Performance Evaluation of Adaptive Backoff Mechanism of Random Access Procedures in NB-IoT

Anas Al Bakri

A Thesis
in
The Department
of
Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science (Electrical Engineering) at
Concordia University
Montreal, Quebec, Canada

September 2020

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Anas Al Bakri

Entitled: Performance Evaluation of Adaptive Backoff Mechanism of Random Access Procedures in NB-IoT

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. D. Qiu

_____ External Examiner
Dr. C. Assi (CIISE)

_____ Internal Examiner
Dr. D. Qiu

_____ Supervisor
Dr. M.K. Mehmet Ali

Approved by: _____
Dr. Y.R. Shayan, Chair
Department of Electrical and Computer Engineering

_____ 20___ _____
Dr. Mourad Debbabi, Interim Dean,
Gina Cody School of Engineering and
Computer Science

Abstract

Performance Evaluation of Adaptive Backoff Mechanism of Random Access Procedures in NB-IoT

Anas Al Bakri

Narrow Band Internet of Things (NB-IoT) is a promising radio technology that was standardized by 3GPP in 2016 for connecting a massive number of low-cost, low-power and delay-tolerant devices to the Internet of Things (IoT) with a small bandwidth of 180 KHz. Prior studies on the random access of NB-IoT have investigated its performance under various conditions and considering several mechanisms, however, the backoff mechanism as a method for contention resolution was ignored mostly either for simplification or in favor of other mechanisms while few works considered it under certain restrictions. This thesis proposes a comprehensive analytical model that evaluates the performance of random access procedures under the assumption of adaptive backoff mechanism that doubles the backoff window size for each new transmission attempt to achieve a better contention resolution.

The proposed model allows arrival of packets with different rates to each device in the network which is more practical in cases where there are noticeable differences in specifications between connected devices such as in layered, packet loss intolerant and clustered networks. The system is modeled using discrete-time analysis with First-Come-First-Served (FCFS) user queues with infinite buffers. The analysis has led to derivations of probability of successful packet transmission, mean packet delay, utilization, and probability of packet discarding. The results have shown the advantages of the adaptive backoff mechanism in improving the performance of the system compared to constant backoff. Additionally, results have demonstrated the trade-off between packet discarding probability and mean packet delay where higher number of attempts allows for almost zero packet loss probability at the cost of slightly higher delay. A simulation system was implemented which verified the accuracy of the analytical model. The results of this thesis can be helpful in designing new NB-IoT communication networks.

Acknowledgement

I would like to convey my sincere and deep gratitude to my thesis supervisor Prof. Mustafa Mehmet Ali for his valuable comments and continuous support toward completion of my graduate program.

Also, I must express my deepest appreciation to my mother and my father who have always believed in me and encouraged me to pursue my graduate degree.

Likewise, I am very grateful to my family and friends. This success would not have been achievable without their continuous support and motivation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

- **NB-IoT** Narrow Band Internet of Things
- **NPRACH** Narrowband Physical Random Access Channel (Physical sublayer)
- **RACH** Random Access Channel (MAC sublayer)
- **NPUSCH** Narrowband Physical Uplink Shared Channel
- **3GPP** The Third Generation Partnership Project
- **BS** Base Station
- **UE** User Equipment
- **HARQ** Hybrid Automatic Repeat Request
- **RA** Random Access
- **NPRACH-Periodicity** The periodicity between two successive Random-Access periods.
- **ra-Response-WindowSize** Random Access Response Window size
- **RAP** Random Access Preamble or Msg1
- **RAR** Random Access Response or Msg2
- **CE** Coverage Enhancement Level
- **Preamble-TransMax-CE:** Maximum Number of Preamble Attempts
- **LTE** Long-Term Evolution
- **WSN** Wireless Sensor Network
- **mMTC** Massive Machine-Type Communication
- **LTE-M** Long-Term Evolution for Machines radio technology

- **CCCH** Common Control Channel
- **DCCH** Dedicated Control Channel
- **DTCH** Dedicated Traffic Channel
- **RA-RNTI** Random-Access Radio Network Temporary Identity
- **TA** Timing Advance Command
- **PDU** Protocol Data Unit

List of Symbols

N	Total Number of Users in the System
T	Slot Time
λ	Packet arrival rate per slot for any user
λ_i	Packet arrival rate of the i^{th} user.
λ_T	Total packet arrival rate from all users
ω_j	Maximum number of backoff slots for the j^{th} Transmission
Q	Maximum number of allowed transmissions
n_i	Number of packets in the user's queue at the service completion of the i^{th} packet
η	Number of active users
P_s	Probability of access success of a packet during a slot
γ	Probability of collision during a slot
P_0	Probability of having an empty queue
B_i	Number of backoff slots for the i^{th} Transmission
Y	Service time in number of slots
m	Number of transmission attempts
a_i	Number of arrivals during the service time of the i^{th} packet
q_j	Number of arrivals in the j^{th} slot
c	Number of arrivals during the slot where a packet has arrived to an idle queue
$P(z)$	PGF of the population of the system at steady-state
$\bar{n}$	Average population of the system at steady-state
$\bar{d}$	Mean delay of a packet with equal arrival rates

$\bar{d}_i$	Mean delay of a packet in the ith user queue.
ρ	Utilization of a user with equal arrival rates
ρ_i	Utilization of the ith user
ρ_T	Total system utilization of all users or overall load
γ_k	Probability of collision for k busy queues
β_k	Backoff parameter for k busy queues.
b_i	Mean backoff duration of a user for the i^{th} attempt
$H_i(z)$	PGF of the distribution of the number of packets of the i^{th} user queue at steady-state
l_i	Number of packet arrivals to the i^{th} user during a slot.

Chapter 1

Introduction and Literature Review

1.1 Chapter Overview

This chapter prefaces the thesis by giving a brief introduction of NB-IoT and its features and applications, presenting the proposed model, describing the existing literature related to the model, and summarizing the contribution of the thesis compared to previous works. The organization of the thesis's chapters and sections is presented at the end of the chapter.

1.2 Introduction to NB-IoT

The problem of connecting a massive number of low-cost, low-power devices through a wide area network to achieve the requirements of the future Internet of Things (IoT) has been a topic of interest for several years now on the research field. Previously, the low throughput devices which are mostly simple sensors and actuators were connected to networks of high bandwidths along with other more complex devices, thus, using much more resources than they actually need. Several technologies have been developed as a solution to this problem such as Long Range (LORA) and Long Term Evolution Category-M1 (Cat-M1).

Narrow Band Internet of Thing (NB-IoT) is a Low Power Wide Area Network (LPWAN) radio technology standardized by the Third Generation Partnership Project (3GPP) on Release 13 [1]. NB-IoT obtained its name from its very low bandwidth of 180 KHz. It was developed to provide a wider coverage area and deep indoor penetration while maintaining very low power requirements that allows for a much longer battery life of up to 10 years. NB-IoT inherited many of the specifications of the legacy Long-Term Evolution technology (LTE) and it runs on licensed spectrum.

1.3 NB-IoT Features

NB-IoT has several features that lets it stand out compared to other LPWAN technologies. Some of the important features of NB-IoT are:

- Wide coverage and deep indoor penetration: achieving a wide coverage for IoT devices was always a challenge as it comes mainly at the cost of a higher power or resource usage. NB-IoT, however, was able to achieve a wide coverage area while maintaining the lowest power usage compared to other technologies thanks to the repetition scheme which will be explained in the next section.
- Low power: NB-IoT implemented several techniques to minimize the power consumption of the IoT users. Power Saving Mode (PSM) enables the device to enter a deep-sleep mode after notifying the base station where the device is unreachable for a longer time and thus saves more power than in IDLE mode [2]. PSM is particularly good for devices with periodic transmissions. Another power saving technique is extended Discontinuous Reception (eDRX) which increases the sleep cycle of the device in IDLE mode.
- Longer battery life: one of the main advantages of NB-IoT is the longer battery life thanks to the power saving techniques and the lower complexity requirements of the network. The battery of a normal IoT sensor is expected to live for up to 12.8 years for devices sending 200 bytes daily [3].
- Resource utilization efficiency: connecting a large number of IoT users through a channel of a very low bandwidth of 180 KHz is considered an improvement in channel resource utilization efficiency.
- Existing Infrastructure: Since NB-IoT inherits most of its mechanism from LTE, it can be implemented in the widely-spread existing LTE infrastructure. Additionally, NB-IoT can be deployed in three different frequency bands, LTE in-band, LTE guard band, and standalone deployment.
- Low cost: due to its lower power requirements, simple underlying techniques, and existing infrastructure, NB-IoT modules are expected to cost less than the GSM/GPRS.

LTE-M is considered one of the competing technologies for NB-IoT on the field of LPWA networks as it also benefits from the existing infrastructure of LTE [4]. Table. 1.1 lists some of the differences between NB-IoT and LTE-M.

Table 1.1 Comparison of LTE-M and NB-IoT Capabilities

Description	NB-IoT	LTE-M
Bandwidth	180 KHz to 200KHz	1.08 MHz
Maximum Coupling Loss	164 dB	156 dB
Indoor penetration	Excellent	Good
Coverage	7x the range of LTE Cat 1	4x the range of LTE Cat 1
Cost	Low	Higher
Peak Data Rate	Up to 250 Kbps (UL)	Up to 1 Mbps
Battery Lifetime	+10 years	Up to 10 years
Mobility	Limited	Yes
Frequency deployment	LTE in-band, LTE guard-band and standalone	LTE in-band
Duplex Mode	Half	Full
Latency	1.2 to 10 seconds	50 to 100 ms
Voice	No	Yes (VoLTE)
Power Consumption	Best at low data rate	Best at medium data rate

1.4 Repetition Scheme

When trying to provide high coverage for low power devices, most of the devices which are far from the base station will not be able to connect to it properly especially under bad channel quality due to their low complexity. This problem of coverage was solved thanks to the repetition scheme. When the received power of a certain user is very low, repetition of the same message by the transmitter for multiple times is effective to quell the bad channel quality, especially, when the receiver is able to combine the replicas of those repetitions to get the full message. However, to avoid forcing all devices to repeat their transmissions even when some of the devices have high power level, it was decided by 3GPP standards that in each NB-IoT network, there can be up to 3 coverage enhancement levels (CE) namely, normal coverage, extended coverage and extreme

coverage [5]. Each of these levels will have a unique repetition value such that the higher is the received power, the less repetitions are used by the UE. The corresponding Maximum Coupling Loss MCL is 144/154/164dB. It is important to mention that repetition is not meant to be a contention resolution scheme so it has no effect on whether there will be a collision between different users or not and it does not affect the backoff mechanism which is the scope of our work, therefore, we chose not to implement its effect which requires a power-based analysis, however, its time effect was taken into consideration.

1.5 NB-IoT Applications

NB-IoT is envisioned to revolutionize the industries that depends on data transmission in hard-to-reach locations such as deep underground, warehouses, and rural lands. There is a wide range of NB-IoT applications in various fields of business, industry, and healthcare sectors. NB-IoT can be implemented in smart metering application used by electricity, gas, and water industries as well as environmental monitoring [6-8]. Such meters would usually have a simple design and low transmitting power and they transmit data on a very infrequent basis making NB-IoT ideal for their cases. One of the various types of smart metering applications is the smart parking system and several works have suggested the use of NB-IoT to improve the performance of this system[9-12].

1.6 Proposed Model

Various studies are being done on different topics related to NB-IoT. The topic of interest of this thesis work is the random access procedures of NB-IoT. We concentrate on studying the performance of the random access procedures under the effect of adaptive backoff mechanism of NB-IoT by proposing an analytical model that enables determination of the mean delay of a packet imposed by the random access procedures.

Several works have studied the random access procedures of NB-IoT. In the following section, we will preview some of the important works related to our topic of interest, and in the section after, we will present the main contributions of our work compared to the previous studies.

1.7 Related Work

In this section, we will discuss the works of several papers concerned with the topic of performance evaluation and delay estimation of random-access procedures in the uplink channel of NB-IoT networks.

In [13], throughput of the random-access procedures of uplink channel in NB-IoT network under certain assumptions and restrictions has been studied. It was assumed that the network has a population of N users and that each user can have up to Q packets stored in its buffer. The analysis puts a heavy emphasis on the backoff mechanism of NB-IoT intended for contention resolution, a mechanism that is ignored by some other papers, usually, to simplify the analysis. The study is performed from the perspective of an observed UE and the queue in the buffer of the UE is assumed to be First-In-First-Out (FIFO) queue. According to their analytical model, the first step when a UE wants to initiate a random-access procedure is to choose a delay value from among W values, so that the user that chooses the smallest delay value will be the successful one in sending the preamble to the base station. A discrete-time Markov chain for the FIFO queue of the observed UE is used as a means to calculate important performance measures such as probability of successfully transmitting a preamble and probability of discarding a packet at the steady state of the system as shown in Fig.1.1, and from that, they derived the formulas of the throughput of the system under different scenarios. Each state of the Markov chain consists of two variables, the number of packets in the IoT device's buffer k, and the retransmission number for the head of queue packet r, which represents number of collisions the packet has suffered. A simulation model has been designed to verify the analytical model and was tested under several values of different parameters such as the number of users, packet generation rate and contention window size.

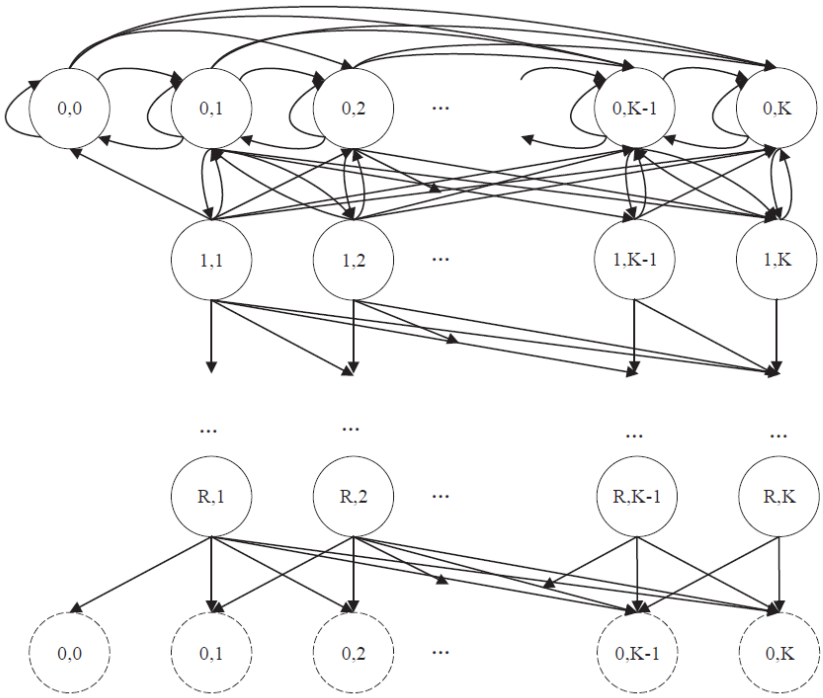

Fig.1.1 The Markov Chain of the system presented in [13]

One of the limitations of the model in [13] is the assumption that the buffer of the UE is limited to *K* packets. Additionally, they assumed that all UEs have the same packet arrival rate. The analytical model presented in [13] assumes that every user, from the start of random-access procedures, has to choose a delay value before sending the preamble and the user can send its preamble immediately once its backoff period has ended, however, according to 3GPP standards, backoff mechanism is used only when there is a collision between two or more preambles. Thus, when a user wants to transmit a preamble, it doesn't need to choose any backoff value, however, it has to wait till the start of the random access channel (RACH) cycle which is called "Nprach_periodicity" in the standards. Additionally, [13] is using a fixed backoff mechanism so that the backoff window size is constant. Furthermore, their analysis does not lead to closed form results, which is based on determining the transition probability matrix of the system and determining the probability distribution of the states from the solution of $\pi = \pi P$. The analytical model presented [13] has motivated us to investigate further in the backoff mechanism of NB-IoT by considering a more general model that is compatible with the 3GPP standards and removing some of the limitations of that model.

An analytical model for Random Access Channel (RACH) has been designed using stochastic geometry in [14]. This paper studies the effect of repetition on preamble transmission and RACH success probability from the power perspective. The objective of repetition is to help the users that experience bad channel conditions transmit their preambles successfully. The numerical results show that repetition improves RACH success probability under light traffic, but it has only small positive effect under heavy traffic at the cost of inefficient channel utilization. While the work in [14] models both repetition and preamble collision, it assumes a constant percentage of active devices to total devices at all times.

Authors in [15] calculated the probability of success of preamble in the random-access procedure, as well the average delay imposed on a packet in the random access procedures of Msg1 and Msg2. They additionally found the expression of normalized flow rate under the effect of different parameter values. The analysis assumes the success probability to depend on the number of active stations and the number of available subcarriers. Using all the parameters described in 3GPP standards such as Ra-ResponseWindowSize which is the window for Msg2, they derived the formulas of probability of success, the average delay, and normalized flow. They obtained the numerical results under various scenarios. The analysis in [15] has ignored several important factors that would affect the final results such as the assumption that number of active terminals and number of subcarriers are constant.

In [16], a simulation model is assumed to incorporate the whole random-access procedure starting from Msg1 till Msg4. A major contribution in this paper is the consideration of Msg3 collisions instead of Msg1 collisions. Several factors are assumed to be taken into consideration in their model such as several traffic models, a radio propagation model, a power SINR-based model for Msg3 and Msg1 transmission, Msg3 modulation and coding scheme and the power ramping step. The several factors included in their model would make it impossible to implement a complete analytical model, thus, they only implemented a simulation model. The numerical results obtained from the simulation shows the percentage of successful, retransmitted, and lost (or discarded) preambles under different traffic scenarios. In [16], despite that the simulation model is assumed to include many factors that play an important rule in real life scenario of any NB-IoT network, the lack of any analysis is a major deficiency.

An optimization of Random-Access Channel in NB-IoT has been done in [17] to configure the various parameters affecting the access success probability under a specified delay constraint. The analysis used a multi-band multi-channel slotted Aloha model, in which the multi-band describes the multiple CE levels of the system (for a maximum of 3 CE levels) and the multi-channel accounts for the various number of subcarriers that can be allocated to RACH channel. The main challenge in their optimization technique is to decide what is the best allocation of the 48 subcarriers of RACH among the 3 available CE levels. The presented model in [17] is one of the most comprehensive models to describe RACH in NB-IoT as it includes all of the main factors that affect RACH success such as the backoff mechanism, the multi-carrier and the multi-CE level system. However, the use of slotted Aloha model might not fit for some applications, particularly, when the traffic is heterogenous with different users having different packet generation rates, priority, or packet-loss tolerance such as in clustered or layered network where every cluster.

1.8 Contribution of the Thesis

Considering the previous works, the main contribution of this thesis can be summarized as follows:

- Developed an analytical model that includes the backoff mechanism of contention resolution of the random access procedures in NB-IoT technology. The model enables determination of the average delay imposed on packets from the random access procedures, the total utilization, and the probability of packets discarding.
- The model incorporates an adaptive backoff model with several backoff window sizes that varies continuously with the number of attempts made by the transmitting terminal.
- The model allows that UE having different packet arrival rates to their buffers compared to most of the previous works which assume a single arrival rate for all users.
- Every UE in the proposed model is assumed to have an infinite buffer for storing packets which makes the results of the analysis general for all cases as opposed to previous analyses that assume finite buffer size.

It is believed that the mentioned points allow for the recognition of this work's contribution compared to the existing literature.

1.9 Organization of the Thesis

The remainder of the thesis is organized as follows.

The chapter 2 describes the MAC layer in NB-IoT and its main features and functions. Then, it explains in details the random access procedures of NB-IoT, their purpose, and their main steps. Finally, it describes the collision process, the contention resolution and the backoff mechanism which is the scope of our work.

Chapter 3 presents the proposed system and explains the system model and the analysis. System specifications and assumptions are defined in system model as well as the adaptive backoff mechanism taken into consideration. The analysis is presented in three sections, discrete service time, and probability of success. In each section, we define all considered parameters and then we derive the required formulas in an organized sequence.

Chapter 4 presents the simulation model and provide the numerical results. It firstly explains the design of the system model with brief explanation of its algorithm. Then we provide the numerical results in graphs showing the delay of the packet and total utilization of all queues as a function of the total arrival rate under different scenarios of various parameters.

Conclusion and future work of the thesis are presented in chapter 5.

.

Chapter 2

MAC Sublayer in NB-IoT

2.1 Chapter overview

NB-IoT is built upon the existing LTE design in order to be coexistent with legacy LTE. Thus, it uses mostly the same medium access control (MAC) procedures used by LTE. Since our analysis is concerned with the Random Access (RA) procedures of NB-IoT network, and particularly, uplink channel, we will give a brief insight about those procedures and about MAC sublayer in general in this chapter.

2.2 MAC Sublayer Functions

Medium Access Control (MAC) is a lower sublayer that interacts directly with the physical sublayer and performs several important functions. One of the main functions of MAC is the random access procedures aimed at organizing the uplink connection and transmissions from User Equipment (UE) devices to the base station and resolving contention between different users.

Some of the other important functions of MAC layer [18] are listed below:

- Forming a single MAC PDU from multiplexing several RLC PDUs for uplink transmission
- Forming several RLC PDUs from demultiplexing the received MAC PDU
- Mapping of transport channels to logical and physical channels
- Hybrid Automatic Repeat Request (HARQ) Operation for error control and correction
- Organizing the discontinuous reception operation aimed at saving power

2.3 Mapping with Logical and Physical Channels

Fig. 2.1 shows the different uplink channels of each layer and their mapping. The transport channels of data link layer are Random Access Channel (RACH) and Uplink Shared Channel (UL-SCH). RACH is dedicated for transmitting preambles which is the first step in random access procedures. RACH is not mapped to any logical channel because preambles are initiated by MAC layer itself when there is a packet to send. Whereas UL-SCH is responsible for data transmission from the user to the evolved Node B (eNB). It is linked to three logical channels, common control channel (CCCH), dedicated control channel (DCCH) and dedicated traffic channel (DTCH).
In physical sublayer of NB-IoT, both uplink and downlink channels have a bandwidth of 180 KHz.

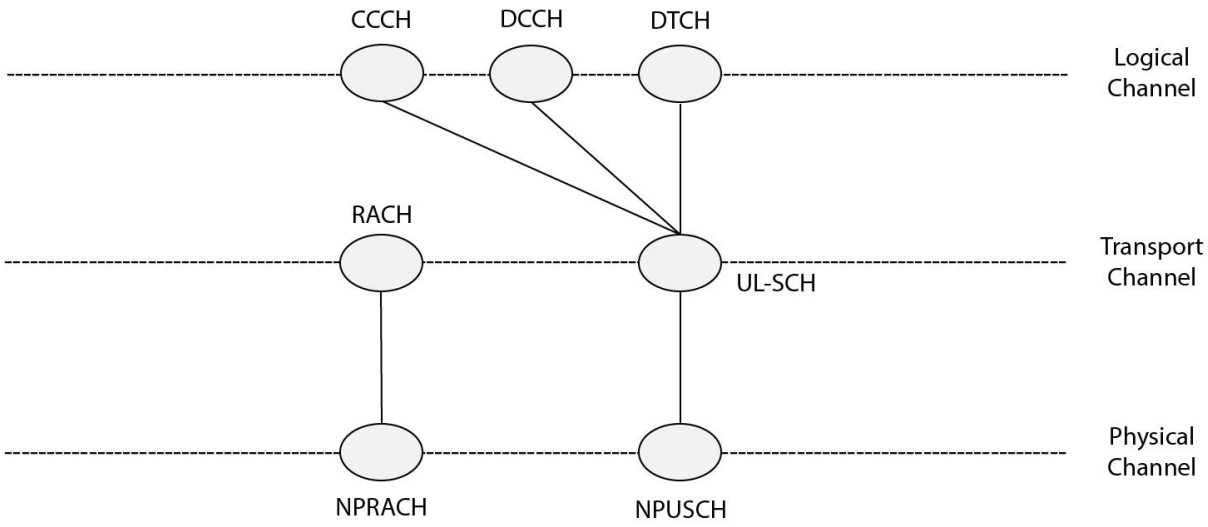

Fig. 2.1. Mapping of different uplink channels

The uplink channel of 180 KHz is divided into 12 sub-carriers each with 15 KHz for Narrowband Physical Uplink Shared Channel (NPUSCH). Whereas, for Narrowband Physical Random-Access Channel (NPRACH), the channel can be of 12, 24, 36 or 48 sub-carriers of 3.75 KHz each [19]. The dedicated number of subcarriers in the RACH resource and other MAC configuration parameters are sent by the RRC sublayer to MAC sublayer. The subcarriers in the RACH can be divided among the available CE levels and there can be up to 3 CE levels. These MAC

configuration parameters are broadcasted from eNB to RRC frequently. When RACH is set to have less than the maximum of 48 subcarriers, the remaining frequency resources will be assigned to UL-SCH. UL-SCH may support the tone spacing of 3.75 KHz (48 subcarriers) in certain cases. In radio frame (time domain), the assignment of uplink resources to NPRACH and NPUSCH is done using the MAC configuration parameter "NPRACH-Periodicity" which denotes the period of time (in ms) between two successive NPRACH resources. Thus, the time axis is practically divided into periods or slots. In each of these slots, there is one NPRACH resource and the remaining part of the slot contains NPUSCH resources. NPRACH-Periodicity is sometimes referred to as RA cycle or RA slot. Fig. 2.2 shows the assignment of NPRACH and NPUSCH in both time domain and frequency domain. In this figure, NPRACH has its maximum frequency resources of 48 subcarriers.

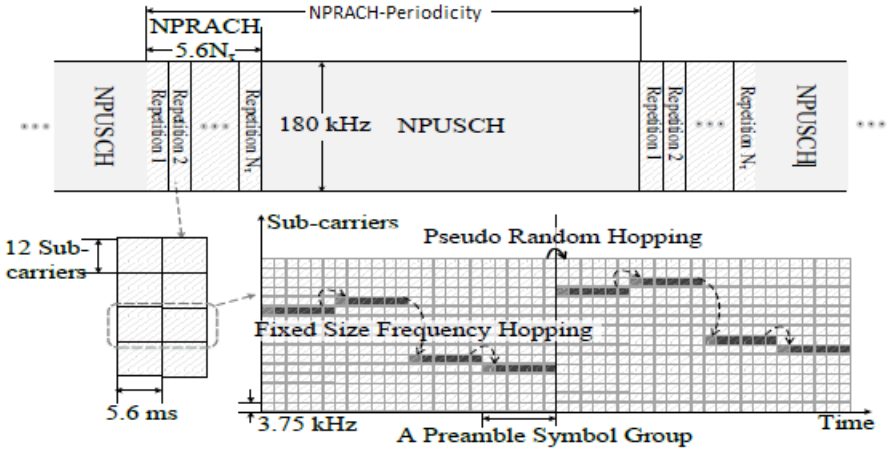

Fig. 2.2. Assignment of NPRACH and NPUSCH [14]

In NPRACH, a preamble is transmitted by the UE multiple times continuously without gaps according to a repetition value denoted by the 3GPP standards as "numRepetitionsPerPreambleAttempt". As explained previously in section 1.4, repetition of the preamble counters the effect of interference resulted from the bad channel quality. As a result, it will improve the base station ability to decode the preamble successfully by combining the replicas of the repeated preambles. Repetition rate is higher when the power level of the UE is low. As seen from Fig. 2.2, the amount of time assigned for NPRACH is equal to number of repetitions multiplied by the transmission time of one preamble (5.6 ms). Therefore, for higher repetition

value, NPRACH will have more time. However, the total slot time (NPRACH-Periodicity) also will be higher so that there are enough time resources for NPUSCH in one slot. We will refer to slot time in our analysis by T.

2.4 Random Access Procedures:

Random access procedures occur when a UE wants to initiate a connection with the base station. It aims to accomplish the uplink channel synchronization, gain uplink grant for RRC connection request, and contention resolution if occurred. When the UE successfully completes its RA procedures, it moves from the "idle" mode to the "connected" mode where it can transmit its unicast data to the BS [20].

The RA procedures as defined in 3GPP standards [18] consist of 4 main steps:

1) Random Access Preamble (RAP or Msg1): This is the first step when UEs initiate RA procedures. In this step, UE sends a preamble sequence to the base station and repeat it with a dedicated repetition number. Every preamble sequence consists of 4 symbol groups sent on 4 different subcarriers with a frequency hopping scheme as shown in Fig. 2.2. Each symbol group comprises a cyclic prefix (CP) and five OFDM symbols. After sending Msg1 and its repetitions, UE starts a timer called random access response window size and denoted in the standards by (*ra_ResponseWindowSize*). If the UEs do not receive a response before the timer expires, the preamble and its repetitions need to be retransmitted.

2) Random Access Response (RAR or Msg2): After the eNB receives the preamble, it sends the RAR to the UEs through the Narrowband Physical Downlink Control Channel (NPDCCH). RAR will give a grant to the UE to send Msg3. This grant consists of assigned resources in both time and frequency domain of NPUSCH. Assignment of grants depends on the load of uplink channel. As there are many message types that uses uplink channel, it is hard to estimate when the grant will be available. Base station uses a specific scheduling mechanism to assign these grants.
   RAR includes the Random-Access Radio Network Temporary Identity (RA-RNTI), the Timing Advance (TA), and a temporary Cell Radio Network Temporary Identifier (C-RNTI). If the UE

successfully receives its designated RAR, it concludes that the main part of RA procedures is completed successfully [20].

3) Scheduled transmission: This transmission carries Msg3. It is transmitted in NPUSCH on the grant given by base station in Msg2. After sending it, UE will wait for another window size to receive response (Msg4) from base station. This window size is denoted in the standards as mac-Contention-ResolutionTimer. If the UE does not receive Msg4 after expiration of the timer, it will need to repeat the RA procedures from the beginning by retransmitting a preamble again. Msg3 contains the RRC connection, reconfiguration requests, Data Volume and Power Headroom Report (DPR). DPR consists of Power Headroom Report (PHR) and Buffer Status Report (BSR). UEs can be allocated with the corresponding channel resources according to DPR.

4) Contention resolution: The contention resolution is also called Msg4. It is sent by the base station to the UE through Narrowband Physical Downlink Shared Channel (NPDSCH). The MAC PDU of Msg4 contains a contention resolution identity. After the UE receive the MAC PDU, it checks if its identity matchers the identity of the MAC PDU. If it matches, the UE consider the contention resolution procedures to be completed and move its RRC status from idle mode to connected mode. After that, the UE begins data transmission. Data transmission takes place in NPUSCH.
The previous steps are summarized in Fig. 2.3.

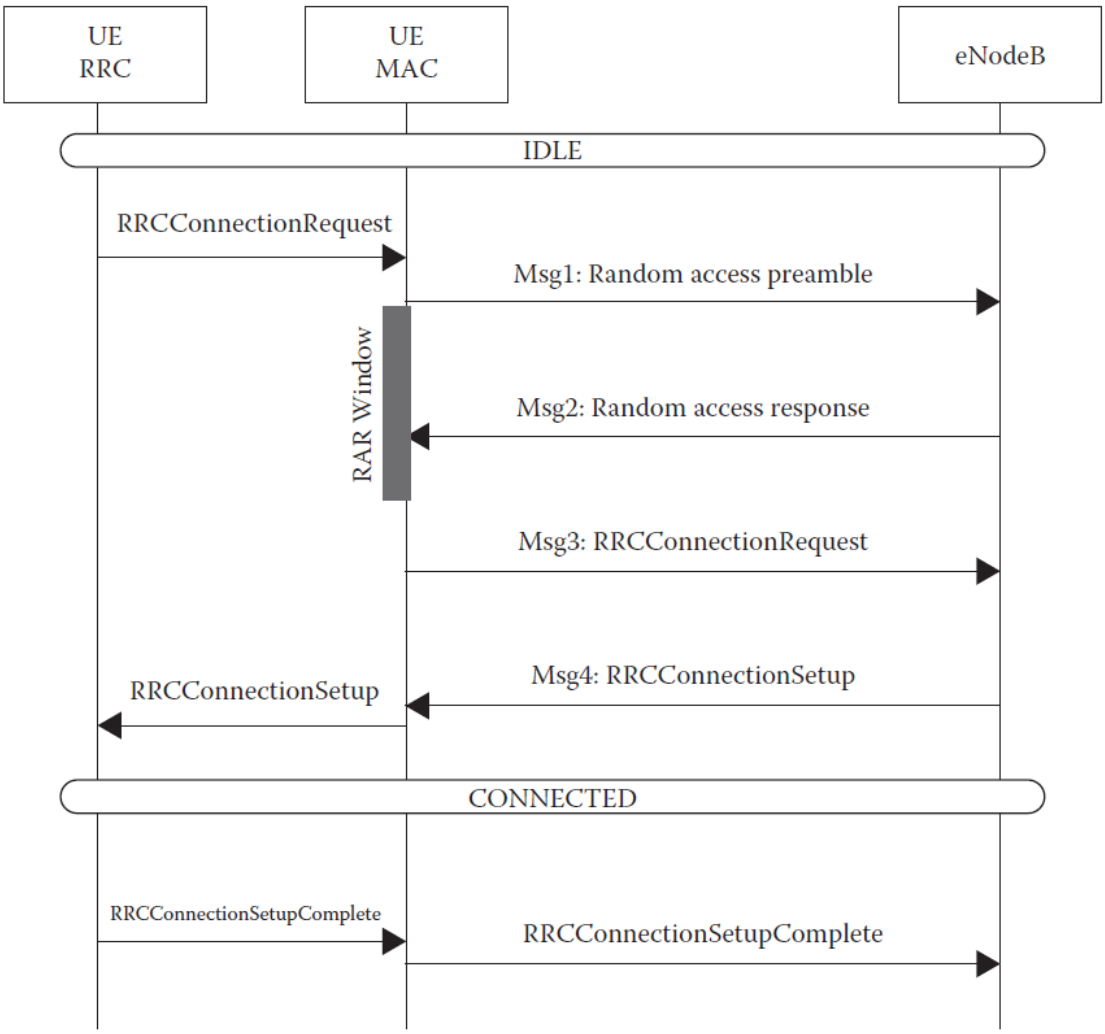

Fig. 2.3. Random Access Procedures [20]

Once a UE enters connected mode, eNB will start scheduling specific grants periodically for the UE to send the data packets. In case of exhaustion of the resources or under heavy traffic, the eNB orders the UE to return to the “idle” mode to free resources for other users. This process can be done also by the eNB to ensure a fair usage of resources among all connecting UEs. If the UE returns to “idle” mode and it has data packets to send, it needs to perform RA procedures again.

As mentioned before, after transmitting the preamble, the UE waits for a RAR window size duration till receiving the RAR. A period of 4ms elapse from the moment of finishing transmission of last preamble till the UE starts the timer of RAR window size. This 4ms period accounts for the time the UE spends to switch from transmission mode to receiving mode. RAR window size is one of the MAC configuration parameters and its value is measured in number of NPDCCH periods denoted in the standards as pp. Where one pp is the time interval between two successive NPDCCH opportunities. pp period can vary from being as small as 4ms and up to 2.3 minutes [21]. Its value is broadcasted by the eNB to the UE. The eNB choice for RAR window size depends on the number of repetitions on the NPDCCH and the CE level of the UE.

RA Procedures can be contention-based or contention-free. Contention-free RA procedures is done only when the eNB orders the UE to perform RA to establish a connection and then receive certain data from the eNB. In all other cases, RA procedures are contention-based.

In contention-based RA, preamble collision happens when two or more UEs existing in the same CE level attempt transmission in the same slot and choose the same subcarrier [14]. The main method of contention resolution stated in the 3GPP standards is the backoff mechanism which is the scope of our analysis. [22] studied another method of contention resolution by detecting the collided preambles.

Backoff mechanism aims to resolve the contention by letting the collided UEs wait for a random period of backoff time before retransmitting another preamble. When each UE chooses a certain backoff time, it is unlikely that two of them will choose the same values, thus, each user will reattempt their transmission at a different time, and as a result, avoiding a new collision.

When the UE is informed of the collision, it picks a random backoff value that is uniformly distributed in the interval range [0, Back-off]. Then, it waits for the chosen backoff time to expire before attempting to transmit again.

As per 3GPP standards, there is a set of values for the backoff window size parameter “Back-off” which represents the maximum backoff time that the UE can wait for. These values are known by

all UEs. When a collision happens, the UE may receive a backoff index message from the base station to inform about which Back-off window size to choose. Table 2.1 lists the set of all Back-off window sizes and their corresponding index as determined by 3GPP standards [19].

Table 2.1 set of values for the backoff parameter "Back-off"

Back-off index	Back-off Window Size in ms
0	0
1	256
2	512
3	1024
5	4096
6	8192
7	16384
8	32768
9	65536
10	131072
11	262144
12	524288

Since UE can only transmit at the start of a RA slot, the actual backoff time will be in number of slots. As seen in Fig. 2.4., when the user picks any backoff time that is less than or equal to $T - X$ (where X is the time till the user is informed of the collision), it can transmit at the next slot directly. Practically, this means that the UE will at least spend one slot time after collision.

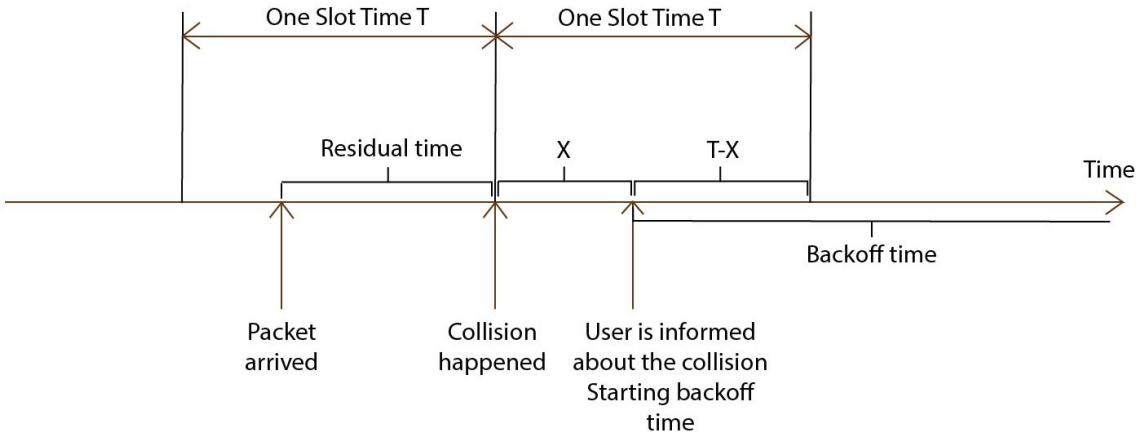

Fig. 2.4. Time Diagram of a case of collision

As eNB can inform the UE of which Back-off value it should chose, several backoff strategies can be considered by the eNB manufacturer such as adaptive backoff or fixed backoff in all cases. Adaptive backoff can be according to several factors such as traffic intensity, delay tolerance or number of transmission attempts.

As we will show in the next section, we study the adaptive backoff with the number of transmission attempts. In this backoff scheme, each time the UE suffers a new collision (the number of transmission attempts increase by one), the "Back-off" value gets doubled. This scheme is similar to the backoff mechanism of CSMA/CA.

The UE can retransmit its preamble up to a maximum number of times denoted in the standards as "preamble-TransMax-CE". This maximum number of transmissions is another parameter of the MAC configuration parameters sent by the RRC sublayer. Similar to the backoff mechanism, the maximum number of transmissions can be fixed or adaptive. Table 2.2 lists the set of possible values of some of the MAC configuration parameters.

Table 2.2 Set of values of MAC configuration parameters

Name in 3GPP Standards	Meaning	Set of Values

numRepetitions-PerPreambleAttempt	Number of repetitions of the preamble	{1, 2, 4, 8, 16, 32, 64, 128}
preambleTransMax-CE	Maximum number of transmission attempts	{3, 4, 5, 6, 7, 8, 10, 20, 50, 100, 200 }
nprach-Periodicity	Slot time or the time between two successive NRACH resources	{40, 80, 160, 240, 320, 640, 1280, 2560} ms
nprach-NumSubcarriers	Number of subcarriers in the NRACH resource	{12, 24, 36, 48}
ra-ResponseWindowSize	RAR window size	{ pp2, pp3, pp4, pp5, pp6, pp7, pp8, pp10 }

From the table above, high values of nprach-Periodicity do not accommodate the lower values of window sizes shown in Table 2.1, therefore, it is practical to take into consideration the higher values of backoff window sizes when having a high RA slot time.

Chapter 3

System Model and Analysis

3.1 Introduction

We study the backoff mechanism of random-access procedures of NB-IoT system with the purpose of finding probability of access success and the mean delay imposed on a packet by the RA procedures in different traffic scenarios and under the effect of several NB-IoT MAC layer parameters.

We consider a single cell of an NB-IoT network with number of nodes where each node will be a user equipment (UE) device. The UE devices taken into consideration in our analysis might be a gateway for several sensors in a wireless sensor network (WSN), a head of a cluster in a clustered NB-IoT network, or an ordinary NB-IoT device with higher arrival rate than the average NB-IoT device. All of the aforementioned devices will allow queueing of the MAC sublayer PDU packets in their buffers. In layered and clustered networks, only gateways and cluster heads will perform random access procedure with the base station.

3.2 System Model

The objective of the analysis is determining the PGF of the probability distribution of the number of packets in a user queue in the steady-state and from there obtaining the mean packet delay. In this section, the system model under consideration will be described.

3.2.1 Assumptions

For simplification, we consider one CE level to exist in the system. We assume that there is one subcarrier for preamble transmission so that only one user can transmit its preamble successfully during a slot. As shown in Fig. 3.1, it is assumed that whether or not transmission of Msg1 is

successful, it will be resolved during the NPUSCH period following transmission of this message. Furthermore, if Msg1 transmission is successful, transmission of data packet will take place during that NPUSCH period.

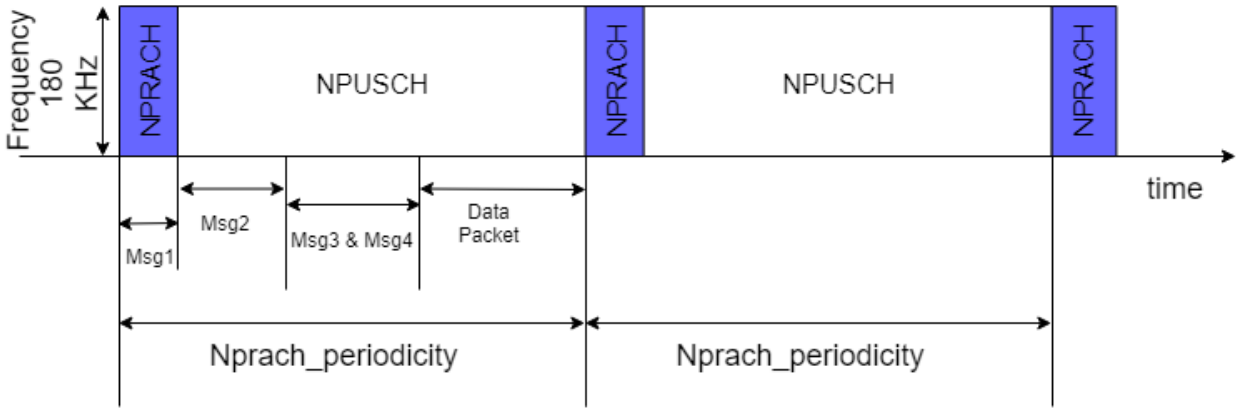

Fig. 3.1. Uplink physical channels in NB-IoT

The UE devices will be referred to as users. It will be assumed that the cell is serving a total of N users. Arrival of packets to each user will be according to a Poisson process with a different arrival rate. It is assumed that every user has an infinite buffer for storing packets. The users serve the packets in their queues according to first-come-first-served (FCFS) service discipline. Thus, the amount of time that a packet spends in the system consists of its waiting time in the queue plus its service time at the head-of-line (HOL). When the packet reaches HOL, the user initiates a random-access procedure by transmitting a random-access preamble (RAP), it is allowed only at the start of Nprach-Periodicity or RA slot. Duration of time slot will be denoted by T. To ensure the balance of the proposed queuing model and a fair usage of resources among all users, it is assumed that the user will initiate a random-access procedures for each packet in its buffer.

Successful RA procedures and data packet transmission are assumed to take one slot. After the user transmits a packet successfully, it starts RA procedures of the next packet in its buffer if it has any. When a packet is transmitted, it may experience a collision, then it has to backoff before retransmission. To determine the PGF of user queue length, probability distribution of service time is needed, which depends on the back-off mechanism.

3.2.2 Backoff Mechanism

In this subsection, we will describe the back-off mechanism. As mentioned in the previous chapter, backoff mechanism is used to resolve the contention between collided users. For our model, collision happens when two or more users attempt preamble transmission during the same slot. All of the colliding users will retransmit their preambles after a random backoff time. We assume that the user is informed of the collision before the start of the next slot. Therefore, the backoff window size starts from one slot which corresponds to the slot where the collision happened.

The adaptive backoff mechanism under study doubles the window size of the backoff time each time a new collision happens. Users choose backoff times according to uniform distribution in the range $[1, \omega_j]$ where ω_j is the maximum backoff window size in number of slots in the j^{th} transmission.

The set of values of the backoff window sizes defined by 3GPP standards (shown previously in Table 2.1) are in milliseconds but in practice backoff times will be expressed in slots. The Table 3.1 presents backoff window sizes in terms of number of slots for two different slot durations.

Table 3.1 Backoff window sizes in number of slots for two slot durations

Backoff window size in ms	Backoff window size in slots (slot duration= 40ms)	Backoff window size in slots (slot duration= 80ms)
256	7	4
512	13	7
1024	26	13
4096	103	52
8192	205	103
16384	410	205
32768	820	410
65536	1639	820
131072	3277	1639
262144	6553	3277

524288	13107	6553

Maximum number of allowed transmission attempts for a packet until its success is denoted as Q. In the system, Q is considered to be constant such that it cannot be changed while the system is running. Thus, different values of Q in the results will be shown in several scenarios. When Q exceeds 12 which is the number of window sizes, window size will not double any more and this maximum window size will be used in all of the remaining attempts.

When the number of unsuccessful transmission attempts of a packet exceeds Q, the packet will be discarded, and the user will attempt transmission of a new packet if its queue is not empty.

3.3 Analysis of Packet Service Time

In this section, we will determine probability distribution of packet service time in number of slots. The packet service terminates either by successful transmission or discarding of a packet. Let us introduce the following notations:

B_i : duration of the backoff time in number of slots for the i^{th} transmission.

m : number of transmissions of a packet.

Y : service time of a packet in number of slots.

$B_i(z)$: PGF of the probability distribution of the random variable B_i.

$Y(z)$: PGF of the probability distribution of the random variable Y.

P_s : Probability of successfully transmitting a packet during a slot

B_i is uniformly distributed random variable that represents the number of backoff slots a packet needs to wait in its i^{th} transmission. $B_i \in [1\ \omega_i]$ where ω_i is the maximum number of slots the packet can wait for the i^{th} transmission. Thus, probability distribution of B_i is given by,

$$Prob(B_i = k) = \frac{1}{\omega_i}, \qquad 0 \leq k \leq \omega_i \tag{3.1}$$

From the definition of PGF,

$$B_i(z) = E[z^{B_i}] = \sum_{k=0}^{\omega_i} z^k Prob(B_i = k) \tag{3.2}$$

Then,

$$B_i(z) = \frac{1}{\omega_i} \sum_{k=0}^{\omega_i} z^k \tag{3.3}$$

Which results in,

$$B_i(z) = \frac{z}{\omega_i} \frac{1 - z^{\omega_i}}{1 - z} \tag{3.4}$$

Next, we determine the discrete service time of a packet.

$$Y = 1 + \sum_{i=1}^{m} B_i \tag{3.5}$$

Where one outside the summation corresponds to the slot for the transmission time of the packet. The probability generating function of service time of a packet is given by,

$$Y(z) = E[z^Y] = E[z^{1+\sum_{i=1}^{m} B_i}] \tag{3.6}$$

In the above formula, we have two random variables namely, m and B_i, so first let us determine the conditional PGF of $Y(z)$ when the number of transmissions m is constant.

$$Y(z|m = k) = E\left[z^{1+\sum_{i=1}^{k} B_i}\right] \tag{3.7}$$

$$Y(z|m = k) = zE\left[z^{\sum_{i=1}^{k} B_i}\right] \tag{3.8}$$

Since random backoff times are independent of each other,

$$Y(z|m=k) = z\prod_{i=1}^{k} B_i(z) \tag{3.9}$$

Next, we uncondition the last formula with respect to the number of transmissions,

$$Y(z) = \sum_{k=1}^{Q} Y(z|m=k)Prob(m=k) \tag{3.10}$$

Assuming that probability of successful transmission is same in all transmissions and independent of the other transmissions. Then, probability distribution of number of transmissions, Pr (m), is given by,

$$\Pr(m=k) = \begin{cases} P_s(1-P_s)^{k-1}, & for \quad k = 1, \dots, Q-1 \\ P_s(1-P_s)^{Q-1} + 1 - \sum_{j=1}^{Q-1} P_s(1-P_s)^{j-1}, & for\ k = Q \end{cases} \tag{3.11}$$

Where P_s is the probability of successfully transmitting a packet during a slot. P_s value will be found in section 3.5. Q is the maximum number of transmissions and probability of having Q transmissions include both cases of either success or the discarding case.
We note that the probability of discarding a packet is given by,

$$Pr(packet\ discarding) = 1 - \sum_{j=1}^{Q} P_s(1-P_s)^{j-1} \tag{3.12}$$

Next, we determine the unconditioned PGF of the service time $Y(z)$:

$$Y(z) = \sum_{k=1}^{Q-1} zP_s(1-P_s)^{k-1} \prod_{i=1}^{k} B_i(z) + zPr(m=Q)\prod_{i=1}^{Q} B_i(z) \tag{3.13}$$

From the PGF of discrete service time in equation (3.13), we can obtain the mean service time. Firstly, we will determine the derivative of $Y(z)$,

$$
\begin{aligned}
Y'(z) &= \frac{dY(z)}{dz} \\
&= \sum_{k=1}^{Q-1} P_s\,(1-P_s)^{k-1} \prod_{i=1}^{k} B_i(z) + \sum_{k=1}^{Q-1} z P_s (1-P_s)^{k-1} \sum_{j=1}^{k} B'_j(z) \prod_{\substack{i=1 \\ i\neq j}}^{k} B_i(z) \\
&\quad + \Pr(m=Q) \prod_{i=1}^{Q} B_i\big(q(z)\big) + z \Pr(m=Q) \sum_{j=1}^{Q} B'_j(z) \prod_{\substack{i=1 \\ i\neq j}}^{Q} B_i(z)
\end{aligned}
\tag{3.14}
$$

The mean packet service time, $E[Y]$, is determined by substituting z=1 in the above. Noting that at z=1, $B_i(z)$ terms become one,

$$
\begin{aligned}
E[Y] &= Y'(z)|_{z=1} \\
&= 1 + \sum_{k=1}^{Q-1} P_s (1-P_s)^{k-1} \sum_{j=1}^{k} B'_j(1) + \Pr(m=Q) \sum_{j=1}^{Q} B'_j(1)
\end{aligned}
\tag{3.15}
$$

3.4 Derivation of the PGF of the User Queue Length Distribution

In this section, we derive the PGF of the user queue length distribution by embedding at the departure points of the packets. It is assumed that the arrival of the packets to each user queue is according to a Poisson process with the same arrival rate. Initially, it will be assumed that all users have the same arrival rate and in a later section results will be extended to the case of different rates. For this, let us introduce the following notation,

λ : arrival rate of packets per slot to a user queue.

n_i : number of packets in a user's queue at the time of service completion of the ith packet.

$P_i(z)$: PGF of the distribution of number of packets at the ith imbedded point.

a_i : number of packet arrivals during the service time of the ith packet.

q_j : number of packet arrivals during the jth slot.

c : number of packet arrivals during a slot that at least one packet has arrived.

Y : service time of a packet in number of slots.

ρ : utilization factor or probability of having a busy user queue.

From the definition of q_j, since arrival of the packets is according to a Poisson process and the number of arrivals in different slots are independent of each other,

$$Prob(q_j = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad for\ k = 0,1,2,\cdots \tag{3.16}$$

$$Prob(c = k) = \frac{Prob(q_j = k)}{1 - Prob(q_j = 0)}, \quad for\ k = 1,2,\cdots \tag{3.17}$$

The PGF of probability distribution of q_j is given by,

$$q_j(z) = E[z^{q_j}] = e^{-\lambda(1-z)} \tag{3.18}$$

Defining $C(z)$ also as the PGF of c,

$$C(z) = \mathrm{E}[z^c] = \sum_{k=1}^{\infty} z^k \frac{e^{-\lambda}(\lambda)^k}{(1-e^{-\lambda})k!} = \frac{e^{-\lambda(1-z)} - e^{-\lambda}}{(1-e^{-\lambda})} \tag{3.19}$$

From its definition, a_{i+1} is given by,

$$a_{i+1} = \sum_{j=1}^{Y} q_j \tag{3.20}$$

Next, let us determine the equation describing the system at the embedded points. Thus, we will express number of packets in the system at the (i+1)st imbedded point to the number of packets at the i^{th} embedded point. There are two cases to be considered,

i) In this case, the user queue is busy following the departure of the i^{th} packet. From Fig. 3.2, the equation describing the system is given by,

$$n_{i+1} = n_i - 1 + a_{i+1} \ , \qquad for \ n_i > 0 \tag{3.21}$$

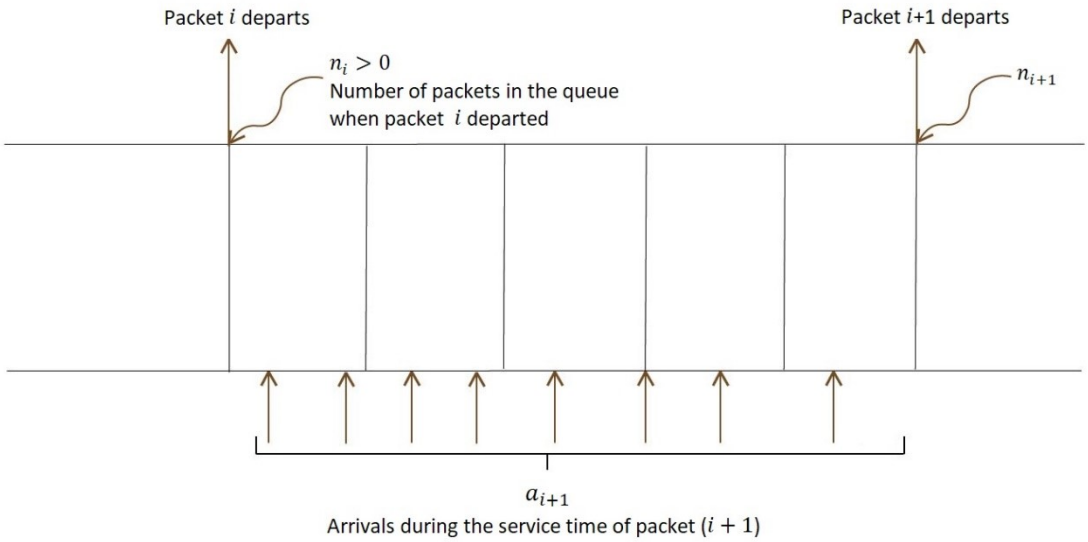

Fig. 3.2: Markov Chain Imbedded points when ($n_i > 0$)

ii) In this case, user queue is idle following the i^{th} departure. From Fig. 3.3, the equation describing the system is given by,

$$n_{i+1} = c + a_{i+1} - 1 \quad , \qquad for \ n_i = 0 \tag{3.22}$$

Where c is defined as the number of packet arrivals during the slot where the $(i+1)^{st}$ packet has arrived. As such, c must be greater than or equal to one.

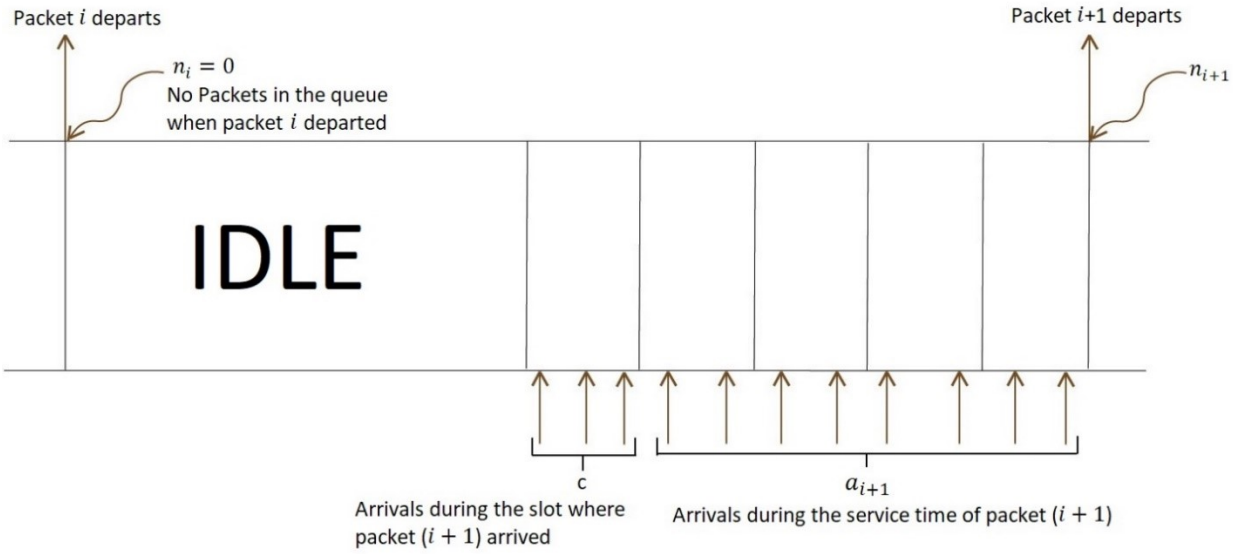

Fig. 3.3: Markov Chain Imbedded points when ($n_i = 0$)

The two equations (3.21) and (3.22) describing the system can be rewritten jointly as follows,

$$n_{i+1} = n_i - 1 + a_{i+1} + [1 - u(n_i)]c \tag{3.23}$$

Where $u(n_i)$ is the unit step function,

$$u(n_i) = \begin{cases} 0 & when\ n_i = 0 \\ 1 & when\ n_i > 0 \end{cases} \tag{3.24}$$

Next, let us define PGF of the distribution of number of packets at the i^{th} embedded point as follows,

$$P_i(z) = E[z^{n_i}] = \sum_{k=0}^{\infty} z^k Prob(n_i = k) \tag{3.25}$$

Next, raising both sides of equation (3.23) to the power of z and taking expectation gives,

$$P_{i+1}(z) = E\left[z^{n_i - 1 + a_{i+1} + (1-u(n_i))c}\right] \tag{3.26}$$

The number of arrivals during the service time of a packet is independent of the number of packets already in the system, therefore,

$$P_{i+1}(z) = E\left[z^{n_i - 1 + (1-u(n_i))c}\right] E[z^{a_{i+1}}] \tag{3.27}$$

Since number of arrivals during service times are independent identically distributed (i.i.d), we can define the PGF of the packet arrivals during a service time as follows,

$$A(z) = E[z^{a_{i+1}}] \tag{3.28}$$

Then, using equation (3.20), we can find the PGF of the arrivals during a service time,

$$A(z) = Y(z)|_{z=q(z)} \tag{3.29}$$

Now, we can rewrite $P_{i+1}(z)$,

$$P_{i+1}(z) = E\left[z^{n_i - 1 + (1-u(n_i))c}\right] A(z) \tag{3.30}$$

Next from the definition of expectation,

$$P_{i+1}(z) = A(z)\left\{\sum_{k=1}^{\infty} z^{k-1} \Pr(n_i = k) + E[z^{-1+c}] \Pr(n_i = 0)\right\} \tag{3.31}$$

Taking expectation with respect to *c*, we get,

$$P_{i+1}(z) = A(z)\left\{\sum_{k=1}^{\infty} z^{k-1} \Pr(n_i = k) + z^{-1} C(z) \Pr(n_i = 0)\right\} \tag{3.32}$$

Where $C(z)$ is given by (3.19). Completing the summation in the above to a PGF, we will have,

$$P_{i+1}(z) = A(z)\{z^{-1}[P_i(z) - \Pr(n_i = 0)] + z^{-1}C(z)\Pr(n_i = 0)\} \tag{3.33}$$

Next let us define the following notation for steady-state,

$$\begin{aligned} n &= \lim_{i\to\infty} n_i \\ P_k &= \Pr(n = k) \\ P(z) &= \lim_{i\to\infty} P_i(z) \end{aligned} \tag{3.34}$$

Taking limit of equation (3.33) as $(i \to \infty)$, we will have,

$$P(z) = A(z)\{z^{-1}C(z)P_0 + z^{-1}(P(z) - P_0)\} \tag{3.35}$$

Finally, we can solve the above equation for $P(z)$,

$$P(z) = \frac{A(z)P_0[C(z) - 1]}{z - A(z)} \tag{3.36}$$

Next, we determine the unknown P_0 from the normalization condition,

$$P(z)|_{z=1} = 1 = \frac{A(1)P_0[C(1) - 1]}{1 - A(1)} = \frac{0}{0} \tag{3.37}$$

As seen above, normalization condition results in 0/0 indeterminacy, to solve this issue, we need to apply L'Hopital's rule to obtain the result. Following application of L'Hopital's rule and simplification, we determine P_0,

$$P_0 = 1 - \rho = \frac{1 - A'(1)}{C'(1)} \tag{3.38}$$

In the above we also note that P_0 is the compliment of the utilization factor ρ. Substitution of the expression for P_0 in (3.36) determines PGF of the probability distribution of the number of packets in a user queue as a function of the probability of successful packet transmission, P_s, at the steady-state.

Next, we will determine average number of packets in a user's queue at an embedded point at the steady-state, $\bar{n}$,

$$\bar{n} = \frac{dP(z)}{dz}\Big|_{z=1} = \frac{\{A'(z)P_0(C(z)-1) + A(z)P_0(C'(z))\}\{z - A(z)\} - \{1 - A'(z)\}\{A(z)P_0(C(z)-1)\}}{(z - A(z))^2}\Bigg|_{z=1} \tag{3.39}$$

As seen from the above derivative, we have the term $(z - A(z))^2$ in the denominator. When substituting at z=1, this term will become zero, similarly, the numerator will tend to zero as well, resulting in a 0/0 indeterminacy. Thus, we need to apply L'Hopital's rule twice which results in the following result,

$$\bar{n} = \frac{dP(z)}{dz}\Big|_{z=1} = \frac{\{2C'(1)A'(1)P_0 + P_0C''(1)\}(1 - A'(1)) + \{A''(1)C'(1)P_0\}}{2(1 - A'(1))^2} \tag{3.40}$$

Thus, we can see that we only need to find $A'(1)$, $A''(1)$, $C'(1)$ and $C''(z)$. Next, we show derivation of these quantities,

To find $A'(z)$, we can use the average of Y obtained in equation (3.15) as shown in the following:

$$E[A] = A'(z)|_{z=1} = q'(1)Y'|_{z=q(1)} = \lambda Y'(1) \tag{3.41}$$

A similar approach can be followed to obtain $A''(1)$ as follows,

$$A''(z)|_{z=1} = \lambda^2 Y'(1) + \lambda^2 Y''(1) \tag{3.42}$$

Now, we will get the derivatives at $z = 1$ for the remaining functions we have, namely, $B_i(z)$, $q(z)$ and $C(z)$. We note that substituting $z = 1$ at any of these functions will have a result of one due to normalization rule.

$$B'_i(z)|_{z=1} = \frac{d}{dz}\left\{\frac{z}{\omega_i}\frac{1-z^{\omega_i}}{1-z}\right\}|_{z=1} = \left(\frac{\omega_i + 1}{2}\right) \tag{3.43}$$

$$q'(z)|_{z=1} = \frac{d}{dz}\left\{e^{-\lambda(1-z)}\right\}|_{z=1} = \lambda \tag{3.44}$$

$$C'(z)|_{z=1} = \frac{d}{dz}\left\{\frac{e^{-\lambda(1-z)} - e^{-\lambda}}{1-e^{-\lambda}}\right\}|_{z=1} = \frac{\lambda}{1-e^{-\lambda}} \tag{3.45}$$

$$B_i''(z)|_{z=1} = \frac{d2}{dz^2}\left\{\frac{z}{\omega_i}\frac{1-z^{\omega_i}}{1-z}\right\}|_{z=1} = \frac{1}{6}\left(2{\omega_i}^2 + 3\omega_i + 1\right) \tag{3.46}$$

$$q''(z)|_{z=1} = \frac{d2}{dz^2}\left\{e^{-\lambda(1-z)}\right\}|_{z=1} = \lambda^2 \tag{3.47}$$

$$C''(z)|_{z=1} = \frac{d2}{dz^2}\left\{\frac{e^{-\lambda(1-z)} - e^{-\lambda}}{1-e^{-\lambda}}\right\}|_{z=1} = \frac{\lambda^2}{1-e^{-\lambda}} \tag{3.48}$$

Finally, we obtain the average packet delay through Little's result,

$$\bar{d} = \frac{Average\ queue\ length}{Arrival\ rate\ of\ packets} = \frac{\bar{n}}{\lambda} \tag{3.49}$$

Simplifying the above equation with equations (3.40 to 3.48) we will get $\bar{d}$ as a function of λ, $Y'(1)$ and $Y''(1)$ as follows,

$$\bar{d} = Y'(1) + \frac{1}{2} + \frac{\lambda[Y'(1) + Y''(1)]}{2[1 - \lambda Y'(1)]} \tag{3.50}$$

All of these previous results, including the average population of the system at steady-state, are functions of probability that transmission of a packet is successful, P_s , which is unknown.

3.5 Analysis of Probability of Success

In this section, we will derive the formula of probability of success P_s. To determine the probability of success, we used the backoff analysis of CSMA/CA model in [23], as the two have similar backoff mechanisms. The CSMA/CA analysis in [23] assumes that user queues are saturated with packets, thus, users always have packets to transmit. In the system under consideration, user queues are not always busy, therefore, number of users in contention is a random variable. To adopt the analysis in [23], it will be assumed that the number of users with a packet to transmit is constant and then the result will be unconditioned with respect to this variable.

In [23], it has been shown that the overall average backoff time and probability that a user will suffer a collision during its transmission are related to each other. It is assumed that the overall backoff times are exponentially distributed. Let β_k denote parameter of this exponential distribution and γ_k probability of collision when there are k number of users that have packets to transmit, then, from [23], these two variables are related by the following two equations,

$$G(\gamma_k) = \beta_k = \frac{1 + \gamma_k + \gamma_k^2 + \cdots + \gamma_k^{Q-1}}{b_1 + b_2\gamma_k + b_3\gamma_k^2 + \cdots + b_Q\gamma_k^{Q-1}} \tag{3.51}$$

$$\Gamma(\beta_k) = \gamma_k = 1 - e^{-(k-1)\beta_k} \tag{3.52}$$

Where b_i is the mean backoff duration of a user for the i^{th} attempt in number of slots and Q is the maximum number of packet transmissions. Since the backoff durations are uniformly distributed, with i^{th} backoff time $B_i \in [1\ \omega_i]$, then,

$$b_i = \frac{\omega_i + 1}{2} \tag{3.53}$$

Simultaneous solution of equations (3.51, 3.52) determines probability of collision, γ_k, as a function of the number of contending packets *k*. The number of contending packets will be given by the number of busy queues. Next let γ denote the unconditional probability of collision and η the number of user queues that are busy at the steady-state, then,

$$\gamma = \sum_{k=1}^{N} \gamma_k \frac{Prob(\eta = k)}{1 - Prob(\eta = 0)} \tag{3.54}$$

We assume that probability of user queues being busy are independent of each other. Since each of the user queues is busy with probability ρ, probability mass function of η is given by,

$$Prob(\eta = k) = \binom{N}{k} \rho^k (1-\rho)^{N-k}, \quad k = 0,1, \dots, N \tag{3.55}$$

Finally, the probability of successfully transmitting a packet in a slot P_s is given by,

$$P_s = 1 - \gamma \tag{3.56}$$

From the above it may be seen that P_s is a function of ρ. From (3.38), ρ is given by,

$$\rho = 1 - \frac{1 - A'(1)}{C'(1)} \tag{3.57}$$

Substituting for $A'(1)$ and $C'(1)$ from (3.41, 3.45) respectively,

$$\rho = 1 - \frac{\left(1 - e^{-\lambda}\right)[1 - \lambda E[Y]]}{\lambda} \tag{3.58}$$

Where $E[Y]$ is given by equation (3.15) which is a function of P_s. From simultaneous solution of (3.56, 3.58) P_s and ρ may be determined.

3.6 Analysis of Unequal Arrival Rates

In the previous sections, we assumed that all users have the same packet arrival rate. In this section, we will generalize the previous analyses to find the mean packet delay and probability of success in the case of different arrival rate for each user. We note that service time of a packet in all user queues will be same. To find mean packet delay, we first need to determine the PGF of each user's queue length at steady-state.

For this, let us introduce the following notation,

λ_i : arrival rate of packets per slot to the i^{th} user queue.

l_i : number of packet arrivals to the i^{th} user during a slot.

$A_i(z)$: PGF of the distribution of number of packets arrivals to the i^{th} user during a service time.

$C_i(z)$: PGF of the distribution of number of packets arrivals to the i^{th} user during a slot that at least one packet has arrived.

$H_i(z)$: PGF of the distribution of the number of packets of the i^{th} user queue at steady-state.

ρ_i : utilization factor of the i^{th} user.

Firstly, we define l_i as the number of packet arrivals to the i^{th} user during a slot. Since arrival of the packets is according to a Poisson process and the number of arrivals in different slots are independent of each other,

$$Prob(l_i = k) = \frac{e^{-\lambda_i}{\lambda_i}^k}{k!}, \quad for\ k = 0,1,2,\cdots \tag{3.59}$$

As number of arrivals during slots are independent identically distributed (i.i.d), we can find the PGF of the above number of arrivals during a slot as follows,

$$L_i(z) = E[z^{l_i}] = e^{-\lambda_i(1-z)} \tag{3.60}$$

From the definition of $A_i(z)$ and similar to equation (3.29),

$$A_i(z) = Y(z)|_{z=L_i(z)} \tag{3.61}$$

Where $Y(z)$ is given by equation (3.13).
Similar to equation (3.19), we can find $C_i(z)$ as per its definition,

$$C_i(z) = \frac{e^{-\lambda_i(1-z)} - e^{-\lambda_i}}{(1 - e^{-\lambda_i})} \tag{3.62}$$

Similar to equation (3.58), we can get the utilization factor for the i^{th} user ρ_i,

$$\rho_i = 1 - \frac{\left(1 - e^{-\lambda_i}\right)[1 - \lambda_i E[Y]]}{\lambda_i} \tag{3.63}$$

Then, from equation (3.36), we can get the PGF of the probability distribution of the number of packets of the i^{th} user queue at steady-state $H_i(z)$,

$$H_i(z) = \frac{A_i(z)(1 - \rho_i)[C_i(z) - 1]}{z - A_i(z)} \tag{3.64}$$

We denote the average number of packets in the i^{th} user queue at steady-state by $\overline{H}_i$. From the above formula we can get $\overline{H}_i$ as,

$$\overline{H}_i = \frac{dH_i(z)}{dz}|_{z=1} \tag{3.65}$$

Then, we can finally obtain the mean packet delay of the i^{th} user $\bar{d}_i$ using Little's result,

$$\bar{d}_i = \frac{\bar{H}_i}{\lambda_i} \tag{3.66}$$

Assuming that probability of success is independent and the same for all transmissions, the probability of success formula shown in equation (3.56) is valid for the case of different arrival rates. However, equation (3.55) needs to be rewritten as each user in this case has its unique utilization factor and the formula will differ. For that, the PGF of the distribution of the number of users with busy queues $\eta(z)$ is obtained as follows,

$$\eta(z) = \sum_{k=0}^{N} Pr(\eta = k) z^k \tag{3.67}$$

Since the probability of having a busy queue for each user is independent of one another, and considering that each user performs a Bernoulli trial with probability ρ_i, the above PGF can be found in another way

$$\eta(z) = \prod_{i=1}^{N} (\rho_i z + 1 - \rho_i) \tag{3.68}$$

By simplifying both equations (3.67) and (3.68), then equating the coefficients of each z term in both equations, we can obtain $Pr(\eta = k)$ for all possible k values as a function of the utilization factors $\rho_1, \rho_2, \dots, \rho_N$. We note that we have a total of N+1 unknowns, which are $\rho_1, \rho_2, \dots, \rho_N$, P_s. These unknowns may be determined from simultaneous solution of equations (3.56) and (3.63).

3.7 Conclusion

In this chapter the proposed model for NB-IoT system has been presented and analyzed. The analysis resulted in the derivation of the PGF of the probability distribution of the number of packets in user queues and mean packet delay.

Chapter 4

Numerical and Simulation Results

4.1 Introduction

The analysis has been tested numerically under different traffic scenarios and the results are shown in the following sections in graphs. Additionally, the system has been implemented for simulation to compare the analytical results with the results from simulation to measure the accuracy of the assumptions taken in the analysis like the assumption of independency of probability of success between different transmissions. The following section will explain the simulation approach while the next section will show the numerical results.

4.2 Simulation Model

This section briefly describes the simulation model that has been used to test the accuracy of the analysis. A discrete-event simulation model has been implemented using Matlab and it follows MAC procedures in NB-IoT. As in the analytical model, simulation considers one channel for all users, the network has N users each with a queue of infinite capacity for storing packets and the time is slotted. Each user serves its packets according to FCFS service discipline.

New packets are being generated according to a Poisson process during each slot. Following generation, the packet chooses one of the available users according to a uniform distribution and it enters its queue. The information about each packet in its queue is stored including its arrival time. The service time of a packet begins when it reaches to the head of its queue. The transmission of a packet is successful if its user is the only user contending during that slot.

If more than one user transmits a packet in the same slot, a collision happens. The colliding users will increase their transmission attempt number by one and they will choose randomly a backoff period. The backoff period is chosen from a specified range according to a uniform distribution. Each time the transmission number increases, the range of backoff period is doubled. As per the

analysis, transmission time of each packet is assumed to be one slot time. Delay of each packet is found by subtracting the time of departure slot from arrival slot.

The simulation results of each iteration are taken from averaging the results of 4000 packets ignoring the first 100 packets as the system initially will not be at the steady-state. Each simulation sample shown in the graphs is the average of 100 iterations.

4.3 Numerical and Simulation Results

Next, we present numerical and simulation results and compare them. The average packet delay, probability of packet discarding, utilization, probability of successfully transmitting a packet, and the average service time have been plotted as a function of total packet arrival rate for different number of users and system parameters. Results will be presented for equal as well as unequal arrival rates to the users. In the case of equal arrival rates, λ and ρ will denote user arrival rate and utilization. The total arrival rate and utilization of the system, respectively, are given by,

$$\lambda_T = \lambda_1 + \lambda_2 + \cdots + \lambda_N \tag{4.1}$$

$$\rho_T = \rho_1 + \rho_2 + \cdots + \rho_N \tag{4.2}$$

Utilization of the system is also referred to as total utilization. Let us define maximum backoff window size vector for a maximum of Q transmissions as follows,

$$W_Q = \{\omega_1 \ldots . \omega_i \ldots . \omega_Q\} \tag{4.3}$$

Next, numerical results are presented for different cases,

Case 1: $\boldsymbol{N}$**=5,** $\boldsymbol{\lambda_T = N\lambda}$**,** $\boldsymbol{\rho_T = N\rho}$**,** **Q=5,** $\boldsymbol{W_5 = \{0\ 7\ 13\ 26\ 51\}}$

For this case, Fig.s 4.1, 4.2, and 4.3 show average packet delay, utilization of the system, and probability of packet discarding as a function of total packet arrival rate both for analysis and simulation. It may be seen that numerical and simulation results are very close to each other, which confirms accuracy of the analysis. From Fig. 4.1, average packet delay increases with increasing

packet arrival rate. From Fig. 4.2, the utilization of the system approaches to one at the total packet arrival rate of 0.26 packets/slot. From Fig. 4.3, probability of packet discarding remains under 0.001 even under heavy traffic loading. Highest delay result shown in Fig. 4.1 under heavy traffic is around 5 slots. Even considering a high slot time of 1280ms (which is utilized in real life scenarios when number of repetitions is very high in deep indoor areas), the average packet delay would still not exceed the latency limit of 10 seconds.

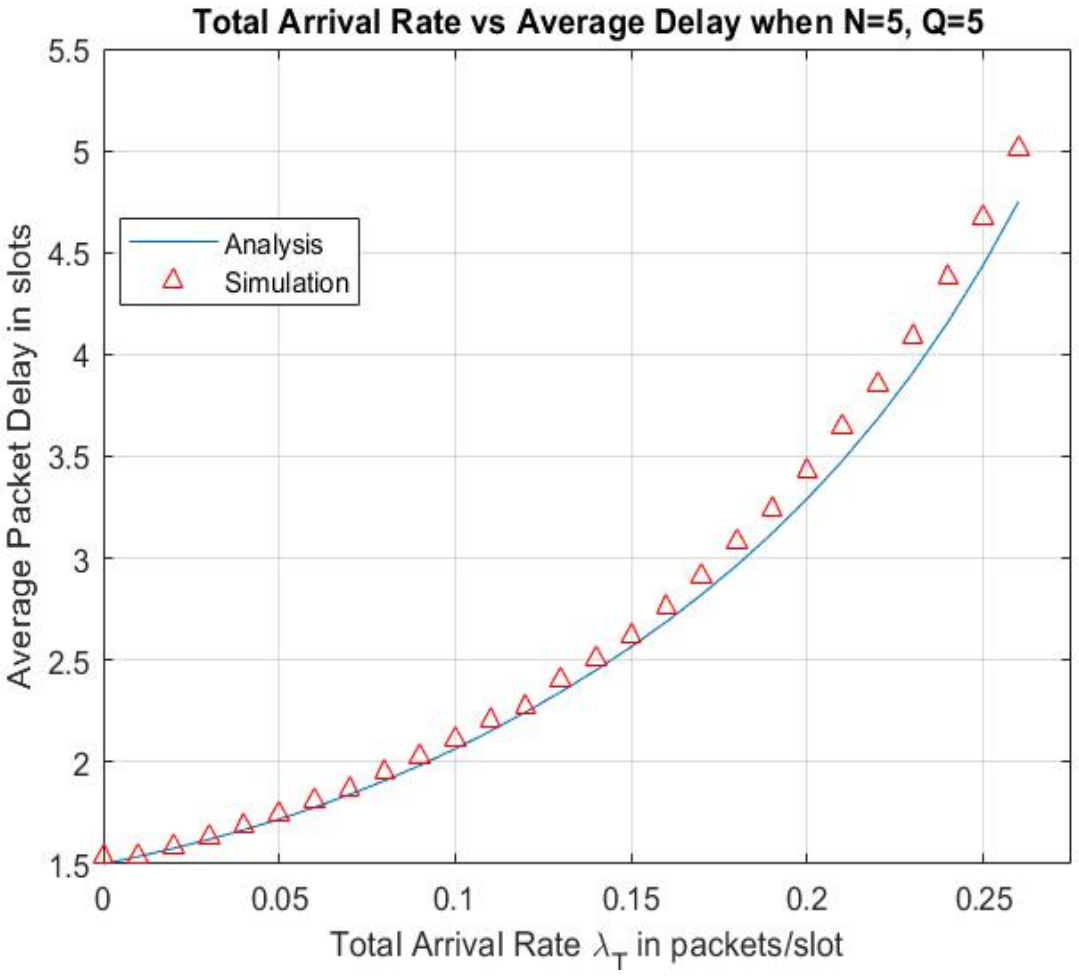

Fig. 4.1. Average packet delay as a function of the total packet arrival rate for N=5 , Q=5 with equal arrival rates.

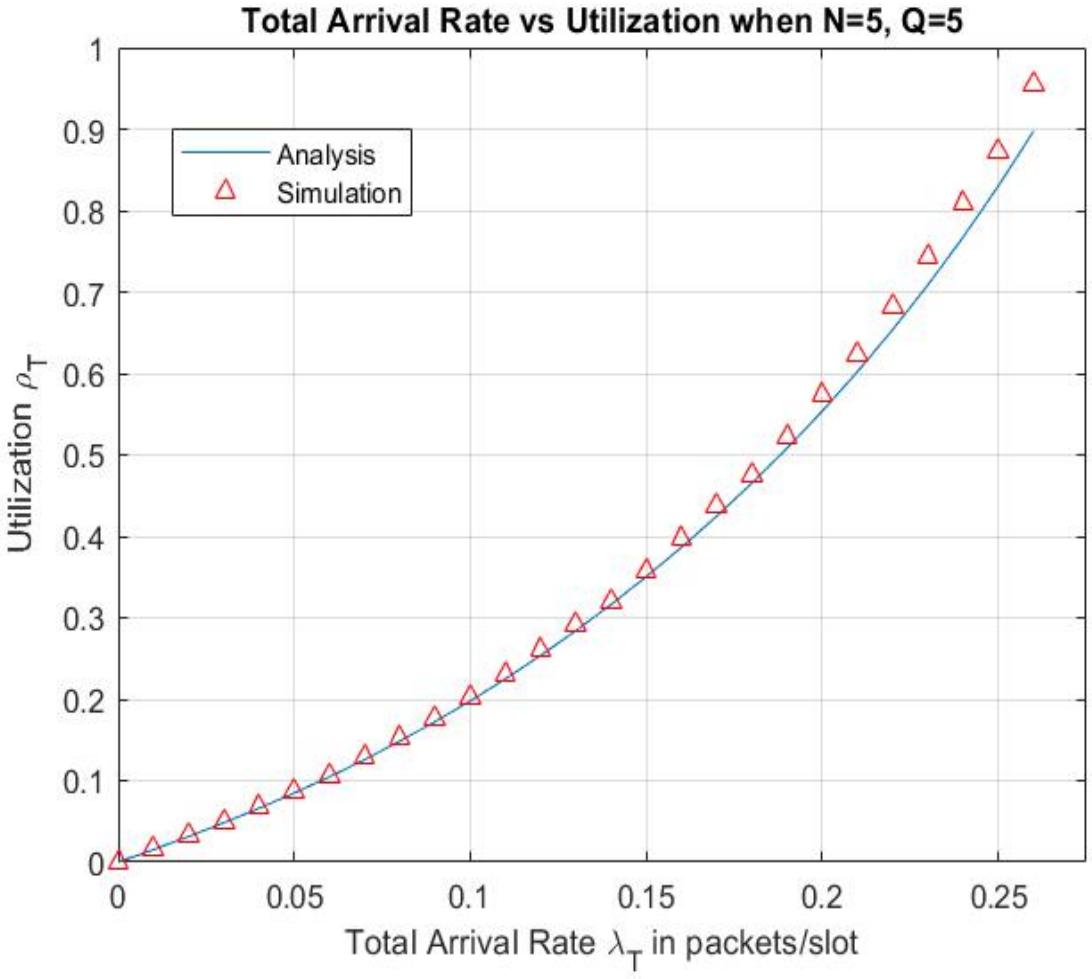

Fig. 4.2. Total system utilization as a function of the total packet arrival rate for N=5 , Q=5 with equal arrival rates.

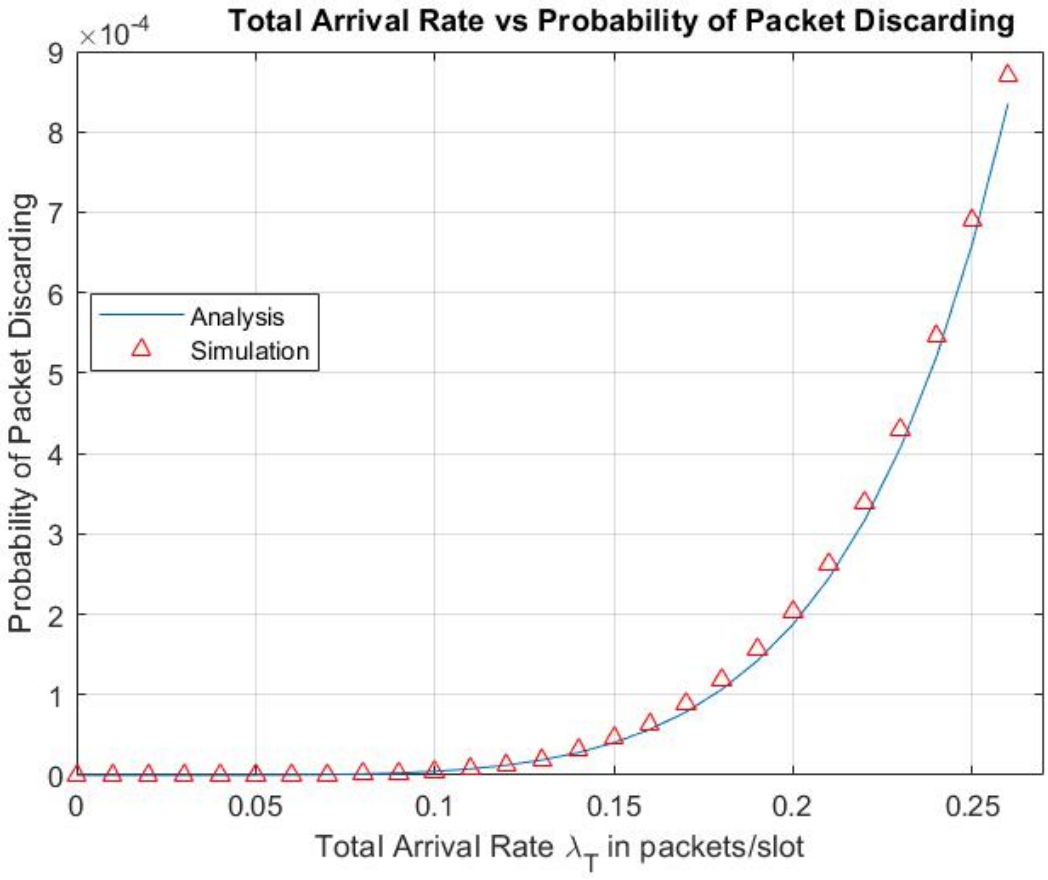

Fig. 4.3. Probability of packet discarding as a function of the total packet arrival rate for N=5 , Q=5 with equal arrival rates.

Case 2: $\boldsymbol{N = 10,\ \lambda_T = N\lambda,\ \rho_T = N\rho,\ Q = 5, W_5 = \{0\ 7\ 13\ 26\ 51\}}$

For this case, Fig.s 4.4, 4.5, and 4.6, show average packet delay, utilization of the system, and probability of packet discarding as a function of total packet arrival rate both for analysis and simulation. As may be seen, again, numerical and simulation results agree with each other. Compared to the previous case, average packet delay is higher at any total packet arrival rate because there are more users contending for the channel which results in a higher number of collisions. From Fig. 4.6, probability of packet discarding also has increased because it depends on the collisions.

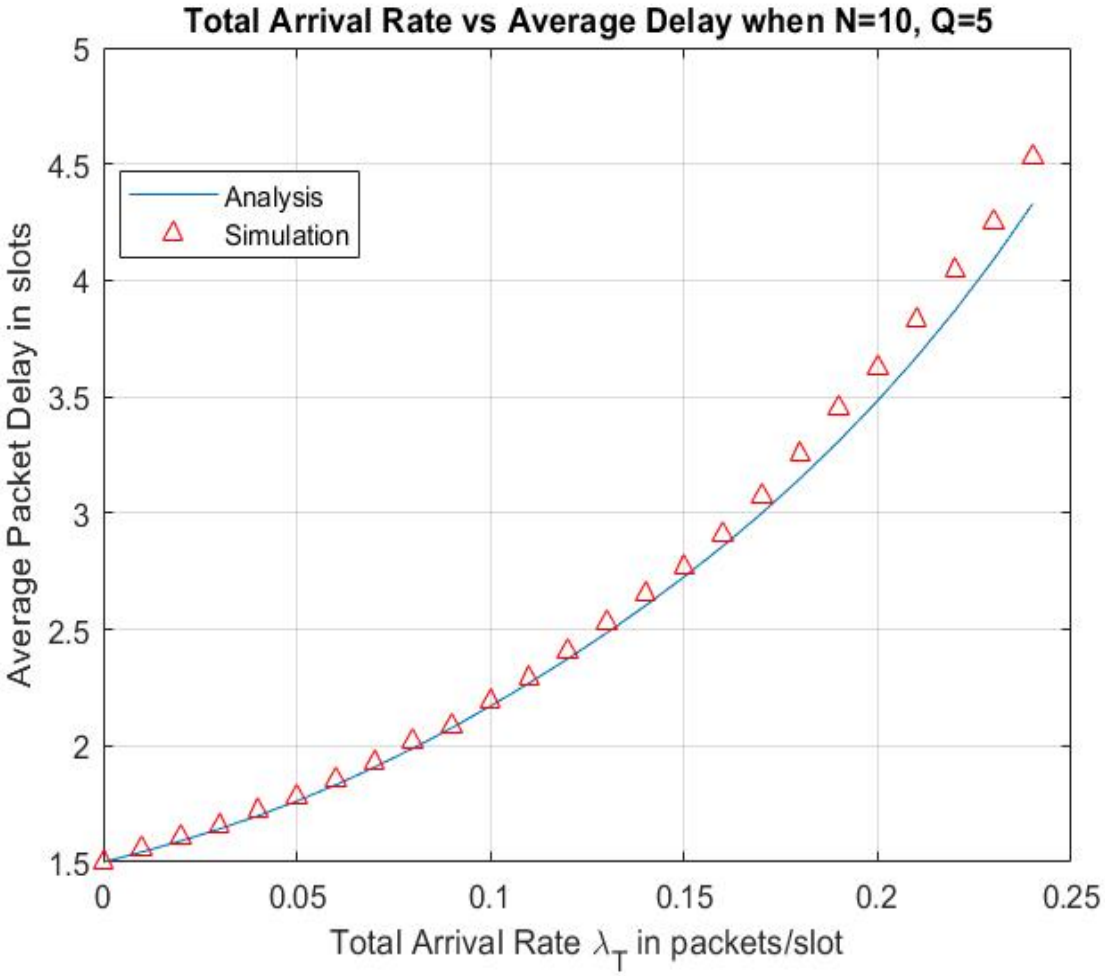

Fig. 4.4. Average packet delay as a function of the total packet arrival rate for N=10 , Q=5 with equal arrival rates.

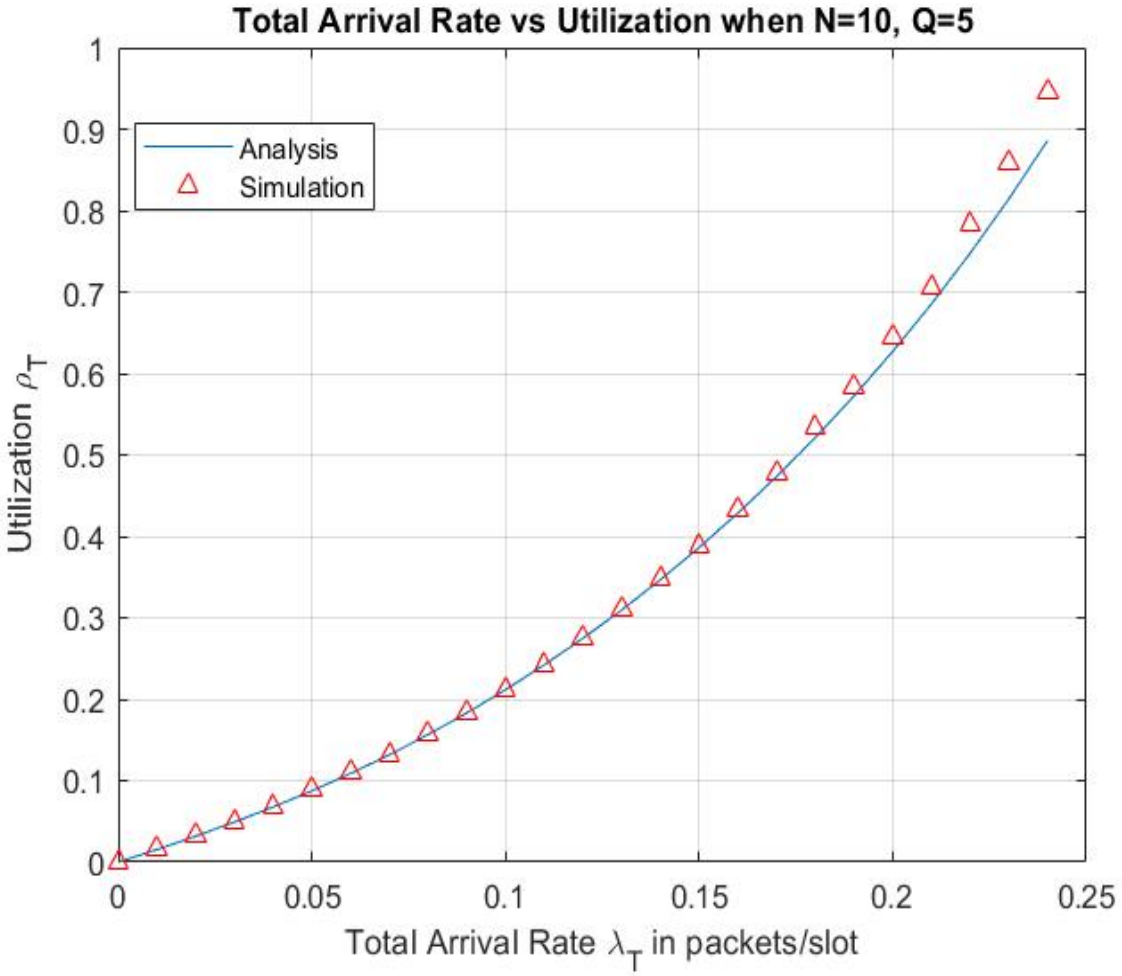

Fig. 4.5. System utilization as a function of the total packet arrival rate for N=10 , Q=5 with equal arrival rates.

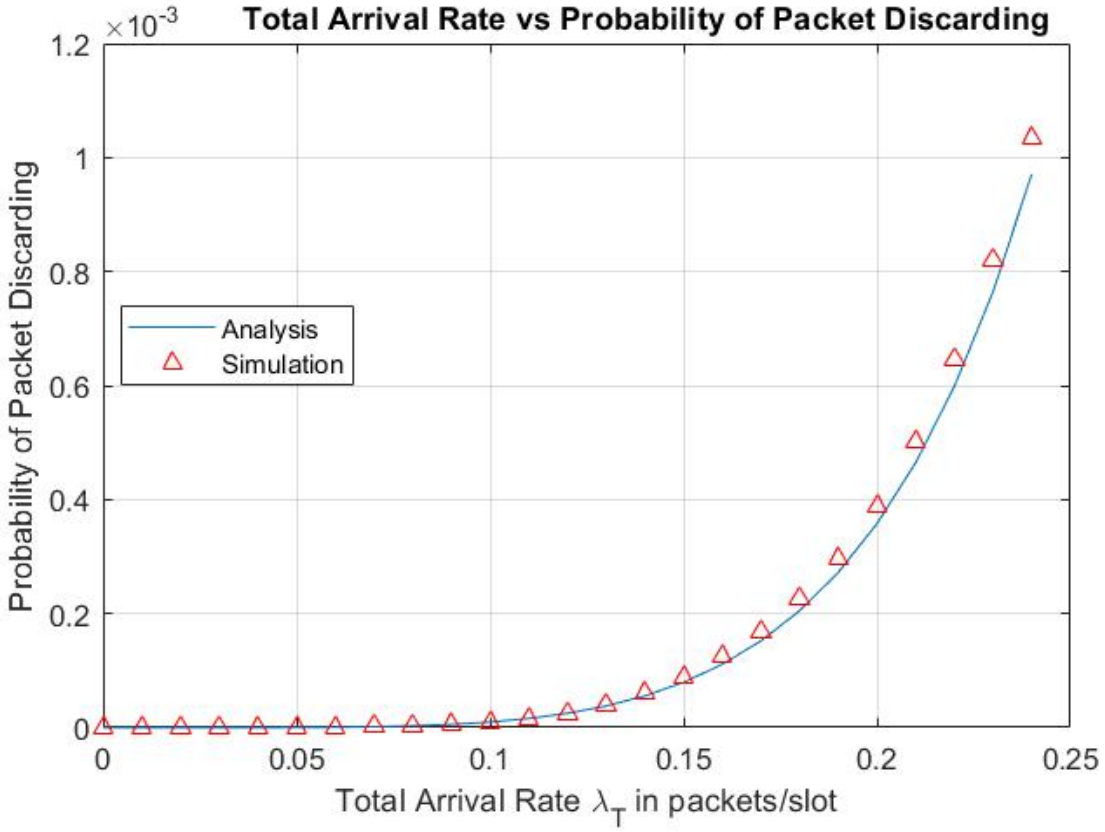

Fig. 4.6. Probability of packet discarding as a function of the total packet arrival rate for N=10 , Q=5 with equal arrival rates.

Additionally, Fig.s 4.7 and 4.8 show the analytical results of probability of successfully transmitting a packet and the mean service time as a function of the total packet arrival rate for both case 1 and case 2. As seen from both figures, increasing the number of users in case 2 resulted in a higher mean service time and a lower probability of success.

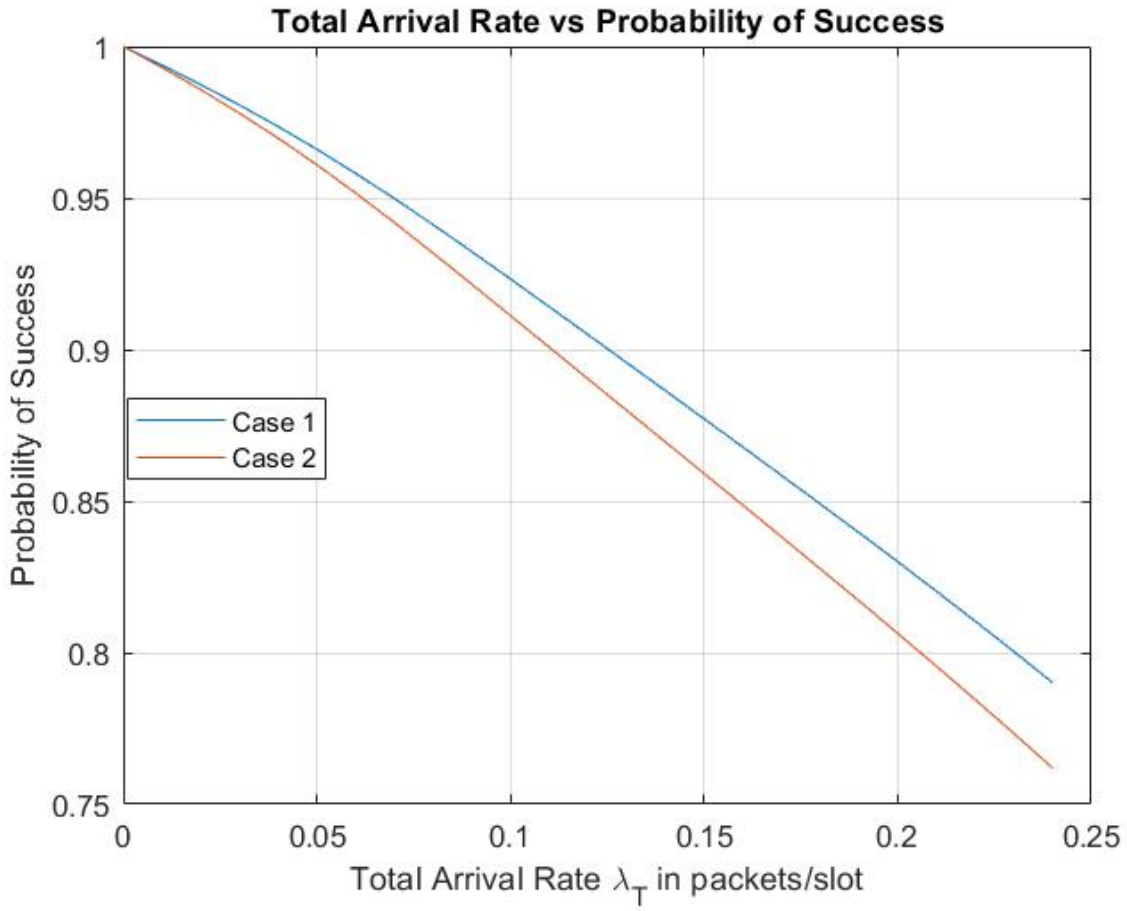

Fig. 4.7. Probability of successfully transmitting a packet as a function of the total packet arrival rate for case 1 and case 2 with equal arrival rates.

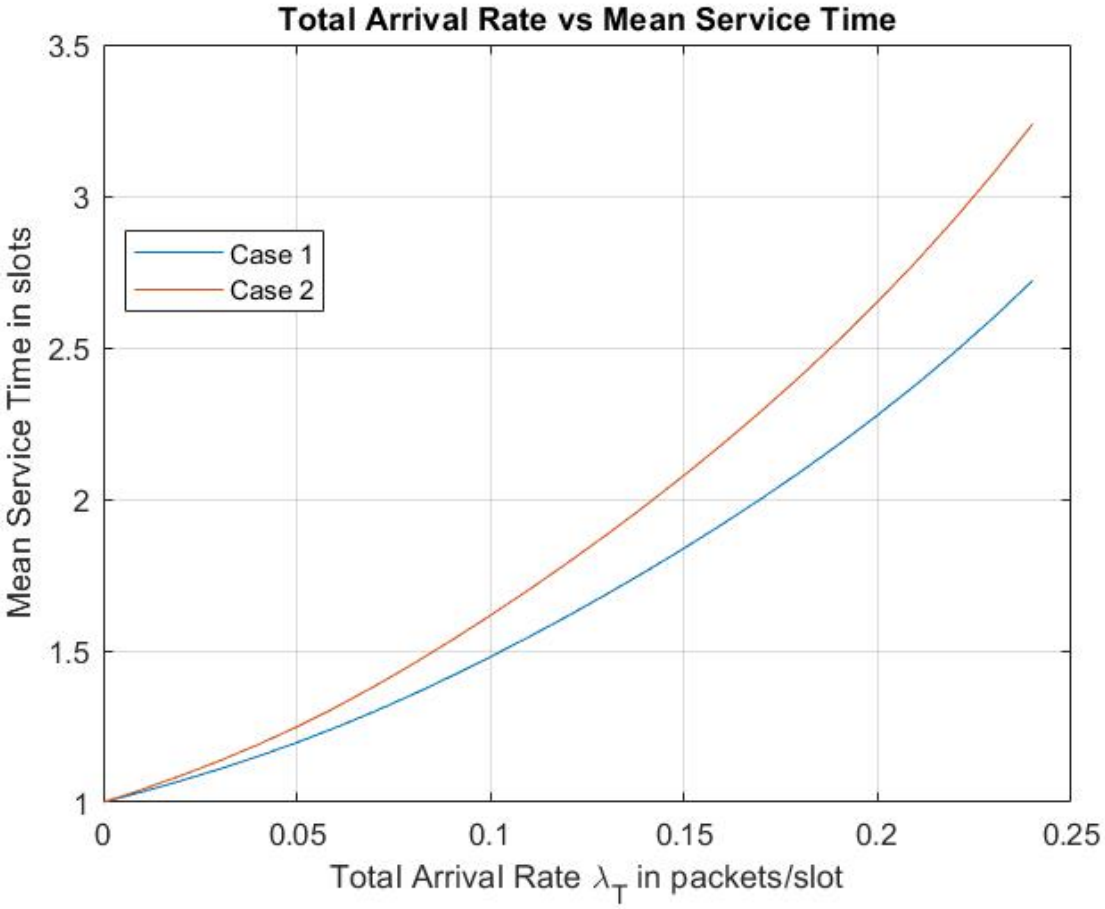

Fig. 4.8. Mean service time as a function of the total packet arrival rate for case 1 and case 2 with equal arrival rates.

Case 3: Adaptive vs. Constant Backoff mechanism, $\boldsymbol{\lambda_T = N\lambda}$, $\boldsymbol{\rho_T = N\rho}$, $\boldsymbol{Q = 5}$

To compare the performance of adaptive backoff mechanism with the fixed backoff mechanism that sets a constant backoff window size, Fig.s 4.9, 4.10, and 4.11 show the analytical results of average packet delay, mean service time, and probability of packet discarding for both adaptive backoff scheme and constant backoff scheme considering $\boldsymbol{N = 5}$. The backoff window sizes of adaptive and constant backoff schemes, respectively, are shown below:

$$\boldsymbol{W_5 = \{0\ 7\ 13\ 26\ 51\}}$$

$$\boldsymbol{W_5 = \{0\ 13\ 13\ 13\ 13\}}$$

In Fig. 4.9, under light loading, constant backoff mechanism uses larger window size than adaptive mechanism which results in higher delay. On the other hand, under heavy loading constant backoff mechanism saturates earlier than adaptive mechanism. From Fig. 4.11, adaptive backoff has a

lower packet discarding probability results compared to constant backoff. Choosing a lower fixed window size for constant backoff will lead to a lower delay results, however, the discarding probability would be higher.

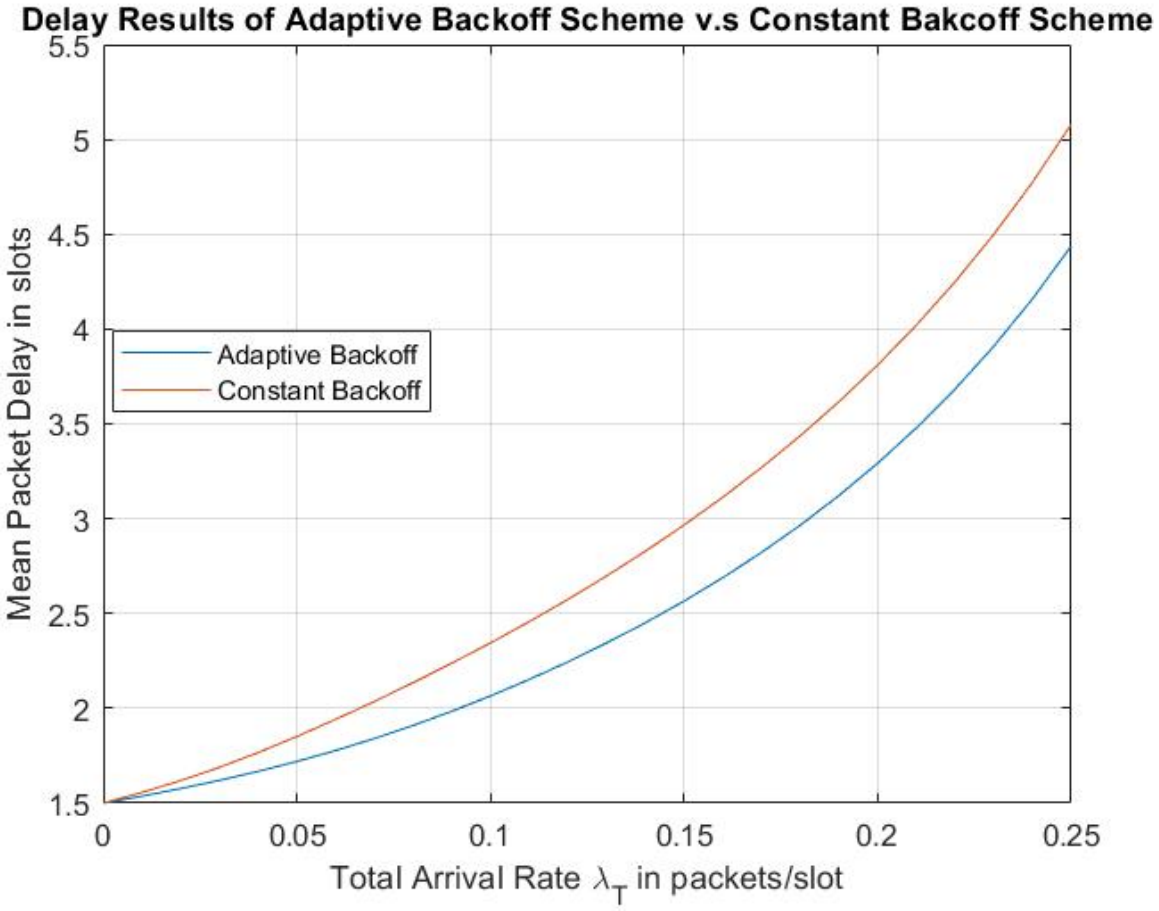

Fig. 4.9. Average packet delay as a function of the total packet arrival rate for N=5 , Q=5 with equal arrival rates and with adaptive backoff vs constant backoff.

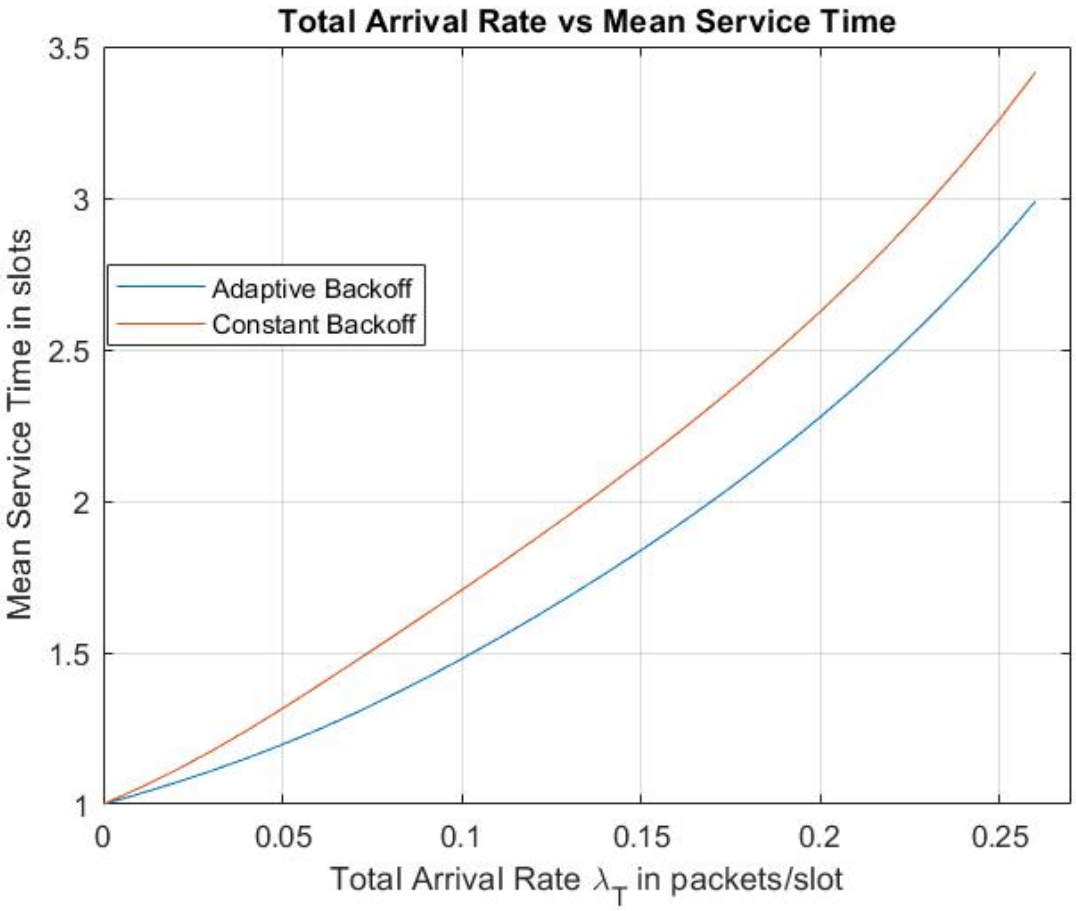

Fig. 4.10. Mean service time as a function of the total packet arrival rate for N=5 , Q=5 with equal arrival rates and with adaptive backoff vs constant backoff.

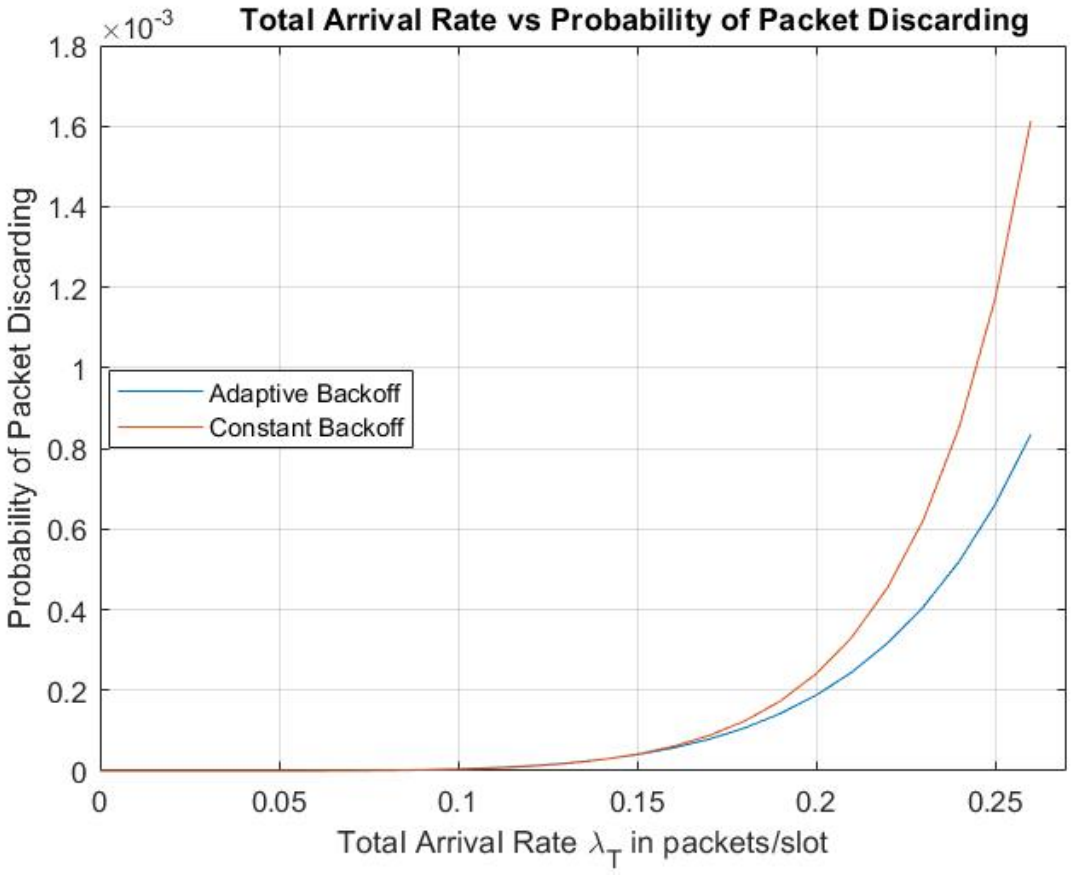

Fig. 4.11. Probability of packet discarding as a function of the total packet arrival rate for N=5 , Q=5 with equal arrival rates and with adaptive backoff vs constant backoff.

Additionally, the comparison between adaptive and constant backoff is shown for the case of high number of users. For that, Fig.s 4.12, and 4.13 plot the average packet delay and probability of packet discarding as a function of the total packet arrival rate for $\boldsymbol{N = 100}$ for both adaptive and constant backoff.

As can be seen in Fig. 4.12, under light traffic, mean delay results for adaptive backoff are still lower than those of constant backoff. However, under heavy traffic, the mean delay results for adaptive backoff are higher than the delay results of constant backoff because of the higher window sizes of the adaptive backoff for the third and fourth attempts. Therefore, when having a higher number of users and under heavy traffic, adaptive backoff mechanism will have a higher delay results compared to constant backoff however, the packet discarding probability for adaptive backoff will be lower.

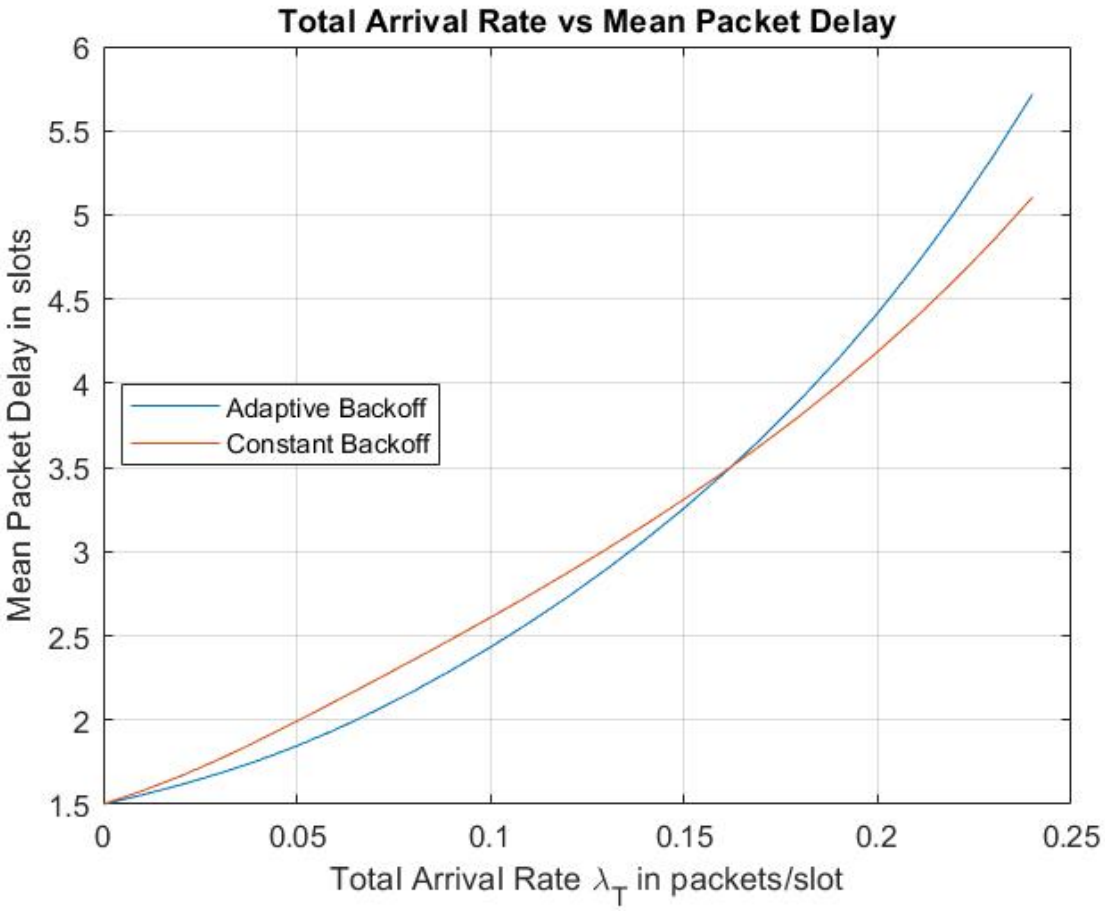

Fig. 4.12. Mean Packet Delay as a function of the total packet arrival rate for N=100 , Q=5 with equal arrival rates and with adaptive backoff vs constant backoff.

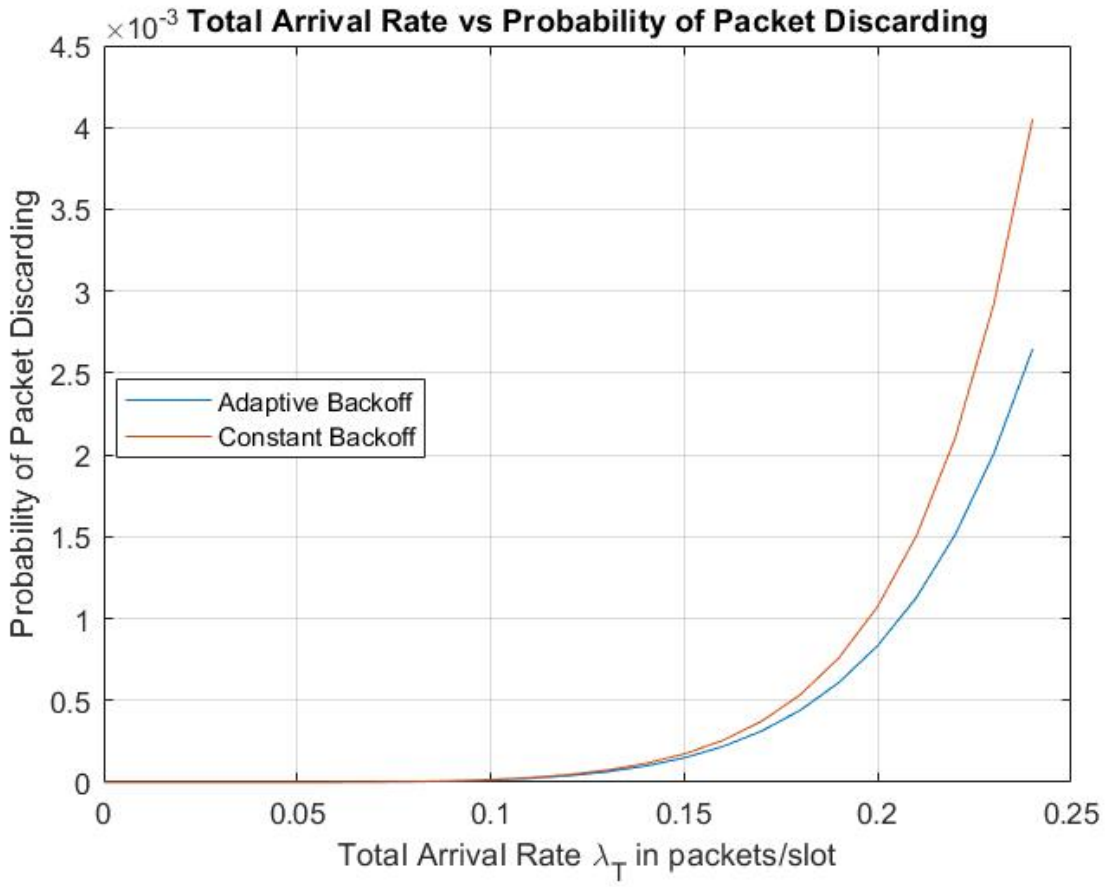

Fig. 4.13. Probability of packet discarding as a function of the total packet arrival rate for N=100 , Q=5 with equal arrival rates and with adaptive backoff vs constant backoff.

Case 4: $\boldsymbol{N = 10,\ \lambda_T = N\lambda,\ \rho_T = N\rho,}$

In this case, we consider the impact of the maximum number of transmissions allowed, Q, on the performance of the system. The results for two values are determined with the following maximum window sizes for $\boldsymbol{Q = 3}$ and $\boldsymbol{Q = 7}$,

$\boldsymbol{W_3 = \{0\ 7\ 13\}}$

$\boldsymbol{W_7 = \{0\ 7\ 13\ 26\ 51\ 103\ 206\}}$

It is noted that $Q = 3$ is the minimum possible value for the maximum number of allowed transmissions in 3GPP standards. Fig.s 4.14, 4.15 and 4.16 plot the average packet delay, total system utilization and probability of packet discarding as a function of the total packet arrival rate for $Q = 3$ and 7, respectively. From Fig. 4.14, average packet delay has increased with increasing the value of maximum number of allowed transmissions Q. On the other hand, probability of packet discarding almost drop to zero with higher value of Q as a packet is allowed more

transmissions before getting discarded. Thus, increasing Q can help in improving the QoS packet loss requirement at the cost of a slightly higher delay, particularly, in light and medium traffic.

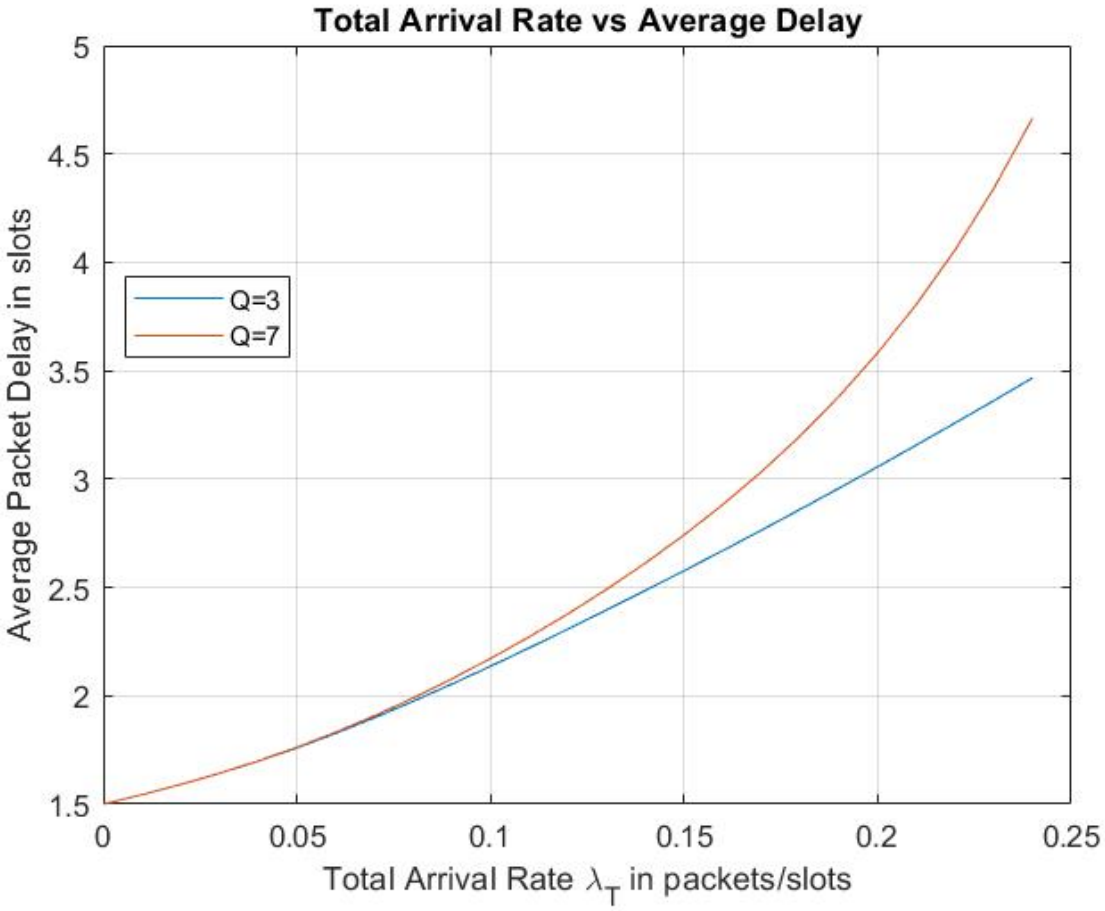

Fig. 4.14. Average packet delay as a function of the total packet arrival rate for N=10 , with equal arrival rates and Q as a parameter.

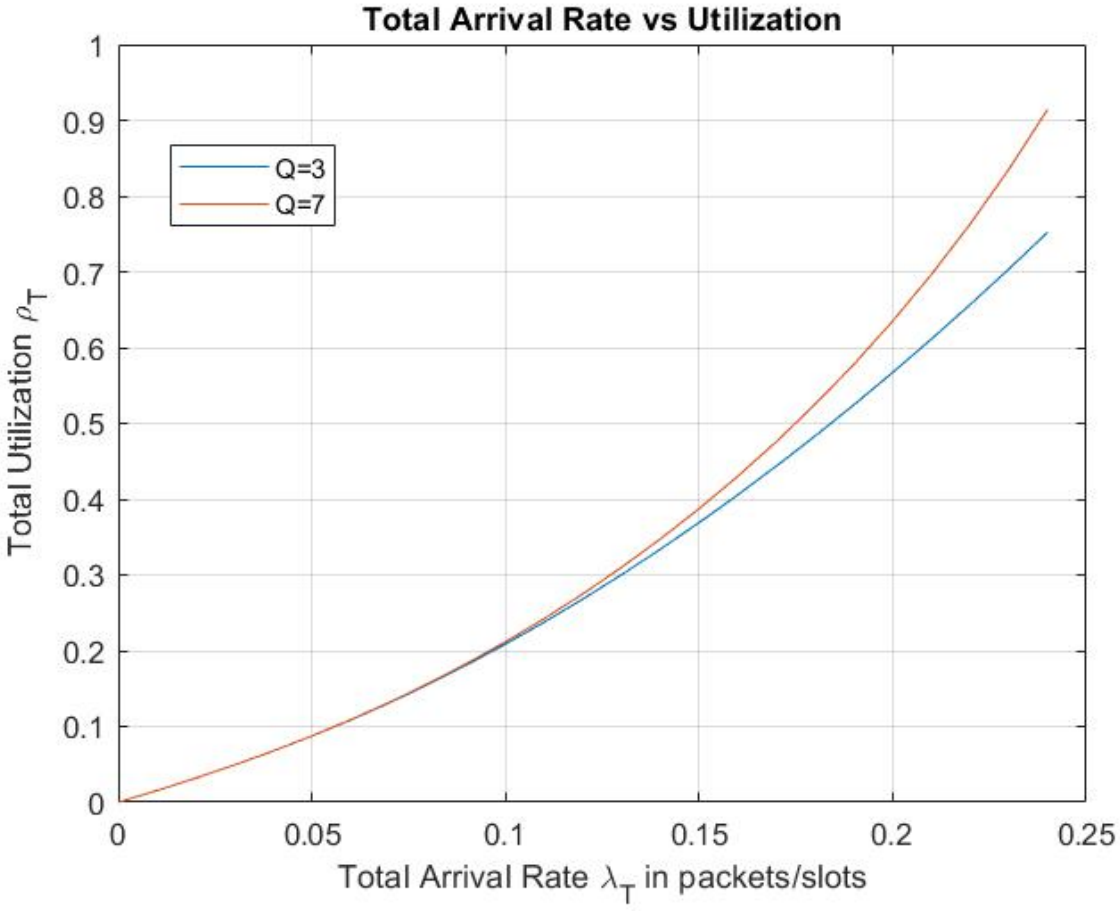

Fig. 4.15. Total system utilization as a function of the total packet arrival rate for N=10 , with equal arrival rates and Q as a parameter.

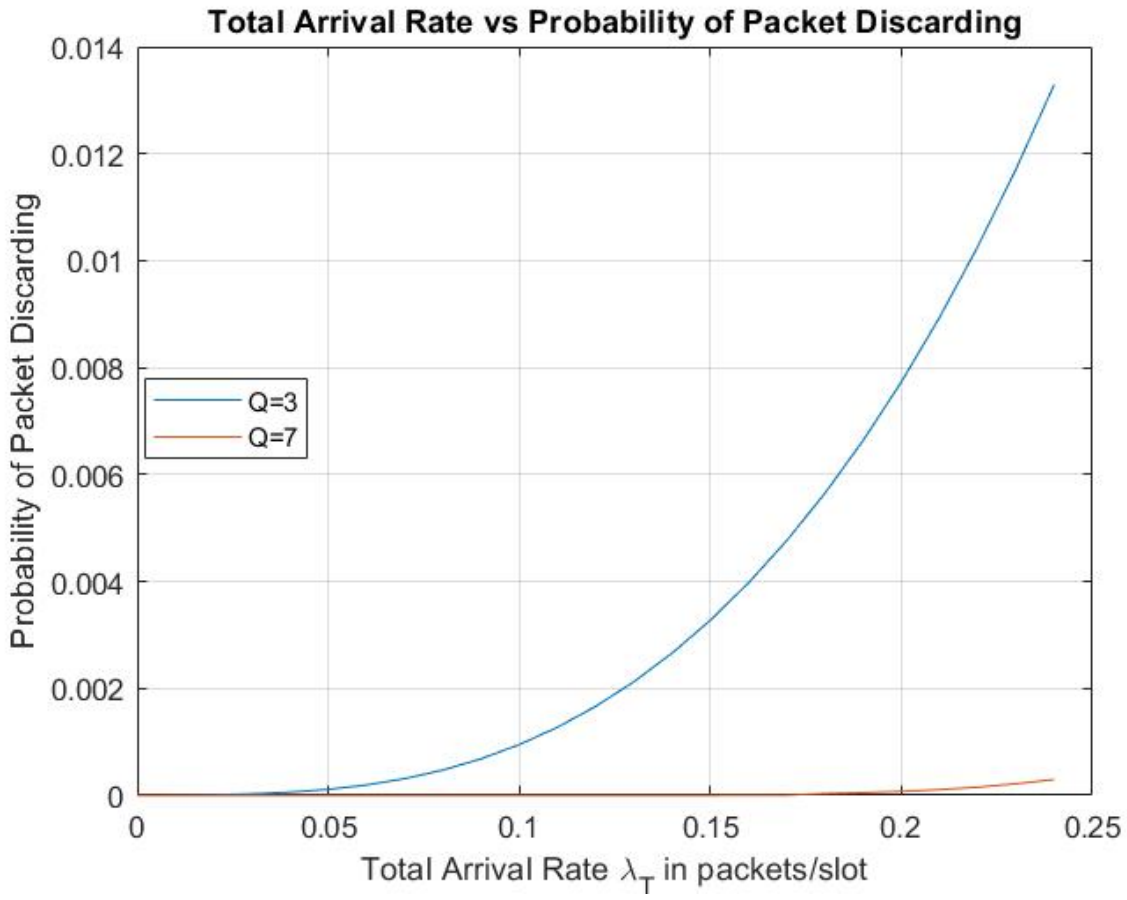

Fig. 4.16. Probability of packet discarding as a function of the total packet arrival rate for N=10 , with equal arrival rates and Q as a parameter.

In real life scenarios, when packet loss requirement is more important or cannot be tolerated, it is advisable to choose a higher value for the maximum number of attempts Q. Whereas, when latency requirement is more important, a small probability of packet discarding can be acceptable to achieve a lower mean packet delay and better latency results. This also could be helpful in countering the negative effect of increasing the number of users. For example, in a network that is working under high traffic and utilization is close to one, then, several new users may attempt to connect to the network. To accommodate the new users without letting the network to exceed the balance condition, it is better to choose a slightly lower Q value to reduce utilization.

Case 5: $\boldsymbol{Q = 3,\ \lambda_T = N\lambda,\ \rho_T = N\rho, W_3 = \{0\ 7\ 13\}}$

In this case, we consider the impact of increasing the number of users on the packet discarding probability. The results will be provided for two values of number of users $\boldsymbol{N = 100}$ and $\boldsymbol{N = 200.}$ Fig. 4.17 plots the probability of packet discarding as a function of total packet arrival

rate. As seen from the figure, under high traffic the probability of packet discarding exceeds 1% which is unreliable. Therefore, when having a high number of users, it is recommended to choose a higher value of Q.

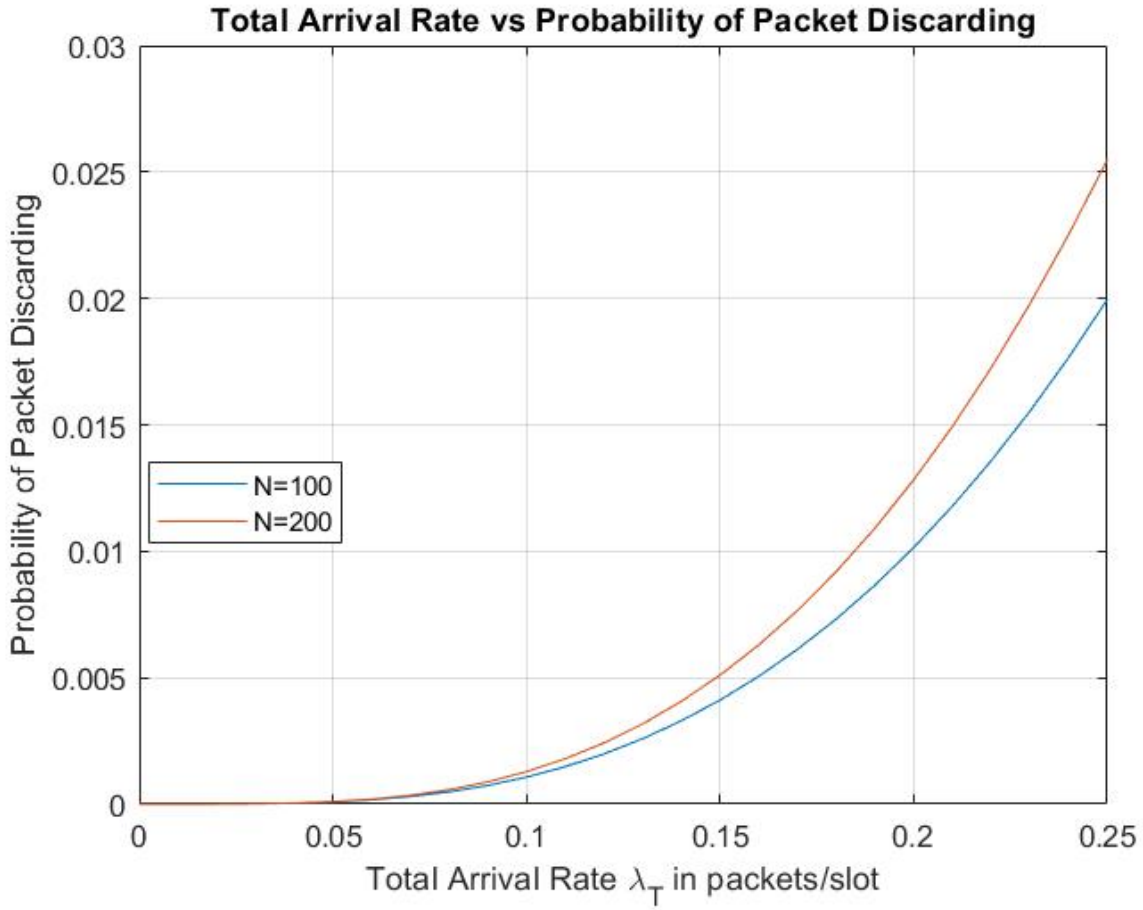

Fig. 4.17. Probability of packet discarding as a function of the total packet arrival rate for Q=3 , with equal arrival rates and N as a parameter.

Case 6: $N = 10,\ \lambda_T = N\lambda,\ \rho_T = N\rho\ , Q = 6$

In 3GPP standards, the first transmission of a packet takes place immediately without backoff and only collided packets have to go through backoff before retransmission. This approach has the disadvantage that under heavy traffic a single user may dominate the channel for long periods of time, not giving transmission opportunity to the other users. We would like to compare the results of system with and without backoff in the first attempt of a packet. We will denote this scheme as the "first attempt backoff scheme" and the scheme where backoff is done only when there is

collision as "collision backoff scheme". Maximum window size vectors for the two cases are given below,

$$\boldsymbol{W_6 = \{0\ 7\ 13\ 26\ 51\ 103\}}$$

$$\boldsymbol{W_6 = \{7\ 13\ 26\ 51\ 103\ 206\}}$$

For this case, Fig.s 4.18 and 4.19 show average packet delay, utilization of the system as a function of total packet arrival rate. As may be seen, introduction of backoff in the first attempt resulted in increased packet delay and system utilization.

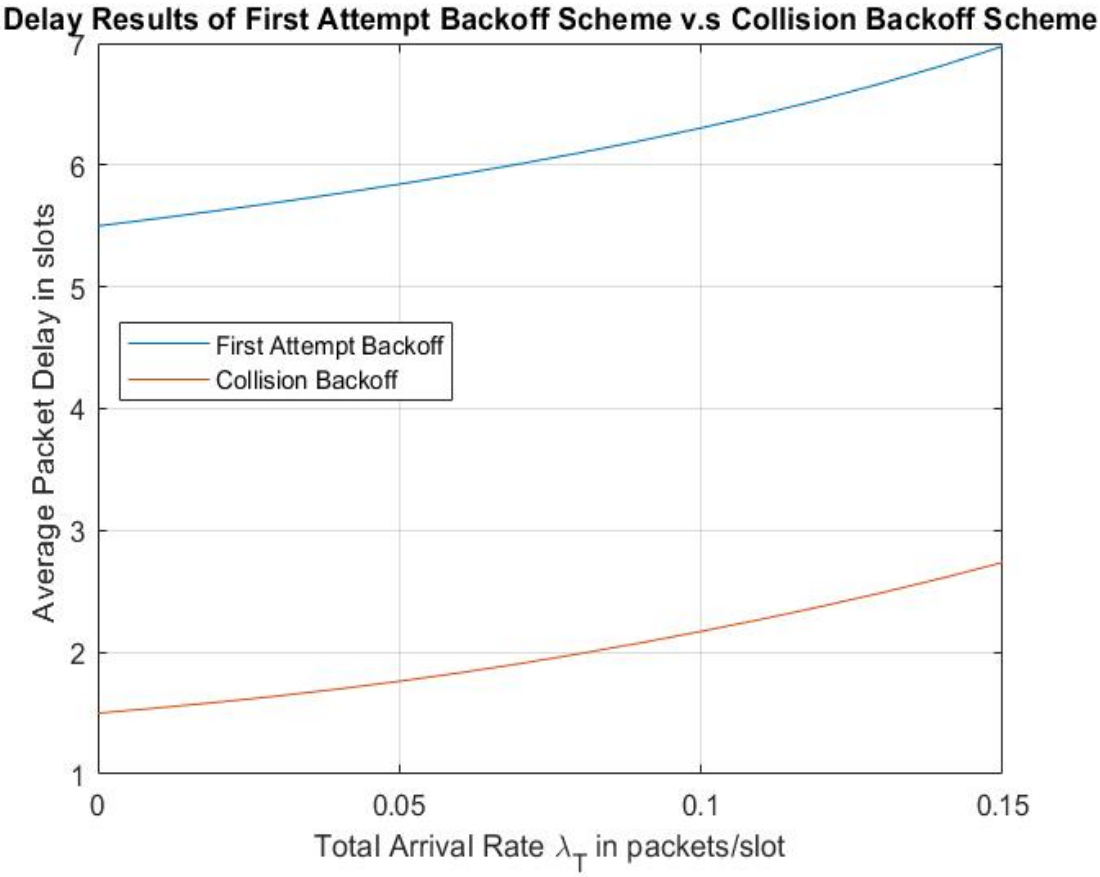

Fig. 4.18. Average packet delay as a function of the total packet arrival rate for N=10 , Q=6 with equal arrival rates and with/without backoff in the first attempt.

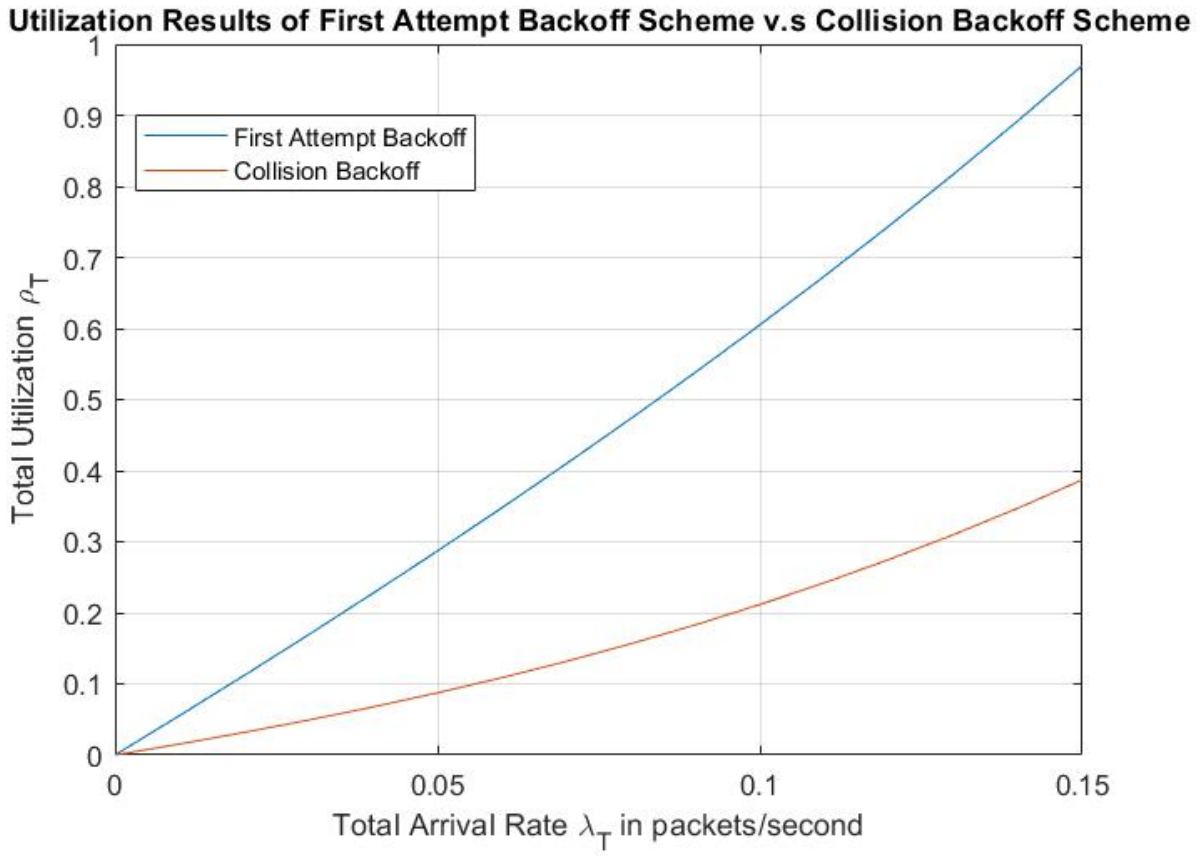

Fig. 4.19. Total system utilization as a function of the total packet arrival rate for N=10 , Q=6 with equal arrival rates and with/without backoff in the first attempt.

Case 7: $\boldsymbol{N = 10,\ \lambda_T = N\lambda,\ \rho_T = N\rho,\ Q = 6}$

In this case, again, results are compared for with/without backoff in the first attempt for the following maximum backoff window size vectors,

$$\boldsymbol{W_6 = \{0\ 7\ 13\ 26\ 51\ 103\}}$$

$$\boldsymbol{W_6 = \{3\ 7\ 13\ 26\ 51\ 103\}}$$

As may be seen, compared to the previous case maximum backoff window size of the first attempt has been reduced. For this case, Fig. 4.20 and Fig. 4.21 show average packet delay and system utilization as a function of total packet arrival rate. From Fig. 4.20, average packet delay with backoff is higher than without backoff in the first attempt but it has been reduced compared to the previous case.

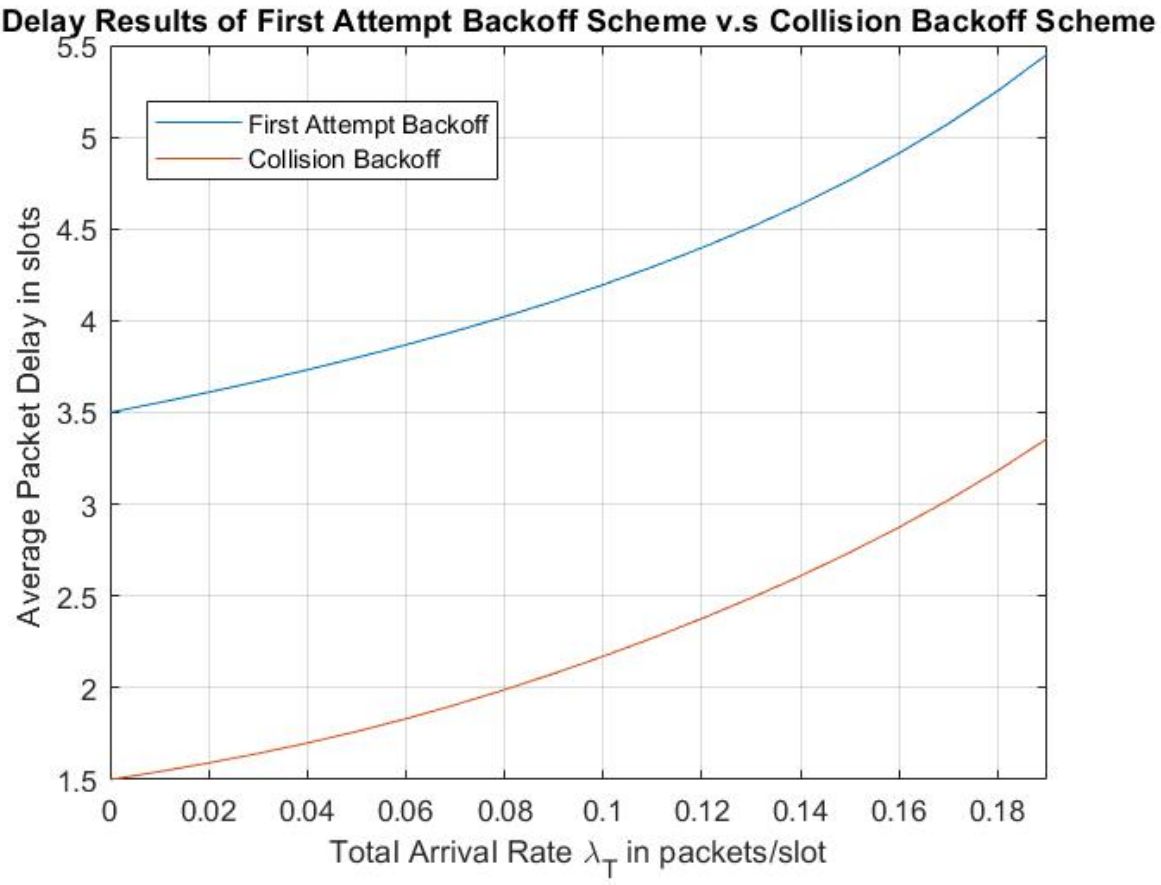

Fig. 4.20. Average packet delay as a function of the total packet arrival rate for N=10 , Q=6 with equal arrival rates and with/without backoff in the first attempt.

Therefore, a good implementation of first attempt backoff scheme is to use it with devices that tolerate high latency while other devices use the collision backoff scheme. This way, there will be two classes of devices and each has its service time. Thus, the good effect of reducing the number of collisions will still be effective but only devices that tolerate high latency will experience higher delay values.

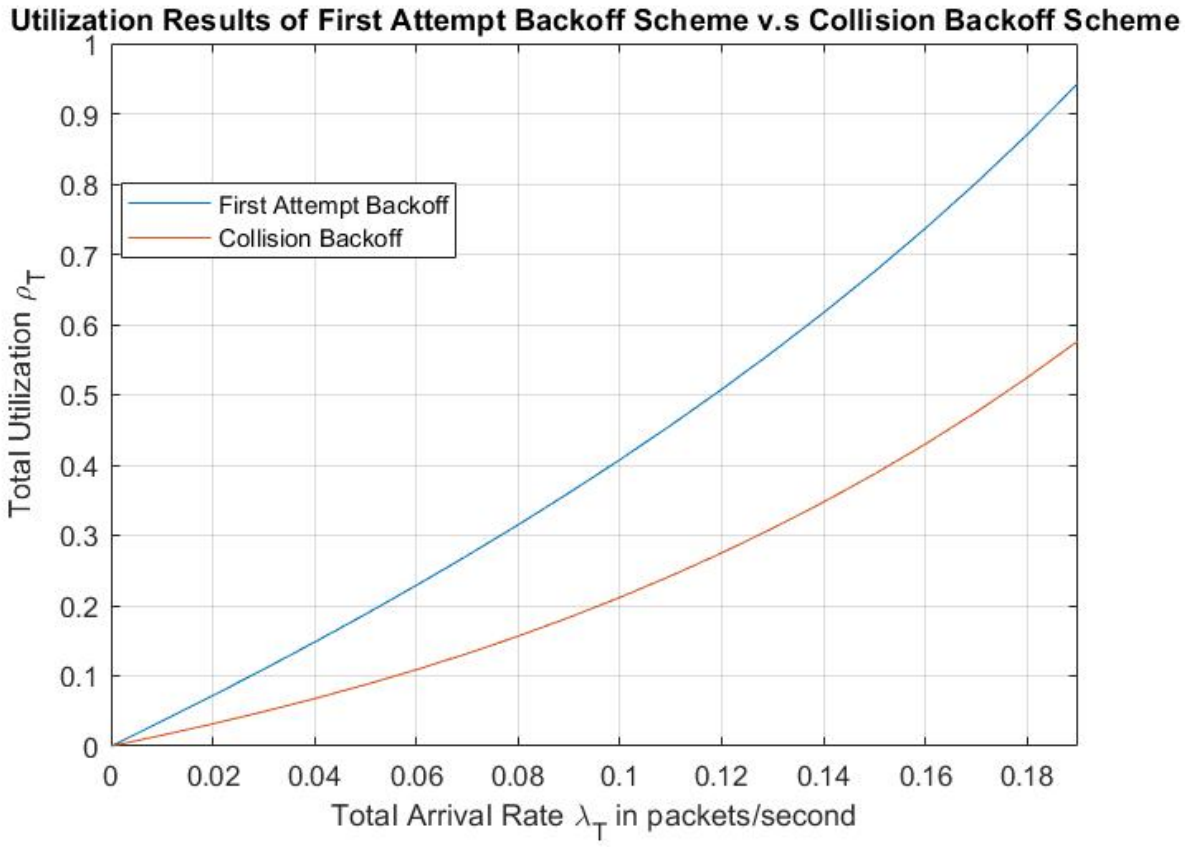

Fig. 4.21. System utilization as a function of the total packet arrival rate for N=10 , Q=6 with equal arrival rates and with/without backoff in the first attempt.

Case 8: N=5, Q= 6

One of the features provided by our model compared to previous models is the realistic consideration that every user may have its unique arrival rate. In this case, we consider a system with unequal user arrival rates. The users have been assigned unequal portions of the total arrival rate as shown below,

$$\boldsymbol{\lambda_1 = \frac{\lambda_T}{15}}$$

$$\boldsymbol{\lambda_2 = 2\lambda_1 = \frac{2\lambda_T}{15}}$$

$$\boldsymbol{\lambda_3 = 3\lambda_1 = \frac{\lambda_T}{5}}$$

$$\boldsymbol{\lambda_4 = 4\lambda_1 = \frac{4\lambda_T}{15}}$$

$$\boldsymbol{\lambda_5 = 5\lambda_1 = \frac{\lambda_T}{3}}$$

We will show the results of this case considering first attempt backoff scheme and collision backoff scheme. The maximum window size vectors for the collision backoff scheme and first attempt backoff scheme, respectively, are shown below,

$$\boldsymbol{W_6 = \{0\ 7\ 13\ 26\ 51\ 103\}}$$

$$\boldsymbol{W_6 = \{3\ 7\ 13\ 26\ 51\ 103\}}$$

Fig.s 4.22 shows the analytical and simulation results for mean packet delay as a function of total packet arrival rate for the collision backoff scheme. Additionally, Fig. 4.23 shows the analytical results of utilization for the system and for every user. From Fig. 4.22, analytical and simulation results agree with each other for the case of different arrival rates too.

Fig.s 4.24 and 4.25 shows the average packet delay and system utilization as a function of total packet arrival rate for the first attempt backoff scheme. From Fig 4.24, the mean delay result under the highest possible total arrival rate of 0.24 packets/slot is 5.6 slots which is slightly above its corresponding result of the collision backoff scheme which is 5.5 slots as shown in Fig. 4.22. The first attempt backoff scheme can keep a low rate of collision, near to zero packet loss probability, and almost similar delay results to the collision backoff scheme under heavy traffic loading. Therefore, it can be seen that the first attempt backoff scheme is more effective to use in the case of unequal arrival rates under heavy load in comparison with collision backoff scheme.

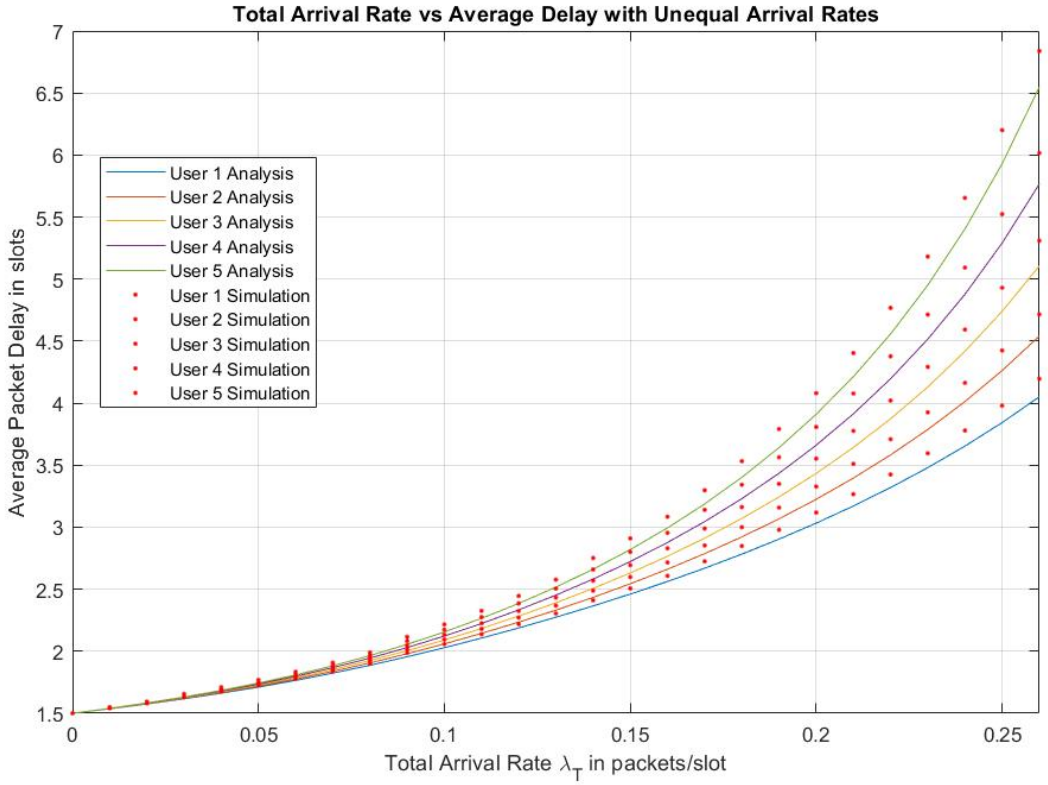

Fig. 4.22. Average packet delay for all users as a function of the total packet arrival rate for N=5 , Q=6 with unequal arrival rates under collision backoff scheme.

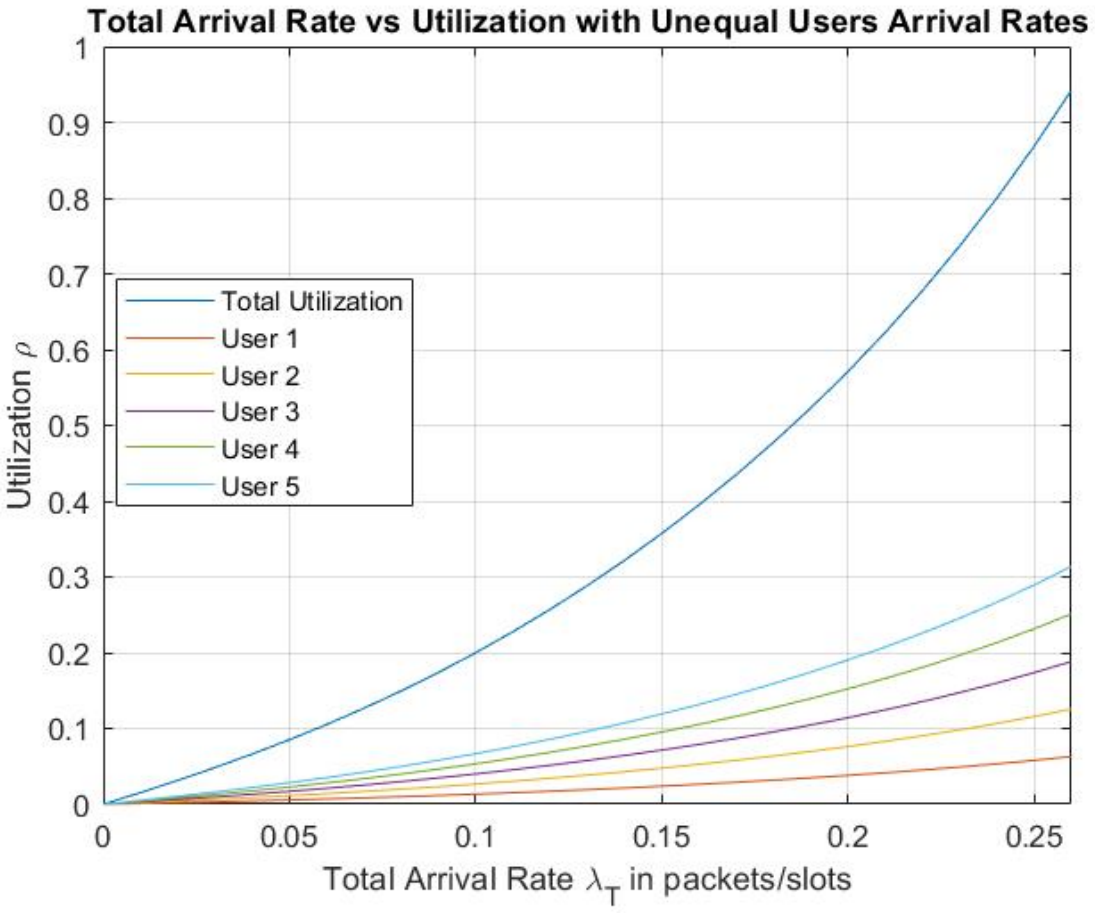

Fig. 4.23. System and users' utilizations as a function of the total packet arrival rate for N=5 , Q=6 with unequal arrival rates under collision backoff scheme.

As seen from the figures, the user with the highest arrival rate has the higher delay and utilization results and is the most contributing user towards increasing the total utilization to approach one at high traffic. On comparison, a user with a lower arrival rate such as user 1, would still experience lower delay even if he is combined with other users who have much higher arrival rates. Generally, users who have lower arrival rates will have a delay result that is close to their service time.

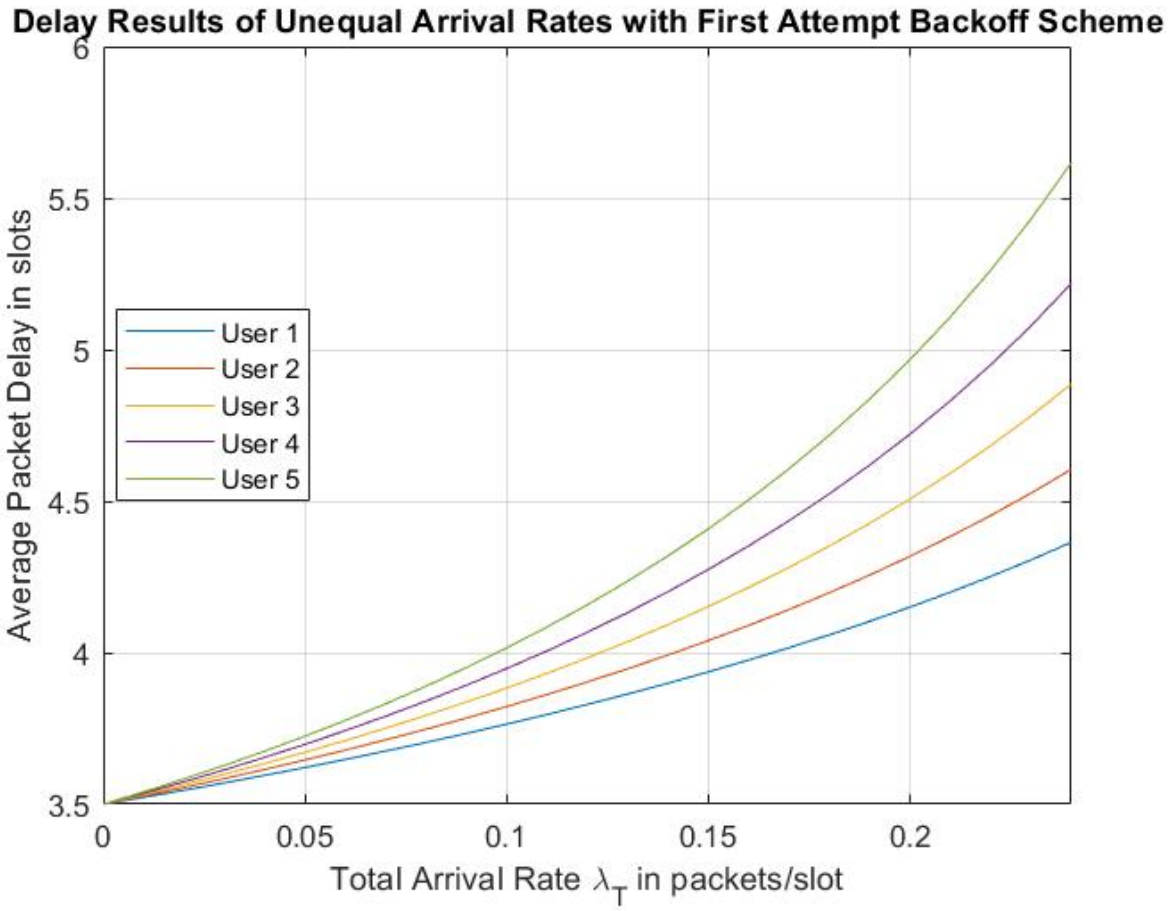

Fig. 4.24. Average packet delay for all users as a function of the total packet arrival rate for N=5 , Q=6 with unequal arrival rates under first attempt backoff scheme.

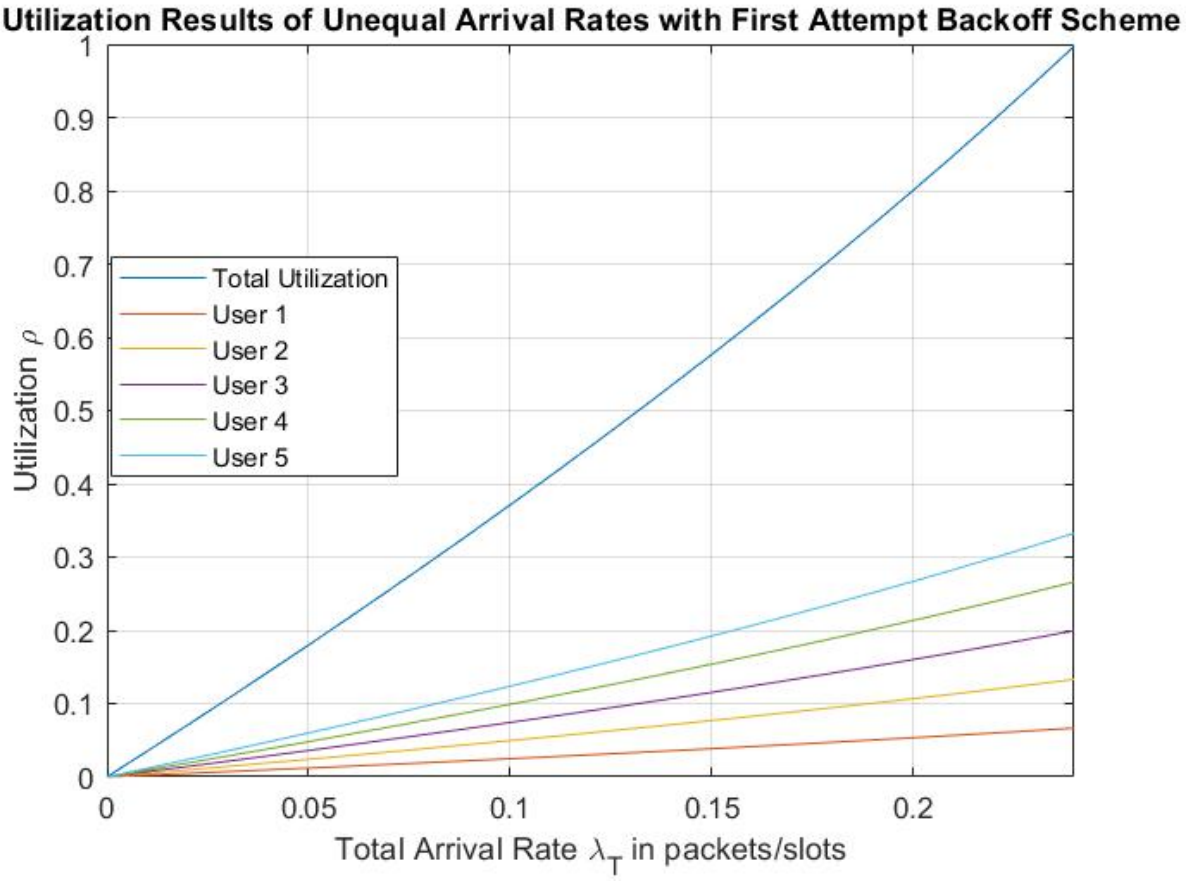

Fig. 4.25. System and users' utilizations as a function of the total packet arrival rate for N=5 , Q=6 with unequal arrival rates under first attempt backoff scheme.

4.4 Conclusion:

This chapter presents numerical results for the analytical model developed in the previous chapter. Numerical results are shown for average packet delay, system utilization and packet discarding probability as a function of the traffic load for different number of users in the system and various system parameters. Simulation results show that numerical and simulation results have a good agreement. Adaptive backoff mechanism is shown to have a good performance compared to the fixed backoff. Results have demonstrated the trade-off between packet discarding probability and mean packet delay where higher number of attempts allows for almost zero packet loss probability at the cost of a slightly higher delay.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

As the implementation of NB-IoT technology in different industries are increasing, several challenges are arising such as minimizing latency, managing mobility, improving resource utilization, and enhancing random access procedures. Prior works have investigated several aspects of the current random access procedures of NB-IoT but they either ignored the backoff mechanism or included it within certain limitations and assumptions. The research in this thesis tried to determine the impact of backoff mechanism on the performance of random access procedures by designing a comprehensive analytical model that describes the random access of NB-IoT with adaptive backoff.

The proposed model assumes arrival of packets to each user with different rates and infinite user queues. Additionally, the system considers an adaptive backoff mechanism that doubles the window size each time the user suffers a new collision to achieve better contention resolution. The system was modeled as a discrete-time system. Due to the similarities in backoff mechanism, the probability of success successful packet transmission was derived from an analytical model of CSMA/CA. Then, the analysis derived the of PGF of the probability distribution of the number of packets in user queues, the mean packet delay and probability of packet discarding.

The accuracy of the proposed analytical model was verified through simulation system. To get a better insight on the effect of different parameters on the system, numerical results were shown for different scenarios. The results have shown good agreement between the analytical and simulation model. Adaptive backoff was proven to be effective as a contention resolution mechanism. Comparing the results of different scenarios, it was found that increasing the number of attempts reduce the packet discarding probability at the cost of a slightly higher delay. Additionally, results show that under different arrival rates, only the users with high traffic would suffer a higher delay. This research advances existing results in the literature on performance evaluation of random access procedures of NB-IoT. It shows that adaptive backoff improves the performance compare

to constant backoff mechanism, it enables performance evaluation of the system with heterogeneous user traffic and it obtains results that have closed forms.

5.2 Future Work

The proposed model and results can be used as a guideline for understanding the behaviour of the network under the adaptive backoff mechanism when designing future NB-IoT networks. There are several options to improve on the current analytical model in the future work. A repetition scheme along with signal-to-noise ratio and power analysis can be incorporated into the existing model to check how the performance of the system would react under the effect of both mechanisms. Although, the repetitions are developed mainly to enhance the coverage and does not have a direct effect on the collision, however, the probability of access success would be affected by the number of repetitions, and as a result, other performance evaluation measurements would be affected.

References

[1] Evolved universal terrestrial radio access (E-UTRA): physical channels and modulation. 3GPP, TS 36.211 v.13.2.0, Release 13, 2016.

[2] M. Chafii, F. Bader and J. Palicot, "Enhancing coverage in narrow band-IoT using machine learning", *IEEE Wireless Communications and Networking Conference (WCNC)*, 2018. Available: 10.1109/wcnc.2018.8377263.

[3] D. Guohua and Y. Junhua, “Research on nb-iot background, standard development, characteristics and the service,” Mobile Communication, vol. 40, no. 7, pp. 31–36, 2016.

[4] J. Lee and J. Lee, "Prediction-Based Energy Saving Mechanism in 3GPP NB-IoT Networks", *Sensors*, vol. 17, no. 9, pp. 2008, 2017. Available: 10.3390/s17092008.

[5] R. Mozny, P. Masek, M. Stusek, K. Zeman, A. Ometov and J. Hosek, "On the Performance of Narrow-band Internet of Things (NB-IoT) for Delay-tolerant Services", *42nd International Conference on Telecommunications and Signal Processing (TSP)*, 2019. Available: 10.1109/tsp.2019.8768871.

[6] K.Lin, M.Chen, J.Deng, M.Hassan, and G.Fortino, “Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings,” *IEEE Transactions on Automation Science and Engineering*, vol.13, no.3, pp.1294–1307,2016.

[7] D. Xiong, Y. Chen, X. Chen, M. Yang and X. Liu, "Design of Power Failure Event Reporting System Based on NB-IoT Smart Meter", *International Conference on Power System Technology (POWERCON)*, 2018.

[8] Z. Yan and H. Gang, "Design of Intelligent Water Metering System for Agricultural Water Based on NB-IOT", *IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 2019.

[9] J. Shi, L. Jin, J. Li and Z. Fang, "A smart parking system based on NB-IoT and third-party payment platform", *17th International Symposium on Communications and Information Technologies (ISCIT)*, 2017. Available: 10.1109/iscit.2017.8261235.

[10] Y. Xing, J. Li and X. Wang, "Research and Design of Parking Detector Based on NB-IoT and Geomagnetism", *IEEE 2nd International Conference on Information and Computer Technologies (ICICT)*, 2019. Available: 10.1109/infoct.2019.8711174.

[11] M. Praveen and V. Harini, "NB-IOT based smart car parking system", *International Conference on Smart Structures and Systems (ICSSS)*, 2019. Available: 10.1109/icsss.2019.8882847.

[12] X. Lin, B. Qin, F. Zheng, X. Chen, C. Huang and R. Pan, "Application Research of NB-IoT Technology Based on Fog Computing in Intelligent Parking System", *IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 2019.

[13] Y. Sun, F. Tong, Z. Zhang and S. He, "Throughput Modeling and Analysis of Random Access in Narrowband Internet of Things", *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1485-1493, 2018. Available: 10.1109/jiot.2017.2782318.

[14]N. Jiang, Y. Deng, M. Condoluci, W. Guo, A. Nallanathan and M. Dohler, "RACH Preamble Repetition in NB-IoT Network", *IEEE Communications Letters*, vol. 22, no. 6, pp. 1244-1247, 2018. Available: https://ieeexplore.ieee.org/document/8258982

[15] G. Baracat and J. Brito, "NB-IoT Random Access Procedure Analysis", *IEEE 10th Latin-American Conference on Communications (LATINCOM)*, 2018. Available: https://ieeexplore.ieee.org/document/8613207

[16] Martin, R. Leal, A. Armada and A. Duran, "NBIoT Random Access Procedure: System Simulation and Performance", *Global Information Infrastructure and Networking Symposium (GIIS)*, 2018. Available: https://ieeexplore.ieee.org/document/8635738

[17] R. Harwahyu, R. Cheng, C. Wei and R. Sari, "Optimization of Random Access Channel in NB-IoT", *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 391-402, 2018. Available: https://ieeexplore.ieee.org/document/8239592

[18] "Medium Access Control (MAC) protocol specification", 3GPP Std. 36.321, Jul. 2018, v15.2.0.

[19] Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation, 3GPP Std. TS 36.211, Rev. Release 14, Sept. 2016. Available: https://www.etsi.org/deliver/etsi_ts/136200_136299/136211/13.02.00_60/ts_136211v130200p.pdf

[20] H. Fattah, "5G LTE Narrowband Internet of Things (NB-IoT)", 2018. Available: 10.1201/9780429455056.

[21] L. Feltrin et al., "Narrowband IoT: A Survey on Downlink and Uplink Perspectives", *IEEE Wireless Communications*, vol. 26, no. 1, pp. 78-86, 2019. Available: 10.1109/mwc.2019.1800020.

[22] J. Hwang, C. Li and C. Ma, "*Efficient Detection and Synchronization of Superimposed NB-IoT NPRACH Preambles*", IEEE Internet of Things Journal, vol. 6, no. 1, pp. 1173-1182, 2019. DOI: 10.1109/jiot.2018.2867876

[23] A. Kumar, J. Kuri and D. Manjunath, "Communication Networking: An Analytical Approach" *(The Morgan Kaufmann Series in Networking)*. Morgan Kaufmann Publishers, 2004.