DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING

MASTER'S THESIS

ON THE PERFORMANCE OF MACHINE-TYPE COMMUNICATIONS NETWORKS UNDER MARKOVIAN ARRIVAL SOURCES

Author	Fahad Rahim Qasmi
Supervisor	Dr. Hirley Alves
Second Examiner	Prof. Matti Latva-aho

May 2018

Qasmi F. (2018) On the Performance of Machine-Type Communications Networks under Markovian Arrival Sources. Department of Communications Engineering, University of Oulu, Oulu, Finland. Master's thesis, 52 p.

ABSTRACT

This thesis evaluates the performance of reliability and latency in machine type communication networks, which composed of single transmitter and receiver in the presence of Rayleigh fading channel. The source's traffic arrivals are modeled as Markovian processes namely Discrete-Time Markov process, Fluid Markov process, Discrete-Time Markov Modulated Poisson process and Continuous-Time Markov Modulated Poisson process, and delay/buffer overflow constraints are imposed. Our approach is based on the reliability and latency outage probability, where transmitter not knowing the channel condition, therefore the transmitter would be transmitting information over the fixed rate. The fixed rate transmission is modeled as a two state Discrete time Markov process, which identifies the reliability level of wireless transmission. Using effective bandwidth and effective capacity theories, we evaluate the trade-off between reliability-latency and identify QoS requirement. The impact of different source traffic originated from MTC devices under QoS constraints on the effective transmission rate are investigated.

Keywords: Markov arrivals, QoS constraints, Effective bandwidth, Effective capacity, Optimum transmission rate, Machine-to-machine communication, Traffic Models

TABLE OF CONTENTS

FOREWORD

The focus of this thesis is to study On the Performance of Machine-Type Communications Networks under Markovian Arrival Sources. This research was carried out at Center for wireless communication (CWC) as part of High5 project and was partially supported by Finnish Funding Agency for Technology and Innovation (Tekes), Bittium Wireless, Keysight Technologies Finland, Kyynel, MediaTek Wireless, Nokia Solutions and Networks. I would like to express my sincere gratitude to Dr. Hirley Alves for his guidance, immense knowledge and flexible supervision whenever needed. His comments, suggestions, and vision have been very valuable during my work. I would like to thank Prof. Matti Latva-aho for giving me the opportunity to join the research group in the CWC. I would like to thank Mohammad Shehab for continuously guiding me and helping me out with technical difficulties during my thesis. Finally I would like to thank my parents and wife who supported me all the time and above all to God who led me to this achievement.

ABBREVIATIONS AND SYMBOLS

CSI	Channel State Information
C_E	Effective Capacity
C_E^*	Optimum Effective Capacity
B_E	Effective Bandwidth
EEE	Effective Energy Efficiency
FB	Finite Blocklength
LoS	Line of Sight
LTE	Long Term Evolution
max	maximize
NBP	Non-empty Buffer Probability
PDF	Probability Density Function
CDP	Cumulative Distribution Function
QoS	Quality Of Service
SNR	Signal to Noise Ratio
s.t	subject to
UR	Ultra Reliable
URC	Ultra Reliable Communication
5G	fifth mobile generations
4G	forth mobile generations
V2V	vehicle-to-vehicle communion
IoT	internet of things
CMMPP	Coupled Markov Modulated Poisson Process
DTMS	Discrete Time Markov Source
FMS	Fluid Markov Source
MMPP	Markov Modulated Poisson Process
DTMMPP	Discrete Time Markov Modulated Poisson Process
CTMMPP	Continuous Time Markov Modulated Poisson Process
MTC	Machine Type Communication
MTD	Machine Type Devices
HTC	Human Type Communication
mMTC	Massive Machine Type Communication
URLLC	Ultra Reliability Low Latency communication
eMMB	Extreme Mobile Broadband
FIFO	First in First out
$\Pr$	probability
$A(t)$	Time accumulated arrival process
$R(k)$	Time accumulated service process
Q	stationary queue length in steady state
q	overflow threshold that indicates the tolerance of maximum queue length
θ	delay exponent
ζ	the probability of non-empty buffer
D	queueing delay in steady state buffer
d	delay threshold
$a(\theta)$	effective bandwidth of arrival process
$\mathbf{G}$	transition rate matrix of Markov chain

$\mathbf{J}$	irreducible and aperiodic transition probability matrix of Markov chain
$\mathbf{\Lambda}$	the diagonal matrix of arrival rate in the different states
$sp(.)$	spectral radius of input matrix
$\mu(.)$	represents the maximum real eigenvalue of input matrix
π	stationary distribution of discrete time Markov chain
λ_i	arrival rate in i^{th} state
$\overline{\lambda}$	average arrival rate
$\overline{\lambda}_{max}$	maximum average arrival rate
p_{11}	probability of staying in OFF state
p_{22}	probability of staying in ON state
p_{12}	transition probabilities from OFF state to ON state
p_{21}	transition probabilities from ON state to OFF state
α	transition rate from OFF to ON state
β	transition rate from ON to OFF state
$\mathrm{P_{ON}}$	average ON state probability
s	single parameter for source burstiness
C_E^{opt}	maximize effective capacity
$C(i)$	Shannon capacity
D_{max}	maximum delay
$E[]$	expectation of
$Pr()$	probability of
P_c	power dissipated in circuit
P_{max}	maximum transmission power
P_{nb}	non-empty buffer probability
P_{out_delay}	delay outage probability
$P_t(\rho)$	power consumption
$Q(x)$	Gaussian Q-function
e	exponential Euler's number
$\lvert h\rvert^2$	fading coefficient
$\log$	natural logarithm to the base e
$\log_2$	logarithm to the base 2
m	fading parameter
r	fixed transmission rate
$r*$	optimum transmission rate
$p_z(z)$	fading parameter probability density function
$\mathbf{w}$	additive while Gaussian noise vector
$\mathbf{x}_i$	transmitted signal vector of node
$\mathbf{y}_i$	received signal vector of node
$\mathbf{n}_i$	signal noise
$\mathbf{h}_i$	channel fading coefficient
$\mathbf{z}_i$	square-envelop of Rayleigh fading block coefficients

1. INTRODUCTION

Recent trends in research and development suggest that the Fifth Generation (5G) of mobile network may bring technology evolution in the form of expanding broadband capacity, mobility and cloud services. This evolution does not seem to be standalone but almost every industry seem to have a major impact and may need a re-definition of their business models [1]. 5G is not just extension of the 4G technologies, therefore it not only focuses on the enhanced coverage, connectivity, data rates and spectral efficiency but also addresses critical and massive traffic generated by machine type devices. Such devices operate and communicate with little or no human interaction and is called Machine Type Communication (MTC) [2]. The 5G use cases can be categorized into three types of communication in terms of the requirement and their objectives [3]:

- Extreme Mobile Broadband (eMBB): It provides high data rate with low latency communication. Moreover, it also provides extreme coverage with uniform connectivity and data rate over the edge of cell. However the performance degrades as the number of devices increase.

- Ultra Reliability Low Latency communication (URLLC): The use case is applied for real time applications where latency requirement is extremely low. This mode performs two main functions i.e., Ultra-reliability that can extremely ensure that the communication services with low probability of failure in the presence of stringent delay requirement and latency has been defined as the delay experience by the packet from transmitter to the receiver. This type of communication is used in mission critical application applications, such as vehicle-to-vehicle communion (V2V), remote surgery and critical link between industrial processes.

- Massive MTC (mMTC): It is used when a large number of devices (approximately hundreds of thousand devices per square kilometer) are deeply installed within a cell. It can be used to cover monitoring, automation and infrastructure of buildings, smart agriculture, logistics, tracking and fleet management. The devices are normally used to collect and forward information in both real as well as non-real time modes. The main objective for this type of communication is to provide ubiquitous connectivity in order to reduce software and hardware complexity. As machine installed within the coverage area may have different requirements as per data rate, latency, reliability, number of device and connectivity technologies therefore the requirements may also vary according to the nature of application.

The Internet of Things (IoT) promises huge market growth with expected 50 billion-connected devices by 2020 [4]. Although some of the IoT applications are suitable for wired or short-range radio but increasingly cellular networks are becoming more appropriate in order to fulfill the requirements of MTC applications in terms of mobility, coverage, ecosystem, diversity and ease of deployment. Therefore a massive number of machines are expected in the future cellular networks; and as a result, it will not only generate new revenue but will also have deep implications on the end-to-end network architecture. Here decreasing the power consumption and deployment cost are

the primary requirements in welcoming the migration from high data rate network to MTC optimized and low cost networks [5].

MTC has attracted much interest in the recent years. It is important to study massive MTC in the light of new scenarios; however, one of the major bottlenecks in experimentations of the scenarios is that the number of machines is far below than what is expected for future growth. Hence a realistic definition of traffic models and reference scenario is required which can be used to validate the current and future network scenarios [6].

Traffic model is a stochastic process that matches the behavior of physical quantities of measured data traffic [7]. Current cellular network is based on standard traffic model which is designed and optimized for typical behavior of human subscribers. It is most typically used in day and evening for phone calls, sending SMS/MMS, video streaming, video calls and downloading possibly large volumes of data. The massive traffic generated by the machines was not taken in consideration when designing the current cellular network traffic models. However the traffic behavior of MTC is quite different from the HTC one. The main traffic properties of the MTC model are given as [8]:

- MTC Traffic is mostly uplink dominant.
- MTC traffic generation frequency is typically all around the day i.e., 24 hours while HTC traffic mostly flow duration is day or evening time but not at night mostly.
- MTC traffic is homogenous in nature (i.e. all machine running same application behave similarly), however HTC traffic is heterogeneous.
- MTC traffic is bursty (suddenly the volume of data flow increase in response to trigger of certain events).
- MTC traffic Raw and aggregated packets (i.e. combine traffic generated from multiple sources in to a single packet, which is sent to the specific node gateway.)
- MTC is coordinate (i.e simultaneous access attempts from many machine reacting to the same events), while HTC is uncoordinated.
- MTC uses short as well as small number of packets.
- MTC traffic is real as well as non-real time, periodic and event driven.
- MTC QoS requirement is different from HTC (i.e. different reliability and latency requirements).

Therefore new traffic models are needed so to capture the behavior of massive MTC traffic and also provide the adequate communication services for both types of traffic communication with required Quality of Service (QoS) [9].

The traffic models are mainly classified into source and aggregated traffic models. The source traffic models capture the traffic pattern of individual users or sources, hence it is not feasible to model the traffic generate by large amount of autonomous

machine simultaneously because due to heterogeneous and uncoordinated characteristics of the traffic. It can be said that the source traffic models were designed for human generated traffic; which are based on Poisson arrival rates; however, it is not applicable in the case of MTC applications as Poisson distribution usually fails to capture the burstiness and multimodality of the real traffic sequence [10]. While aggregated traffic models capture the traffic properties as a group of users or networks therefore, mostly aggregated traffic models are suitable for MTC network and having homogeneous and coordinated characteristics of the traffic, where the number of machines assigned to one server are called aggregator and are used for capturing the traffic patterns [10].

As discussed above, source and aggregated traffic model are the main categories of traffic models. The aggregated models can be further divided into sub-models namely correlate (with high synchronization) and uncorrelated (with no -synchronization). The sub-models have been used in many applications but have often suffered from limited precision, no spatial correlation and poor flexibility. On the other hand, source traffic models have proved to be more precise along with having high computation complexity. Three state semi-Markov model is the example of source traffic model that has been used in many areas of applications. Nonetheless, the main disadvantage for this model is the independence of sources from each other without temporal and spatial correction that often requires incorporate of additional processes. Coupled Markov Modulated Poisson Process is effectively negotiation between source and aggregated traffic model. This model captures the spatial and temporal correction with Poisson process as source traffic model [10].

The communication traffic for various MTC services are often categorized for example, as periodic, event and payload exchange basis [11]:

- Periodic update (PU): This type of traffic is usually generated by a machine and this is used to periodically transmit the status report for updating the center unit. PU is non-real time traffic having constant data size. The server as per the requirements of an application reconfigures these parameters. For example, humidity sensor may send data to the server after every10 minute duration.

- Event Driven (ED): In this type of traffic, a machine is triggered by an event and correspondingly it sends the data to the server. The event may be the result of certain threshold's exceed value or by the server to send required data to application. This type of traffic is real time with variable time and data size in uplink and downlink direction. For example, temperature sensor in an industrial plant sends data when the values exceed threshold.

- Payload Exchange (PE): This is another type of traffic generated data by a machine that is sent to the central unit in the response of PU or ED traffic type. It is non-real time, variable data size and mostly uplink dominates with large amount of data is exchange between server and machine. For example, temperature sensor sent PE by taking and sending a picture of current situation after ED traffic was generated for informing that excessed threshold value.

In many MTC use cases, Quality of Service (QoS) is the primary requirement for acceptable performance and efficiency. For instance, ED in MTC application, latency and reliability are the key metrics of QoS that should not exceed specific threshold.

Wireless channel change due the changes in environment and multipath fading. These random changes may lead to variations in strength of received signals, which affects the QoS requirement of the MTC type applications.

Hence, source characteristics, reliable transmission rate are also time varying. In this case time varying sever is required in the queueing system. Therefore, effective capacity is link layer model that can ensure QoS in time varying wireless channel. In this context effective capacity is defined as the maximum constant arrival rate that a given time varying service process can support while providing statistical QoS guarantees. It is derived from the large deviation theory and incorporates the statistical QoS constraint by capturing the decay rate of the buffer occupancy probability for the queue length. Effective capacity of wireless channels mainly focuses on constant arrival rates in the analysis of throughput. Here we are particularly interested in using Markovain source models including discrete time Markov, Markov fluid and Markov modulated Poisson sources with effective capacity to conduct throughput analysis of random and bursty source traffic pattern [12, 13, 14, 15].

Recently effective capacity of wireless communication has been attracted much attention to estimate reliability, latency, security, energy efficiency and power control [16, 17, 18, 18, 19, 20]. For instance, in [21], authors consider fixed rate transmission technique and the effective capacity evaluate energy efficiency under QoS constraints. The authors used fixed-rate transmission is modeled as a two-state (ON/OFF) discrete-time Markov chain. In [22] the authors considered fixed-rate transmission is modeled as a two-state (ON/OFF) continuous-time Markov chain and effective capacity analyses energy efficiency with Markov arrival under QoS constraint. In [23], the authors considered effective capacity when the transmitter and receiver know the instantaneous channel gain, and derived the optimal power and rate adaptation technique under QoS constraint which maximize the throughput.

1.1. Thesis contribution

In this thesis, we evaluate the performance of MTC network, which composed of single transmitter and receiver point-to-point link. Using effective capacity, we design the reliable and latency aware wireless communication link layer model. By estimating the impact of random and bursty arrival of traffic, we incorporate Markovain source arrival processes with reliable wireless communication model as well.

Different from [22], [24] that the fixed-rate transmission which modeled as a two-state (ON/OFF) discrete-time Markov chain and effective capacity analyses reliability and latency with Markovian arrival process under QoS constraint.

In [21], the authors evaluating the impact of the constant source arrival traffic in the design of effective energy efficiency (EEE) model. We build our contribution based upon the reliability and latency framework proposed in [21]. This framework help us to determine the level of reliability and latency in each transmission rate. We also build upon contribution [25] to incorporate different arrival source traffic so we estimated the performance of reliability of network. We consider ON-OFF state channel model to investigate throughput analysis of fixed and optimum transmission rate of random and bursty arrivals of sources. The optimum transmission rate is the one that maximizes the system throughput. In [21], reliability and latency trade-off has been analyzed

under effective capacity have mainly focus on fixed constant source arrival rates. In this thesis, we take into the account stochastic nature of source information flow and then we investigate the impact of the randomness and burstiness on the reliability and latency of the wireless link. We assume that the data sources are modeled as Markovian source models namely

- Discrete-Time Markov Source.
- Fluid Markov source.
- Discrete-Time Markov Modulated Poisson process.
- Continuous-Time Markov Modulated Poisson process.

Additionally, it is assume that the time varying channel condition are not known the transmitter therefore the transmitter would be transmitting information over the fixed rate. Fixed rate transmission over the Rayleigh fading channel is modeled as a discrete time Markov process. Using these modeling assumptions the contribution of this thesis can be further detailed as follows:

- The introduction of general framework for performance analysis of the effective rate using effective capacity and incorporating with Markovian arrival models.
- Throughput expressions are provided and subsequently reliability-latency metric is identified for the Markovian source models discussed above. Overall, a framework is provided to study reliable transmission rate in the presence of random arrival with QoS constraints.
- We propose an ON-OFF adaptive fixed rate transmission scheme for which we provided an optimal transmission rate expression in closed-form, which maximizes the throughput of the system.
- The impact of randomness, burstiness, probability of buffer overflow, and channel fading on the optimal transmission rate is evaluated via numerical and analytical analysis.

Finally, from this thesis, we have managed to write the following paper:

- F. Qasmi, M. Shehab, H. Alves, and M. Latva-aho, "Optimum Transmission Rate in Fading Channels with Markovian Sources and QoS Constraints," in 2018 International Symposium on Wireless Communication Systems, Lisbon, Portugal, August 2018.

1.2. Thesis Outline

The rest of the thesis is organized as follows: in Chapter 2, we define queuing constraints, effective bandwidth, effective bandwidth of Markovian Arrival Source Models and effective capacity. In Chapter 3, we obtain a close form expression of effective capacity for point-to-point link in Rayleigh fading channel, which maximize the maximum average arrival rate. After that, in Chapter 4, we analyze the effects of randomness, burstiness, throughput and delay outage probability to the reliable transmission rate. Finally, we state the thesis conclusion and suggested future work in Chapter 5.

2. PRELIMINARIES

Most of MTC uses cases have diverse QoS requirements and traffic characteristics. QoS guarantees in wireless links have been studied mostly relying on simple though accurate channel models. Physical channel models do not characteristerize the wireless channel in term of QoS metrics but mainly estimate the fluctuation in frequency, phase and amplitude. To use these models for required QoS constraint we need first to estimate the parameters from physical models and then extract QoS metrics. This two step approach may become too complex which may lead to inaccurate results. Therefore, it is important to model a wireless channel in term of data rate, delay and delay violation probability [24].

We have to identify the probabilistic properties of flow of arrival data and service time to achieved guarantee QoS requirement, by doing so we are able to identify QoS parameters, efficiently pertain resource allocation and work with a tractable analytical framework.

2.1. Queuing constraint

The data generated by random sources is stored in a First in First out (FIFO) buffer at the transmitter before transmission. In general, statistical QoS constraints would be applied with the purpose of restricting buffer overflow probability, which is defined as

$$\lim_{q\to\infty} \frac{\log \Pr\{Q \geq q\}}{q} = -\theta, \tag{1}$$

where Q represents the stationary queue length in steady state, θ represents the decay rate of the tail distribution of the queue length, and q is overflow threshold that indicates the tolerance of maximum queue length. Notice that $\log$ denotes the natural logarithm throughout this thesis.

In the case of large value of q, (1) is approximated as [28]

$$\Pr\{Q \geq q\} \approx \zeta e^{-\theta q}, \tag{2}$$

where $\zeta = \Pr\{Q > 0\}$ is the probability of non-empty buffer, which can be approximated by the ratio of the constant arrival rate to the average service rate. From (2), it can be seen that in the cases of larger value of the threshold, the overflow probability may have an exponential decay with rate controlled by θ. Here we can consider that θ is non-negative scalar value which can also be called as QoS exponent. Therefore it can be implied that the larger value of θ ($\theta \to \infty$) would mean strict limitations on the buffer overflow probability which indicate that stringent QoS constraint has been imposed, thus the system can not tolerate delay. On the other hand, for loose QoS requirement, the value of θ is small ($\theta \to 0$), which may lead to lower buffer overflow probability therefore system can tolerate larger delays [28].

Figure 2.1 shows the buffer overflow probability as function of θ. It is observed that as the value of QoS exponent θ is increased the buffer overflow probability is decreased which indicates stringent QoS is applied in buffer. $\Pr\{Q \geq q\}$ is equal to overflow probability limit. However, delay may occur in the transmission system because of

long wait of data in the buffer in steady state, therefore buffer overflow probability can also be characterized as delay violation probability which is given as (3) [29].

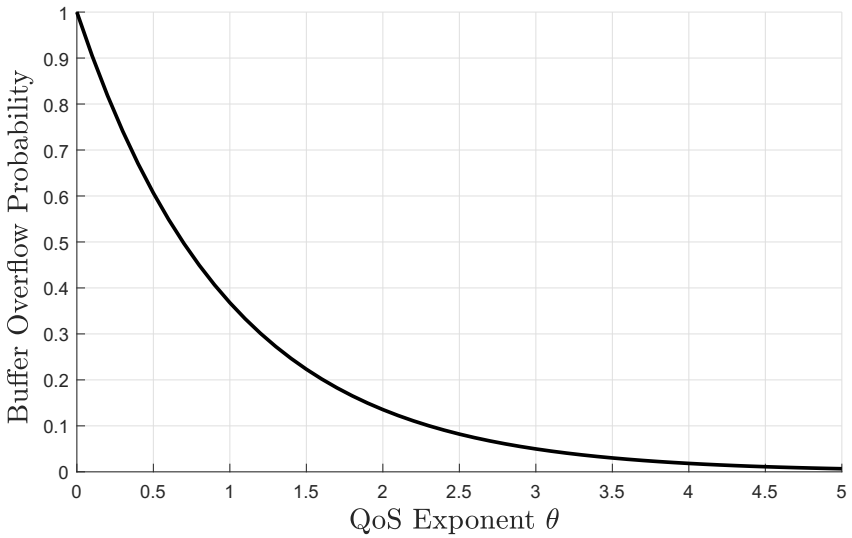

Figure 2.1: Buffer overflow probability as a function of the QoS exponent θ, which represents the queuing constraints imposed on the buffer.

$$\Pr\{D \geq d\} \approx \zeta e^{-\theta \, B_E \, d}. \tag{3}$$

where D is the queueing delay in steady state buffer, d is delay threshold , and B_E is effective bandwidth of arrival process.

2.2. Effective Bandwidth

The effective bandwidth provides means to characterize a minimum constant service rate required to support the random arrival of data in buffer under some statistical QoS requirements, such as buffer violation probability.

Let $\{a(k), k = 1, 2, \cdots\}$ shows the series of random arrival rate non negative variables, and it is represented by time accumulated arrival process as a

$$A(t) = \sum_{k=1}^{t} a(k). \tag{4}$$

Then the effective bandwidth can be characterized by the asymptotic logarithmic generation function of $A(t)$ expressed as [30]

$$B_E(\theta) = \lim_{t \to \infty} \frac{1}{\theta t} \log \mathbb{E}\{e^{\theta A(t)}\}. \tag{5}$$

2.3. Effective Bandwidth of Markovian Arrival Source Models

Herein we evaluate the B_E expressions of four distinct ON-OFF Markovian arrival sources, their probability of being active (ON state) $\mathrm{P_{ON}}$, which is used to the calculate average arrival rate $\overline{\lambda}$ when buffer in the steady state, and then maximum average arrival rate $\overline{\lambda}_{max}$ of Markovian sources that can satisfying QoS requirement, when effective bandwidth of arrival process is equal to the effective capacity of service process.

2.3.1. Discrete Time Markov Sources (DTMS)

In this section, source data arrival is modeled as a discrete in time. We assume irreducible and aperiodic transition probability matrix of Markov chain and that is denoted by **J**, arrival rate is represented as a λ_i in i^{th} state, and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_n\}$ is the diagonal matrix of arrival rate in the different states. Then, the effective bandwidth of discrete source model is described by [31]

$$B_E(\theta) = \frac{1}{\theta}\log[\mathrm{sp}(e^{\theta\mathbf{\Lambda}}\mathbf{J})] \tag{6}$$

where $\mathrm{sp}(.)$ represents the spectral radius of input matrix and the stationary distribution π of the discrete time Markov chain, which are determined by the following equation

$$\begin{aligned} \pi\mathbf{1} &= 1, \\ \pi\mathbf{J} &= 1 \end{aligned} \tag{7}$$

where $\pi = [\pi_1, \pi_2, \cdots \pi_n]$ and $\mathbf{1} = [1, \cdots, 1]^T$.

As an example, and for easy of mathematical tractability, consider a two state (ON-OFF) model as depicted in figure 2.2.

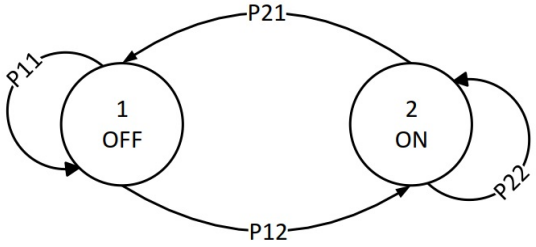

Figure 2.2: The ON-OFF state discrete Markov source Model.

We assume that λ bits arrive in ON state and no arrival of bits in the OFF state. Then the transition probability matrix **J** for a two state discrete time source model is defined as

$$\mathbf{J} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{8}$$

Hence, using above transition probability matrix **J** in (8) into (6), the effective bandwidth for this ON-OFF Markov model can be derived as [31],

$$B_E(\theta) = \frac{1}{\theta}\log\left(\frac{1}{2}\left(p_{11} + p_{22}e^{\theta\lambda} \ + \sqrt{(p_{11} + p_{22}e^{\theta\lambda})^2 + 4(p_{11} + p_{22} - 1)e^{\theta\lambda}}\right)\right), \tag{9}$$

where p_{11} determines the probability of staying in OFF state, while p_{22} identifies the probability of being in the ON state. The transition probabilities from one state to another are denoted by $p_{21} = 1 - p_{22}$ and $p_{12} = 1 - p_{11}$.

As illustrated in Figure 2.2, $\mathrm{P_{ON}}$ is the probability of ON state in the steady state regime, which is used to the calculate average arrival rate. Substituting transition probability matrix $\mathbf{J}$ in (8) into (7), we obtain the ON state probability as

$$\mathrm{P_{ON}} = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}. \tag{10}$$

In addition, average arrival rate is defined as

$$\overline{\lambda} = \lambda \, \mathrm{P_{ON}} = \lambda \, \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \, , \tag{11}$$

which is equal to the departure rate when the buffer is in steady state [32].

2.3.2. Markov Fluid Source (FMS)

In this section, the mechanism of data arrival is modeled as continuous in time, again we assume a two state model. The irreducible transition rate matrix of Markov chain is denoted by $\mathbf{G}$, the arrival rate is represented as λ_i in i^{th} state, and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_n\}$ is the diagonal matrix of arrival rate in the different states. Then, the effective bandwidth of continuous source model is expressed as [33], [34]

$$B_E(\theta) = \mu(\mathbf{\Lambda} + \frac{1}{\theta}\mathbf{G}) \tag{12}$$

where $\mu(.)$ represents the maximum real eigenvalue of input matrix and the stationary distribution π of the continuous time Markov chain can be determined from

$$\begin{aligned} \pi\mathbf{1} &= 1, \\ \pi\mathbf{G} &= 0 \end{aligned} \tag{13}$$

where $\pi = [\pi_1, \pi_2, \cdots \pi_n]$ and $\mathbf{0} = [0, \cdots, 0]^T$.

As in the previous case, it is assumed bits arrive at rate λ during ON state and no arrival of bits in the OFF state. Then, the transition rate matrix $\mathbf{G}$ for a two state continuous time source model is expressed as follows

$$\mathbf{G} = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}. \tag{14}$$

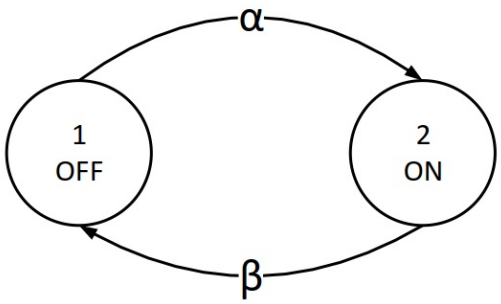

Figure 2.3: The ON-OFF state fluid Markov source Model.

While Figure 2.3 illustrates this model and its transition rates matix $\mathbf{G}$, α denotes the transition rate from OFF to ON state and β is the transition rate from ON to OFF state. Using above transition rate matrix G in (14) into (12), the effective bandwidth of two state (ON-OFF) Markov model is expressed through [25]

$$B_E(\theta) = \frac{1}{2\theta}\left[\theta\lambda - (\alpha+\beta) + \sqrt{(\theta r - (\alpha+\beta))^2 + 4\alpha\theta\lambda}\right], \tag{15}$$

then, we substitute generator rates matrix $\mathbf{G}$ in (14) into (13), attain the probability of ON state, in steady state as

$$\mathrm{P_{ON}} = \frac{\alpha}{\alpha+\beta}\,, \tag{16}$$

further, the average arrival rate is

$$\overline{\lambda} = \lambda\,\mathrm{P_{ON}} = \lambda\,\frac{\alpha}{\alpha+\beta}\,. \tag{17}$$

2.3.3. Discrete Time Markov Modulated Poisson Process (DTMMPP)

Next, we discuss that source data arrival in buffer is modeled as a Poisson process, whose intensity is controlled by discrete-time Markov chain. For instance, λ_i is the intensity of Poisson arrival process in the i^{th} state in Markov chain. Therefore source arrival is modeled as Markov modulated Poisson process. We assume irreducible and aperiodic transition probability matrix of Markov chain and that is denoted by $\mathbf{J}$, and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_n\}$ is the diagonal matrix of the Poisson arrival rate in the different state. Afterwards, the effective bandwidth of discrete Markov Modulated Poisson Process is expressed as [33], [34]

$$B_E(\theta) = \frac{1}{\theta}\log[\mathrm{sp}(e^{(e^{\theta}-1)\mathbf{\Lambda}}\mathbf{J})] \tag{18}$$

where $\mathrm{sp}(.)$ represents the spectral radius of input matrix and the stationary distribution π of the discrete time Markov chain, which are determined by the following equation

$$\begin{aligned} \pi\mathbf{1} &= 1, \\ \pi\mathbf{J} &= 1 \end{aligned} \tag{19}$$

where $\pi = [\pi_1, \pi_2, \cdots \pi_n]$ and $\mathbf{1} = [1, \cdots, 1]^T$.

As an example, and for easy of mathematical tractability, consider a two state (ON-OFF) model as depicted in Figure 2.2.

Discrete Time Markov chain follows the same structure as in described in Section 2.3.1, however, we opt to repeat the steps and notation for easy of readability and completeness. Thus, we assume that λ bits arrive in ON state and no arrival of bits in the OFF state. Then the transition probability matrix $\mathbf{J}$ for a two state discrete time source model is defined as

$$\mathbf{J} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{20}$$

As in previous Sections, with the transition probability matrix $\mathbf{J}$ with simple two state (ON-OFF) model. We assume that the no arrival in OFF state, while λ bits is intensity of the Poisson arrival process in the ON state. Therefore assuming the same generator matrix $\mathbf{J}$ as in (20) into (18), we can express the effective bandwidth as

$$B_E(\theta) = \frac{1}{\theta}\log\left(\frac{p_{11}+p_{22}e^{\lambda(e^{\theta}-1)}}{2} + \frac{\sqrt{\left(p_{11}+p_{22}e^{\lambda(e^{\theta}-1)}\right)^2 - 4(p_{11}+p_{22}-1)e^{\lambda(e^{\theta}-1)}}}{2}\right), \tag{21}$$

where p_{11} determines the probability of staying in OFF state, while p_{22} identifies the probability of being in the ON state. The transition probabilities from one state to another are denoted by $p_{21} = 1 - p_{22}$ and $p_{12} = 1 - p_{11}$ as discussed as well for the DTMS [25]. The average arrival rate in bps is again given by

$$\overline{\lambda} = \lambda\, \mathrm{P}_{\mathrm{ON}} = \lambda\, \frac{1-p_{11}}{2-p_{11}-p_{22}}. \tag{22}$$

Note that there are similarities between expressions (9) and (21); however, their behavior differs considerably due to the randomness introduced by the Poisson process.

2.3.4. Continuous Time Markov Modulated Poisson Process (CTMMPP)

Herein, we extend the model analyzed in Section 2.3.3 by modeling the source's data arrival as a Poisson process, whose intensity changes according to the continuous time values in the Markov chain. For example, λ_i is the intensity of Poisson arrival process in the i^{th} state in Markov chain. Thus source arrival is modeled as Markov modulated Poisson process. We assume irreducible transition rate matrix is denoted by $\mathbf{G}$ of Markov chain, and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_n\}$ is the diagonal matrix of the Poisson arrival rate in the different state. Afterwards, the effective bandwidth of continuous Markov Modulated Poisson Process is expressed as [33], [34]

$$B_E(\theta) = \frac{1}{\theta}\mu((e^{\theta}-1)\mathbf{\Lambda} + \mathbf{G}) \tag{23}$$

where $\mu(.)$ represents the maximum real eigenvalue of input matrix. Moreover and the stationary distribution π of the continuous time Markov chain can be determined from

$$\begin{aligned} \pi\mathbf{1} &= 1, \\ \pi\mathbf{G} &= 0 \end{aligned} \tag{24}$$

where $\pi = [\pi_1, \pi_2, \cdots \pi_n]$ and $\mathbf{0} = [0, \cdots, 0]^T$.

With transition rate matrix $\mathbf{G}$ with simple two state (ON-OFF) model. We assume that the no arrival in OFF state, while λ denotes the intensity of the Poisson arrival process in ON state. Therefore, the same transition rate matrix $\mathbf{G}$ into (24), the effective bandwidth of two state (ON-OFF) Markov model is expressed as

$$\mathbf{G} = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}. \tag{25}$$

Note that, α shows the transition rate from OFF to ON state and β is the transition rate from ON to OFF state. Using above transition rate matrix $\mathbf{G}$ in (25) into (23), the effective bandwidth of two state (ON-OFF) Markov model is expressed as given here [25]

$$B_E(\theta) = \frac{1}{2\theta}\left[\,(e^{\theta}-1)\lambda - (\alpha+\beta)\right] + \frac{1}{2\theta}\sqrt{(\,(e^{\theta}-1)\lambda - (\alpha+\beta))^2 + 4\alpha(e^{\theta}-1)\lambda}. \tag{26}$$

We substitute then the transition rate matrix $\mathbf{G}$ in (25) into (24), as in the previous case, attain the $\mathrm{P_{ON}}$ as

$$\mathrm{P_{ON}} = \frac{\alpha}{\alpha+\beta}\,, \tag{27}$$

further, the average arrival rate is

$$\overline{\lambda} = \lambda\,\mathrm{P_{ON}} = \lambda\,\frac{\alpha}{\alpha+\beta}. \tag{28}$$

We further note that if the transition rate β = 0, then the $\mathrm{P_{ON}} = 1$. In this case, CTMMPP model specializes to a pure Poisson source with intensity λ, and the effective bandwidth of this source is given by

$$B_E(\theta) = \frac{1}{\theta}(e^{\theta}-1)\lambda \tag{29}$$

It is note that effective bandwidth expressions in (15) and (26) are some similarities, due to the randomness behavior of CTMMPP multiplicative factor $(e^{\theta}-1)$ in (26) gives great difference in performance.

Figure 2.4, where we examine the impact of QoS constraint to the effective bandwidth for different Markovian source arrival Models assuming fixed values of source burstiness $\mathrm{P_{ON}}$, and arrival rate λ. As θ increase QoS requirement becomes too stringent, then the effective bandwidth tends to increase and can be easily proved that discrete Markovain models have larger variation as compared to the Fluid process at tight QoS constraints.

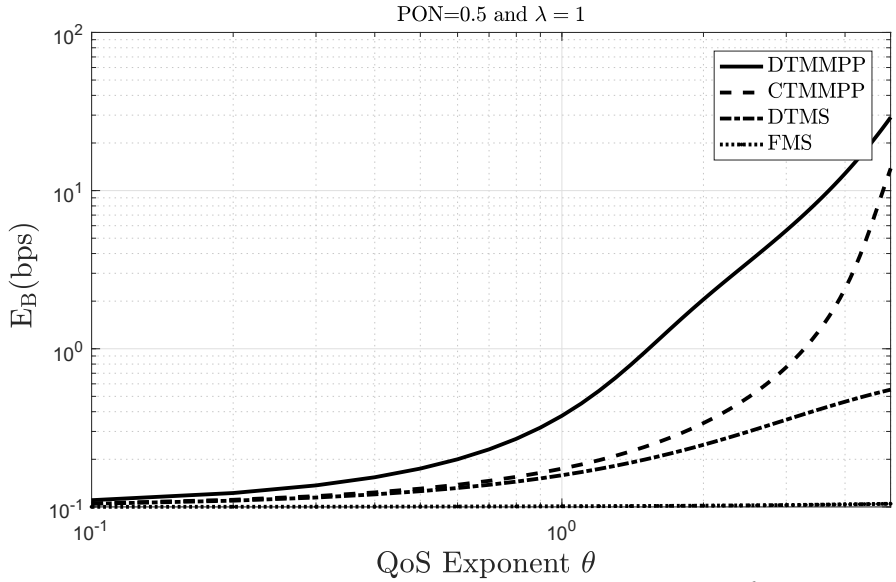

Figure 2.4: Effective bandwidth as a function of QoS exponent θ for different Markovian source models.

2.4. Effective capacity

Effective capacity C_E is a dual concept of effective bandwidth, which uses to ensure the QoS requirement in a wireless communication link. Effective capacity states the maximum constant arrival rate that can support a time varying service process, such as wireless fading channel, while satisfying QoS constraint is specified by QoS exponent θ [28, 35], which is defined in Section 2.1.

Let $\{v(k), k = 1, 2, \cdots\}$ denote the discrete-time stationary and ergodic stochastic services process and

$$S[t] = \sum_{k=1}^{t} v(k), \tag{30}$$

be the time accumulated service process.

Then, the effective capacity is defined as

$$C_E(\theta) = -\lim_{t\to\infty} \frac{1}{\theta t} \log \mathbb{E}\{e^{-\theta S[t]}\}. \tag{31}$$

Usually performance measures of B_E and C_E behave under the queue length, because if the queue length is finite or short then the lower arrival rate is expected, which may lead data loss (with high probability) when the queue is full. Hence, system with limited queue length normally requires more energy despite having and worst performance [21].

At this point, we aim to introduce the concept of C_E, but we remark that further details of the C_E for the system model under analysis shall be given in Chapter 3.

2.5. Throughput with Markovian Source Models

In what follows, we define the throughput of wireless fading channel. Throughput is becomes quite important as well as challenging when the account for random data arrivals and, in special, when we want to impose the QoS constraints, such as buffer overflow probability. Given the arrival models described in Section 2.3, without loss of generality we assume that a rate of arrival λ during ON state, or otherwise zero during OFF state. For instance λ_i is the arrival rate in the i^{th} state and stationary distribution π of the Markov process, then the n-state source Markov chain average arrival rate becomes quite simple to define in the following manner,

$$\overline{\lambda} = \sum_{i=1}^{n} \pi_i \, \lambda_i, \tag{32}$$

which is equal to the depature rate when the queue is in steady state.

Now, we are interested to find as a maximum average arrival rate of Markovian sources that can support fading channel and which can satisfying the QoS requirement (2). As pointed out in [25], the QoS requirement is satisfied when effective bandwidth of arrival process is equal to the effective capacity of service process service process in other words

$$B_E(\theta) = C_E(\theta). \tag{33}$$

Therefore, the maximum arrival rate that can support fixed rate transmissions at given signal to noise ratio (SNR) and θ.

$$\overline{\lambda}_{max} = \lambda_(\theta)\ \mathrm{P_{ON}}. \tag{34}$$

Next, we take the into account the service process (namely effective capacity) and the equality in (33). Then, we define the maximum average arrival rate for each of the arrival sources described in the previous sections, as we shall see next.

2.5.1. Discrete Markov Sources

In this case, two states ON-OFF Markov discrete source model, in order to determine maximum average arrival rate in term of $C_E(\theta)$, we substituting the effective bandwidth formulation of discrete time Markov source in (9), we can express (33) in the following equivalent form:

$$\frac{1}{\theta}\log\left(\frac{1}{2}\left(p_{11}+p_{22}e^{\theta\lambda}\ +\sqrt{(p_{11}+p_{22}e^{\theta\lambda})^2+4(p_{11}+p_{22}-1)e^{\theta\lambda}}\right)\right)=C_E. \tag{35}$$

Further simplifying (35), then we obtain

$$p_{11}+p_{22}e^{\theta\lambda}+\sqrt{(p_{11}+p_{22}e^{\theta\lambda})^2-4(p_{11}+p_{22}-1)e^{\theta\lambda}}=2e^{\theta C_E}. \tag{36}$$

Now, exchanging the first two terms on the left-hand side to the right-hand side and then taking the square of both sides, we have

$$(p_{11}+p_{22}e^{\theta\lambda})^2-4(p_{11}+p_{22}-1)e^{\theta\lambda}=\left(2e^{\theta C_E}-p_{11}-p_{22}e^{\theta\lambda}\right)^2. \tag{37}$$

After further, shifting the second term on the left-hand side with the term on the right-hand side, we obtain

$$(p_{11}+p_{22}e^{\theta\lambda})^2-\left(2e^{\theta C_E}-p_{11}-p_{22}e^{\theta\lambda}\right)^2=4(p_{11}+p_{22}-1)e^{\theta\lambda}. \tag{38}$$

$$\left(2p_{11}+2p_{22}e^{\theta\lambda}-2e^{\theta C_E}\right)2e^{\theta C_E}=4(p_{11}+p_{22}-1)e^{\theta\lambda}. \tag{39}$$

Now, rearrangements, we have

$$(p_{11}+p_{22}-1-p_{22}e^{\theta C_E})e^{\lambda\theta}=p_{11}e^{\theta C_E}-e^{2\theta C_E}. \tag{40}$$

After solving (40) for λ, we obtain maximum ON state arrival rate as [25]

$$\lambda_{max}=\frac{1}{\theta}\log\left(\frac{e^{2\theta C_E(\theta)}-p_{11}e^{\theta C_E(\theta)}}{(1-p_{11}-p_{22})+p_{22}e^{\theta C_E(\theta)}}\right). \tag{41}$$

Therefore, using (34), we expresses the maximum average arrival rate in term of QoS exponent, effective capacity fading channel and state transition probabilities [25]

$$\overline{\lambda}_{max}=\frac{\mathrm{P_{ON}}}{\theta}\log\left(\frac{e^{2\theta C_E(\theta)}-p_{11}e^{\theta C_E(\theta)}}{(1-p_{11}-p_{22})+p_{22}e^{\theta C_E(\theta)}}\right). \tag{42}$$

Let us further simplify the source model thus $p_{11} = 1 - s$ and $p_{22} = s$. In this case, source characteristics describe in single parameter s. It is notice that $\mathrm{P_{ON}} = s$ and s can be seen as a measure the burstiness of the source. The smaller the value of s, the less frequently the data arrives and more bursty source becomes. On the other hand if $s = 1$ source is ON all the time and we have constant arrivals. Furthermore, with the above choice of p_{11} and p_{22}, the expression for the maximum average arrival rate simplifies to

$$\overline{\lambda}_{max} = \frac{s}{\theta} \log \left(\frac{e^{\theta C_E(\theta)} - (1-s)}{s} \right). \tag{43}$$

2.5.2. Markov Fluid Sources

For Fluid Markov source, in order to obtain maximum average arrival rate in term of $C_E(\theta)$, we insert the effective bandwidth expression of Fluid Markov source as in (15) and simplify (33) following similar steps in as the DTMS detailed in Section 2.5.1,

$$(\theta\lambda - (\alpha+\beta) - 2\theta C_E(\theta))^2 = (\theta\lambda - (\alpha+\beta))^2 + 4\alpha\theta\lambda. \tag{44}$$

After solving (44) for λ, we obtain maximum ON state arrival rate as

$$\lambda_{max} = \frac{\theta C_E(\theta) + \alpha + \beta}{\theta C_E(\theta) + \alpha} C_E(\theta). \tag{45}$$

Therefore, using (34), we expresses the maximum average arrival rate in term of QoS exponent, effective capacity fading channel and source transition rate [25].

$$\overline{\lambda}_{max} = \mathrm{P_{ON}} \frac{\theta C_E(\theta) + \alpha + \beta}{\theta C_E(\theta) + \alpha} C_E(\theta). \tag{46}$$

2.5.3. Discrete Time Markov Modulated Poisson Sources

Following the same rationale as in the previous cases, in order to determine the maximum average arrival rate in terms of $C_E(\theta)$, we insert the effective bandwidth expression in (21) into (33) and obtain

$$\left(p_{11} + p_{22} e^{\lambda(e^\theta - 1)} - 2e^{\theta C_E} \right)^2 = \left(p_{11} + p_{22} e^{\lambda(e^\theta - 1)} \right)^2 - 4(p_{11} + p_{22} - 1) e^{\lambda(e^\theta - 1)}. \tag{47}$$

After solving the above equation for λ, we obtain maximum average arrival rate as

$$\overline{\lambda}_{max} = \frac{\mathrm{P_{ON}}}{(e^\theta - 1)} \left[\log \left(\frac{e^{2\theta C_E(\theta) - p_{11} e^{\theta C_E(\theta)}}}{(1 - p_{11} - p_{22}) + p_{22} e^{\theta C_E(\theta)}} \right) \right]. \tag{48}$$

2.5.4. Continuous Time Markov Modulated Poisson Sources

In this section, similarly as for pervious source models, we find the maximum average arrival rate of two state ON and OFF CTMMPP source model by incorporating (26) into (33) and expressing (33) as

$$\Big((e^{\theta}-1)\lambda-(\alpha+\beta)-2\theta C_E\Big)^2=\Big((e^{\theta}-1)\lambda-(\alpha+\beta)\Big)^2+4\alpha(e^{\theta}-1)\lambda. \tag{49}$$

We can simplify above equation and solve for maximum Poisson arrival intensity in the ON state to obtain the maximum average arrival rate as

$$\overline{\lambda}_{max}=\mathrm{P_{ON}}\frac{\theta[\theta C_E(\theta)+\alpha+\beta]}{(e^{\theta}-1)\theta C_E(\theta)+\alpha}C_E(\theta). \tag{50}$$

2.6. Comparative View of Source Models and Performance Level

In this section we compare all source models discussed in this thesis. The discrete time model described state transitions using transition probability matrix where transition between states occur in discrete time steps and Markov fluid model use transition rate matrix where rate is the time Markov chain spend continuous in one state. The state holding time is geometric distribution in discrete time and exponential distribution in fluid Markov sources. In discrete Markov and Markov fluid source models, arrival rate are assumed to be constant in any state, whereas Markov Modulated Poisson process have different intensity of rate in each state.

In two state ON/OFF discrete time and Markov fluid models are easily quantified when source with constant arrival using ON state probability $\mathrm{P_{ON}}=1$. On the other hand, when $\mathrm{P_{ON}}=1$ in ON/OFF MMPP sources reduces to pure Poisson arrival process.

Herein, we introduce Table 1, where we summarize the main equations that will be employed in the next chapters, namely Maximum average arrival rate and probability of ON state. Notice that the model are split in to discrete and continuous cases.

Table 1: Maximum average arrival rate and $\mathrm{P_{ON}}$ formula of Markovian source models

	Models	Maximum average arrival rate $\overline{\lambda}_{max}$	$\mathrm{P_{ON}}$
Discrete	DTMS	$\frac{\mathrm{P_{ON}}}{\theta}\log\left(\frac{e^{2\theta C_E(\theta)}-p_{11}e^{\theta C_E(\theta)}}{(1-p_{11}-p_{22})+p_{22}e^{\theta C_E(\theta)}}\right)$	$\frac{1-p_{11}}{2-p_{11}-p_{22}}$
	DTMMPP	$\frac{\mathrm{P_{ON}}}{(e^{\theta}-1)}\left[\log\left(\frac{e^{2\theta C_E(\theta)}-p_{11}e^{\theta C_E(\theta)}}{(1-p_{11}-p_{22})+p_{22}e^{\theta C_E(\theta)}}\right)\right]$	$\frac{1-p_{11}}{2-p_{11}-p_{22}}$
Continuous	FMS	$\mathrm{P_{ON}}\frac{\theta C_E(\theta)+\alpha+\beta}{\theta C_E(\theta)+\alpha}C_E(\theta)$	$\frac{\alpha}{\alpha+\beta}$
	CTMMPP	$\mathrm{P_{ON}}\frac{\theta[\theta C_E(\theta)+\alpha+\beta]}{(e^{\theta}-1)\theta C_E(\theta)+\alpha}C_E(\theta)$	$\frac{\alpha}{\alpha+\beta}$

3. EFFECTIVE CAPACITY IN FIXED RATE WIRELESS TRANSMISSION

The proposed work is based on the following system model: assume a single transmitter and receiver with point-to-point link. Moreover, it is assumed that the data generated by the source divided into frames. These frames are first stored in the buffer before transmitting it as illustrated in Figure 3.1. We consider Rayleigh block fading channel model. Hence, fading coefficient varies independently from one frame to another. The discrete time channel input output relation of the i^{th} symbol duration is assumed as

$$C[i] = \log_2(1 + \mathrm{SNR}z[i]), \tag{51}$$

$$y[i] = h[i]x[i] + n[i] \quad i = 1, 2, \ldots., \tag{52}$$

where x_i and y_i are the channel input and output respectively. $n[i]$ is the zero mean, circularly symmetric, complex Gaussian random noise which is assumed to form an independent and identical distributed sequence (i.i.d) and is normalized to the unity. Finally h_i is the channel fading coefficient and it is stationary and ergodic discrete

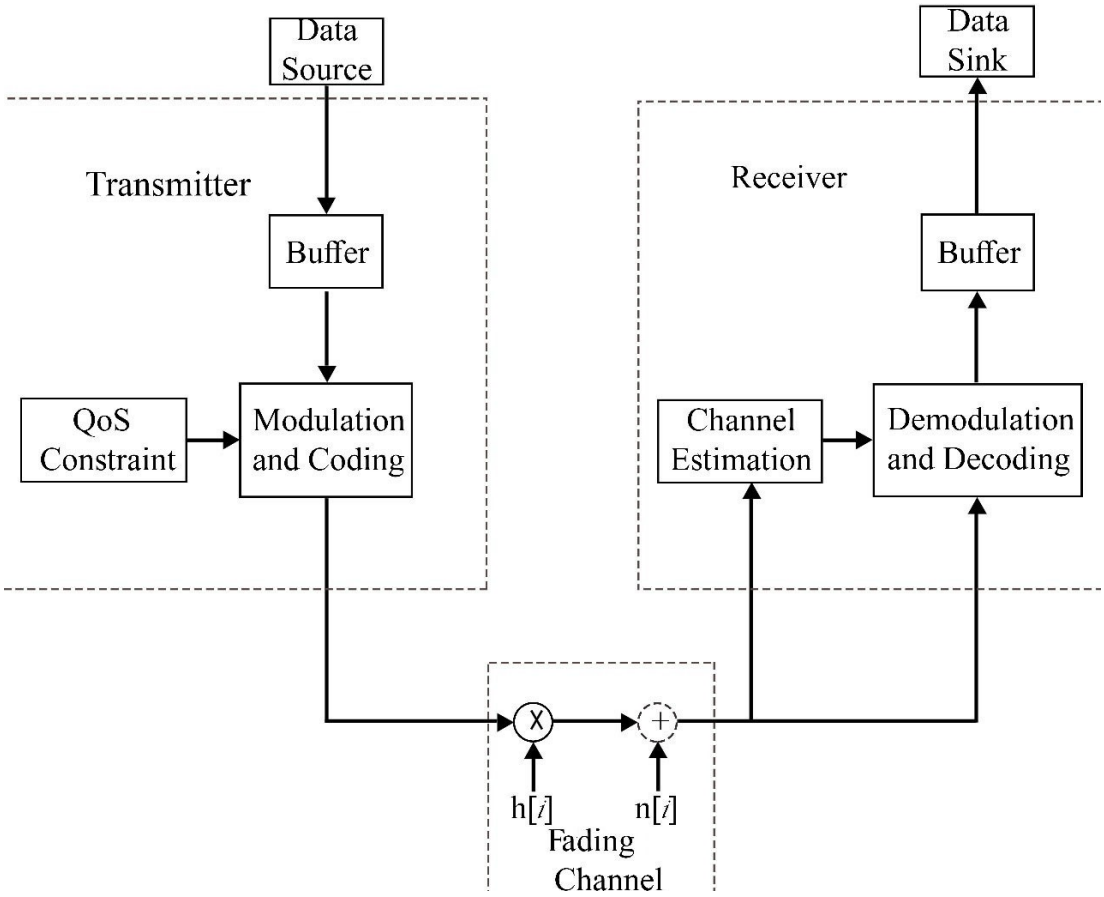

Figure 3.1: The general system model.

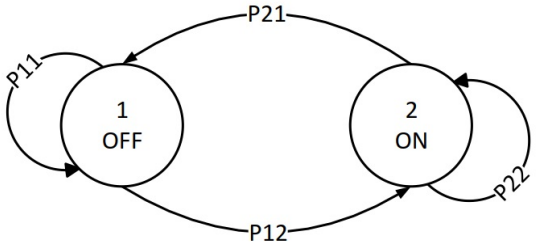

Figure 3.2: ON-OFF state transition model.

process, thus $z[i] = |h[i]|^2$ is square-envelop of Rayleigh fading block coefficients, which are exponentially distributed with mean 1.

In this work, we assume that the receiver is able to estimate its own channel h_i, whereas the transmitter does not know this information, therefore the transmitter would be transmitting information over the fixed rate r bps. Therefore, this behavior can be modeled as a two-state Markov chain such that when $r < C$ then the channel is considered in ON state and reliable communication is accomplished at the rate r. While $r \geq C$ then the channel is considered in OFF state so we can not achieve reliable communication at the rate of r. We assume unitary bandwidth. Figure 3.2 illustrates this process, and we note p_{ij}, where $i, j \in \{1, 2\}$ transition probabilities, which are a function of the SNR and target transmission rate r as we shall see next.

3.1. Formulation of Effective capacity in fixed rate wireless transmission

As discuss in previous Section 2.4, the effective capacity is for a given QoS exponent is obtained from

$$C_E(\theta) = -\lim_{t\to\infty} \frac{1}{\theta t} \log\{e^{-\theta S[t]}\} \stackrel{\text{def}}{=} -\frac{\Lambda(-\theta)}{\theta}, \tag{53}$$

where

$$s[t] \triangleq \sum_{k=1}^{t} R[k], \tag{54}$$

is the time accumulated service process and $\{R[k], k = 1, 2, \cdots\}$ shows the discrete time stationary and ergodic stochastic service process. Therefore for fixed rate transmission $R[k] = r$, hold true when the channel is considered in ON state and otherwise 0 in OFF state. $R[k] = r^*$, is true for channel is consider in ON state of optimum transmission rate. The formulation of C_E for two state model is by considering (9) and noting that $p_{11} + p_{22} = 1$. We formulate the effective capacity for a given statistical QoS constraint θ, as [21]

$$C_E(\mathrm{SNR}, \theta) = -\frac{\Lambda(-\theta)}{\theta} \tag{55}$$

$$= -\frac{1}{\theta} \log_e(p_{11} + p_{22} e^{-\theta r}) \tag{56}$$

$$= -\frac{1}{\theta} \log_e(1 - P\{z > \Psi\}(1 - e^{-\theta r})) \tag{57}$$

where $\Psi = \frac{2^r-1}{\mathrm{SNR}}$, which is distributed as the probability density function of z is equal to $p_z(z) = e^{-z}$ and $P\{z \geq \Psi\}$ cumulative distribution function which is equal to the $p_{11} = p_{21} = P\{z \leq \Psi\} = \int_0^{\Psi} p_z(z)dz \quad \text{and } p_{22} = p_{12} = P\{z > \Psi\} = \int_{\Psi}^{\infty} p_z(z)dz.$

Therefore

$$P\{z > \Psi\} = (1 - P\{z \leq \Psi\}) \tag{58}$$

$$= (1 - \int_0^{\Psi} p_z(z)\, dz) \equiv e^{-\Psi} \tag{59}$$

then plugging it into (36), we have the effective capacity for fixed rate transmission as follows

$$C_E(\mathrm{SNR}, \theta) = -\frac{1}{\theta} \log(1 - (1 - (1 - e^{-\Psi}))\,(1 - e^{-\theta r})) \tag{60}$$

$$= -\frac{1}{\theta} \log(1 - e^{-\Psi}\,(1 - e^{-\theta r})) \tag{61}$$

$$= -\frac{1}{\theta} \log(1 - e^{-\frac{2^r-1}{\mathrm{SNR}}}\,(1 - e^{-\theta r})) \tag{62}$$

3.2. Maximization of Effective Capacity

We are interested investigate which characteristics (SNR, QoS exponent and transmission rate) maximize the effective capacity. From Figure 3.3 maximum average arrival rate $\overline{\lambda}_{max}$ as a function of C_E for different source Markov models when $\theta = 1$, $\mathrm{P_{ON}} = 0.5$ and $r = 1$. Notice that throughout this thesis the SNR values are considered in linear scale, unless stated otherwise. We clearly observe from the figure that the maximum average arrival rate $\overline{\lambda}_{max}$ of Markovian source models are monotonically increasing function of C_E. Therefore if we maximize C_E which subsequently maximize throughput. Thus, we examine effective capacity to any characteristics at a time keeping all others parameters are kept constant. Figure 3.3 also compares maximum average arrival rate $\overline{\lambda}_{max}$ for different source arrivals assuming fixed values of source burstiness $\mathrm{P_{ON}}$, queueing constraint θ and transmission rate. It is observe from figure that as the value of C_E increases maximum average arrival rate of DTMMPP and CT-MMPP increases less sharply as compared to FMS, this is because of fixed value of transmission rate.

In a wireless communication system, it is generally presumed that by increasing SNR performance of the system is increase. However, Figure 3.4 shows the effective capacity, as in (62) as a function of SNR for fixed value of transmission rate r. We consider SNR values are in linear. It is clearly observed that the effective capacity is increasing logarithmically when SNR increases. However, effective capacity gradually slows its improvement in larger values of SNR. The increase in SNR produces drastic boost at certain point in effective capacity due to the fixed transmission rate which may lead to waste of scarce resources.

In Figure 3.5 we examine the impact of QoS constraints to the effective capacity for fixed value of SNR and transmission rate. It is observed that when there is no queueing constraint imposed i.e. $\theta = 0$ the effective capacity converges to the ergodic capacity

of the fading channel, thus maximum effective capacity can be sustained at this point. Further, notice that as QoS exponent increases effective capacity gradually decreases. The figure also confirms the fact that C_E is monotonically decreasing in θ where the C_E degrades with the increase of delay constraint θ and this effect appears to be more severe for tight delay constraints.

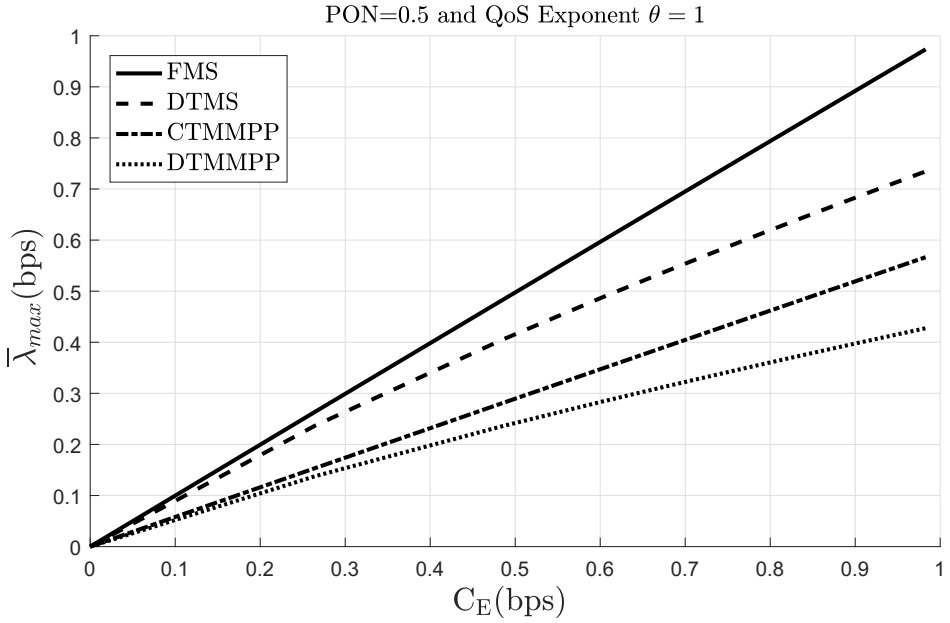

Figure 3.3: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of effective capacity of different Markovian source models.

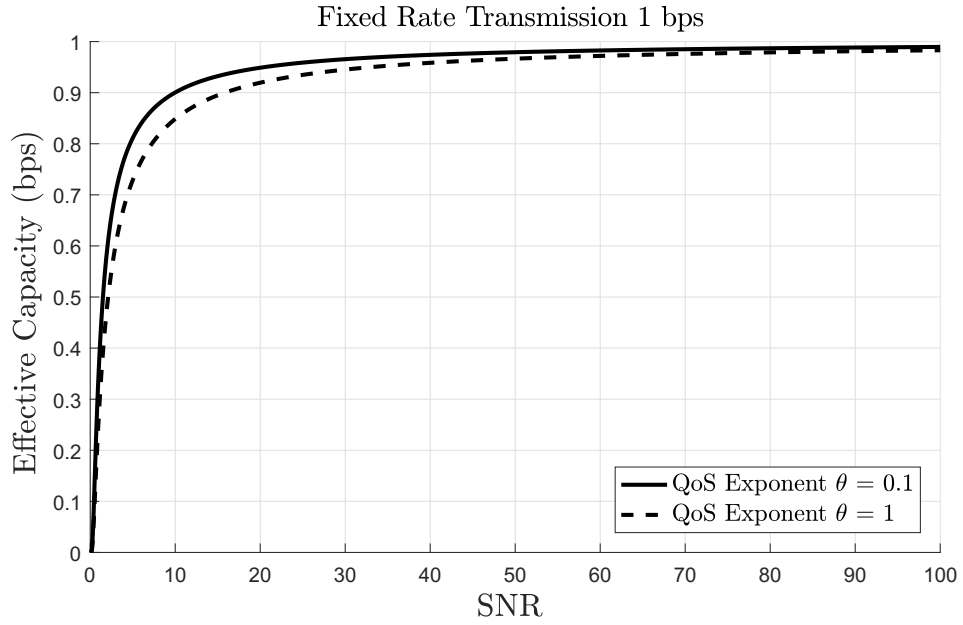

Figure 3.4: Effective capacity as a function of SNR with fixed transmission rate $r = 1$ bps and $\theta \in \{0, 0.1\}$.

Figure 3.6 shows the effective capacity as a function of transmission rate r for different value of SNR with fixed QoS constraint θ. It is clearly visible that the effective capacity increases as transmission rate increases, but after certain limit the effective capacity gradually start decreasing due to fixed value of SNR. If one increase the transmission rate it will degrade the performance of effective capacity. For example, when $\mathrm{SNR} = 10$ and $\theta = 1$ then the maximum effective capacity can be support by optimum transmission rate is 1.7 bps, if we further increase the transmission rate then the performance of effective capacity is degraded, whereas 3 bps is maximum transmission rate required achieving optimum effective capacity when $\mathrm{SNR} = 100$ and $\theta = 1$.

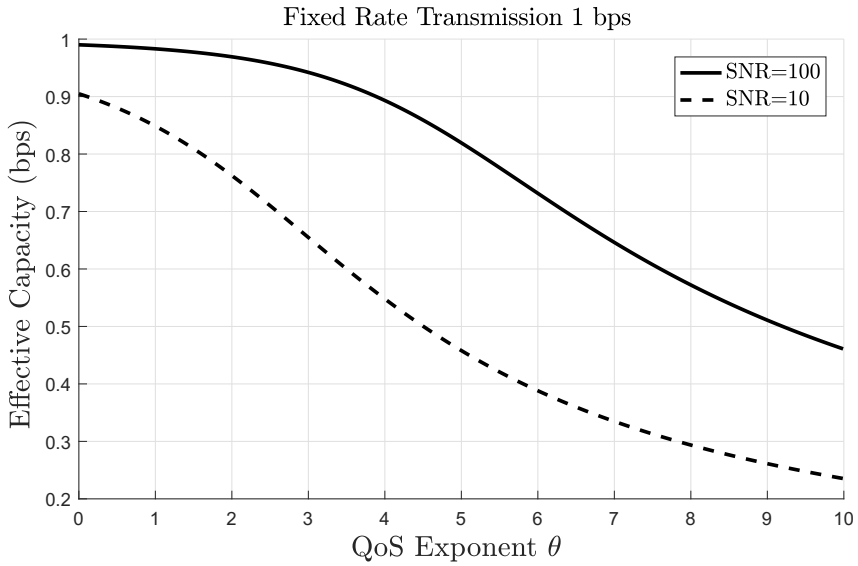

Figure 3.5: Effective capacity as a function of QoS exponent with fixed rate transmission $r = 1$ bps.

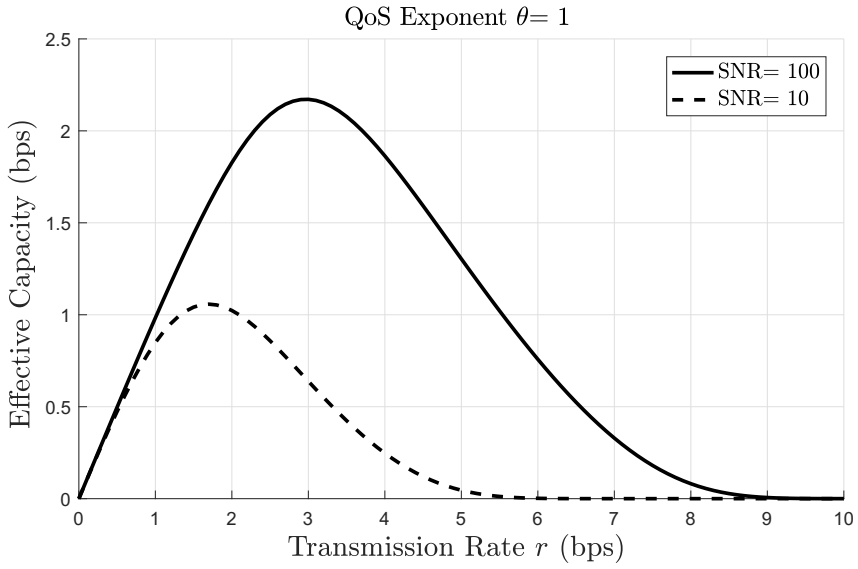

Figure 3.6: Effective capacity as a function of fixed rate transmission for different values of SNR.

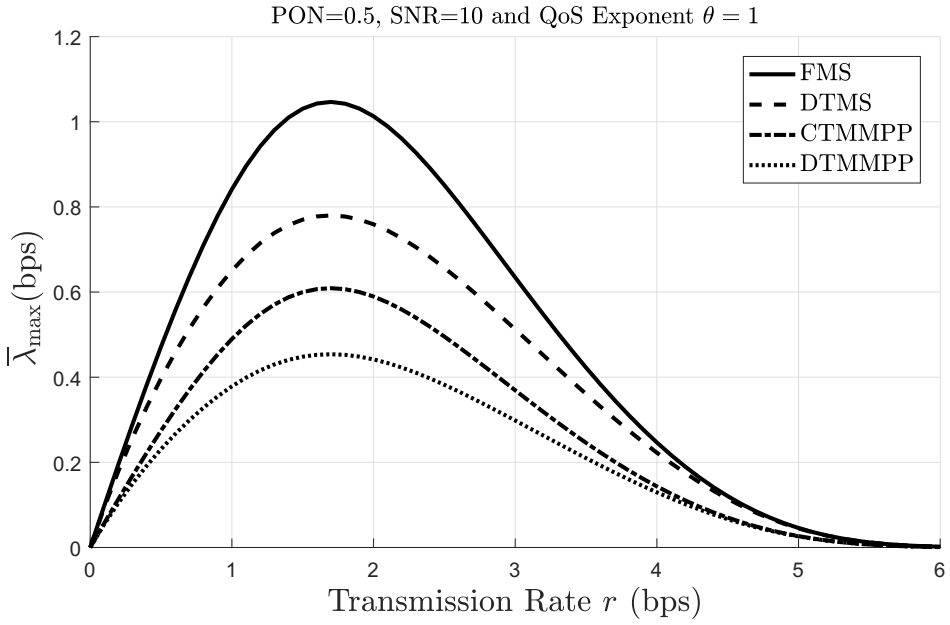

Figure 3.7: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of transmission rate r with different Markov source Models.

Therefore, efficient use of transmission rate boosts the performance of communication system. We conclude optimizing the effective capacity with respect to the transmission rate, allows for high link throughput while allowing larger arrival rates as we shall see next section.

Now we formulate the optimization problem to maximize C_E subject to constraint on transmission rate. The C_E maximization problem of a system with delay outage probability requirement falls under the constraint of r greater than zero. Hence C_E constraint can be mathematically expressed as

$$C_E^{opt}(\mathrm{SNR},\theta) = \max_{r\geq 0}\left\{-\frac{1}{\theta}\log_e(1-e^{-\frac{2^r-1}{\mathrm{SNR}}}\,(1-e^{-\theta r}))\right\}. \tag{63}$$

Figure 3.7 shows maximum average arrival rate $\overline{\lambda}_{max}$ as a function of transmission rate with $\mathrm{SNR} = 10$ for different Markovian source models. It is clearly seen that $\overline{\lambda}_{max}$ is a quasi concave because its upper contour set is convex and since (66) is second derivative of (62) which gives $\frac{\partial^2 EC}{\partial r^2}|_{r=r^*} \leq 0$ negative value of C_E for all practical SNR, it means that C_E is quasi concave in r, as follows.
Therefore taking first derivative of (62) and equating it to zero yields the optimum value of r.

$$\frac{\partial C_E(\mathrm{SNR},\theta)}{\partial r} = \frac{\theta\,\mathrm{e}^{-\Psi}\,\mathrm{e}^{-r\,\theta} + \frac{\log(2)\,2^r\,\mathrm{e}^{-\Psi}\left(\mathrm{e}^{-r\,\theta}-1\right)}{\mathrm{SNR}}}{\theta\,\left(\mathrm{e}^{-\Psi}\,\left(\mathrm{e}^{-r\,\theta}-1\right)+1\right)} = 0, \tag{64}$$

$$\frac{\mathrm{SNR}\,\theta \mathrm{e}^{-\Psi}\mathrm{e}^{-r\theta} + 2^{r}\mathrm{e}^{-\Psi}\log(2)\left(\mathrm{e}^{-r\,\theta}-1\right)}{\mathrm{SNR}\,\theta\left(\mathrm{e}^{-\Psi}\,\left(\mathrm{e}^{-r\,\theta}-1\right)+1\right)} = 0, \tag{65}$$

$$\frac{\partial^2 C_E(\mathrm{SNR},\theta)}{\partial r^2} = \frac{\left(\theta\,\mathrm{e}^{-2^{r}-1\,\mathrm{SNR}}\,\mathrm{e}^{-\,r\,\theta} + \frac{2^{r}\,\mathrm{e}^{-\Psi}\log(2)\left(\mathrm{e}^{-r\,\theta}-1\right)}{\mathrm{SNR}}\right)^{2}}{\theta\left(\mathrm{e}^{-\Psi}\,\left(\mathrm{e}^{r\,\theta}-1\right)+1\right)^{2}} - \frac{\theta^{2}\,\mathrm{e}^{-\Psi}\,\mathrm{e}^{r\,\theta}}{\theta\,\left(\mathrm{e}^{-\Psi}\,\left(\mathrm{e}^{r\,\theta}-1\right)+1\right)}$$

$$+ \frac{\frac{2^{2^{r}}\,\mathrm{e}^{-\Psi}\log(2)^{2}\left(\mathrm{e}^{r\,\theta}-1\right)}{\mathrm{SNR}^{2}} - \frac{2^{r}\,\mathrm{e}^{-\Psi}\log(2)^{2}\left(\mathrm{e}^{r\,\theta}-1\right)}{\mathrm{SNR}} + \frac{2.\,2^{r}\;\theta\,\mathrm{e}^{-\Psi}\,\mathrm{e}^{r\,\theta}\log(2)}{\mathrm{SNR}}}{\theta\,\left(\mathrm{e}^{-\Psi}\,\left(\mathrm{e}^{r\,\theta}-1\right)+1\right)}. \tag{66}$$

After some algebraic manipulation, we have

$$\mathrm{e}^{-\Psi}\left[\mathrm{SNR}\,\theta\mathrm{e}^{-r\theta} + 2^{r}\log(2)\left(\mathrm{e}^{-r\,\theta}-1\right)\right] = 0. \tag{67}$$

Since $\mathrm{e}^{-\Psi}$ is not zero, then we obtain

$$\mathrm{SNR}\,\theta\mathrm{e}^{-r\theta} + 2^{r}\log(2)\left(\mathrm{e}^{-r\,\theta}-1\right) = 0. \tag{68}$$

Further simplifying (68), reduces to

$$(\mathrm{SNR}\,\theta + 2^{r}\log(2)) = \frac{2^{r}\log(2)}{\mathrm{e}^{-r\theta}}. \tag{69}$$

Then, by taking logarithm in the both sides, we have

$$\log\left(\mathrm{SNR}\,\theta + 2^{r}\log(2)\right) = \log(2^{r}\log(2)) + r\theta. \tag{70}$$

After further simplifying (70), we get

$$\frac{1}{\theta}\log\left(1 + \frac{\mathrm{SNR}\,\theta}{2^{r}\log(2)}\right) = r. \tag{71}$$

3.3. Impact of Optimum Transmission Rate to the Effective Capacity

In this section, we investigate the behaviour transmission rate to the effective capacity. In earlier section we use we examine the impact of fixed transmission rate to the effective capacity. Next we consider fixed and optimum transmission rate for comparing effective capacity as a function of SNR, secondly, effective capacity as a function of QoS exponent θ.

Figure 3.8, we examine the impact of optimum transmission rate to the effective capacity in high SNR regime where we plot effective capacity as a function of SNR for fixed and optimum transmission rate with different value of θ. The optimum transmission rate is the rate at which maximum effective capacity is achieved with using

efficient use of resources. The red line shows effective capacity with optimum transmission rate and black lines show the effective capacity with fixed rate transmission. It is clearly observed that higher effective capacity can be achieved when using optimum transmission rate as compared to fixed transmission rate. Moreover, notice that effective capacity of fixed transmission rate does not increase at high SNR regime, whereas optimum transmission rate maximize effective capacity and consequently. One can generalize that is the effective capacity is equal to average arrival rate the minimal latency can be achieved using optimal resources.

Figure 3.9, where we examine the impact of QoS constraint to the effective capacity for fixed and optimum transmission rate when $\mathrm{SNR} \in \{10, 100\}$. It is clearly observed that optimum transmission rate provides higher C_E as compared to the fixed transmission rate. We further observed that when there is no queueing constraint imposed i.e. $\theta = 0$, the effective capacity is equal to the ergodic capacity of the block fading channel. It is further noticing that as QoS exponent increase effective capacity gradually decrease. The figure also confirms the fact that C_E is monotonically decreasing in θ, where the C_E degrades with the increase of delay constraint θ and this effect appears to be more severe for tight delay constraints (larger value of θ). For instance, as we can see figure 3.8 that optimum transmission rate of $\mathrm{SNR} = 100$ is 3 bps, which gives us maximum 2.1 bps value of C_E. It is easy verify from the figure that optimum transmission rate of $\mathrm{SNR} = 100$ gives us same maximum C_E i.e. 2.1 bps when $\theta = 1$.

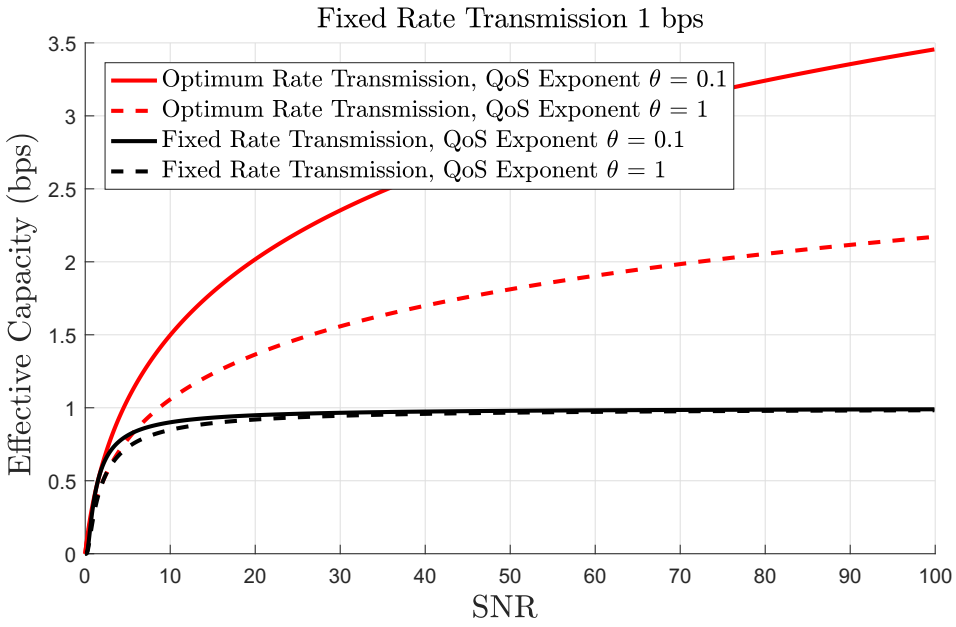

Figure 3.8: Effective capacity as a function of SNR for fixed and optimum transmission rate when queueing constraint $\theta \in \{0.1, 1\}$.

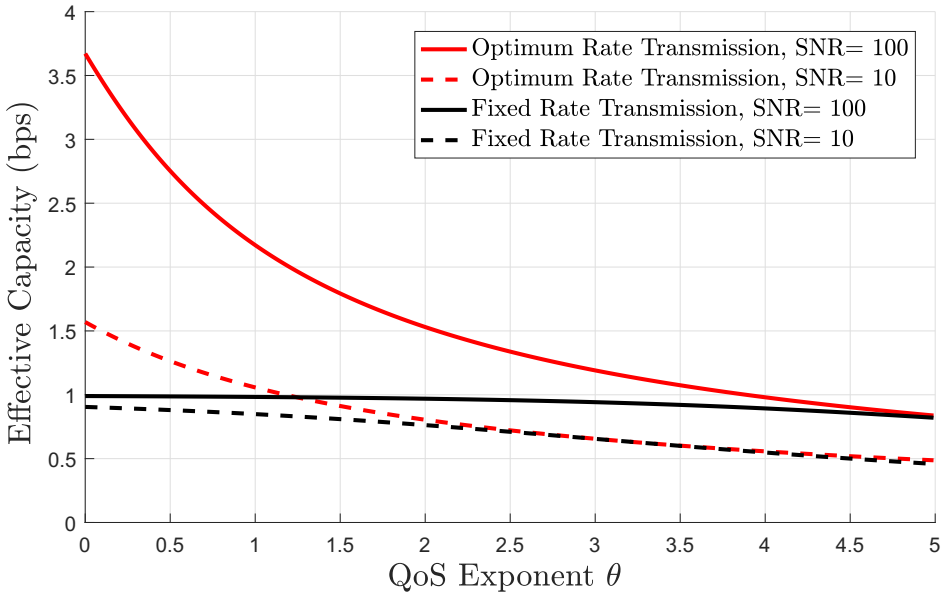

Figure 3.9: Effective capacity as a function of queueing constraint θ for fixed and optimum transmission rate when $\mathrm{SNR} \in \{10, 100\}$.

4. ANALYSIS OF MARKOVIAN ARRIVAL SOURCE MODELS

4.1. Analysis under Throughput

As discussed in the previous chapter, the rise of transmission rate boosts the effective capacity but now we are interested to investigate the impact of burstiness and randomness to the optimum effective capacity. Therefore, we consider ON-OFF Markovian arrival source models described in Section 2.3 and how they are affect by optimal effective capacity. Maximum average arrival rate is a throughput metric that can captures the effects of randomness and burstiness of Markovian sources when channel responses is timing varying in the presence of QoS constraints. The QoS constraints are satisfied when the effective bandwidth of arrival process is equal to the effective capacity of the service process service process as discuss in section .

The smaller value of $\mathrm{P_{ON}}$ identified that when data arrive less frequently, which increase arrival rate λ with certain rate to cope with non-decreasing throughput with the same departure rate in ON state and due to this arrival rate, source becomes bursty which reduced maximum average arrival rate of the system. On the other hand, we achieve maximum maximum average arrival rate when $\mathrm{P_{ON}} = 1$ thus constant arrival rate. As θ increase QoS requirement becomes too stringent, then the maximum average arrival rate of the system tends to decrease.

4.1.1. Discrete Markov Sources

In this section, we consider maximum average arrival rate $\overline{\lambda}_{max}$ of two-state (ON/OFF) discrete Markov sources which is described in Section 2.5.1, that can be supported by the fading channel while satisfying the statistical QoS limitations given in the form in (2).

Figure 4.1 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as the function of effective capacity for different value of $\mathrm{P_{ON}}$ when QoS exponent $\theta = 1$. It is clearly seen that $\overline{\lambda}_{max}$ is monotonically increasing function of C_E, the maximizing C_E will maximize $\overline{\lambda}_{max}$ consequently. We further easy verify that when $\mathrm{P_{ON}} = \frac{1-p_{11}}{2-p_{11}-p_{22}} = 1$ or equivalently $p_{22} = 1$, (42), simplifies to $\overline{\lambda}_{max}\ (\mathrm{SNR}, \theta)$ = $C_E(\mathrm{SNR}, \theta)$. Hence source is always in ON state, then the constant arrival of data in buffer. Therefore maximum average arrival rate is equal to effective capacity. As $\mathrm{P_{ON}}$ decreases, the arrival rate in ON state λ needs to increase with certain rate to cope with non-decreasing throughput with the same departure rate. Hence, the throughput degraded as the value of $\mathrm{P_{ON}}$ decrease and smaller average arrival rates are supported for given effective capacity. It is concluded that discrete time Markov bursty sources rigorously decrease throughput at high SNR regime. For example, when source has maximum average arrival rate is 0.5 bps and data arrives frequently i.e. $\mathrm{P_{ON}} = 1$ with queueing constraints is $\theta = 1$, then system need 0.5 bps effective capacity to satisfied the required QoS constraints. On the other hand, when data arrives less frequently means $\mathrm{P_{ON}} = 0.4$ with source has maximum average arrival rate is 0.5 bps then system needs 0.7 effective capacity to fulfil the same QoS requirements.

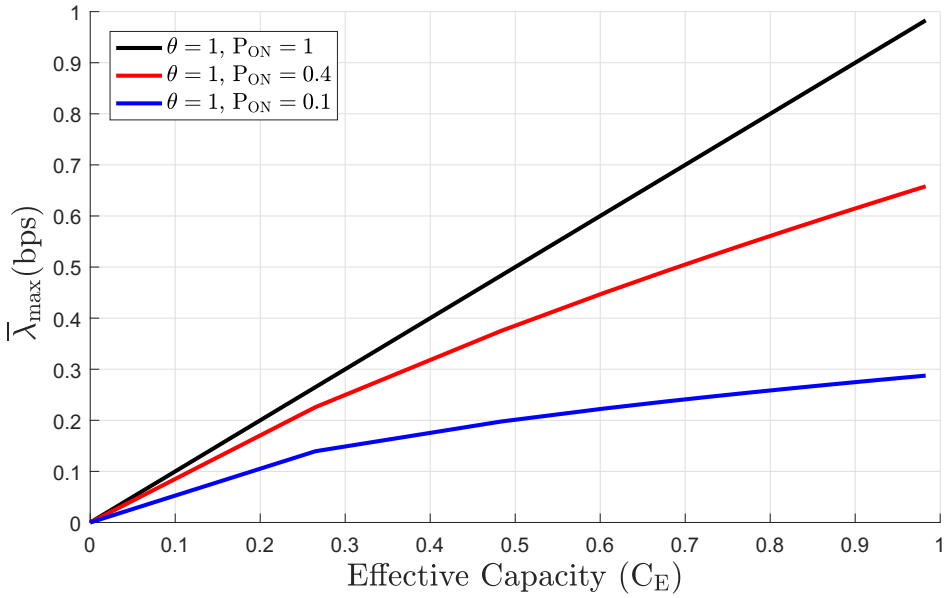

Figure 4.1: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of effective capacity for different $\mathrm{P_{ON}}$ and QoS exponent $\theta = 1$.

Figure 4.2 shows maximum average arrival rate $\overline{\lambda}_{max}$ as a function of SNR for different value of $\mathrm{P_{ON}}$ and θ. It is also interesting to notice from (42) that the arrival rate in the ON state, which is given by $\overline{\lambda} = \frac{\overline{\lambda}_{max}}{\mathrm{P_{ON}}}$, the smaller value of $\mathrm{P_{ON}}$ identified that when data arrive less frequently, which increase arrival rate in ON state and due to this arrival rate source becomes bursty. It is also observed that lower value of $\mathrm{P_{ON}}$ give us reduced throughput caused by burstiness of source. On the other hand, we achieve maximum throughput when $\mathrm{P_{ON}} = 1$ thus constant arrival rate. For instance with $\theta = 0.1$, $\mathrm{P_{ON}} = 0.4$ and maximum average arrival data rate is $\overline{\lambda}_{max} = 2.5$ bps then system needs high SNR for transmitting data in Rayleigh fading channel to satisfy the QoS constraint. Overall the throughput diminishes with increasing of θ and with decreasing of $\mathrm{P_{ON}}$.

Figure 4.3 depicts the maximum average arrival rate $\overline{\lambda}_{max}$ as a function of QoS exponent θ for discrete Markov source with different $\mathrm{P_{ON}}$ values when $\mathrm{SNR} = 10$. We notice that higher maximum average arrival rate can be sustained for low value of QoS exponent, hence different value of $\mathrm{P_{ON}}$ do not affect the $\overline{\lambda}_{max}$ when QoS exponent $\theta = 0$. As θ increase QoS requirement becomes too stringent, then the maximum average arrival rate of the system tends to decrease, because of fixed transmission rate. This reduction in $\overline{\lambda}_{max}$ is more severe for more bursty sources (e.g when $\mathrm{P_{ON}} = 0.1$). Hence more bursty sources support larger arrival rates which requires careful design in order to avoid buffer overflows.

Next, we investigate the impact of source burstiness on the maximum average arrival rate $\overline{\lambda}_{max}$, when QoS exponent $\theta = 1$ and $\mathrm{SNR} = 10$, as shown in Figure 4.4. We observe that the $\overline{\lambda}_{max}$ is the increasing function of $\mathrm{P_{ON}}$. Notice that as $\mathrm{P_{ON}}$ approach to 1, then source achives constant arrival rate which gives maximum throughput. At this point the maximum average arrival rate is equal to the effectice capacity. Lower values

of $\mathrm{P_{ON}}$ gives us larger average arrival rate which increases the burstiness of the source, therefore throughput is degraded. Figure 4.4 also confirm the result from Figure 4.3 that when source has constant arrival of data i.e. $\mathrm{P_{ON}} = 1$ with QoS constraint $\theta = 1$ and arrival rate 1.1 bps then system needs $\mathrm{SNR} = 10$ to to fulfil the QoS requirements defined in value of θ.

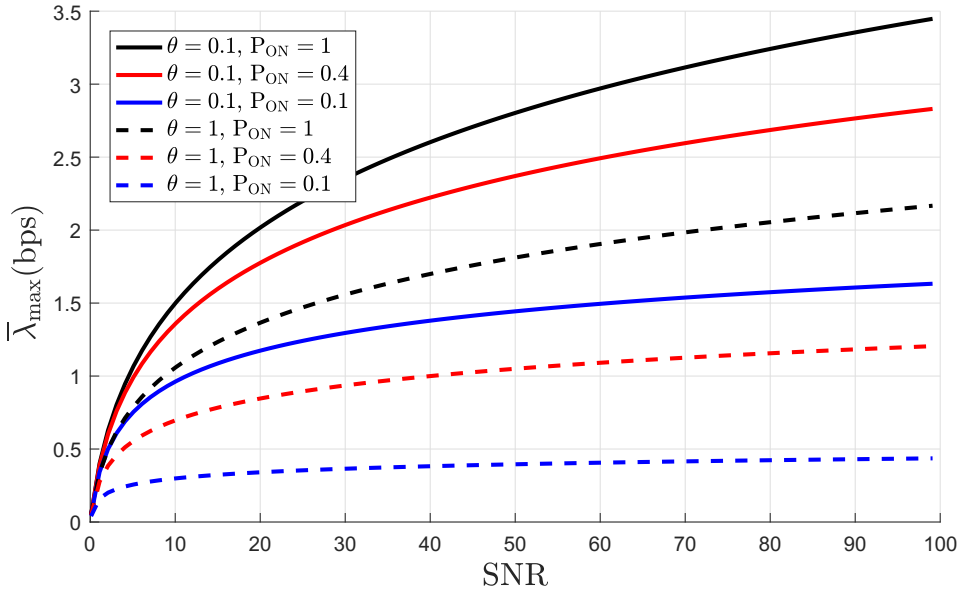

Figure 4.2: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of SNR (linear-scale) for different value of QoS exponent θ and $\mathrm{P_{ON}}$.

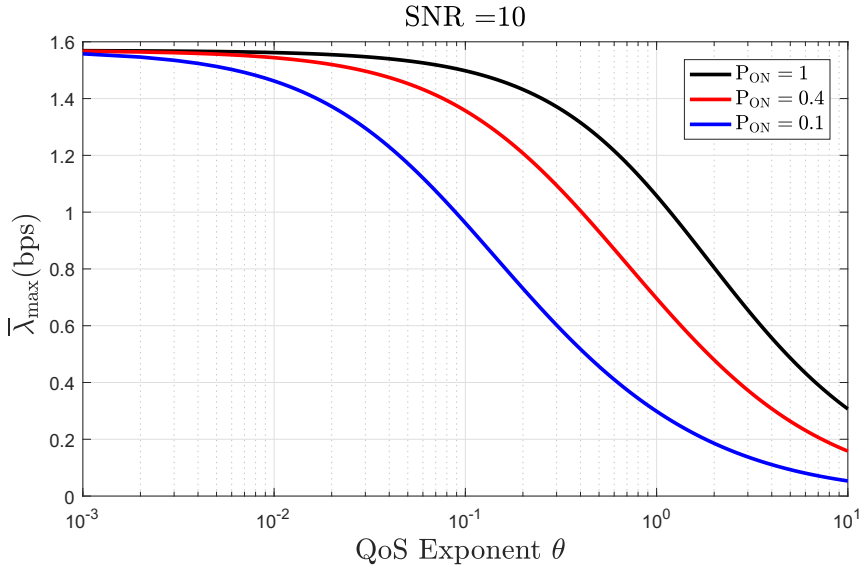

Figure 4.3: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of QoS exponent θ for different $\mathrm{P_{ON}}$ values when $\mathrm{SNR} = 10$.

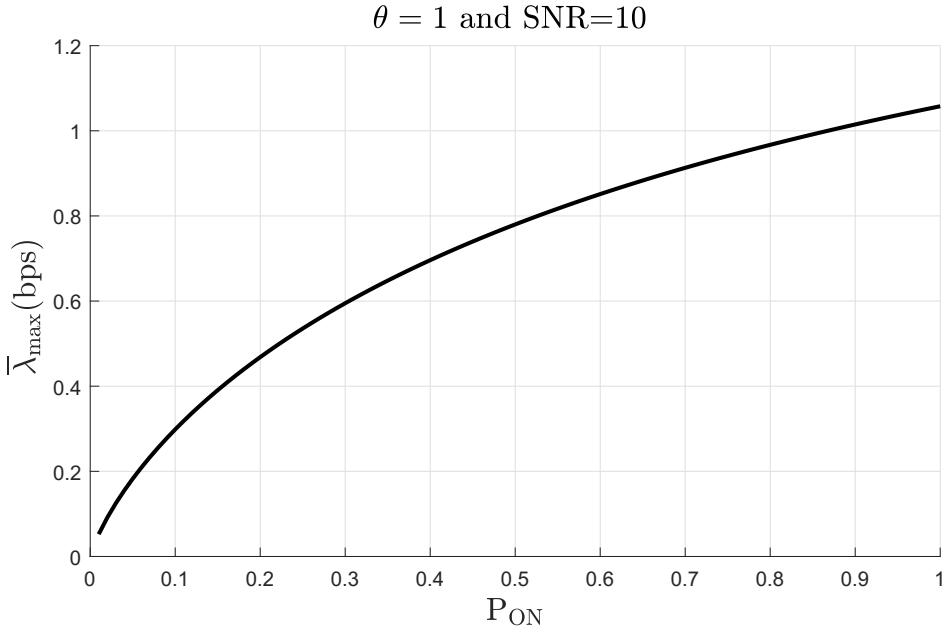

Figure 4.4: Maximum average arrival rate as a function of $\mathrm{P_{ON}}$ of discrete Markov source with fixed value of SNR and θ.

4.1.2. Markov Fluid Sources

In the following, we go through similar steps as in the previous subsection and maximum average arrival rates $\overline{\lambda}_{max}$ of Markov fluid sources which is described in Section 2.5.2 in (46).

Figure 4.5 illustrates the maximum average arrival rate $\overline{\lambda}_{max}$ as the function of effective capacity for different value of $\mathrm{P_{ON}}$ when QoS exponent $\theta = 10$. It is again seen that $\overline{\lambda}_{max}$ is monotonically increasing function of C_E, so if we maximize C_E will maximize $\overline{\lambda}_{max}$ consequently. As previously discussed when the source is always in ON state $\mathrm{P_{ON}} = \frac{\alpha}{\alpha+\beta} = 1$, where α shows the transition rate from OFF to ON state and β is the transition rate from ON to OFF state. Thus (46) simplifies $\overline{\lambda}_{max}\ (\mathrm{SNR}, \theta)$ = $C_E(\mathrm{SNR}, \theta)$ where the maximum average arrival rate is equal to effective capacity. As $\mathrm{P_{ON}}$ decreases, the arrival rate in ON state λ needs to increase with certain rate to cope with non-decreasing throughput with the same departure rate. Hence, the throughput diminishes as $\mathrm{P_{ON}}$ decrease and smaller average arrival rates are supported for given effective capacity. For Markov fluid sources, ON state probability is not the sole indicator of burstiness. Having low α and β values also indicates that source is more bursty as the transition between ON and OFF states becomes less frequent. Hence, OFF state can be more persistent. When α and β are large, state transitions occur more rapidly, leading to lower required SNR levels. Therefore, Markov fluid source throughput is less afftected by burstiness when compared to discrete time Markov source at high SNR regime.

Figure 4.6 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as a function of SNR when $\theta \in \{1, 10\}$. We further notice that maximum average arrival rate diminishes with increasing θ and decreasing $\mathrm{P_{ON}}$. It is also observed from (46) that at high SNR bursti-

ness affects less the throughput of devices with mild QoS constraint as compared to stringent QoS constraint. It is also observed that as $\mathrm{P_{ON}}$ diminishes, throughput degrades faster in high SNR regime with increasing value of θ. For example, with $\theta = 1$, $\mathrm{P_{ON}} = 0.4$ and maximum average arrival data rate is 2 bits/s then system needs high SNR for transmitting data in Rayleigh fading channel to satisfy the QoS constraint.

Figure 4.7 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as a function of QoS exponent θ for Markov fluid source. We set $\mathrm{SNR} = 10$ and keep α = β, the ON probability $\mathrm{P_{ON}}$ is fixed at 0.5, while average duration of ON and OFF states changes as the value of α and β changes. We notice that higher value of maximum average arrival rate can be sustained for low value of QoS exponent θ, hence at this point different values of α and β do not affect the $\overline{\lambda}_{max}$, when QoS exponent $\theta = 0$. As θ increase QoS requirement becomes too stringent, the throughput of the system tends to zero, this reduction in $\overline{\lambda}_{max}$ is more severe for more lower values of α = β because higher α , β values leads to shorter the transition rate from ON to OFF. Therefore ON state probability is not the sole indicator of burstiness, lower α and β values also indicates that source is bursty. As an outcome of this fact gives us higher throughput, consequently leading to higher arrival rates, which may lead to buffer overflow, thus careful design is needed.

In order to investigate the impact of source burstiness on the maximum average arrival rate, we plot the maximum average arrival rate $\overline{\lambda}_{max}$ as a function of $\mathrm{P_{ON}}$ for fluid Markov source with QoS exponent $\theta = 1$ and $\mathrm{SNR} = 10$, as depicted in figure 4.8. We observe that the $\overline{\lambda}_{max}$ is the increasing function of $\mathrm{P_{ON}}$. It is also notice that as $\mathrm{P_{ON}}$ approach to 1, the source becomes constant arrival rate which gives maximum efficiency. Thus lower values of $\mathrm{P_{ON}}$ render us larger average arrival rate which increases the burstiness in the sources, therefore throughput is degraded.

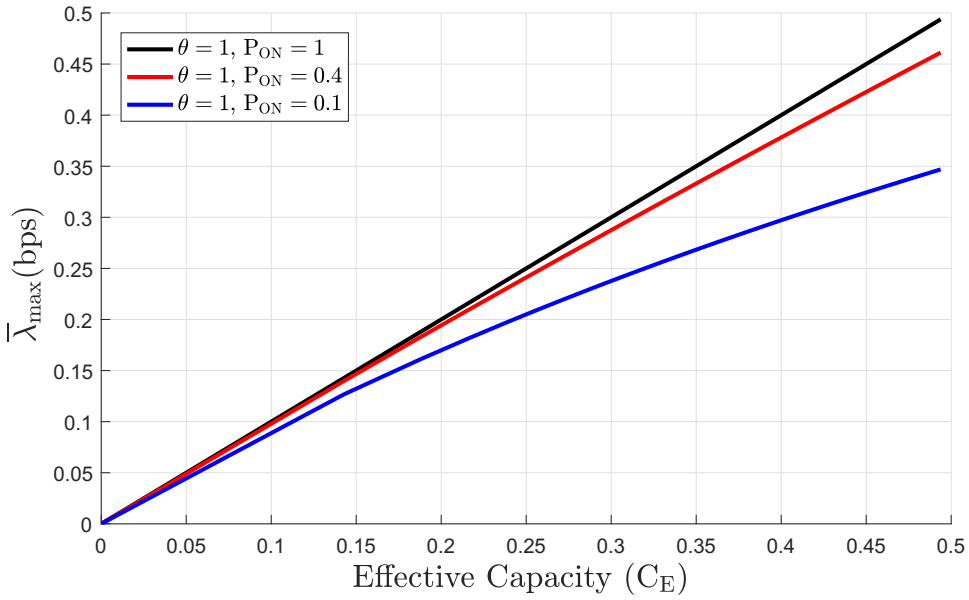

Figure 4.5: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of effective capacity for different value of $\mathrm{P_{ON}}$ and QoS exponent $\theta = 1$.

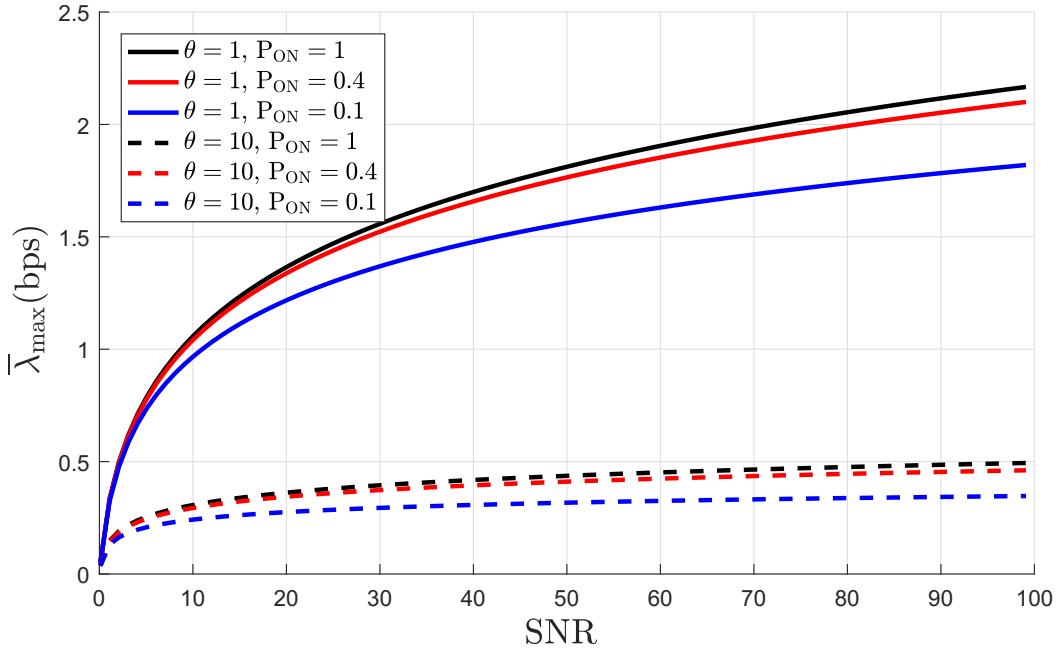

Figure 4.6: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of SNR for different value of QoS exponent and $\mathrm{P_{ON}}$.

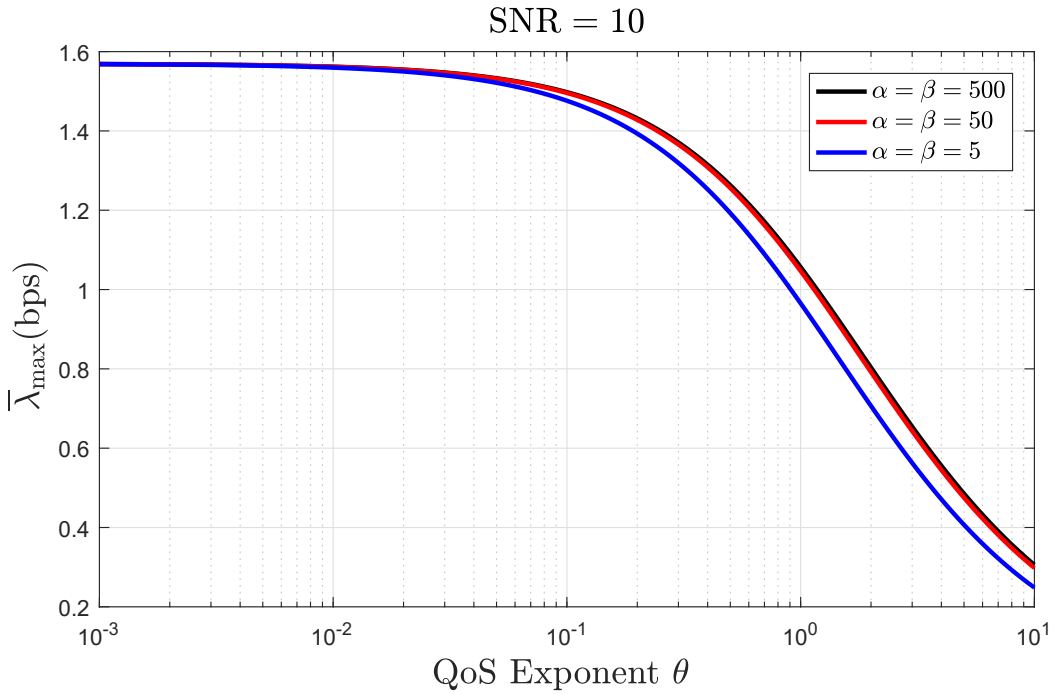

Figure 4.7: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of QoS exponent θ for different values of α and β when $\mathrm{SNR} = 10$.

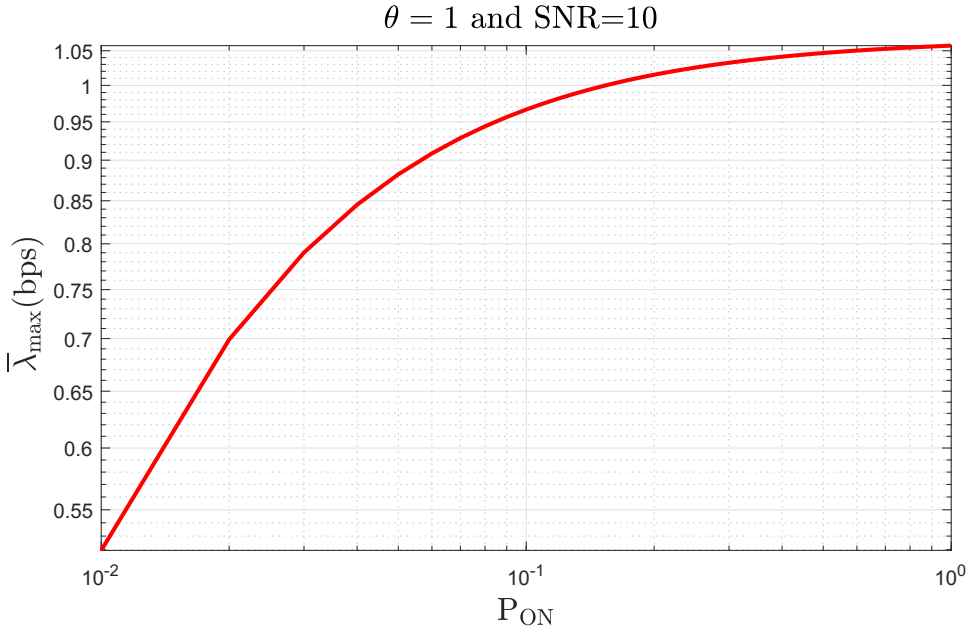

Figure 4.8: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of $\mathrm{P_{ON}}$ of Fluid Markov source with $\mathrm{SNR} = 10$ and $\theta = 1$.

4.1.3. Discrete Time Markov Modulated Poisson Sources

The Maximum average arrival rates $\overline{\lambda}_{max}$ of ON/OFF Discrete Time Markov Modulated Poisson Sources which is described in Section 2.5.3 in (48). It is interesting to observe that the throughput with the DTMMPP source is almost identical to that with the Markov Discrete Time Markov source model, save only for the multiplicative factor $\frac{\theta}{e^{\theta}-1}$ in (42). Hence, the throughput is generally smaller with DTMMPP sources and decreases fast with θ. This can be attributed to the much more randomness we experience with an DTMMPP source.

As in the previous cases, Figure 4.9 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as the function of effective capacity for different value of $\mathrm{P_{ON}}$ when QoS exponent $\theta = 1$. It is clearly seen that $\overline{\lambda}_{max}$ is monotonically increasing function of C_E, so if we maximize C_E will maximize $\overline{\lambda}_{max}$ consequently. As $\mathrm{P_{ON}}$ decreases, the arrival rate λ needs to increase with certain rate to cope with non-decreasing throughput with the same departure rate. Hence, the throughput is less affected by variations $\mathrm{P_{ON}}$ and also smaller average arrival rates are supported for given effective capacity.

Figure 4.10 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as a function of SNR for different value of QoS exponent and $\mathrm{P_{ON}}$. It is verifies from results that source becomes bursty in smaller value of $\mathrm{P_{ON}}$, implies that data arrive less frequently therefore source burstiness increase, which reduced the throughput of the system. It is also illustrated in Figure 4.9 that the throughput of DTMMPP is more effect by burstiness at high SNR regime as compared to others Markov sources. It is conclude that throughput diminishes with increasing θ and decreasing $\mathrm{P_{ON}}$. For instance with $\theta = 1$, $\mathrm{P_{ON}} = 1$ and average arrival rate is 1.2 bps for $\mathrm{SNR} = 80$, then required QoS constraints is satisfied.

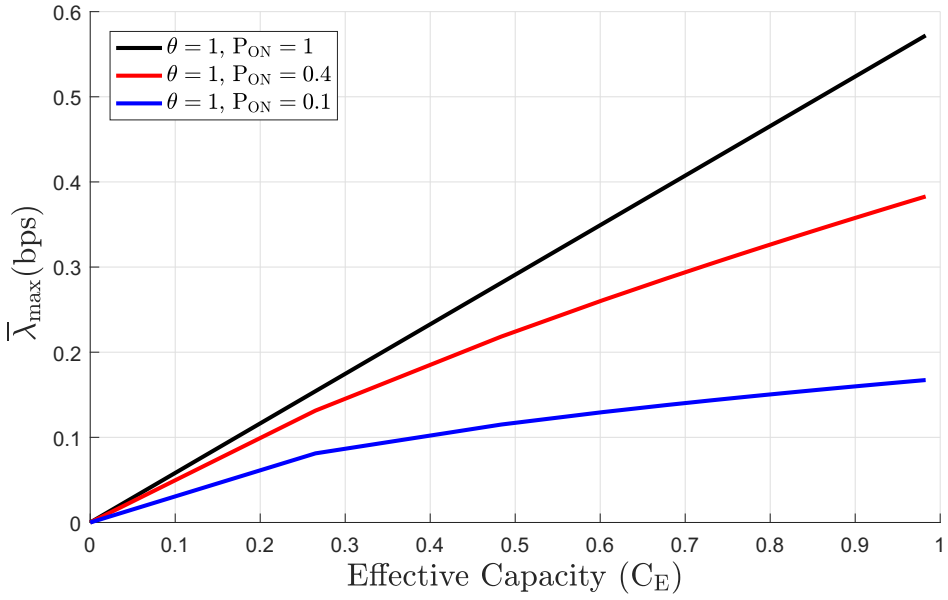

Figure 4.9: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of effective capacity for different value of $\mathrm{P_{ON}}$ and QoS exponent $\theta = 1$.

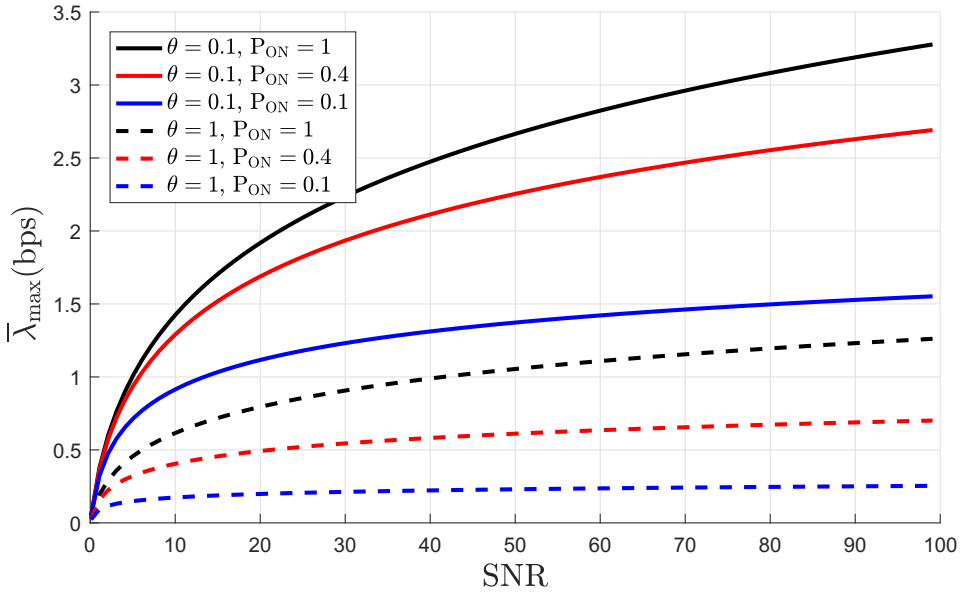

Figure 4.10: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of SNR for different value of QoS exponent θ and $\mathrm{P_{ON}}$.

Next, Figure 4.11 shows the maximum average arrival rate as a function of QoS exponent θ for Markov modulated Poisson source with different $\mathrm{P_{ON}}$ values. We set $\mathrm{SNR} = 10$, we notice that higher value of maximum average arrival rate can be sustained for low value of QoS exponent θ, hence at this point different values of $\mathrm{P_{ON}}$

do not affect the $\overline{\lambda}_{max}$, when QoS exponent $\theta = 0$. As θ increase QoS requirement becomes too stringent, the throughput of the system tends to zero, this reduction in $\overline{\lambda}_{max}$ is more severe for more bursty sources (e.g when $\mathrm{P_{ON}} = 0.1$). Hence more bursty sources support larger arrival rates which requires careful design in order to avoid buffer overflows.

In order to investigate the impact of source burstiness on the maximum average arrival rate, we plot the maximum average arrival rate as a function of $\mathrm{P_{ON}}$ for DTMMPP source for optimum transmission rate with QoS exponent and $\mathrm{SNR} = 10$, as depicted in Figure 4.12. We observe that the $\overline{\lambda}_{max}$ is the increasing function of $\mathrm{P_{ON}}$.

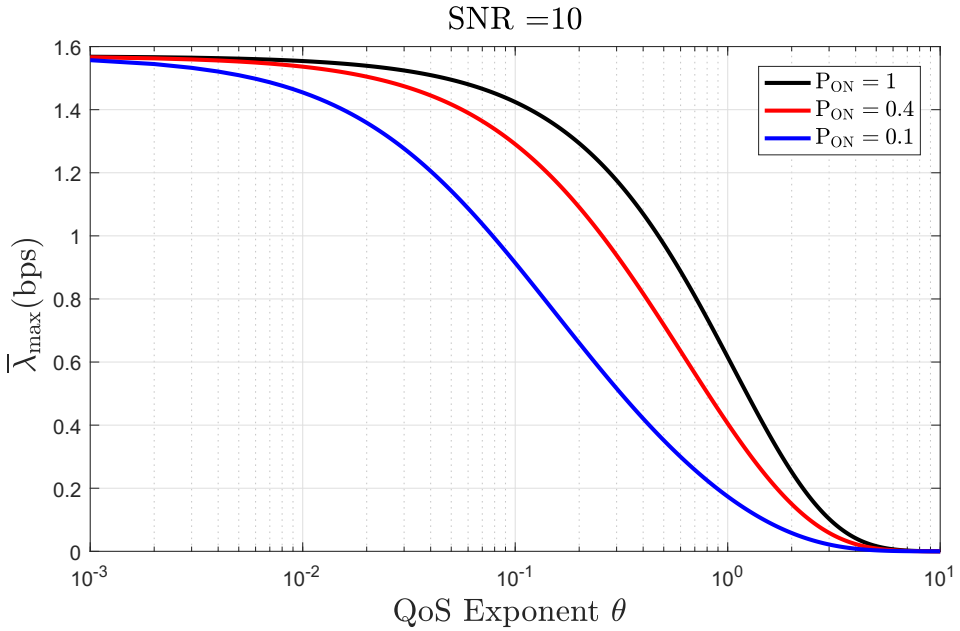

Figure 4.11: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of QoS exponent θ for different values of $\mathrm{P_{ON}}$ when $\mathrm{SNR} = 10$.

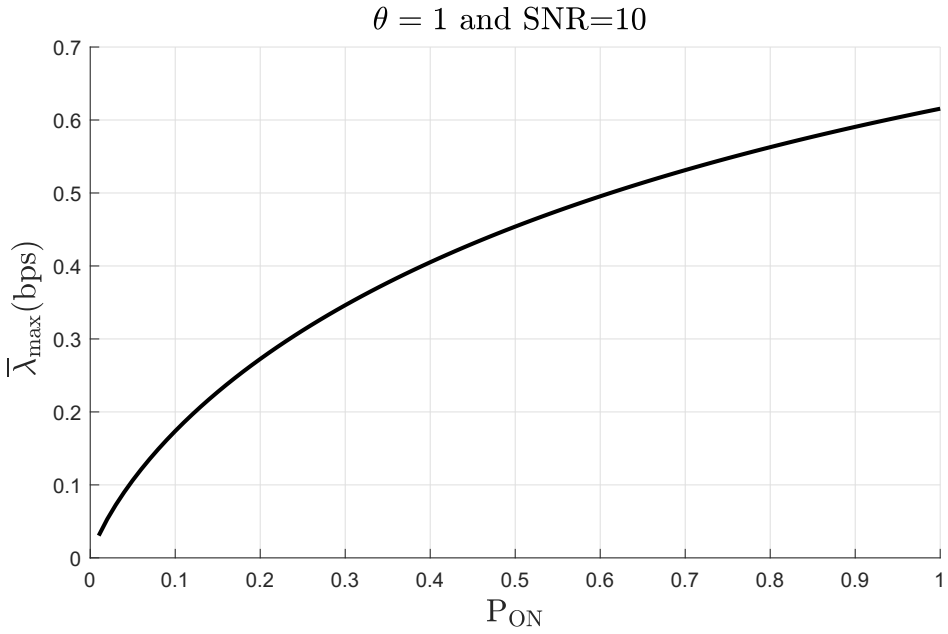

Figure 4.12: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of $\mathrm{P_{ON}}$ of DTMMPP with fixed value of SNR and θ.

4.1.4. Continuous Time Markov Modulated Poisson Sources

The Maximum average arrival rates $\overline{\lambda}_{max}$ of ON/OFF Continuous Time Markov Modulated Poisson Sources which is described in Section 2.5.4 in (50). It is interesting to observe that the throughput with the CTMMPP source is almost identical to that with the Markov fluid source model, save only for the multiplicative factor $\frac{\theta}{e^{\theta}-1}$ in (46). Hence, the throughput is generally smaller with CTMMPP sources and decreases fast with θ. This can be attributed to the much more randomness we experience with an CTMMPP source with respect to the previous Markov models.

Figure 4.13 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as the function of effective capacity for different value of $\mathrm{P_{ON}}$ when QoS exponent $\theta = 1$. It is also interesting to notice from (50) that $\overline{\lambda}_{max}$ is monotonically increasing function of C_E, so if we maximize C_E will maximize $\overline{\lambda}_{max}$ consequently. As $\mathrm{P_{ON}}$ decreases, the arrival rate λ needs to increase with certain rate to cope with non-decreasing throughput with the same departure rate. Hence, the throughput is less affected by variations $\mathrm{P_{ON}}$ and also smaller average arrival rates are supported for given effective capacity.

Figure 4.14 shows the maximum average arrival rate as a function of SNR for different value of QoS exponent and $\mathrm{P_{ON}}$. It is verifies from results that source becomes bursty in smaller value of $\mathrm{P_{ON}}$, implies that data arrive less frequently therefore source burstiness increase, which reduced the throughput of the system. It is also illustrated in Figure 4.13 that the throughput of CTMMPP is less effect by burstiness at high SNR regime as compared to others Markov sources. It is conclude that throughput diminishes with increasing θ and decreasing $\mathrm{P_{ON}}$. For instance with $\theta = 1$, $\mathrm{P_{ON}} = 1$ and average arrival data rate $\overline{\lambda} = 1.2$ bps if system support $\mathrm{SNR} = 80$ for transmitting data in Rayleigh channel then required QoS constraints is satisfied.

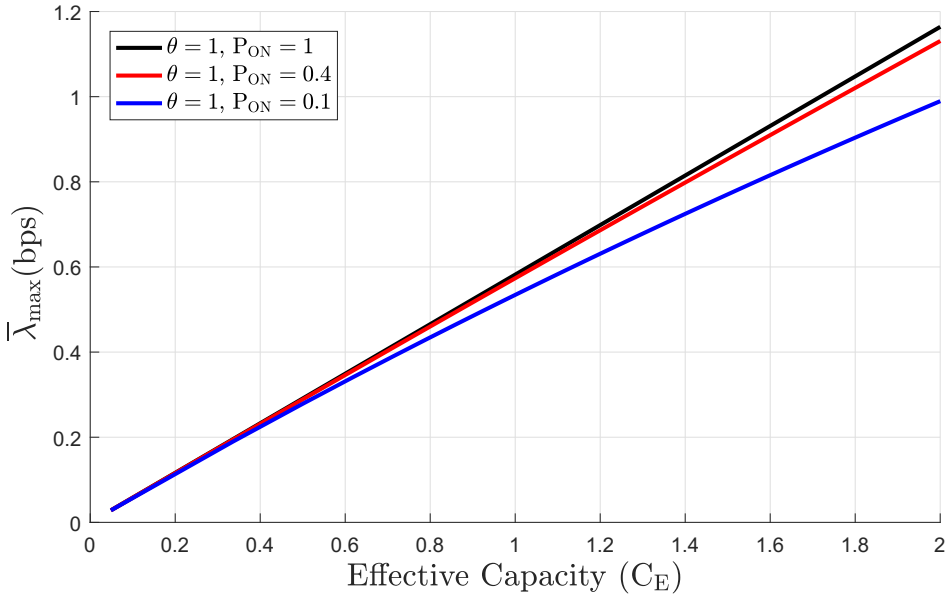

Figure 4.13: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of effective capacity for different value of $\mathrm{P_{ON}}$ and QoS exponent $\theta = 1$.

Next, Figure 4.15 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as a function of QoS exponent θ for Markov modulated Poisson source with different of α and β values. We set $\mathrm{SNR} = 10$, we notice that higher value of maximum average arrival rate can be sustained for low value of QoS exponent θ, hence at this point different values of α and β do not affect the $\overline{\lambda}_{max}$, when QoS exponent $\theta = 0$. As θ increase QoS requirement becomes too stringent, the throughput of the system tends to zero, this reduction in $\overline{\lambda}_{max}$ is more severe for more lower values of $\mathrm{P_{ON}}$ because higher α values leads to larger the transition rate from OFF to ON state. As in the previous cases, this outcome leads to high arrival rates, which may cause buffer overflow.

In what follows Figure 4.16 shows the maximum average arrival rate $\overline{\lambda}_{max}$ as a function of $\mathrm{P_{ON}}$ for Markov modulated Poisson source with QoS exponent $\theta = 1$ and $\mathrm{SNR} = 10$. We observe that the λ_{avg} is the increasing function of $\mathrm{P_{ON}}$, as $\mathrm{P_{ON}}$ increase maximum average arrival rate $\overline{\lambda}_{max}$ is also increase. Thus lower values of $\mathrm{P_{ON}}$ render us larger average arrival rate which increases the burstiness in the sources, therefore throughput is degraded.

The basic feature of a Fluid Markov process is to characterize the traffic as a continuous data of input with a finite arrival rate. By capturing the rate changes at the input, the models analyzes the different events. Because of the characterization of traffic, the fluid modes are analytically tractable and easier to simulate. CTMMPP, uses an underlying Markov process that determines the rate of the sources. At any instant, the current state of the underlying Markov process determines the arrival rate of the inputs.

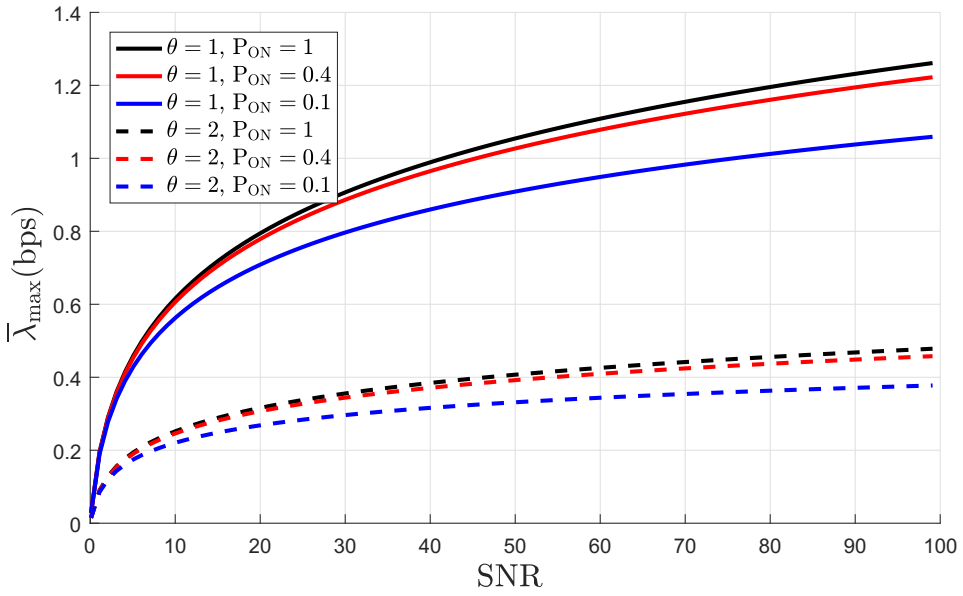

Figure 4.14: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of SNR for different value of QoS exponent and $\mathrm{P_{ON}}$.

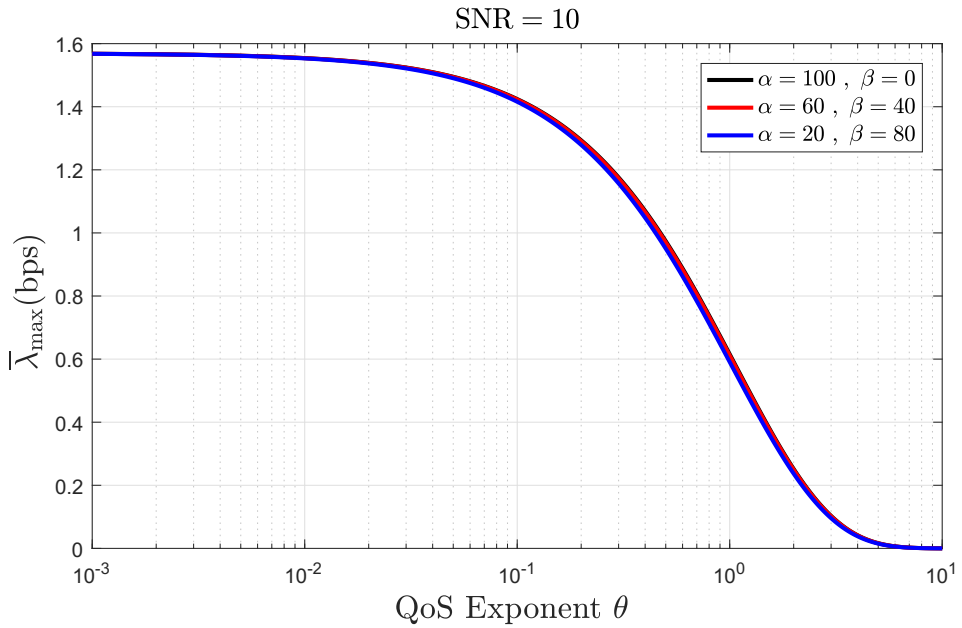

Figure 4.15: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of QoS exponent θ for different values of α and β when $\mathrm{SNR} = 10$.

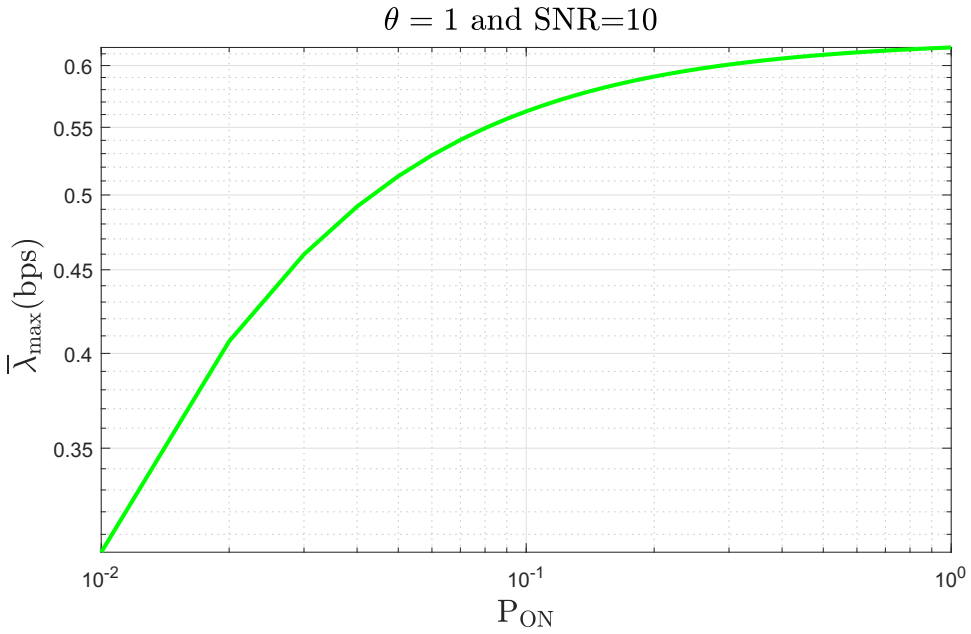

Figure 4.16: Maximum average arrival rate $\overline{\lambda}_{max}$ as a function of $\mathrm{P_{ON}}$ of CTMMPP with fixed value of SNR and θ.

4.1.5. Impact of Optimum Effective capacity to the Effective Bandwidths of Markov Source Models

Next, we aim to identify the different values of arrival rate and $\mathrm{P_{ON}}$ that can support by effective bandwidth, while satisfying the statistical QoS requirement given in the

form in (2). The values effective bandwidth is equal to optimum effective capacity C_E^*, when the data arrivals are and channel response are random under QoS constraint.

For determine optimum effective capacity, we substitute SNR values in (71) into (63) to achieve optimum transmission rate and optimum effective capacity.

In Figure 4.17, we plot the arrival rate levels vs. $\mathrm{P_{ON}}$ curves that required to support given C_E^* for different value of $\mathrm{SNR} \in \{1, 10, 100\}$, when QoS exponent $\theta = 1$. It is noticed that when the source is always ON i.e, $\mathrm{P_{ON}} = 1$, then the maximum arrival rate equal to C_E^*. We also illustrate that as $\mathrm{P_{ON}}$ diminishes, arrival rate in ON state needs to increase with a certain level in order to keep average arrival non-decreasing. However, with same departure rate it is difficult to keep throughput non-decreasing, when QoS constraints are imposed. Therefore, higher arrival rate is required to achieve C_E^* when the source becomes bursty (i.e $\mathrm{P_{ON}} < 1$). We also observe that CTMMPP has more tolerance as compare to DTMS and FMS under the arrival of bursty/random source. For instance with $\mathrm{SNR} = 10$ and $\mathrm{P_{ON}} = 0.2$, DTMS provide $C_E^* = 1.057$ bps with arrival rate of 2.34 bps while, CTMMPP supports an arrival rate of 1.02 bps which is closer to C_E^*. Hence, DTMS and FMS are more affected by the bursty sources as they require a significant adaption of the arrival rate to guarantee QoS when the source is more bursty.

There are some resemblances between DTMS/FMS in the analysis of their throughput performance at high SNR regime and due to high level of variation in CTMMPP sources burstiness penalize the performance, and therefore we achieved lower throughput in CTMMPP sources.

4.1.6. Analysis under Delay Violation Probability

In the previous section, we discussed the impact of burstiness to the optimum effective capacity. Therefore, we investigate the impact of reliability latency trade-off considering source characteristics. We consider statistical delay guarantees for delay sensitive traffic generated by MTD and include the impact of optimum transmission rate as designed in in Chapter 3 at Section 3.1

It is observed from earlier discussion that delay violation probability as function of θ, as the value of QoS exponent θ is increased the delay violation probability is decreased. Delay violation probability of the system where large and small values of θ correspond to fast and slow decaying rates indicating stringent and loose QoS requirements, respectively.

In what follows, we are interested to evaluate effects of latency, thus delay violation probability is showed against SNR for different QoS constraints θ, as illustrated in figure 4.18. We notice that delay violation probability decreases logarithmically with the increase in SNR. As SNR increases, corresponding optimum transmission rate also grow. Therefore, improved channel capacity decrease the delay violation probability in the presence of constant arrival rate and fixed QoS requirements. This degradation in delay violation probability is more severe when stringent QoS requirement are imposed. For example, if we set $\mathrm{SNR} = 10$ and $\theta = 0.1$ then there is 88% possibility that data experience more latency degradation as per required QoS guarantees due to the relax QoS constraint. On the other hand, if we impose QoS requirement $\theta = 1$ then 35% delay violation probability.

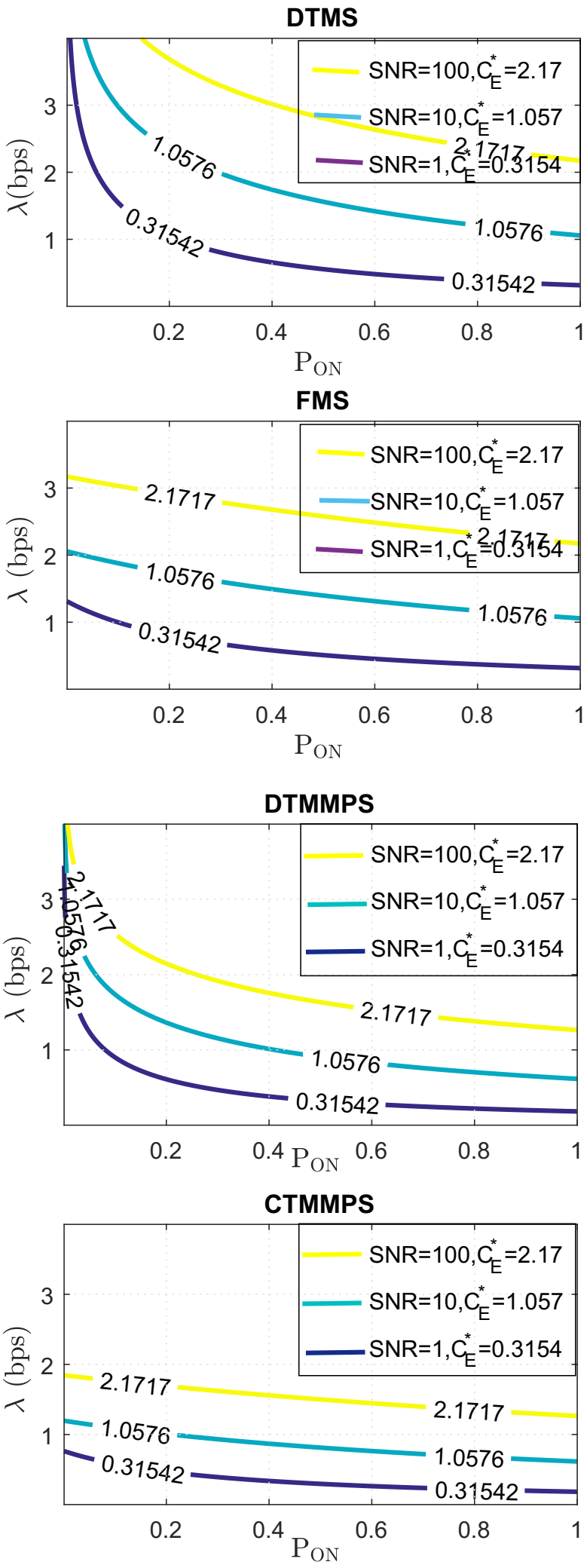

Figure 4.17: Arrival rate vs P_{ON} for different C_E^* of Markovian sources.

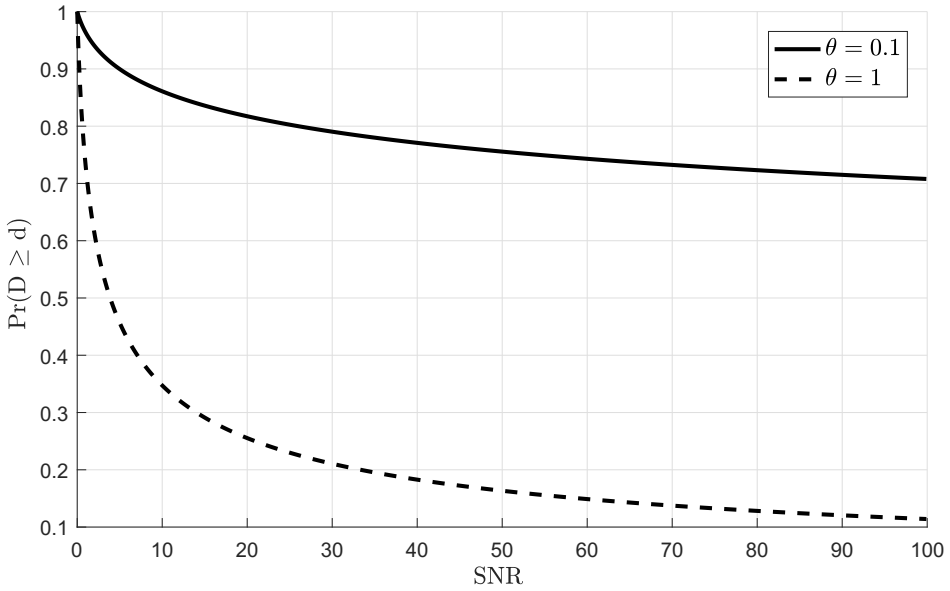

Figure 4.18: Delay Violation Probability as a function of SNR for different values of QoS constraints.

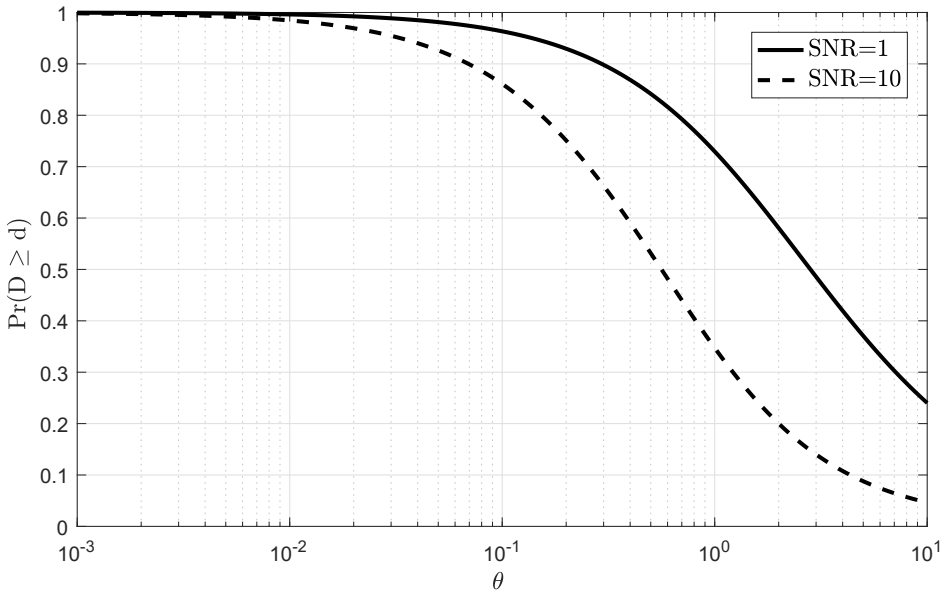

Figure 4.19: Delay Violation Probability as a function of QoS constraints θ for different values of Markovain sources.

Next, Figure 4.19 shows the delay violation probability as a function of QoS exponent for different SNR values. We observed that higher delay violation occurs when no queueing constraint are imposed. As the value of θ increases in fixed transmission rate, the stringent delay constraint are imposed, it means waiting time of data in buffer

is reduced therefore it decrease delay violation probability drastically. This reduction in $\mathrm{Pr(D \geq d)}$ more severe for higher values of SNR. For instance when $\theta = 1$ and $\mathrm{SNR} = 0$ then there is 70% chance that the data experience more delay as per specified latency requirements, where as 35% violation chance if we operate system on $\mathrm{SNR} = 10$.

In Figure 4.20 delay violation probability is plotted as a function of $\mathrm{P_{ON}}$ for fixed arrival rate $\lambda = 0.5$ bps and fixed QoS constraint $\theta = 1$. We notice that delay violation probability decreases exponentially with the increase in $\mathrm{P_{ON}}$. Burstiness is measured from the average arrival of data in ON state. It is observed that more bursty sources degrade the maximum average arrival rate, which directly rises delay violation probability. It is clearly noticed that CTMMPP sources tolerate low delay violation probability in the presence of bursty sources as compared DTMS and FMS.

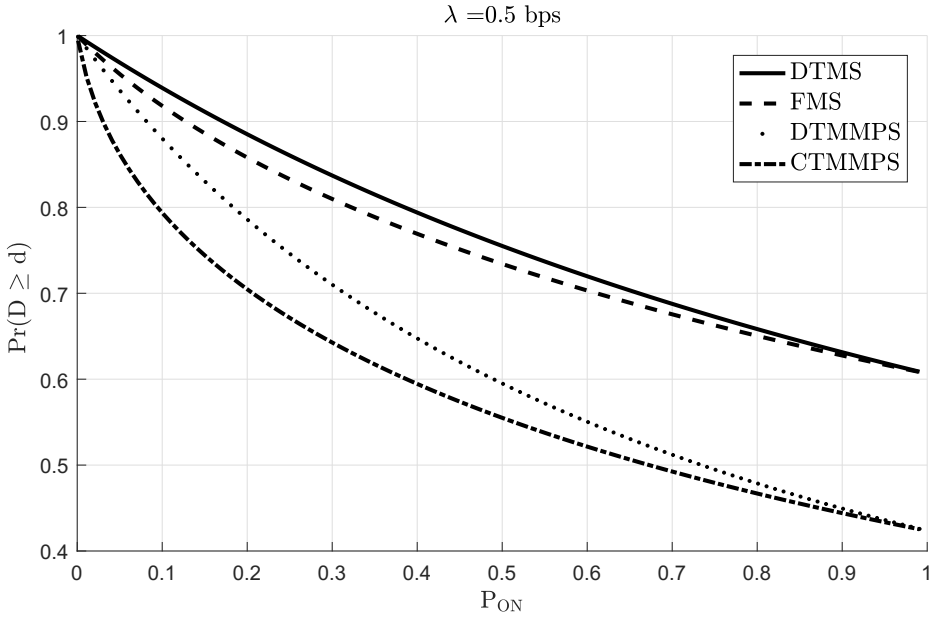

Figure 4.20: Delay Violation Probability as a function of source burstiness $\mathrm{P_{ON}}$ for different Markov source models.

5. CONCLUSION

In this work, we formulated effective transmission rate model to achieve adequate reliability and latency requirement in single point-to-point machine type devices. We incorporated Markov source models to investigate their performance over sources arrival traffic and Rayleigh fading channel. The source, buffer, and channel characteristics have major impact on the performance of model when certain QoS constraints are imposed. Moreover, we introduced a throughput metric that captures the behaviour of sources burstiness and channel condition in the design of effective transmission link when certain level of reliability and latency is required. The results showed that CT-MMPP models is more robust to bursty sources than FMS while DTMS is the least robust. Moreover, increased source burstiness and stringent QoS requirement all need an increase in SNR gain to fulfill reliable communication.

In this work, we presented a detailed analysis of the performance of delay constrained on fixed rate transmission of single point-to-point MTD, when source arrival of data is bursty and random. First, we provided a literature review for queueing constraint, effective bandwidth, effective capacity, Markovian sources and throughput in Chapter 2. When considering Rayleigh fading channels, we obtained a closed form for the optimum C_E, when transmitter does not know the channel condition, therefore the transmitter uses fixed rate, which is a key contributions of the thesis. Moreover, we investigated the effect of SNR and QoS constraints on optimum C_E in Chapter 3. Our analysis was not limited only to determine the throughput through the maximum average arrival rate that can be supported by fading channel while satisfying QoS constraints, but we explored the impact of randomness and burstiness to optimum effective capacity and delay violation probability as well, which are also another key contribution of this thesis and as far as we are aware, such analysis has not yet been reported in the literature. In Chapter 4 we introduce a detailed numerical analysis and investigated the impact of the randomness and burstiness on the reliability and latency of the wireless link, given four distinct Markovian arrival models as described in Chapter 2.

As future work, we are looking investigate the Reliability trade-off into optimization, and besides rate allocation, we aim to investigate power allocation as well. It would be of high interest to move our analysis from fixed rate to adaptive rate transmission as in [26] for performance evaluation under Markovian arrival sources. However, full CSI required at the transmitter, which imposes challenges for MTC type networks and thus needs further investigation and careful design of CSI acquisition. Moreover, we aim to formulate effective transmission rate which is based on two-state continuous-time Markov chain channel model as in [22] (rather than discrete Markov chain model currently used in this thesis). Furthermore, we aim to investigate reliability-latency trade-off, rate allocation and power allocation. Moreover, incorporating Markovian sources to investigate the impact of different types of traffic in the optimal fixed and adaptive rate protocols. Furthermore, we aim to do traffic fitting via real data traces into analytical models, which allows characterize MTC traffic and its impact on the network performance [27, 5].

6. REFERENCES

[1] Ericsson, "5G systems," *Ericsson White Paper*, January 2017.

[2] W. H. Hsu, Q. Li, X. H. Han, and C. W. Huang, "A hybrid IoT traffic generator for mobile network performance assessment," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, June 2017, pp. 441–445.

[3] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov 2014, pp. 146–151.

[4] Y.Morioka, "LTE for Mobile Consumer Devices," *ETSI Workshops on Machine to Machine standardization*, 2011.

[5] N. Nikaein, M. Laner, K. Zhou, P. Svoboda, D. Drajic, M. Popovic, and S. Krco, "Simple Traffic Modeling Framework for Machine Type Communication," in *International Symposium on Wireless Communication Systems*, Aug 2013, pp. 1–5.

[6] M. Laner, P. Svoboda, N. Nikaein, and M. Rupp, "Traffic Models for Machine Type Communications," in *International Symposium on Wireless Communication Systems*, Aug 2013, pp. 1–5.

[7] A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 82–89, Jul 1997.

[8] E. Soltanmohammadi, K. Ghavami, and M. Naraghi-Pour, "A Survey of Traffic Issues in Machine-to-Machine Communications Over LTE," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 865–884, Dec 2016.

[9] C. Kalalas and J. Alonso-Zarate, "Efficient Cell Planning for Reliable Support of Event-Driven Machine-Type Traffic in LTE," in *2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–7.

[10] E. Grigoreva, M. Laurer, M. Vilgelm, T. Gehrsitz, and W. Kellerer, "Coupled Markovian Arrival Process for Automotive Machine Type Communication traffic modeling," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[11] O. Al-Khatib, W. Hardjawana, and B. Vucetic, "Traffic modeling for Machine-to-Machine (M2M) last mile wireless access networks," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 1199–1204.

[12] Y. Li, M. C. Gursoy, and S. Velipasalar, "Throughput of Hybrid-ARQ Chase Combining with ON-OFF Markov Arrivals under QoS Constraints," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[13] G. Ozcan, M. Ozmen, and M. C. Gursoy, "QoS-driven energy-efficient power control with Markov arrivals and finite-alphabet inputs," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2769–2773.

[14] M. Ozmen and M. C. Gursoy, "Energy efficiency in multiple-antenna channels with markov arrivals and queueing constraints," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 799–803.

[15] S. Akin and M. C. Gursoy, "Performance Analysis of Cognitive Radio Systems under QoS Constraints and Channel Uncertainty," *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 2883–2895, September 2011.

[16] Y. Jin, X. Xu, Y. Wang, and X. Tao, "Multi-QoS mobile services guaranteed resource allocation with effective capacity," in *2017 IEEE/CIC International Conference on Communications in China (ICCC)*, Oct 2017, pp. 1–6.

[17] Y. Gu, Q. Cui, Y. Chen, W. Ni, X. Tao, and P. Zhang, "Effective Capacity Analysis in Ultra-Dense Wireless Networks With Random Interference," *IEEE Access*, vol. 6, pp. 19 499–19 508, 2018.

[18] J. Li, Y. Ding, Q. Ye, N. Zhang, and W. Zhuang, "On Effective Capacity and Effective Energy Efficiency in Relay-assisted Wireless Networks," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2017.

[19] G. Ozcan and M. C. Gursoy, "Optimal Power Control for Fading Channels with Arbitrary Input Distributions and Delay-Sensitive Traffic," *IEEE Transactions on Communications*, pp. 1–1, 2018.

[20] D. Qiao and M. C. Gursoy, "Statistical Delay Tradeoffs in Buffer-Aided Two-Hop Wireless Communication Systems," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4563–4577, Nov 2016.

[21] D. Qiao, M. C. Gursoy, and S. Velipasalar, "The impact of QoS constraints on the energy efficiency of fixed-rate wireless transmissions," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5957–5969, December 2009.

[22] M. Ozmen and M. C. Gursoy, "Energy Efficiency of Fixed-Rate Transmissions with Markov Arrivals under Queueing Constraints," *IEEE Communications Letters*, vol. 18, no. 4, pp. 608–611, April 2014.

[23] J. Tang and X. Zhang, "Quality-of-Service Driven Power and Rate Adaptation over Wireless Links," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 3058–3068, August 2007.

[24] L. Musavian and Q. Ni, "Effective Capacity Maximization With Statistical Delay and Effective Energy Efficiency Requirements," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3824–3835, July 2015.

[25] M. Ozmen and M. C. Gursoy, "Wireless Throughput and Energy Efficiency With Random Arrivals and Statistical Queuing Constraints," *IEEE Transactions on Information Theory*, vol. 62, no. 3, pp. 1375–1395, March 2016.

[26] L. Musavian and Q. Ni, "Effective capacity maximization with statistical delay and effective energy efficiency requirements," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3824–3835, July 2015.

[27] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Large-scale measurement and characterization of cellular machine-to-machine traffic," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1960–1973, Dec 2013.

[28] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, July 2003.

[29] Q. Du and X. Zhang, "Statistical QoS provisionings for wireless unicast/multicast of multi-layer video streams," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 420–433, April 2010.

[30] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[31] Cheng-Shang, *Performance Guarantees in Communication Networks*. springer-Verlag,London, U.K., 2000.

[32] C.-S. Chang and T. Zajic, "Effective bandwidths of departure processes from queues with time varying capacities," Apr 1995, pp. 1001–1009 vol.3.

[33] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 329–343, Jun 1993.

[34] G. Kesidis, J. Walrand, and C. S. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 424–428, Aug 1993.

[35] A. Balasubramanian and S. L. Miller, "The effective capacity of a time division downlink scheduling system," *IEEE Transactions on Communications*, vol. 58, no. 1, pp. 73–78, January 2010.