 'the effect of capture on performance of multichannel slotted ALOHA systems"', IEEE Transactions on Communications **41**, 10, 1433–1435 (1993).

APPENDIX A

EXPECTED NUMBER OF UNSUCCESSFUL UES $E[Z_n]$

Proceeding from (3.8), we evaluate the conditional expectation of the number of unsuccessful UEs Z_n given the number of UEs X_n sending a preamble in slot n, as

$$\begin{aligned} E[Z_n|X_n] &= \sum_{i=1}^{X_n} E\left[1_{\{\exists j\in\{1,\ldots,X_n\},j\neq i:\ \alpha_i=\alpha_j\}}\right] && \text{(A.1)}\\ &= \sum_{i=1}^{X_n} [1 - P\left(\forall j \in \{1,\ldots,X_n\},\right. \\ &\qquad\qquad \left. j \neq i : \alpha_i \neq \alpha_j\right)] && \text{(A.2)}\\ &= \sum_{i=1}^{X_n} \left[1 - \prod_{j=1,j\neq i}^{X_n} P\left(\alpha_i \neq \alpha_j\right)\right] && \text{(A.3)}\\ &= X_n \left(1 - \left(\frac{O-1}{O}\right)^{X_n-1}\right), && \text{(A.4)} \end{aligned}$$

whereby in the last step we substituted $P\left(\alpha_i \neq \alpha_j\right) = (O-1)/O$ as there are $O-1$ preambles (out of the total of O preambles) that are not equal to a given (fixed) preamble and the UEs select the preambles independently. We note that X_n follows approximately a Poisson distribution (with mean x). To see this, observe from (3.3) and the illustration in Fig. 3.1 that X_n is a sum of random fractions of Poisson random variables. From (A.4), we evaluate $E[Z_n] = E[E[Z_n|X_n]]$ as follows:

$$\begin{aligned} E[Z_n] &= E\left[X_n \left(1 - \left(\frac{O-1}{O}\right)^{X_n-1}\right)\right] && \text{(A.5)}\\ &\approx \sum_{k=1}^{\infty} \frac{x^k}{k!}\,\mathrm{e}^{-x} \cdot k \left(1 - \left(\frac{O-1}{O}\right)^{k-1}\right) && \text{(A.6)}\\ &= x\left[1 - \mathrm{e}^{-x}\exp\left(x\frac{O-1}{O}\right)\sum_{k=0}^{\infty}\frac{1}{k!}\left(x\frac{O-1}{O}\right)^{k}\right. \\ &\qquad\qquad \left.\exp\left(-x\frac{O-1}{O}\right)\right] && \text{(A.7)}\\ &= x\left[1 - \mathrm{e}^{-x/O}\right], && \text{(A.8)} \end{aligned}$$

whereby the summation in (A.7) is over the probability mass function of a Poisson random variable with mean $x(O-1)/O$, i.e., gives one.

APPENDIX B

PROPERTIES OF $h(t)$

In this appendix, we analyze the right-hand side of the balance equation (3.11), i.e.,

$$h(t) := g(\mathrm{e}^{-t}) := \frac{1-(1-\mathrm{e}^{-t})^W}{\mathrm{e}^{-t}}, \tag{B.1}$$

which represents the ratio of steady-state success probability after W transmission attempts to success probability in the first attempt. We readily verify that $h'(t) = 0$ has no solutions: Indeed, abbreviating $y = \mathrm{e}^{-t} \in (0, 1]$ we have

$$h'(t) = g'(\mathrm{e}^{-t})(-\mathrm{e}^{-t}) \tag{B.2}$$

Clearly,

$$\begin{aligned} g'(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\frac{1-(1-y)^W}{y} && \text{(B.3)} \\ &= \frac{W(1-y)^{W-1}y - [1-(1-y)^W]}{y^2}. && \text{(B.4)} \end{aligned}$$

The numerator of (B.4) has no zeros for $y \in (0, 1)$. In fact, this numerator is negative for all $y \in (0, 1)$. Thus h' is positive for all $t > 0$, showing that h is a strictly increasing function starting at $h(0) = 1$ and ending at $h(\infty) = W$.

We next show that $h(t)$ has precisely one convex and one concave piece. Specifically, we show that the equation $h''(t) = 0$ has exactly one solution for $t > 0$, which we will call inflexion point t_0. This shows that h has exactly one convex piece (for arguments in $[0, t_0]$) and one concave piece (for the arguments in $[t_0, \infty)$).

In order to analyze the equation $h''(t) = 0$, note that

$$h'(t) = g'(y)(-y), \qquad h''(t) = g''(y)(-y)^2 + g'(y)y. \tag{B.5}$$

So that $h''(t) = 0$ has a solution for $t > 0$ if and only if

$$g''(y)y + g'(y) = 0 \tag{B.6}$$

has a solution for $y \in (0, 1)$. After some simplifications, and setting $z := 1 - y$, this is equivalent to

$$0 = p(z) := 1 - W(W-1)z^{W-2} + W(2W-3)z^{W-1} - (W-1)^2 z^W. \tag{B.7}$$

This is now a polynomial in z of degree W. It can be seen easily that $p(0) = 1$, $p(1) = 0$, $p'(0) = 0$, $p'(1) = 0$, $p'((W-2)/(W-1)) = 0$, and $p''(1) = -W(W-1) < 0$. Thus, $p(z) = 0$ has exactly one solution in $(0, 1)$ (which we denote by z_0) and, going back, so has $h''(t) = 0$, at

$$t_0 := \log \frac{1}{1-z_0}. \tag{B.8}$$

APPENDIX C

ASYMPTOTICS OF ρ_1 FOR LARGE W

In this appendix, we derive the asymptotics of the load boundary ρ_1 for large transmission attempt limits W (with $Wy \to 0$) given in (4.7). We first approximate (4.5) with the Taylor expansion

$$(1-y)^W = 1 - Wy + \frac{W(W-1)}{2}y^2 + \mathcal{O}((Wy)^3) \tag{C.1}$$

to obtain

$$\begin{aligned} \rho_1 &= \frac{t_1 \mathrm{e}^{-t_1}}{1-[1-W\mathrm{e}^{-t_1}+\mathcal{O}((W\mathrm{e}^{-t_1})^2)]} && \text{(C.2)} \\ &\sim \frac{t_1}{W}. && \text{(C.3)} \end{aligned}$$

We proceed to express t_1 asymptotically in terms of W, whereby we omit in the following the subscript 1 to avoid clutter. We approximate (4.4) with the Taylor expansions (C.1) and

$$Wy(1-y)^{W-1} = Wy[1-(W-1)y] + \mathcal{O}((Wy)^3), \tag{C.4}$$

to obtain after algebraic simplifications and recalling that $y = e^{-t}$,

$$-(t+1)\mathrm{e}^{-(t+1)} = -\frac{2\mathrm{e}^{-1}}{W-1}. \tag{C.5}$$

The solution $t = t(W)$ of (C.5) is a Product-log function, also referred to as Lambert W function Corless *et al.* (1996), whereby the W in the Lambert W function is not to be confused with our notation W for the transmission attempt limit. For the asymptotic behavior of this function, specifically its branch -1, it can be shown that

$$t(W) = \log\frac{W-1}{2\mathrm{e}^{-1}} + \log\log\frac{W-1}{2\mathrm{e}^{-1}} - 1 + o(1), \tag{C.6}$$

where $o(1)$ denotes a term that tends to zero as $W \to \infty$. Inserting (C.6) in (C.3) gives (4.7).

An alternative approach to employing the Lambert W function is to express $t = t(W)$ by defining $\omega = (W-1)/(2e^{-1})$ and

$$t + 1 = \log[\omega\, s] \tag{C.7}$$

with s to be determined. Then, inserting (C.7) in the dominating exponential term $e^{-(t+1)}$, (C.5) becomes

$$(t+1)\frac{1}{\omega\, s} = \frac{1}{\omega} \tag{C.8}$$

i.e., $t + 1 = s$ and thus

$$\begin{aligned} s &= t + 1 = \log[\omega\, s] && \text{(C.9)} \\ &= \log[\omega] + \log[s] && \text{(C.10)} \\ &= \log[\omega] + \log[t+1]\,. && \text{(C.11)} \\ &= \log\omega + \log\log[\omega s] && \text{(C.12)} \\ &= \log\omega + \log\log\omega + o(1). && \text{(C.13)} \end{aligned}$$

This leads to

$$t = s - 1 = \log \omega + \log \log \omega - 1 + o(1), \tag{C.14}$$

which is equivalent to (C.6).