

QUEUING ANALYSIS OF THE ENERGY-DELAY TRADE-OFF IN WIRELESS NETWORKS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

By
Ege Orkun Gamgam
September 2018

Queuing Analysis Of The Energy-delay Trade-off In Wireless Networks

By Ege Orkun Gamgam

September 2018

We certify that we have read this thesis and that in our opinion it is fully adequate,
in scope and in quality, as a thesis for the degree of Master of Science.



Nail Akar(Advisor)

Ezhan Karaşan

Murat Alanyalı

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

QUEUEING ANALYSIS OF THE ENERGY-DELAY TRADE-OFF IN WIRELESS NETWORKS

Ege Orkun Gamgam

M.S. in Electrical and Electronics Engineering

Advisor: Nail Akar

September 2018

Energy-efficiency of wireless communication systems has been an important research topic in recent years. For such a system, a transmission profile is described by the transmission power and the modulation and coding scheme (MCS) to be used for packet transmission. For a given channel condition, higher order MCSs offer higher throughput at the expense of requiring more transmission power. Average power consumption of the system can be reduced by using lower order MCSs at the expense of increased queuing delays. Using this observation, the goal of this study is the development of transmission profile selection policies so as to minimize the average power consumption while meeting a statistical delay constraint for a wireless link. For the purpose of assessing the proposed policies, the system is modeled as an M/M/1 queue where transmission speeds of packets are dynamically selected based on the queuing delay already experienced by them. This setting is shown to give rise to a multi-regime Markov fluid queue model which is used to obtain the waiting time distributions of packets as well as the average power consumption. In the numerical examples, proposed profile selection policies are evaluated for different system parameters using realistic transmission profiles obtained from LTE simulations. A proposed energy-aware profile selection policy is shown to consistently outperform all other proposed policies in terms of energy-efficiency whereas a reasonable performance is also obtained with a simpler-to-implement policy.

Keywords: queuing analysis, adaptive modulation, transmission power control, energy-efficiency, Markov fluid queue.

ÖZET

KABLOSUZ AĞLARDA ENERJİ-GEÇİKME ÖDÜNLEŞİMİNİN KUYRUK SİSTEMİ ANALİZİ

Ege Orkun Gamgam

Elektrik ve Elektronik Mühendisliği, Yüksek Lisans

Tez Danışmanı: Nail Akar

September 2018

Enerji verimliliği, son yıllarda kablosuz iletişim sistemlerinde önemli bir araştırma konusu olmuştur. Bu sistemlerde her bir transmisyon profili, transmisyon güç seviyesi ve adaptif modülasyon ve kodlama yapısı (MCS) olmak üzere iki parametre ile tanımlanır. Belirli bir kanal durumu için yüksek MCS içeren transmisyon profilleri yüksek enerji tüketimi karşılığında daha hızlı servis imkanı sağlamaktadır. Sistemin ortalama güç tüketimi, daha düşük MCS içeren profiller kullanılarak paketlerin kuyrukta bekleme sürelerinin artması dezavantajı ile birlikte azaltılabilir. Bu gözlemden yola çıkarak, bu çalışmanın amacı kablosuz linklerde belirli bir istatistiksel gecikme kısıtlaması altında ortalama enerji tüketiminin minimizasyonu üzerine transmisyon profili seçme politikaları geliştirmektir. Bu politikaların incelenmesi amacıyla sistem, paketlerin transmisyon hızlarının kuyrukta bekleme sürelerine göre dinamik olarak seçildiği bir M/M/1 kuyruk sistemi olarak modellenmiştir. Bu sistemin çok bölgeli bir Markov akışkan kuyruk modeline karşılık geldiği gösterilmiştir. Bu model kullanılarak paketlerin kuyrukta bekleme sürelerinin dağılımları ve sistemin ortalama enerji tüketimi çıkarılmıştır. Önerilen profil seçme politikaları LTE simülasyonlarından elde edilen gerçekçi profiller kullanılarak farklı sistem parameterleri altında sayısal örneklerle değerlendirilmiştir. Önerilen politikalardan enerji farkındalığı ile profil seçimi gerçekleştiren bir politikanın diğer politikalara karşı tutarlı bir şekilde üstünlük sağladığı gösterilmiştir. Bunun yanında donatma kompleksitesi diğer politikalara göre daha az olan bir politikanın önemli miktarlarda enerji kazancı sağlayabildiği gösterilmiştir.

Anahtar sözcükler: kuyruk analizi, adaptif modülasyon, transmisyon güç kontrolü, enerji verimliliği, Markov akışkan kuyruğu.

Acknowledgement

I would like to express my sincere and special gratefulness to my advisor Prof. Nail Akar for his extensive guidance and insight throughout my study. His door was always open whenever I ran into trouble in my research. Without his assistance, this work would never have came to existence.

I would like to thank Prof. Ezhan Karařan and Prof. Murat Alanyalı for agreeing to be on my thesis committee.

I would like to thank The Scientific and Technologies Research Council of Turkey (TÜBİTAK) for funding this study under the project grant EEEAG-115E360.

Finally, I would like to thank my family members, my father Hamza, my mother Zehra and my brother Onur for their continuous support and encouragement throughout my study.

Contents

1	Introduction	1
1.1	Overview	1
1.2	Thesis Contribution	5
1.3	Thesis Outline	6
2	Related Work	7
2.1	Speed Scaling in Queuing Systems	7
2.2	Energy-Efficient Rate Adjustment in 5G Networks	8
2.3	Fluid Queue Models in QoS Analysis in Wireless Links	8
3	Preliminary Study: Multi-Regime Markov Fluid Queues	10
4	System Modeling	14
4.1	Model Description	14
4.2	MRMFQ Model	15

4.3	Performance Measures	18
5	Transmission Profile Selection Policies	20
5.1	Min-Delay Policy (MDP)	21
5.2	Single Threshold Policy (<i>STP</i>)	22
5.3	Proportional Control Policy (<i>PCP</i>)	23
5.4	Energy-efficient <i>PCP</i> (<i>EPCP</i>)	25
5.5	Extended Policies	28
6	Numerical Examples	31
6.1	System Setup	31
6.2	Model Validation	35
6.3	Performance Evaluation	37
7	Conclusions	44

List of Figures

1.1	Power consumption share in cellular networks which is taken from [1].	3
1.2	Physical layer simulation results of an example channel in LTE which is taken from [2].	4
4.1	Sample paths of three processes $A(t)$, $U(t)$, and $X(t)$	16
4.2	State transitions (a) for $X(t) = 0$ and (b) for regime k , $k = 1, \dots, K$.	18
5.1	Service rate selection for the HoL packet in MDP	22
5.2	Service rate selection for the HoL packet in STP	23
5.3	Service rate selection for the HoL packet in PCP	24
5.4	An example of a relatively energy-inefficient profile.	25
5.5	An example of a universal profile set $\mathcal{N}$	26
5.6	Elimination of the relatively energy-inefficient profiles from an ex- ample universal set $\mathcal{N} = \{(1, 1), (2, 3), (6, 4), (7, 7), (11, 10), (13, 14)\}$.	27
6.1	LTE-TDD frame structure	32

6.2	Service rate and transmission power attributes of the profiles used for <i>STP</i> , <i>PCP</i> , and <i>EPCP</i> , when $\alpha_m = 12$ dB and $e_b = 0.02$. . .	35
6.3	Delay violation probability p_v of the particular policy <i>PCP</i> as a function of the parameter T_{PCP} when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms and the maximum attainable SNR $\alpha_m = 12$ dB for both analytical model and simulations.	36
6.4	Average power consumption P of the particular policy <i>PCP</i> as a function of the parameter T_{PCP} when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms and the maximum attainable SNR $\alpha_m = 12$ dB for both analytical model and simulations.	37
6.5	Delay violation probability p_v of the three proposed policies <i>STP</i> , <i>PCP</i> , and <i>EPCP</i> as a function of the parameter T_p when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms, and the maximum attainable SNR $\alpha_m = 12$ dB.	39
6.6	Average power consumption P of the three proposed policies <i>STP</i> , <i>PCP</i> , and <i>EPCP</i> as a function of the parameter T_p when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms, and the maximum attainable SNR $\alpha_m = 12$ dB.	40
6.7	The percentage energy gain G_p as a function of the arrival rate λ for the six proposed policies.	41
6.8	The absolute saved power P_{saved} as a function of the arrival rate λ for the six proposed policies.	41
6.9	The percentage energy gain G_p as a function of the delay bound D_0 for the six proposed policies.	42
6.10	The absolute saved power P_{saved} as a function of the delay bound D_0 for the six proposed policies.	42

- 6.11 The impact of the maximum attainable average SNR α_m on the percentage energy gain G_p for the six proposed policies. 43
- 6.12 The impact of the maximum attainable average SNR α_m on the absolute saved power P_{saved} for the six proposed policies. 43



List of Tables

6.1	BLER $e(\alpha, I_M)$ as a function of the SNR α and MCS index I_M . . .	33
6.2	The throughput $r(\alpha, I_M)$ in bits/PRB as a function of the SNR α and MCS index I_M	34
6.3	Profiles in the set $\mathcal{N}$ for the case $\alpha_m = 12$ dB and $e_b = 0.02$	35

Chapter 1

Introduction

1.1 Overview

In the era of globalization, tremendous demand for data keeps growing each year. It is estimated that there will be 50 billion connected devices in 2020 [3]. Most of these devices will use wireless connectivity due to the advantage of being portable and small. To support this many connected devices, a capacity of 1000 times higher data rates is anticipated to be required [4]. With the emerging 5G technologies such as virtual reality, telemedicine, etc., the aim is also to achieve x1000 higher data rates in these networks. Using today's wireless technologies, this will require a proportional growth of 1000 times more energy consumption. This unmanageable energy demand is raising environmental concerns on CO₂ emissions and electromagnetic pollution which is also acknowledged by the GSM Association in their Mobile's Green Manifesto that demands for a 40% reduction in CO₂ emission per connection by 2020 [5]. In addition to ecological concerns, energy-efficient technologies also allow operators to save on their electricity bills and maintenance costs [5]. All these factors motivated a significant interest in studying and designing energy-efficient communication techniques.

It is well known that base stations (BSs) are one of the major contributors to

the overall power consumption in the context of wireless cellular networks with a share of more than 50% in overall consumption as shown in Fig 1.1 which is taken from [1]. Modeling power consumption in BSs is widely studied in several studies. The study [6] describes a BS consisting of multiple transceivers (TRXs). Each TRX serves to a single antenna element. Power consumption in each TRX is modeled by several factors: RF unit, signal processing unit, losses incurred by feeder, active cooling, mains supply and DC-DC power supply and also a load dependent amplifier term given by P_{out}/P_{max} where P_{out} and P_{max} is the radiated power (called as transmission power throughout the thesis) and the maximum output power at the antenna port, respectively. Reduction of a power ΔP_{out} at the antenna port results in a reduction of $\Delta_p \Delta P_{out}$ in overall input power where Δ_p is defined as the power gradient [7]. For a macro BS setup with $P_{max} = 43$ dbm, 10 MHz bandwidth, 3 sectors per site, 2x2 MIMO configuration and $\Delta_p = 4.7$, the load dependent term $\Delta_p P_{out}$ contributes above 40% to the overall power consumption in the BS [6]. Therefore, it is relatively important to develop energy-efficient transmission strategies by reducing P_{out} in high power BSs which has been studied in [8–10]. Some of these strategies result in reduced delay performance which may be problematic for applications with minimum delay constraint. This concern motivated several studies in which energy-efficiency is investigated under delay constraints to optimize power consumption while ensuring a certain performance [11–13].

In the context of wireless communications, transmission power and modulation and coding schemes (MCSs) can be dynamically selected the latter being referred to link adaptation. In this thesis, we define a transmission profile to be a particular setting of a MCS and a transmission power value. Profiles with higher order MCS (higher service rate) offer higher data rates and better spectral efficiency but they are less resilient to noise and interference, i.e., require more power to overcome noise. Signal-to-Noise ratio (SNR) versus Block Error Rate (BLER) graph for an example channel setup of LTE is illustrated in Fig 1.2 from which it is clear that the required SNR to meet a certain BLER is larger for higher order MCSs [2]. On the other hand, profiles with lower order MCS (lower service rate) typically consume less energy for a given channel condition at the

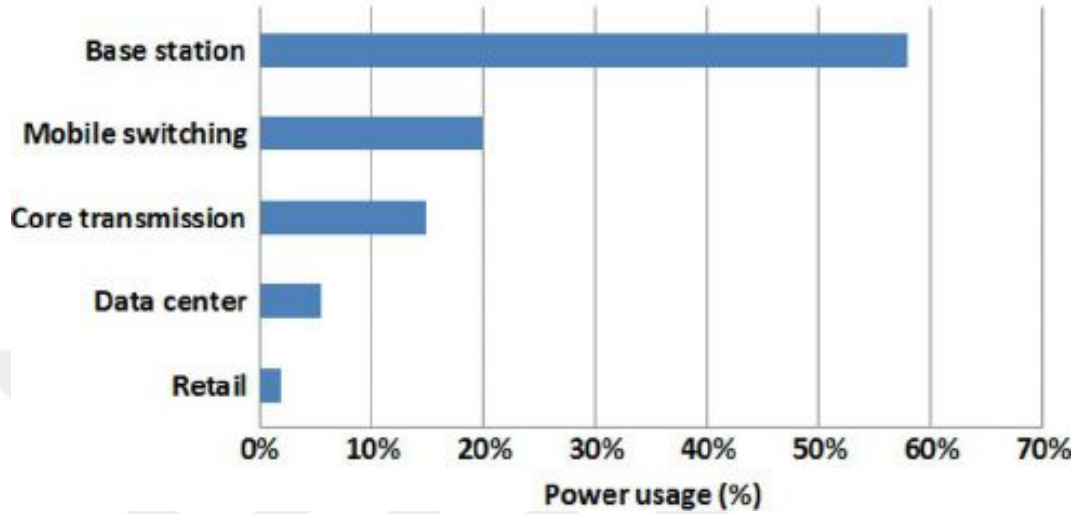


Figure 1.1: Power consumption share in cellular networks which is taken from [1].

expense of increased queuing delays. Therefore, it is important to dynamically choose appropriate profile, i.e., combination of transmission power and MCS to minimize power consumption while satisfying delay constraints.

The objective of minimizing transmission energy under deadline constraints has been studied in [11] over a time-varying stochastic channel in the context of wireless communications. An energy-efficient rate control has been obtained using a decomposition approach and a cumulative curves methodology. Furthermore, a heuristic policy has been then developed for arbitrary packet arrival processes. The study [12] obtains an optimal off-line scheduling with infinitely many transmission profiles for a wireless node operating under delay constraints. Then, an on-line algorithm has also been developed which is shown to, through simulations, perform better than the deterministic scheduler in terms of energy savings. In [14], delay bound violation (packet drop) probability is minimized subject to constraints on delay bound, average power and arrival rate over a Rayleigh fading channel. However, queuing model of a wireless link which uses realistic transmission profiles for the purpose of energy minimization under statistical delay constraints does not exist in literature.

In this thesis, the following setup is considered. The system consists of a wireless transmitter that is capable of dynamically adjusting the transmission

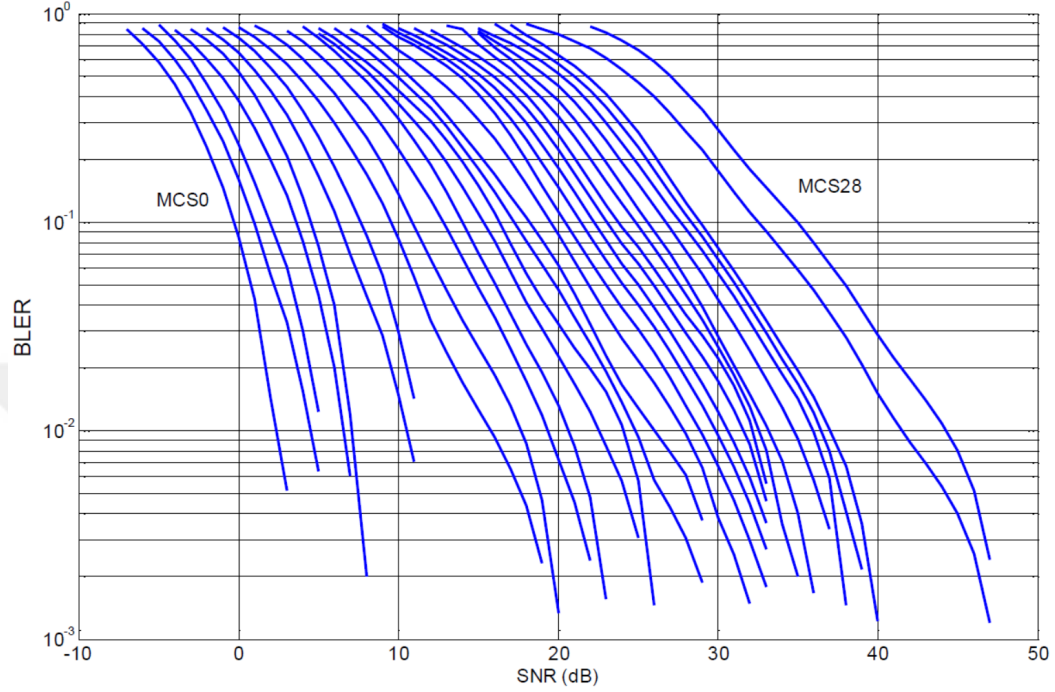


Figure 1.2: Physical layer simulation results of an example channel in LTE which is taken from [2].

power and MCS for each packet destined to a single receiver. Packet arrivals are considered to be stochastic and one packet is transmitted at a time. The transmitter also has an infinite size FIFO buffer to which packets join before being transmitted. A finite set of K transmission profiles is assumed to be available at the transmitter. Each transmission profile consists of two attributes: service rate in packets/sec and transmission power in Watts. A profile to be used for the Head of Line (HoL) packet is selected among the K profiles based on the queuing delay already experienced by that particular packet at the service start time. For ensuring a given delay performance, it is natural to resort to profiles with higher service rates for packets that have already experienced longer queuing delays. However, development of policies that specifies which profile to use and when to use, is an open problem in the literature which is the scope of this thesis.

Within the queuing system of interest, packets sizes are assumed to be exponentially distributed and packet arrivals are assumed to follow a Poisson process.

A transmission profile set with finite size is available to serve packets depending on their waiting time in the queue and a-priori transmission profile selection policy is assumed which gives rise to an M/M/1 queue with delay-dependent service times. If the profile selection is based on number of packets in the queue, performance metrics of interest can be obtained by classical Markov Chain (MC) methods [15]. However, delay-dependence of the profile selection policies motivates a multi-regime Markov fluid queue (MRMFQ) approach for the purpose of modeling the system. The advantage of this approach stems from the lack of need for re-derivation of any integral or differential equations. In this thesis, the well-established theory of MRMFQs is used by which one can use numerically stable and efficient methods that are already available for the solution of MRMFQs. A fast MRMFQ solver based on [16] is employed to solve the model of interest. It is matrix-analytical and its main building blocks are the ordered Schur decomposition and the Sylvester equation. This solver can be used to obtain expressions for performance metrics such as the delay distribution of packets and the average power consumption for the system of interest which is an M/M/1 queue with delay-dependent service times.

1.2 Thesis Contribution

An MRFMQ model based on the study [17] which considers dynamic speed scaling in a completely different context of computer systems is adapted and modified. Average power consumption of the wireless transmitter along with the delay violation probability of packets which is defined as the probability that a packet experiences a queuing delay greater than a given delay threshold, is considered as the performance measure of interest. An MRMFQ model solver that obtains these performance parameters is implemented based on [16]. This solver can be easily adapted to models that consider similar performance metrics in the future.

The main contribution of this thesis is the development of transmission profile selection policies in the context of wireless communication systems. When a policy is given in terms of one or a few parameters, one can repeatedly solve the

MRMFQ model to find the optimum parameter set that minimizes average power consumption while satisfying a given delay constraint. This approach is used for assessing the performance of the proposed transmission profile selection policies. Several policies are proposed and studied using the analytical model. It is shown that substantial energy savings are obtainable with some of the proposed policies for a wide range of systems parameters.

Another contribution is that proposed policies are evaluated using realistic profiles obtained from simulations of Physical Downlink Shared Channel (PDSCH) within the LTE framework. Analytical model is also compared with frame level simulations of the LTE by means of considering the effects that MRMFQ model does not take into account such as byte padding operations. Therefore, it is believed that the current work is expected to find real world applications in similar settings.

1.3 Thesis Outline

The rest of this thesis is organized as follows. Related work is given in Section 2. A brief preliminary study on MRMFQs is provided in Section 3 along with the boundary conditions necessary to solve their steady-state distribution. The system model is described and related MRMFQ model is presented in Section 4. Transmission profile selection policies are developed in Section 5. Numerical examples are presented in Section 6 to assess the performance of the proposed policies using profiles obtained from the LTE simulations. Finally, conclusions and future work items are provided in Section 7.

Chapter 2

Related Work

2.1 Speed Scaling in Queuing Systems

Balancing between energy and performance is adapted in computer systems using speed-scaling which is a method to adjust the speed of the computer systems based on the load [18]. A simple way to accomplish this is the static speed-scaling in which the computer system either goes into sleep mode or employs a single speed depending on the existence of work in the system. More sophisticated rate adjustment method is the dynamic speed scaling in which the speed is adapted all times to the current state [18]. The interaction between the speed scaling and load balancing is studied in [19] where three models are formulated each of which minimizes different performance metrics such as delay experienced by a job or energy usage. The study [20] considers a single server queue with Poisson arrivals and state-dependent service times for which a method for obtaining the minimum achievable average cost is developed where the cost consists of two elements that are motivated by the queue length and the service level chosen. The speed scaling is also considered for energy saving purposes in several studies for wireline technologies such as Ethernet. In [15], authors investigate Adaptive Link Rate for the purpose of minimizing energy in a typical Ethernet link. Several rate adaptation policies are studied and their performance is evaluated with respect

to performance metrics such as mean packet delay.

2.2 Energy-Efficient Rate Adjustment in 5G Networks

As the 5G development and related applications with strict QoS requirements are rapidly advancing, the studies on the energy-efficient communication for 5G networks has been widely conducted in the literature [21–26]. The reference [21] studies power control and rate adjustment methods to maximize effective capacity of the channel under statistical QoS and power constraints. In [22], several scheduling strategies that maximize energy-efficiency by minimizing the number of transmission attempts while meeting a deadline constraint are proposed. An optimization problem is formulated in [23] to optimize energy efficiency metric under maximum power and delay threshold constraints which are defined based on a queuing model of the wireless link. The reference [24] develops centralized and decentralized power control algorithms to maximize energy-efficiency under QoS constraints for 5G wireless technologies. In [25] several Base Station (BS) sleeping schemes are developed to minimize power consumption under rate constraints. An analytical model is also proposed to evaluate the performance metrics, namely power consumption, mean delay and blocking probability. The reference [26] proposes a power control approach based on game theory to jointly optimize QoS and power consumption in wireless links.

2.3 Fluid Queue Models in QoS Analysis in Wireless Links

Several studies consider fluid queue models in wireless channels for the analysis of QoS metrics such as energy-efficiency and packet loss probability [27–33]. The

reference [27] analyzes the packet-loss performance based QoS in a wireless transmitter over a fluid version of Gilbert-Elliot channel model. The traffic burstiness is captured as on-off fluid model and exact packet loss probability is obtained using the proposed fluid model. The same fluid model is also used in [28] where the authors derive distribution of the queuing delay at the transmitter from which delay violation probability is investigated with respect a delay threshold. The reference [29] also adapts the fluid model for obtaining analytical expressions for QoS parameters such as the probability of buffer overflow. The reference [30] studies energy-efficiency of hybrid automatic repeat request (HARQ) schemes under statistical queuing, outage and deadline constraints considering several sources including on-off Markov fluid source. In [31], energy-efficiency and throughput metrics for wireless fading channels are investigated in the presence of Markovian arrival models such as Markov fluid process. The references [32,33] study energy-efficiency and throughput optimization under QoS constraints in wireless links considering several Markovian sources including Markov fluid source.

Chapter 3

Preliminary Study: Multi-Regime Markov Fluid Queues

A Markov fluid queue (MFQ) is defined by two processes $X(t)$ and $Z(t)$ where $X(t)$ is the process that represents the fluid level in the queue and $Z(t) \in \{0, 1, \dots, N\}$ is a Markov Chain that determines the drift, i.e., the rate at which $X(t)$ changes at time t . When $Z(t)$ is in state i , the process $X(t)$ changes by a drift of r_i . Note that it cannot be reduced further if the fluid level is zero at time t , i.e., $X(t) = 0$. For an MFQ with infinite capacity, there is no upper boundary on the fluid level. Mathematically, the following applies when the state is $Z(t) = i$:

$$\frac{dX(t)}{dt} = \begin{cases} \max\{0, r_i\}, & X(t) = 0, \\ r_i, & X(t) > 0. \end{cases} \quad (3.1)$$

MRMFQs are the generalization of MFQs such that the system behavior at a given time t may change with the fluid level [16, 34–36]. Both the modulated process $X(t)$ and the modulating process $Z(t)$ depend on the fluid (buffer) level

at time t . The rest of this chapter briefly summarizes MRMFQs based on the notation used in [16]. The entire buffer is divided into K regimes using the thresholds denoted by $0 = T^{(0)} < T^{(1)} < \dots < T^{(K-1)} < T^{(K)} = \infty$. For an MRMFQ with finite capacity, $T^{(K)}$ equals to some finite value whereas the case of infinite capacity, i.e, $T^{(K)} = \infty$ is considered throughout the thesis. The system is said to be in regime k with a drift of $r_i^{(k)}$, when $T^{(k-1)} < X(t) < T^{(k)}$ and the current state is i . For each regime $1 \leq k \leq K$, a drift matrix and an infinitesimal generator, denoted by $R^{(k)}$ and $Q^{(k)}$ respectively, are defined. $R^{(k)}$ is a diagonal matrix which can be shown as follows:

$$R^{(k)} = \begin{bmatrix} r_0^{(k)} & 0 & \dots & 0 & 0 \\ 0 & r_1^{(k)} & 0 & \vdots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \dots & 0 & r_{N-2}^{(k)} & 0 \\ 0 & 0 & \dots & 0 & r_{N-1}^{(k)} \end{bmatrix},$$

where the diagonal elements are the drift values in regime k . At each boundary $T^{(k)}$, the terms $Q^{(k)}$ and $R^{(k)}$ are also defined as $\tilde{Q}^{(k)}$ and $\tilde{R}^{(k)}$, respectively along with the drift value $\tilde{r}_i^{(k)}$. The joint pdf $f^{(k)}(x)$ is defined for the state k as follows:

$$f^{(k)}(x) = \begin{bmatrix} f_0^{(k)}(x) & f_1^{(k)}(x) & \dots & f_{N-1}^{(k)}(x) \end{bmatrix}. \quad (3.2)$$

Each element $f_i^{(k)}(x)$ in this row is defined as:

$$f_i^{(k)}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} \Pr\{X(t) \leq x, Z(t) = i\}, \quad (3.3)$$

for $0 \leq i < N$. Similarly, a row vector of steady state probability mass accumulations are defined as:

$$c^{(k)} = \begin{bmatrix} c_0^{(k)} & c_1^{(k)} & \dots & c_{N-1}^{(k)} \end{bmatrix}, \quad (3.4)$$

where

$$c_i^{(k)} = \lim_{t \rightarrow \infty} \Pr\{X(t) = T^{(k)}, Z(t) = i\}, \quad (3.5)$$

for $0 \leq i < N$. As shown in [16], the following set of differential equations holds for the joint pdf vector:

$$\frac{d}{dx} f^{(k)}(x) R^{(k)} = f^{(k)}(x) Q^{(k)}. \quad (3.6)$$

Additionally, $S_-^{(k)}$, $S_0^{(k)}$ and $S_+^{(k)}$ denote the set of states with negative, zero and positive drifts in regime k , respectively. Similarly, $\tilde{S}_-^{(k)}$, $\tilde{S}_0^{(k)}$ and $\tilde{S}_+^{(k)}$ denote the set of states with negative, zero and positive drifts at boundary $T^{(k)}$, respectively. Using these definitions, boundary conditions are also provided as follows from [16]:

$$c_i^{(0)} = 0, \quad \forall i \in S_+^{(1)} \quad (3.7)$$

$$c_i^{(k)} = 0, \quad \forall i \in \left(S_+^{(k)} \cap S_+^{(k+1)} \right) \cup \left(S_-^{(k)} \cap S_-^{(k+1)} \right) \quad (3.8)$$

$$c_i^{(k)} = 0, \quad \forall i \in \left(S_-^{(k)} \cap S_+^{(k+1)} \right) \cap \left(\tilde{S}_+^{(k)} \cup S_-^{(k)} \right) \quad (3.9)$$

$$f^{(1)}(0+) R^{(1)} = c^{(0)} \tilde{Q}^{(0)} \quad (3.10)$$

$$f^{(k+1)}(T^{(k)}+) R^{(k+1)} - f^{(k)}(T^{(k)}-) R^{(k)} = c^{(k)} \tilde{Q}^{(k)} \quad (3.11)$$

$$f_i^{(k)}(T^{(k)}-) = 0 \quad \forall i \in S_-^{(k)} \cup \left(\tilde{S}_0^{(k)} \cap \tilde{S}_+^{(k)} \right) \quad (3.12)$$

$$f_i^{(k+1)}(T^{(k)}+) = 0 \quad \forall i \in \left(\tilde{S}_0^{(k)} \cap \tilde{S}_-^{(k)} \right) \cup S_+^{(k+1)} \quad (3.13)$$

$$\left(\sum_{k=1}^K \int_{T^{(k-1)}+}^{T^{(k)}-} f^{(k)}(x) dx + \sum_{k=0}^{K-1} c^{(k)} \right) \mathbf{1} = 1 \quad (3.14)$$

where $\mathbf{1}$ is the column vector with appropriate size. The row vectors of steady-state joint probability distribution (3.2) and the probability mass accumulation (3.4) can be calculated using the boundary conditions given in (3.7)-(3.14) by a matrix-analytical method as proposed in [16]. A linear matrix equation with at most size $N(2K+1)$ needs to be solved to obtain the unknowns. Since the linear

matrix equation is in block tridiagonal form, the computational complexity can be brought down to $\mathcal{O}(N^3K)$ [37]. The constructed system model in next section uses the notations given in this section.



Chapter 4

System Modeling

In this section, the system model will be explained. Then, the related MRMFQ model will be presented. Lastly, performance measures of interest will be obtained.

4.1 Model Description

The system is assumed to be a single server FIFO queue where packet arrival process is Poisson with rate λ packets/sec. A set of transmission profiles indexed by $k = 1, 2, \dots, K$ are assumed to be available to serve a packet. Each transmission profile $k, 1 \leq k \leq K$, is characterized by two properties: service rate and transmit power, denoted by μ_k in packets/sec and P_k in Watts respectively with $\mu_i < \mu_j$ and $P_i < P_j$ when $i < j$ without loss of generality. It is assumed that the profile to be used for any packet transmission is fixed during the service time of that particular packet, i.e, the service profile of any packet can be adjusted only at service start times. Let $D(t)$ denote the delay already experienced by HoL packet at the service start time t . The transmission profile selection for that particular packet is made at time t . The total of K thresholds are defined as $0 = T^{(0)} < T^{(1)} < \dots < T^{(K-1)} < T^{(K)} = \infty$ such that the profile with index

k is selected for the transmission of HoL packet when $T^{(k-1)} \leq D(t) < T^{(k)}$. The motivation behind this selection method is that the packets that experience longer queuing delays should be served with higher service rates in order to meet a statistical delay constraint. Any transmission profile selection policy proposed in next section will be studied within this general system framework.

Let $U(t)$ denote the overall time that HoL packet spent in the system including its transmission time. Note that $U(t) = 0$ when there is no packet in the system at time t . The virtual waiting time $A(t)$ denotes the required time to serve all packets in the system including the transmission time of the last packet. It is obvious that a packet that arrives to the system at time t will be eventually served with the transmission profile k if $T^{(k-1)} \leq A(t) < T^{(k)}$.

Consider an example setup of two thresholds $T^{(1)} = 4$ and $T^{(2)} = 8$ with transmission times 6, 4, and 2 for the regimes 1, 2, and 3, respectively. Suppose the packets arrivals occur at $t = 0, 2, 10, 14$. Resulting sample paths of two processes $A(t)$ and $U(t)$ are given in figures 4.1a and 4.1b, respectively. Note that transmission times are set to those values (6, 4, and 2) for the sake of simplicity.

4.2 MRMFQ Model

A process with only finite drifts can be modeled as a Markov fluid queue. Therefore, an auxiliary process $X(t)$ is also defined, depicted in Fig. 4.1c, in which a drift of -1 replaces the the abrupt downwards jumps in $U(t)$.

Since the sample path of $X(t)$ has only finite drifts, it can be modeled by a modulated process of a MRMFQ with K regimes and $K + 1$ states. The steady-state distribution of the fluid process $(X(t), Z(t))$ can be used to obtain the steady-state distribution of $U(t)$ by censoring out the states that have negative drifts where $Z(t)$ is the background process of the corresponding fluid model. Therefore, the MRFMQ model of $X(t)$ will be described first. Service state,

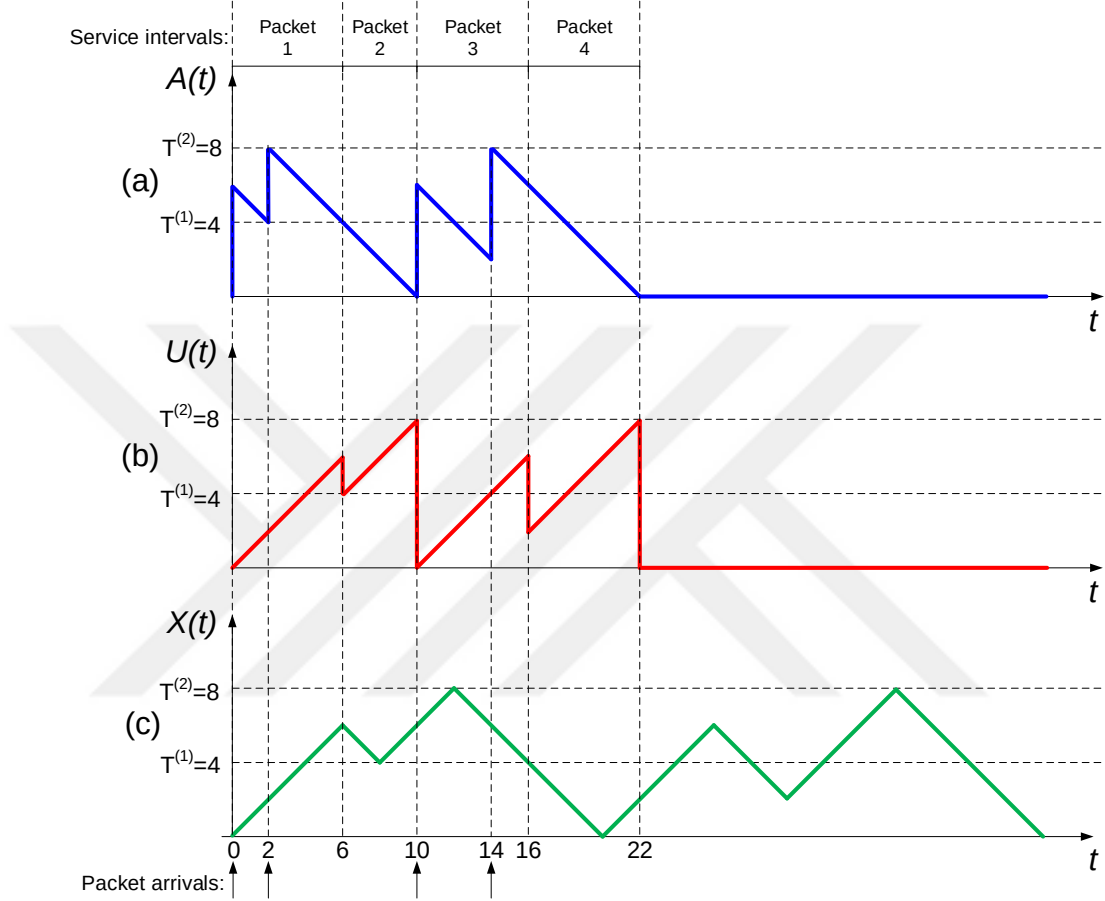


Figure 4.1: Sample paths of three processes $A(t)$, $U(t)$, and $X(t)$.

denoted by I_k for $k = 1, \dots, K$, is defined during which $X(t)$ is increased by a drift of 1 and transmission profile k is used for serving packets. Let $\mathcal{D}$ denote another state to which the system transits into after completing the transmission time of a packet, i.e, finishing transmission of any packet in state I_k . Note that the delay of the new HoL packet has to be reduced by an amount corresponding to its inter-arrival time. Therefore, $X(t)$ is decreased by a drift of -1 in state $\mathcal{D}$ for a random amount of time that follows an exponential distribution with mean $1/\lambda$. Then, the system transits into state I_k if $T^{(k-1)} \leq X(t) < T^{(k)}$ and so on. Moreover, the transmission profile $k = 1$ is selected if a packet arrival occurs when the queue is empty, i.e, $X(t) = 0$. Therefore, only transition that may occur from out of the state $\mathcal{D}$ is into the state I_1 . From these definitions, it is clear that the background process, denoted by $Z(t)$, has a total of $K + 1$ states with I_k for

$i = 1, \dots, K$ and state $\mathcal{D}$. Possible state transitions of $Z(t)$ are demonstrated in Fig. 4.2. Additionally, the infinitesimal generator matrix of regime j , denoted by $Q^{(j)}$, for $j = 1, \dots, K$, can be constructed as follows:

$$\begin{array}{c}
 \\
 I_K \\
 \vdots \\
 I_{j+1} \\
 I_j \\
 I_{j-1} \\
 \vdots \\
 I_1 \\
 \mathcal{D}
 \end{array}
 \begin{bmatrix}
 I_K & \cdots & I_{j+1} & I_j & I_{j-1} & \cdots & I_1 & \mathcal{D} \\
 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
 \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\
 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
 0 & \cdots & 0 & -\mu_j & 0 & \cdots & 0 & \mu_j \\
 0 & \cdots & 0 & 0 & -\mu_{j-1} & \cdots & 0 & \mu_{j-1} \\
 \vdots & & \vdots & \vdots & & \ddots & \vdots & \vdots \\
 0 & \cdots & 0 & 0 & 0 & \cdots & -\mu_1 & \mu_1 \\
 0 & \cdots & 0 & \lambda & 0 & \cdots & 0 & -\lambda
 \end{bmatrix}. \quad (4.1)$$

It is clear that since $X(t)$ increases by a drift of 1 in service state I_k , there may be state transitions from I_k to state $\mathcal{D}$ in regime j for $k \leq j$. The generator at the boundary $j = 1, \dots, K$ is set as follows:

$$\tilde{Q}^{(j)} = Q^{(j+1)}, \quad j = 1, \dots, K. \quad (4.2)$$

Note that the equality $\tilde{Q}^{(0)} = Q^{(1)}$ also holds, however the only transition that occurs at boundary 0 is from state $\mathcal{D}$ to state I_1 . Lastly, drift matrices at regime k , denoted by $R^{(k)}$, are expressed as follows:

$$R^{(k)} = \mathbf{diag}(\mathbf{I}, -1), \quad 1 \leq k \leq K, \quad (4.3)$$

where $\mathbf{I}$ is an identity matrix of appropriate size. Similarly, drift matrices at boundary k , denoted by $\tilde{R}^{(k)}$, are given as follows:

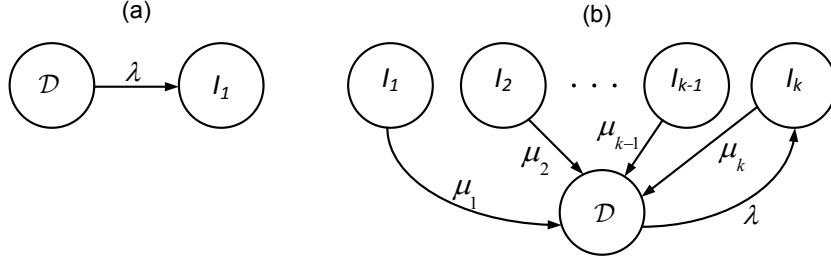


Figure 4.2: State transitions (a) for $X(t) = 0$ and (b) for regime k , $k = 1, \dots, K$.

$$\tilde{R}^{(k)} = \begin{cases} R^{(k+1)}, & 1 \leq k < K, \\ \mathbf{max}(0, R^{(1)}), & k = 0, \end{cases} \quad (4.4)$$

where $\mathbf{max}$ is the element-wise operator. With these definitions, establishment of the related MRFMQ model is finalized.

4.3 Performance Measures

The performance metrics of interest in this thesis are the queuing delay distribution and the average performance consumption, which will be obtained in this section. Note that these metrics can be obtained using steady-state distributions of state $\mathcal{D}$ since the waiting time of a virtual packet arrival will be determined by the process $A(t)$ as a direct consequence of *PASTA* (Poisson Arrivals See Time Averages) property [38]. The steady-state distribution of $A(t)$ can be obtained from the fluid process $(X(t), Z(t))$ by censoring out all the states $I_k, k = 1, \dots, K$ as follows:

$$\lim_{t \rightarrow \infty} \Pr\{A(t) \leq x\} = \lim_{t \rightarrow \infty} \frac{\Pr\{Z(t) = \mathcal{D}, X(t) \leq x\}}{\Pr\{Z(t) = \mathcal{D}\}}. \quad (4.5)$$

The probability that a packet arrival occurs when the queue is empty, denoted by p_0 is expressed as:

$$p_0 = \lim_{t \rightarrow \infty} \Pr\{A(t) = 0\}. \quad (4.6)$$

Also the probability that the transmission profile k is selected for serving the HoL packet, denoted by q_k for $k = 1, \dots, K$, is written as:

$$q_k = \lim_{t \rightarrow \infty} \Pr\{T^{(k-1)} \leq A(t) < T^{(k)}\}, 1 \leq k \leq K. \quad (4.7)$$

Using these definitions, the average power consumption P can be calculated as follows:

$$P = p_0 P_I + \frac{(1 - p_0)}{\sum_{k=1}^K \frac{q_k}{\mu_k}} \sum_{k=1}^K \frac{q_k P_k}{\mu_k}, \quad (4.8)$$

where P_I is the power consumption at the transmitter when idle, i.e, no packet transmission is ongoing. The cumulative distribution of the process $D(t)$ is expressed as below which is equal to the cumulative distribution of $A(t)$ from the PASTA property.

$$F_D(x) = \lim_{t \rightarrow \infty} \Pr\{D(t) \leq x\} = \lim_{t \rightarrow \infty} \Pr\{A(t) \leq x\}. \quad (4.9)$$

Since the constructed MRMFQ model consists of K regimes and $K + 1$ states, the algorithm to obtain the performance metrics of the proposed policies has a computational complexity of only $O(K^4)$ [37]. Therefore, a transmission profile selection policy with numerous profiles can be evaluated very quickly using the established MRMFQ model.

Chapter 5

Transmission Profile Selection Policies

Let $\mathcal{N} = \{U_1, U_2, \dots, U_N\}$ denote a given transmission profile set in a wireless link. Each profile in this set is represented by the power and service rate property pair denoted as $(P^{(i)}, \mu^{(i)})$ with $\mu^{(i)} > \mu^{(j)}$ for $i > j$, respectively. Given the available profile set $\mathcal{N}$, various profile selection policies can be considered to derive the transmission profile set $\mathcal{K}$ for the MRMFQ model. The motivation of any policy should be such that HoL packets with higher $D(t)$ should be transmitted with higher rate and transmit power in order to meet a statistical delay constraint. On the other hand, different approaches can be considered to determine which profiles in the set $\mathcal{N}$ should be used or not. It may be desirable to eliminate some of the profiles by means of obtaining more energy saving or reduced implementation complexity.

The delay violation probability is defined as follows:

$$p_v = \lim_{t \rightarrow \infty} \Pr\{D(t) > D_0\}, \quad (5.1)$$

for a given delay bound D_0 . A transmission profile selection policy needs to

be tuned to satisfy the statistical delay constraint which is given as

$$p_v < \varepsilon, \quad (5.2)$$

for a tolerance parameter ε . Any policy proposed in this study aims to maximize the energy-efficiency under the statistical delay constraint given in (5.2). Selection of profile subset $\mathcal{K} = \{1, 2, \dots, K\} \subseteq \mathcal{N}$ and the thresholds $T^{(1)}, \dots, T^{(K-1)}$ will be defined for each policy such that if the queuing delay $D(t)$ experienced by the HoL packet is within $[T^{(k-1)}, T^{(k)})$, the profile $k \in \mathcal{K}$ with service rate μ_k and power P_k will be used for the transmission of that particular packet.

5.1 Min-Delay Policy (MDP)

The aim of min-delay policy is to minimize the average delay by choosing the single profile with the largest transmission power and service rate for transmission of all packets. Note that this selection is independent of the queuing delay $D(t)$ experienced by the packets. Therefore, in *MDP*, $K = 1$ with

$$(P_1, \mu_1) = (P^{(N)}, \mu^{(N)}). \quad (5.3)$$

Service rate selection for HoL packet in *MDP* is shown in Fig. 5.1. Use of a single regime enables us to obtain the performance metrics of *MDP* using the conventional M/M/1 analysis. Therefore, we can express the average power consumption of *MDP*, denoted by P_{MDP} as follows:

$$P_{MDP} = (1 - \rho)P_I + \rho P^{(N)} \quad (5.4)$$

where $\rho = \lambda/\mu^{(N)} < 1$. *MDP* minimizes the average delay of the system without any concern for saving energy. Therefore, it can be considered as a baseline policy such that the energy gain performance of the other proposed policies will be evaluated relative to *MDP*. In the numerical examples, only the values of

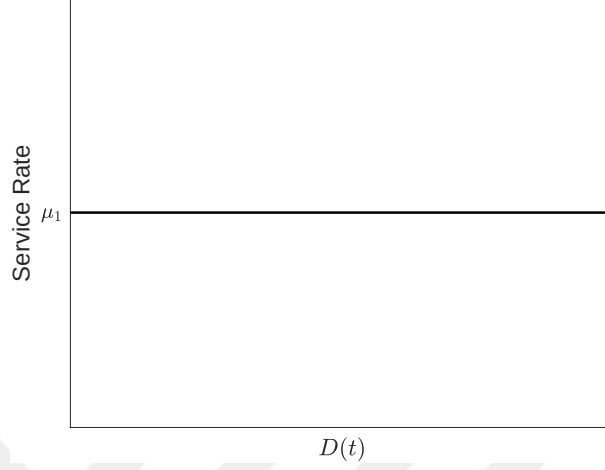


Figure 5.1: Service rate selection for the HoL packet in MDP .

the arrival rate $\lambda < \lambda_M$ will be considered such that MDP satisfies the delay constraint where λ_M is the particular value of λ for which the delay violation probability for MDP equals ϵ , i.e., $p_v = \rho e^{-(\mu^{(N)} - \lambda_M)D_0} = \epsilon$ [38].

5.2 Single Threshold Policy (STP)

STP is a binary rate adjustment policy such that when $D(t)$ is below a single threshold value denoted by T_{STP} , the profile with minimum possible rate is used for HoL packet. When $D(t)$ is above T_{STP} , the maximum possible rate is selected. Therefore, $K = 2$ in STP . Transmission profiles for STP can be represented as:

$$(P_k, \mu_k) = \begin{cases} (P^{(1)}, \mu^{(1)}), & k = 1, \\ (P^{(N)}, \mu^{(N)}), & k = 2. \end{cases} \quad (5.5)$$

Additionally, boundaries for STP apply as follows:

$$T^{(k)} = \begin{cases} 0, & k = 0, \\ T_{STP}, & k = 1, \\ \infty, & k = 2. \end{cases} \quad (5.6)$$

Service rate selection for HoL packet in STP is shown in Fig. 5.2. The MRMFQ model with two regimes and three states can be used to obtain the performance metrics of STP . Optimum value of T_{STP} , denoted by T_{STP^*} , minimizes the average power consumption while satisfying the delay constraint. Employing the particular threshold value T_{STP^*} represents the optimum STP which is denoted by STP^* .

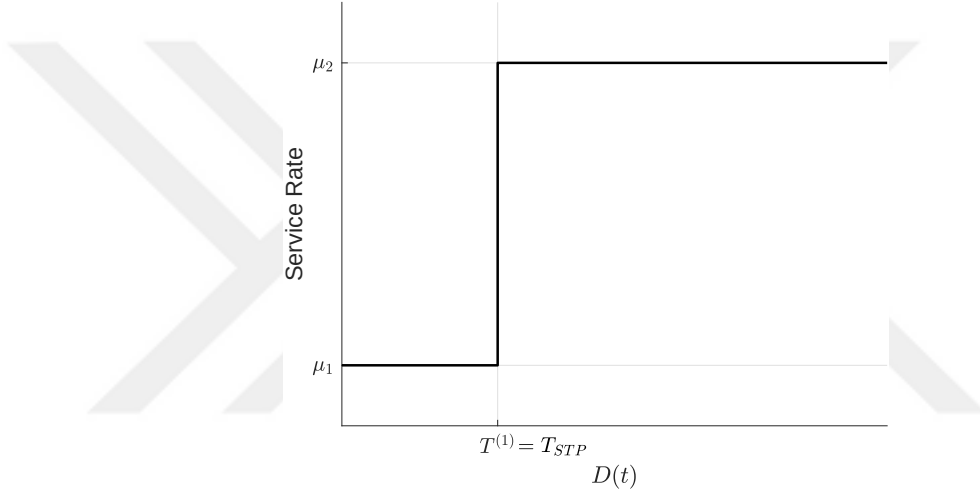


Figure 5.2: Service rate selection for the HoL packet in STP .

5.3 Proportional Control Policy (PCP)

PCP aims to adjust the service rate of the HoL packets by making the profile selection among all of the available profiles in the set $\mathcal{N}$. Therefore, the profile subset $\mathcal{K}$ equals to the universal set $\mathcal{N}$ ($K = N$). Transmission profiles used in PCP can be represented as:

$$(P_k, \mu_k) = (P^{(k)}, \mu^{(k)}), \quad k = 1, 2, \dots, N. \quad (5.7)$$

In PCP , service rate of HoL packet will be selected from the set $\{\mu_1, \mu_2, \dots, \mu_K\}$ such that μ_i will be linearly proportional to $D(t)$ when $0 < D(t) < T_{PCP}$ as shown in Fig. 5.3.

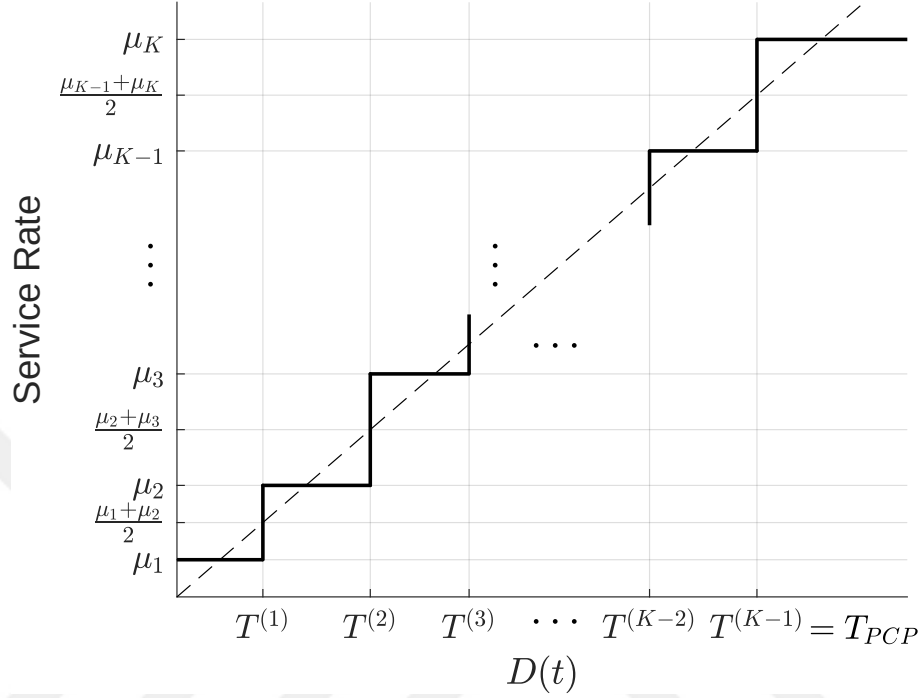


Figure 5.3: Service rate selection for the HoL packet in *PCP*.

The transmission profile with rate μ_1 will be used if the HoL packet has arrived when the queue is empty, i.e. $D(t) = 0$. Lastly, the transmission profile with rate μ_K will be employed if the delay experienced by the HoL packet is greater than the performance parameter of *PCP* which is denoted as T_{PCP} . Mathematically, boundaries of *PCP* can be represented as follows:

$$T^{(k)} = \begin{cases} 0, & k = 0, \\ \frac{T_{PCP}(\mu_k + \mu_{k+1})}{\mu_{K-1} + \mu_K}, & 0 < k < K - 1, \\ T_{PCP}, & k = K - 1, \\ \infty, & k = K. \end{cases} \quad (5.8)$$

MRMFQ model with $N + 1$ states and N regimes can be used to obtain the performance metrics of *PCP*. The particular policy *PCP** with the performance parameter T_{PCP^*} denotes the optimum *PCP* which minimizes the average power consumption while satisfying the delay constraint.

5.4 Energy-efficient PCP (EPCP)

Even though *STP* and *PCP* adjust the service rate for the purpose of saving energy, they do not evaluate the energy efficiency of the transmission profiles in $\mathcal{N}$ while constructing the profile set $\mathcal{K}$. Some profiles in $\mathcal{N}$ may be not as efficient as others in terms of energy. Let the profiles U_f , U_h and U_j are available in $\mathcal{N}$ for the rate adjustment given that $f < h < j$. Suppose the transmission profile U_f is being used for a given time. The strategy of switching to both U_h and U_j results in more service rate and requirement of transmission power since the inequality $f < h < j$ holds. On the other hand, the increase in transmission power per bit may be less for the strategy of switching to U_j than U_h . In that case, it may be more efficient in terms of energy to switch U_j instead of U_h for the purpose of transmitting packets more aggressively. Same idea can also be applied other way around. If the profile U_j is being used for a given time, the strategy of switching to U_f instead of U_h may result in better energy saving performance. Using this motivation, the profile U_h is identified as a relatively energy-inefficient profile if there exists any pair (f, j) such that the following inequality holds:

$$\frac{P^{(h)} - P^{(f)}}{\mu^{(h)} - \mu^{(f)}} > \frac{P^{(j)} - P^{(f)}}{\mu^{(j)} - \mu^{(f)}}. \quad (5.9)$$

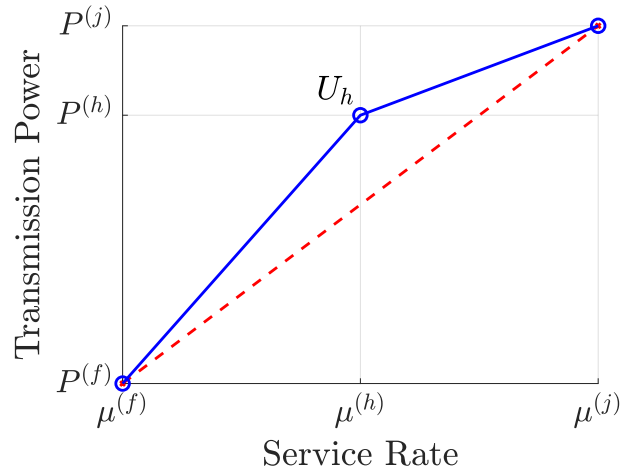


Figure 5.4: An example of a relatively energy-inefficient profile.

An example of a relatively energy-inefficient profile U_h is shown in Fig. 5.4.

EPCP eliminates all of the relatively energy-inefficient profiles from set $\mathcal{N}$. For laying out the motivation behind the idea, consider an example profile set $\mathcal{N} = \{(1, 1), (2, 3), (6, 4), (7, 7), (11, 10), (13, 14)\}$ where service rate and power properties of the profiles are set to those values for the sake of simplicity. This set is depicted in Fig. 5.5. It is clear that the inequality (5.9) holds for $h = 2$, $f = 1$ and $j = 3$ from which the profile U_2 is found to be relatively energy-inefficient. The resulting set obtained by elimination of U_2 is depicted in Fig. 5.6(a). It is clear that the area under the curve is less than it is for the case with no elimination, i.e, the profile set for *PCP*. Moreover, the inequality (5.9) also holds for $h = 1$, $f = 4$ and $j = 5$ which yields U_4 as an another relatively energy-inefficient profile. The area under the curve is further reduced by elimination of U_4 along with U_2 , which is depicted in Fig. 5.6(b).

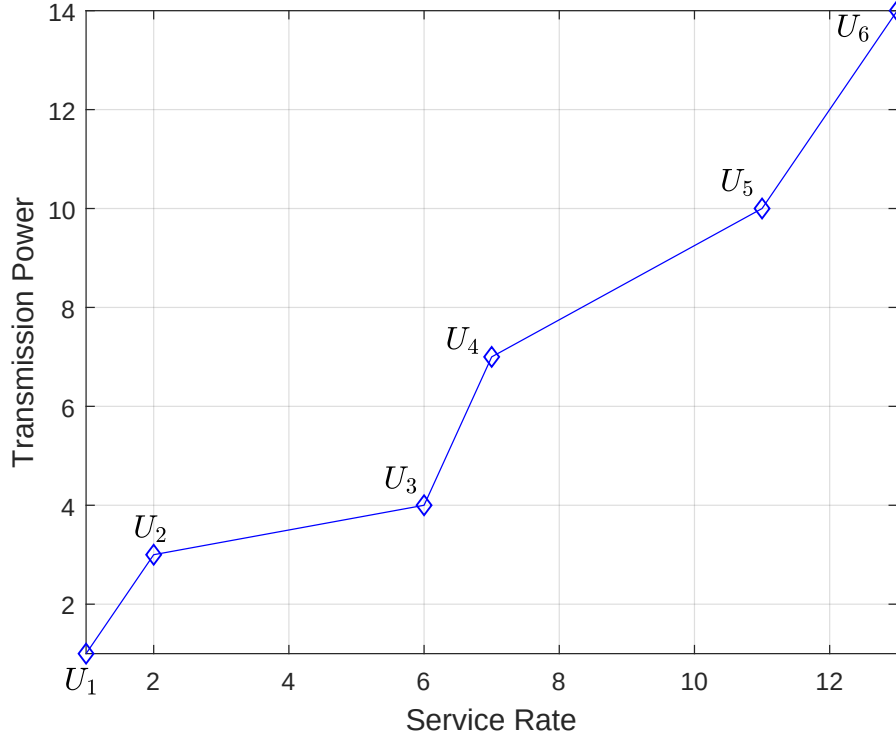
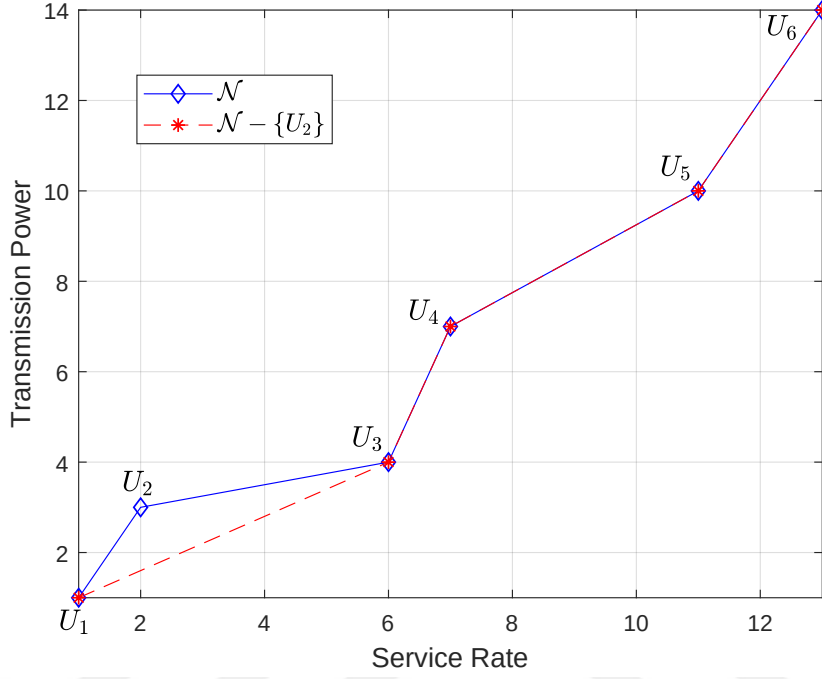
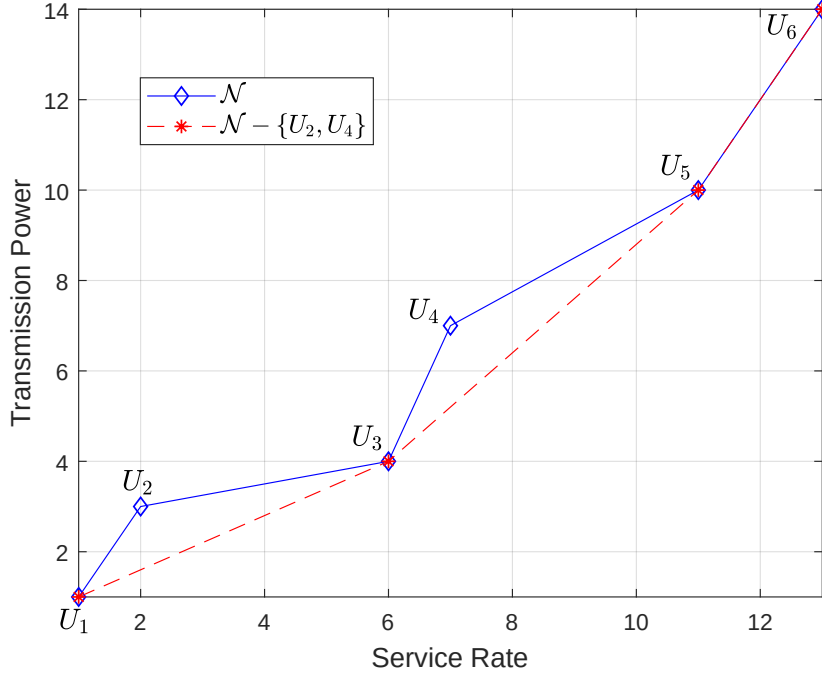


Figure 5.5: An example of a universal profile set $\mathcal{N}$.

Elimination of each relatively energy-inefficient profile results in less area under the given curve. Therefore, we obtain the smallest area when all of the relatively energy-inefficient profiles are eliminated from set $\mathcal{N}$. For this purpose, an energy inefficiency index for a given profile subset $\mathcal{L} \subseteq \mathcal{N}$ denoted by $\Gamma(\mathcal{L})$, is defined



(a) The profile set obtained by elimination of the relatively energy-inefficient profile U_2 .



(b) The profile set obtained by elimination of the relatively energy-inefficient profiles U_2 and U_4 .

Figure 5.6: Elimination of the relatively energy-inefficient profiles from an example universal set $\mathcal{N} = \{(1, 1), (2, 3), (6, 4), (7, 7), (11, 10), (13, 14)\}$.

as:

$$\Gamma(\mathcal{L}) = \sum_{l=2}^L (P^{(l)} + P^{(l-1)})(\mu^{(l)} - \mu^{(l-1)}) \quad (5.10)$$

which is a measure of the area under the given curve. The profile subset $\mathcal{V}$ of size V , which minimizes the energy inefficiency metric, is obtained over all possible subsets containing the profiles U_1 and U_N as follows:

$$\mathcal{V} = \arg \min_{\{U_1, U_N\} \subseteq \mathcal{L} \subseteq \mathcal{N}} \Gamma(\mathcal{L}). \quad (5.11)$$

Lastly, PCP is applied to the particular subset $\mathcal{V}$ which concludes the methodology of EPCP. Similar to *STP* and *PCP*, the particular policy *EPCP*^{*} with the performance parameter T_{EPCP^*} denotes the optimum *EPCP* which minimizes the average power consumption while satisfying the delay constraint.

It is worthwhile to note that all of the proposed policies so far behave exactly same as each other when there is only one available transmission profile in the universal profile set $\mathcal{N}$, i.e, $N = 1$. Additionally, *STP*, *PCP* and *EPCP* behave exactly same as each other when $N = 2$. Moreover, *PCP* and *EPCP* behave identically if there is no relatively energy-inefficient profile in $\mathcal{N}$, i.e, $\mathcal{V} = \mathcal{N}$.

5.5 Extended Policies

The profile to be used in the first regime is always fixed to the minimum service rate profile U_1 for all three proposed policies $p \in \{STP, PCP, EPCP\}$ which will be referred as basic policies hereafter. Each of the basic policies employs the profile U_1 for any packet with $D(t) = 0$ even when the performance parameter T_p is an arbitrarily very small value such as 0.0001. On the other hand, *MDP* employs the maximum service rate profile U_N when $D(t) = 0$. Depending on the difference in the service rate between U_1 and U_N , this may create a significant gap

in the delay performance between MDP and basic policies as will be shown in the numerical examples. In order to avoid this catch, the choice of the first profile can be relaxed, i.e, becomes another performance parameter along with the T_p . This additional performance parameter gives rise to the extended versions of the three basic policies, denoted by $p \in \{STP_e, PCP_e, EPCP_e\}$, respectively.

The methodology described for the each of the three basic policies in the previous section is applied to all the subsets $\{U_m, U_{m+1}, \dots, U_N\} \subseteq \mathcal{N}, m \in \{1, 2, \dots, N-1\}$ indexed by m as the starting profile set. Using two dimensional exhaustive search, the optimum performance parameter pair, denoted by (m_{p^*}, T_{p^*}) , is obtained which minimizes the average power consumption while satisfying the delay constraint. The resulting optimum extended policies are named as STP_e^*, PCP_e^* , and $EPCP_e^*$, respectively.

A percentage energy gain relative to MDP , denoted by G_p , is defined for each of the six proposed policies $p \in \{STP, PCP, EPCP, STP_e, PCP_e, EPCP_e\}$ as follows:

$$G_p = 100 \frac{(P_{MDP} - P_p)}{P_{MDP}}. \quad (5.12)$$

Suppose the average power consumption obtained by employing MDP is relatively low for a given scenario. In that case, even if the percentage energy gain G_p is high, the absolute power savings will also be relatively low. In order to analyze these scenarios, the absolute power savings in Watts at the output of antenna element, denoted by P_{saved} , is also defined for each of the optimum policies as follows:

$$P_{saved} = P_{MDP} - P_p \quad (5.13)$$

where $p \in \{STP^*, PCP^*, EPCP^*, STP_e^*, PCP_e^*, EPCP_e^*\}$. Note that overall absolute saved power in the transmitter depends on several factors such as power supply efficiencies, amplifier efficiency and feeder loss caused by the distance

between the amplifier and the antenna [6]. For the case of BSs, the overall absolute saved power will be proportional to P_{saved} by a factor of Δ_p times the number of antenna pairs per site.



Chapter 6

Numerical Examples

In this section, the proposed transmission profile selection policies will be evaluated in Physical Downlink Shared Channel (PDSCH) of the LTE. In [2], authors conducted physical layer simulations for PDSCH which will be used in this analysis. The construction of $\mathcal{K} \subseteq \mathcal{N}$ will be shown for the basic policies. Then, frame level simulations will be performed to validate the analytical model. Finally, energy gain and absolute saved power performance of the proposed policies relative to the baseline policy MDP will be evaluated for a wide range of system parameters using the analytical model.

6.1 System Setup

Packet sizes are exponentially distributed with mean $\beta = 500$ bytes and packet arrivals follow a Poisson process with rate λ . Power consumption is considered to be 0 in the transmitter when it is idle, i.e., $P_I = 0$. MIMO configuration is assumed to be 2x2 spatial multiplexing. Extended Pedestrian A model with Doppler frequency of 5 Hz (EPA5) is considered as the multipath fading model. A perfect channel estimator is assumed as employed in [2]. LTE-TDD frame structure is considered as shown in Fig. 6.1 where each Physical Resource Block

(PRB) consists of 12 sub-carriers with 15 kHz carrier spacing [2]. Number of PRBs allocated to the wireless link of interest within a sub-frame duration of 1 ms is assumed to be fixed to $N_B = 50$. For other parameters of the physical layer simulation setup, the study [2] can be referred.

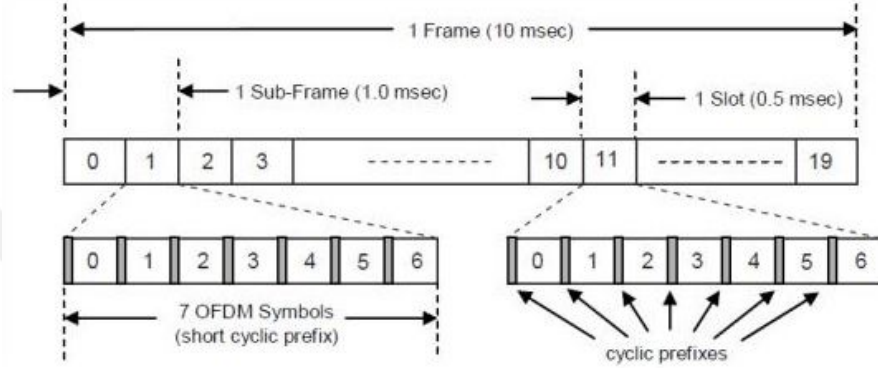


Figure 6.1: LTE-TDD frame structure

For a given average Signal-to-Noise ratio (SNR) at the receiver, denoted by α in dB scale, the block error rate (BLER) and the throughput per PRB in bits are denoted as $e(\alpha, I_M)$ and $r(\alpha, I_M)$ respectively, where I_M is the modulation and coding scheme index. The I_M value which maximizes the throughput while satisfying a target BLER (e_b) is considered to be optimal I_M value, denoted by I_M^* , which is obtained as follows:

$$\tau(\alpha, I_M) = \begin{cases} r(\alpha, I_M), & e(\alpha, I_M) \leq e_b, \\ 0, & \text{otherwise,} \end{cases} \quad (6.1)$$

$$I_M^* = \arg \max_{I_M} \tau(\alpha, I_M). \quad (6.2)$$

Note that when the expression given in (6.1) equals to 0 for a given α , it indicates that there exists no MCS profile that satisfies the target BLER for that particular SNR. The optimal throughput $r^*(\alpha)$ in bits per PRB for a given α is obtained by using the particular MCS index I_M^* :

$$r^*(\alpha) = r(\alpha, I_M^*). \quad (6.3)$$

Physical layer simulation results of the performance metrics $e(\alpha, I_M)$ and $r(\alpha, I_M)$ are tabulated as function of the MCS index I_M and the average SNR at the receiver α (in 1 dB granularity) in Table 6.1 and 6.2 respectively, which are taken from the study [2]. The analytical model does not consider the effect of HARQ retransmissions on the service time distributions of packets. Therefore, a relatively low value of BLER constraint $e_b = 0.02$ is assumed. For this BLER constraint, obtained $r^*(\alpha)$ values using the selection rule (6.1) are given in bold format in Table 6.2. Note that when $\alpha = 1$ db, there exists no MCS that satisfies the target BLER, i.e, $\tau(1, I_M) = 0$ for any I_M .

Table 6.1: BLER $e(\alpha, I_M)$ as a function of the SNR α and MCS index I_M .

α/I_M	0	1	2	3	4	5	6	7	8	9	10	11
1	0.043	0.098	0.154	0.281	0.409	0.536	0.662	0.772	0.881	0.94	1	1
2	0.014	0.056	0.098	0.199	0.3	0.426	0.553	0.678	0.803	0.901	1	1
3	0.005	0.032	0.06	0.133	0.206	0.325	0.444	0.575	0.707	0.831	0.956	0.999
4	0	0.015	0.03	0.081	0.131	0.237	0.343	0.474	0.604	0.735	0.866	0.936
5	0	0.006	0.012	0.044	0.076	0.166	0.256	0.381	0.507	0.635	0.762	0.85
6	0	0	0	0.02	0.04	0.112	0.185	0.299	0.414	0.539	0.663	0.764
7	0	0	0	0.006	0.012	0.07	0.128	0.228	0.328	0.446	0.564	0.671
8	0	0	0	0.001	0.002	0.044	0.086	0.169	0.253	0.363	0.473	0.585
9	0	0	0	0	0.001	0.028	0.055	0.123	0.19	0.289	0.389	0.503
10	0	0	0	0	0	0.014	0.029	0.083	0.136	0.225	0.313	0.428
11	0	0	0	0	0	0.007	0.014	0.054	0.095	0.171	0.247	0.364
12	0	0	0	0	0	0	0	0.033	0.066	0.127	0.188	0.305
13	0	0	0	0	0	0	0	0.023	0.047	0.093	0.14	0.247
14	0	0	0	0	0	0	0	0.017	0.034	0.067	0.1	0.194
15	0	0	0	0	0	0	0	0.012	0.025	0.048	0.071	0.15
16	0	0	0	0	0	0	0	0.009	0.018	0.034	0.051	0.111
17	0	0	0	0	0	0	0	0.006	0.013	0.024	0.036	0.079
18	0	0	0	0	0	0	0	0.004	0.008	0.017	0.025	0.056
19	0	0	0	0	0	0	0	0.002	0.004	0.011	0.018	0.042
20	0	0	0	0	0	0	0	0	0.001	0.007	0.013	0.032

For a given receiver sensitivity and channel condition, α is a function of transmit power, which makes it possible to adjust α by varying the transmit power. Upper limit of the transmit power is constrained by a maximum limit, denoted by P_{max} . We define the maximum attainable average SNR, denoted by α_m , as the α obtained when the transmit power is P_{max} . One can further reduce the transmit power at the expense of reduced throughput as long as the target BLER is satisfied. Mathematically, there must exist at least one I_M value that satisfies the target BLER at any particular transmit power. In line with the simulation results of the study [2] which are presented in 1 dB granularity, the transmission power is reduced by 1 dbm at each step while obtaining the universal profile set $\mathcal{N}$. Therefore, each value of the $\alpha \in \{2, 3, \dots, \alpha_m\}$ yields a different transmission

Table 6.2: The throughput $r(\alpha, I_M)$ in bits/PRB as a function of the SNR α and MCS index I_M .

α/I_M	0	1	2	3	4	5	6	7	8	9	10	11
1	52.98	64.88	74.92	82.03	85.64	81.5	69.58	56.52	33.1	18.98	0	0
2	54.55	67.94	79.94	91.49	101.45	100.69	92.15	79.69	54.69	31.36	0	0
3	55.07	69.62	83.25	98.95	114.98	118.46	114.64	105.19	81.64	53.79	13.94	7.81
4	55.31	70.88	85.96	104.98	125.84	133.89	135.44	130.37	110.15	84.56	42.77	34.54
5	55.34	71.54	87.55	109.16	133.86	146.41	153.42	153.28	137.36	116.66	75.77	67.76
6	55.36	71.99	88.62	111.93	139.12	155.89	168.21	173.63	163.11	147.28	107.48	100.23
7	55.36	72	88.63	113.55	143.22	163.3	179.8	191.31	187.22	177.03	139.33	133.96
8	55.36	72	88.64	114.13	144.67	167.91	188.57	205.86	208.08	203.51	168.37	164.83
9	55.36	72	88.64	114.18	144.81	170.68	194.85	217.43	225.61	227.03	195.31	194.04
10	55.36	72	88.64	114.24	144.96	173.05	200.22	227.33	240.6	247.7	219.44	220.35
11	55.36	72	88.64	114.24	144.96	174.43	203.46	234.4	252.11	264.93	240.69	243.24
12	55.36	72	88.64	114.24	144.96	175.63	206.29	239.68	260.18	278.85	259.28	263.77
13	55.36	72	88.64	114.24	144.96	175.66	206.34	242.11	265.56	289.76	274.92	282.44
14	55.36	72	88.64	114.24	144.96	175.68	206.4	243.77	269.22	298.17	287.56	298.64
15	55.36	72	88.64	114.24	144.96	175.68	206.4	244.86	271.67	304.17	296.75	311.52
16	55.36	72	88.64	114.24	144.96	175.68	206.4	245.7	273.55	308.54	303.33	321.96
17	55.36	72	88.64	114.24	144.96	175.68	206.4	246.37	275.05	311.77	308.07	330.17
18	55.36	72	88.64	114.24	144.96	175.68	206.4	246.92	276.29	314.16	311.42	335.95
19	55.36	72	88.64	114.24	144.96	175.68	206.4	247.43	277.43	315.9	313.61	339.63
20	55.36	72	88.64	114.24	144.96	175.68	206.4	247.83	278.35	317.36	315.47	342.43

profile. Service rate attribute of each profile can be represented as:

$$\mu(\alpha) = \frac{1000r^*(\alpha)N_B}{8\beta}. \quad (6.4)$$

Note that units of $\mu(\alpha)$ is in packets/sec; whereas units of $r^*(\alpha)$ is in bits per PRB. The equation (6.4) is the conversion from bits per PRB to packets/sec between these parameters. Next, transmission power (in dBm) attribute of each profile can be represented as follows:

$$P(\alpha) = P_{max} - (\alpha_m - \alpha). \quad (6.5)$$

In this study, P_{max} is assumed to be 46 dBm which is a typical value for PDSCH. For a specific scenario of $\alpha_m = 12$ dB and $e_b = 0.02$, the transmission power and the service rate attributes of the constructed profiles in $\mathcal{N}$ are listed in Table 6.3. Using the set $\mathcal{N}$, obtained transmission profiles for each of the basic policies are shown in Fig. 6.2. Note that *STP* uses only two profiles with the highest and lowest possible service rates whereas *PCP* uses all of the available profile from the set $\mathcal{N}$. *EPCP*, on the other hand, eliminates all of the relatively energy-inefficient profiles from the set $\mathcal{N}$, i.e, minimizes the area under the curve in Fig. 6.2.

Table 6.3: Profiles in the set $\mathcal{N}$ for the case $\alpha_m = 12$ dB and $e_b = 0.02$.

i	$P^{(i)}$	$\mu^{(i)}$	i	$P^{(i)}$	$\mu^{(i)}$
1	3.98	681.9	7	15.84	1808.4
2	5.01	688.4	8	19.95	1810.2
3	6.30	886	9	25.11	2163.1
4	7.94	1094.3	10	31.62	2543.3
5	10	1107.8	11	39.81	2578.6
6	12.58	1790.2			

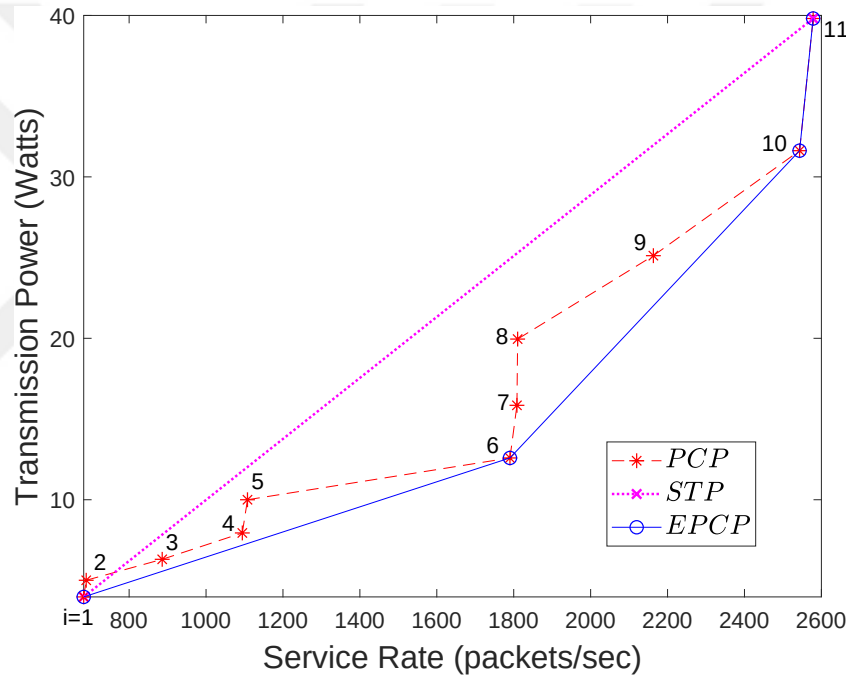


Figure 6.2: Service rate and transmission power attributes of the profiles used for *STP*, *PCP*, and *EPCP*, when $\alpha_m = 12$ dB and $e_b = 0.02$.

6.2 Model Validation

Next, the analytical model is validated by conducting frame level simulations for PDSCH. LTE sub-frame structure is considered such that the required byte paddings are performed to align the size of payload to the nearest Transport Block size given in [39] for $N_B = 50$. Total number of packet arrivals is considered to be 10^7 for each simulation. A specific scenario of $D_0 = 15$ ms, $\alpha_m = 12$ dB, and $\lambda = 1000$ is studied in which the performance metrics of the particular policy *PCP* with respect to its performance parameter T_{PCP} are obtained and depicted

in Fig. 6.3 and Fig. 6.4 respectively for both analytical model and simulations. It can be concluded that the simulations are in line with the analytical results. The average power consumption appears to be slightly higher in simulations for relatively small values of the threshold parameter T_{PCP} as seen in Fig. 6.4. This is the result of using higher order MCS profiles more frequently in which the average power spent for the padded bytes increases due to the fact that higher order MCS profiles require larger Transport Block sizes for a given sub-frame duration of 1 ms. However, the effect of byte padding operation on the power saving performance can be considered from negligible to none depending the system parameters. Therefore, only the analytical model will be used to evaluate the proposed profile selection policies for the rest of the study.

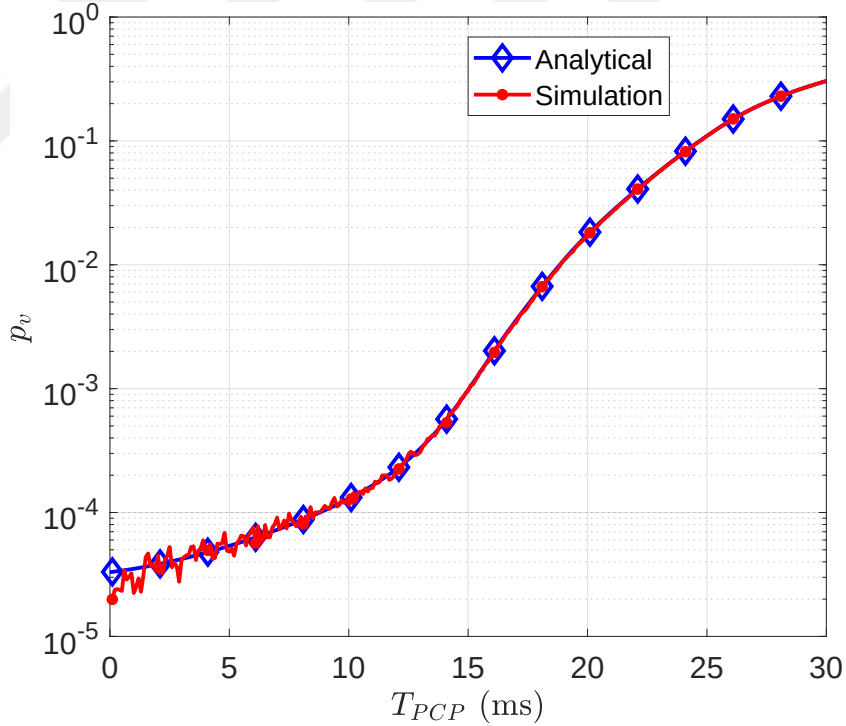


Figure 6.3: Delay violation probability p_v of the particular policy PCP as a function of the parameter T_{PCP} when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms and the maximum attainable SNR $\alpha_m = 12$ dB for both analytical model and simulations.

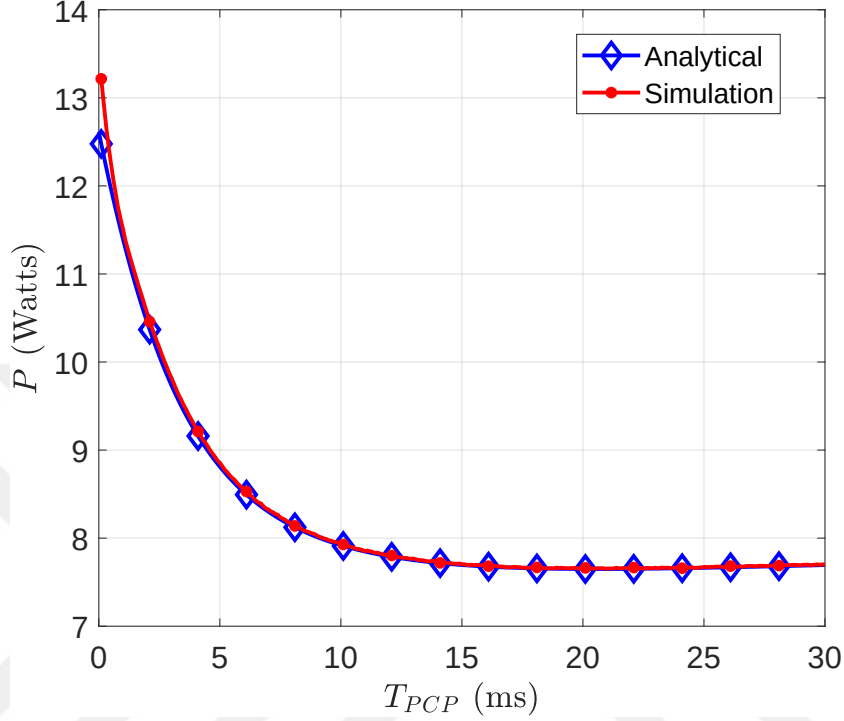


Figure 6.4: Average power consumption P of the particular policy PCP as a function of the parameter T_{PCP} when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms and the maximum attainable SNR $\alpha_m = 12$ dB for both analytical model and simulations.

6.3 Performance Evaluation

In this section, the energy gain and the absolute saved power performances of the proposed policies with respect to the three system parameters λ , α_m and D_0 are investigated using the employed system setup.

In first two examples, the scenario of $D_0 = 15$ ms, $\alpha_m = 12$ dB and the tolerance parameter $\varepsilon = 0.001$ is considered. First, a specific case where $\lambda = 1000$ packets/sec is studied in order to demonstrate the methodology of basic policies. For this case, P_{MDP} can be obtained as 15.44 Watts using (5.4). The delay violation probability and the average power consumption of STP , PCP and $EPCP$ with respect to their performance parameter T_p are depicted in Fig. 6.5 and Fig. 6.6, respectively. In order to find the optimal parameter T_{p^*} , the delay threshold T_p is varied with a step size of 0.1 ms and bounded by a maximum value

of $T_p = 2D_0$ ms. Since all three basic policies behave same as *MDP* when $T_p = 0$ ms, the minimum value of $T_p = 0.0001$ ms is considered in exhaustive search. For a given T_p value, the average power consumption of *EPCP* is much lower than of *PCP* (and also *STP*) which shows that exclusion of relatively energy-inefficient profiles for the rate adjustment results in significant power savings. On the other hand, the delay performance of *PCP* appears to be better than *EPCP* (and also *STP*) which indicates that the delay performance is slightly degraded by elimination of profiles for energy saving purposes. In this example, the optimal delay bound T_{p^*} values are the maximum of T_p values that satisfies the constraint $\epsilon = 0.001$. However, this does not hold in general. The reason is that low indexed transmission profiles may be relatively less energy-efficient than higher indexed profiles in which case higher energy savings may occur at relatively lower values of T_p . An example of this observation can be seen in Fig. 6.6 such that the average power consumption of *PCP* when $T_p = 20$ ms is less than it is for when $T_p = 30$ ms. This the reason why exhaustive search is performed over all possible T_p values instead of the binary search method which would be used if resulting T_{p^*} values were the maximum of T_p values that satisfies the delay violation constraint.

In the second numerical example, performance of the six proposed policies with respect to the parameter λ is evaluated. *MDP* does not satisfy the delay violation constraint for the arrival rate values $\lambda > \lambda_M = 2130.766$. Therefore, the percentage energy gain and the absolute saved power performance of the the six proposed policies are evaluated and depicted for $\lambda < \lambda_M = 2130.766$ in Fig. 6.7 and Fig. 6.8, respectively. For relatively lower values of the arrival rate, all the six proposed policies appear to be serving at the profile with the lowest service rate, i.e., U_1 for majority of the time and percentage energy gain performances are similar with around 62% gain over *MDP*. However, the absolute saved power vanishes for low arrival rates as expected because the average power consumption of *MDP* itself is relatively low for these arrival rates. The absolute saved power peaks for medium to high arrival rates but vanishes eventually, since the profile with highest possible service rate and transmission power, i.e, U_N is mostly used to satisfy the statistical delay constraint in high traffic loads. It is also remarkable that even though $EPCP_e^*$ consistently gives the most absolute saved power as

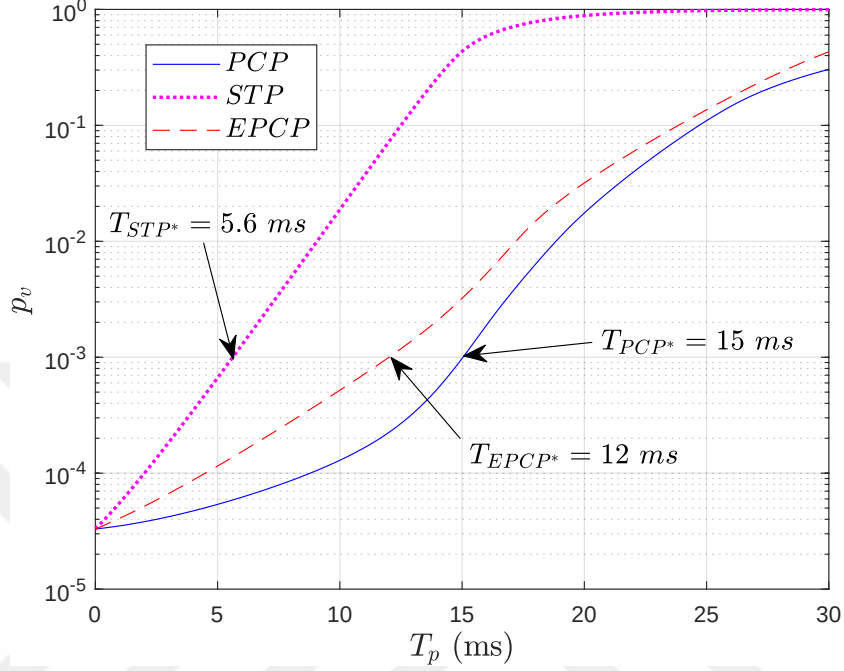


Figure 6.5: Delay violation probability p_v of the three proposed policies STP , PCP , and $EPCP$ as a function of the parameter T_p when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms, and the maximum attainable SNR $\alpha_m = 12$ dB.

seen in Fig. 6.8, the performance of STP_e^* appears to be very reasonable with the advantage of using only two transmission profiles. This is the indication of the need for careful selection of the transmission profiles by considering their energy-efficiency characteristics.

In the third numerical example, the effect of the choice of the delay bound D_0 on the energy gain and the absolute saved power performance of the six proposed policies is investigated. A specific scenario when $\lambda = 1000$ and $\alpha_m = 12$ dB is considered. MDP does not satisfy the delay constraint for $D_0 < 3.776$ ms for this particular scenario. Therefore, percentage energy gain and absolute saved power values of the six proposed policies are depicted for the interval $D_0 > 3.776$ ms in Fig. 6.9 and Fig. 6.10, respectively. It is remarkable that basic policies provide no energy gain for a wide range of D_0 values. Extended policies, on the other hand, still provide significant amount of energy gain in this region which indicates that relaxing the choice of the first profile gives the opportunity to establish energy-efficient communication for systems with relatively strict delay requirements. It

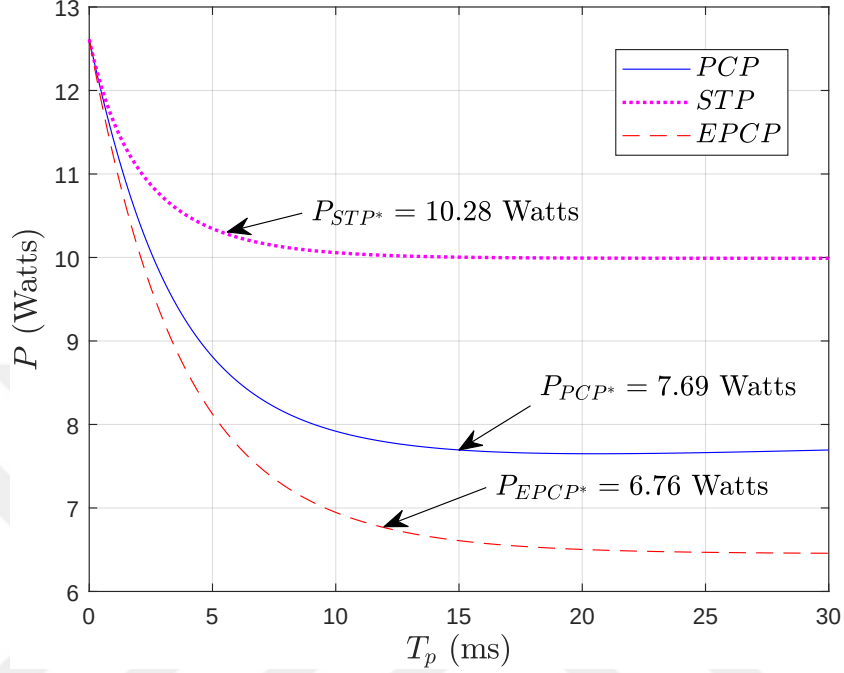


Figure 6.6: Average power consumption P of the three proposed policies STP , PCP , and $EPCP$ as a function of the parameter T_p when $\lambda = 1000$ packets/sec, the delay bound $D_0 = 15$ ms, and the maximum attainable SNR $\alpha_m = 12$ dB.

is also clear that $EPCP_e^*$ consistently outperforms all other policies with STP_e^* being slightly worse just for higher D_0 values.

In the final example, the percentage energy gain and the absolute saved power are obtained with respect to the parameter α_m for six proposed policies when $\lambda = 1000$ and $D_0 = 15$ ms, which is depicted in Fig. 6.11 and Fig. 6.12, respectively. It can be seen that $EPCP_e^*$ gives more energy gain and absolute saved power than any other policy for all values of α_m while STP_e^* follows $EPCP_e^*$ slightly below. The four proposed policies $EPCP^*$, $EPCP_e^*$, STP^* and STP_e^* performs same when $\alpha_m = 7$ dB because their resulting profile subset $\mathcal{K}$ is same for that particular value of α_m . Also note that when $\alpha_m = 20$ dB, the absolute saved power is around 20 Watts which can be considered as a significant amount considering the P_{max} is about 39.8 Watts.

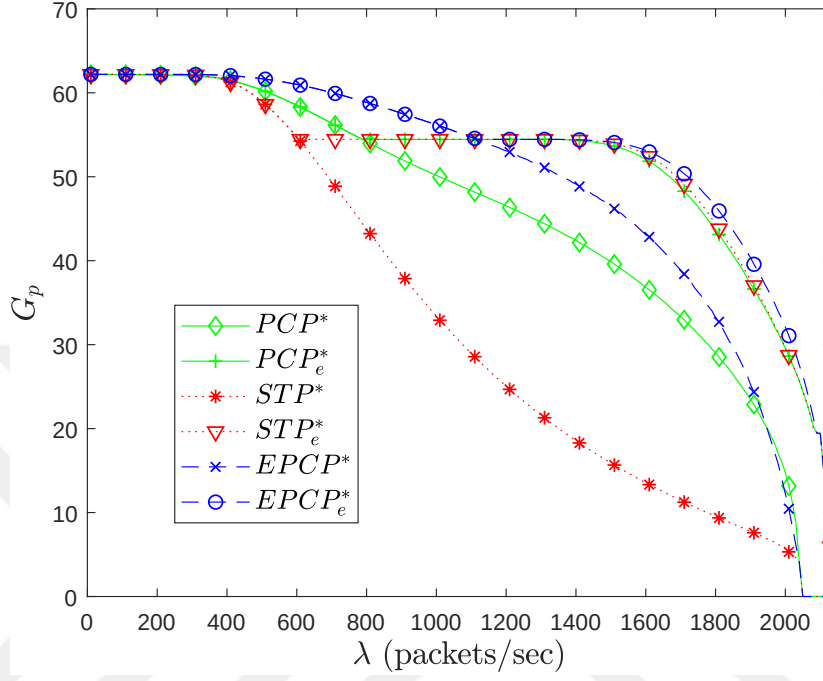


Figure 6.7: The percentage energy gain G_p as a function of the arrival rate λ for the six proposed policies.

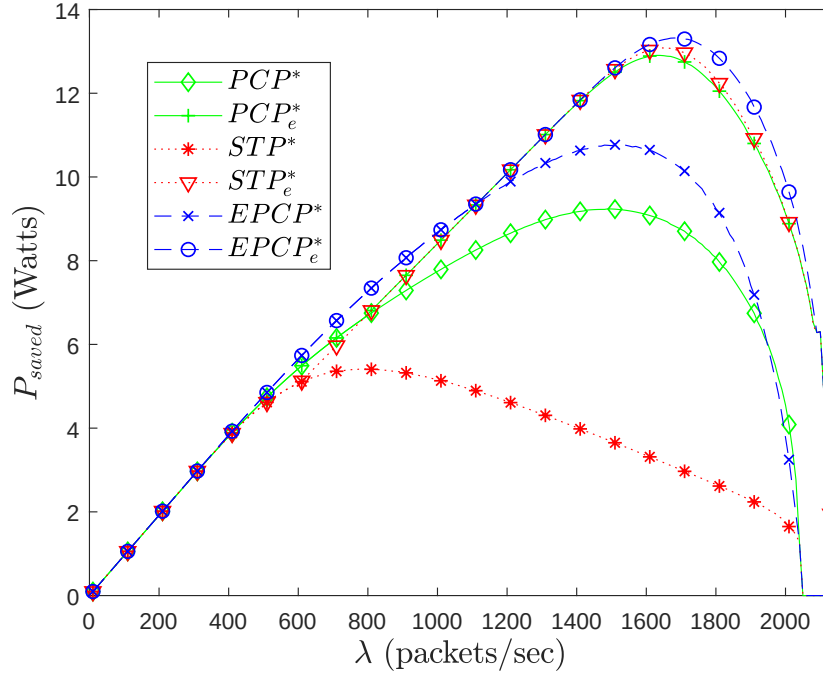


Figure 6.8: The absolute saved power P_{saved} as a function of the arrival rate λ for the six proposed policies.

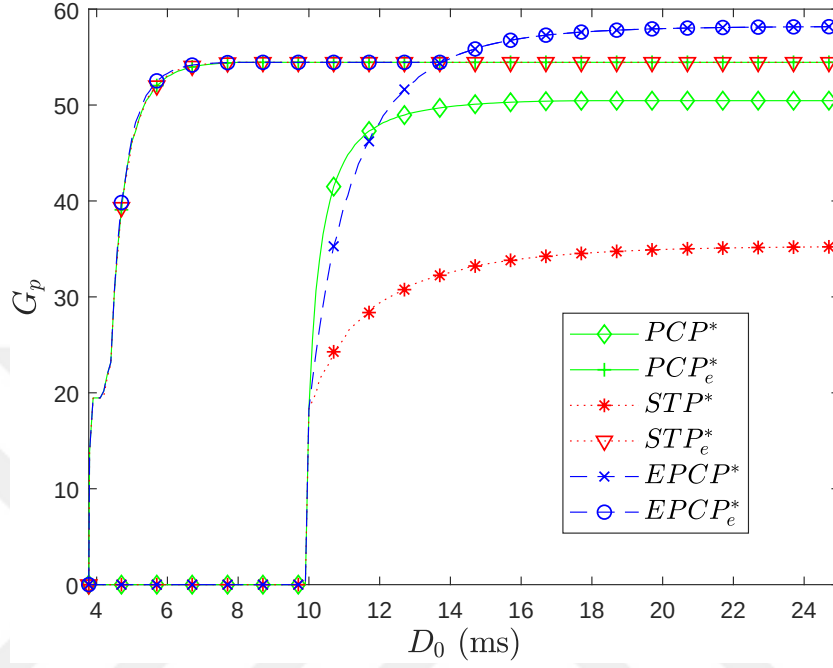


Figure 6.9: The percentage energy gain G_p as a function of the delay bound D_0 for the six proposed policies.

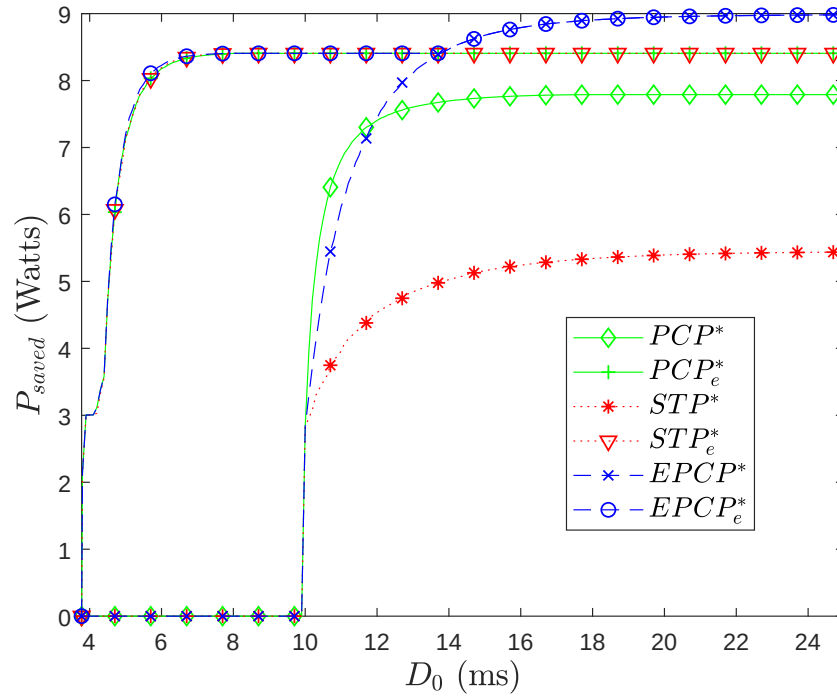


Figure 6.10: The absolute saved power P_{saved} as a function of the delay bound D_0 for the six proposed policies.

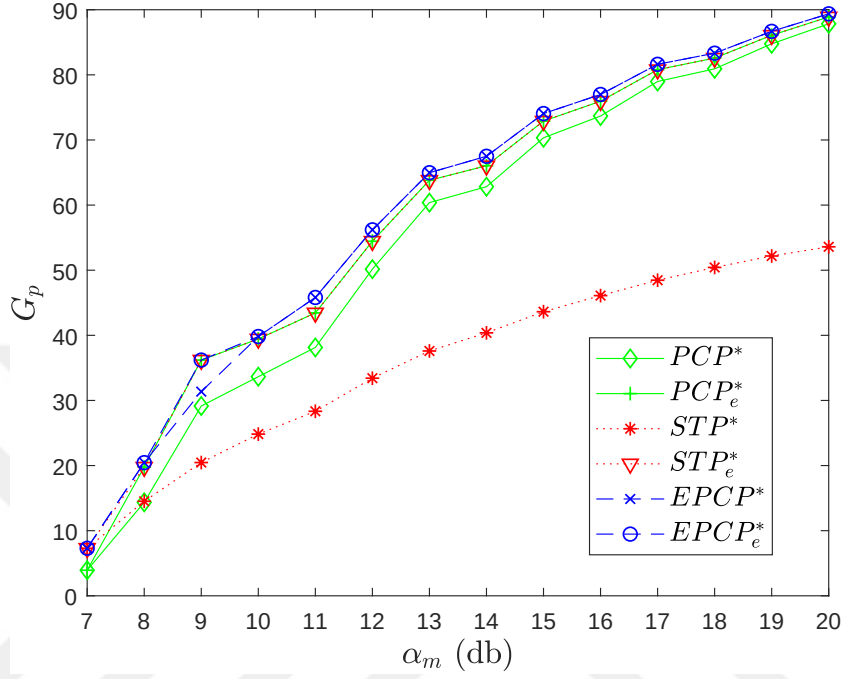


Figure 6.11: The impact of the maximum attainable average SNR α_m on the percentage energy gain G_p for the six proposed policies.

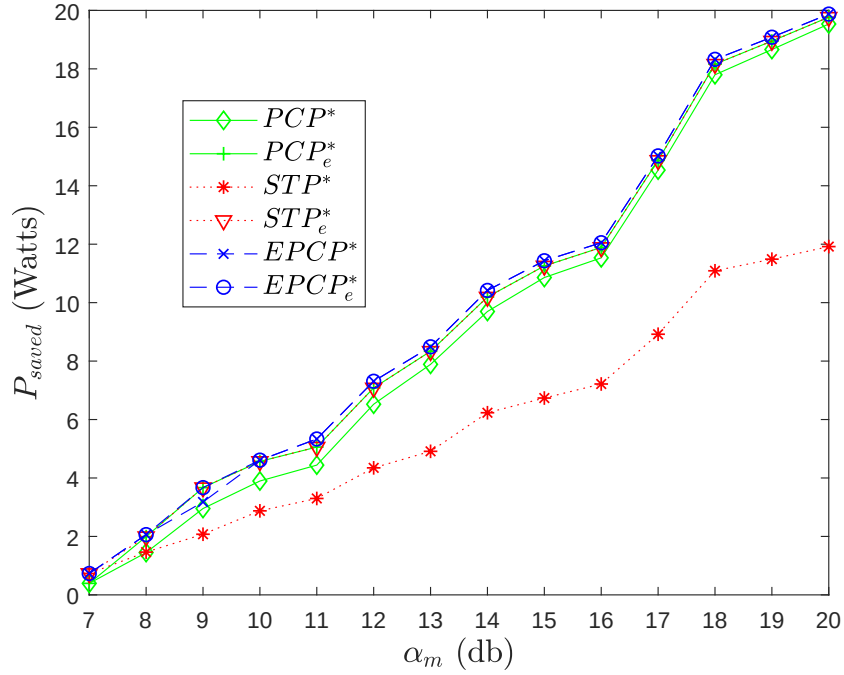


Figure 6.12: The impact of the maximum attainable average SNR α_m on the absolute saved power P_{saved} for the six proposed policies.

Chapter 7

Conclusions

In this thesis, transmission profile selection policies are proposed for the problem of energy minimization under statistical delay constraints over wireless links. The system is first modeled as an M/M/1 queue with delay dependent service times which yields MRMFQ analysis to obtain the steady-state delay distributions as well as the average power consumption. For the purpose of obtaining performance parameters, the constructed MRMFQ model is solved to find the optimum parameter set of interest for a given policy. This framework can easily be adapted to more realistic scenarios in the future. The proposed profile selection policies are studied using realistic transmission profiles obtained from LTE simulations.

Three basic policies along with their extended versions are proposed. Basic versions consider only one performance parameter to optimize, named as delay threshold. If the queuing delay of a HoL packet is greater than this threshold, all policies use the highest possible rate available for the transmission. If it is less than this threshold, *STP* uses the lowest possible rate whereas *PCP* selects transmission rate from the available profile set such that the service rate will be proportional to the delay already experienced by that packet. *EPCP*, on the other hand, first evaluates the energy-inefficiency attribute of all profiles. A novel term, named as relatively energy-inefficient profile, is defined which identifies some of the profiles as inefficient compared to other profiles in terms of energy

saving. Towards the elimination of the relatively energy-inefficient profiles, a performance metric, named as energy-inefficiency metric, is also defined for a given profile set. After elimination of all relatively energy-inefficient profiles, *EPCP* applies same as *PCP* on the set of remaining profiles. Extended versions of each basic policy, denoted by STP_e , PCP_e and $EPCP_e$, respectively, are also proposed for which the choice of the minimum service rate profile is also relaxed along with the delay threshold parameter. Relaxing this choice enables us to obtain significant energy gains under relatively strict delay constraints. It is also concluded that the absolute saved power depends on the traffic load such that it is minimized for low and high loads as expected. However, substantial savings appear to be obtainable for moderate loads. In the numerical examples, it is shown that $EPCP_e$ consistently outperforms all other policies which indicates the importance of considering energy efficiency properties of transmission profiles. Besides, STP_e gives very acceptable gains with the advantage of implementation simplicity whereas its basic version STP shows significantly worse performance compared to all other policies which also yields the importance of optimizing selection on the minimum service rate profile.

Future work includes incorporation of fading channels into the stochastic queuing model and analysis of more general arrival processes and packet size distributions. Using insight from this study, another future work item is the development of online algorithms that dynamically modify their performance parameters based on the feedback information received such as traffic information and channel condition.

Bibliography

- [1] J. Wu, Y. Zhang, M. Zukerman, and E. Kai-Ning Yung, “Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey,” vol. 17, pp. 1–24, 04 2015.
- [2] W.-B. Yang and M. Souryal, “LTE physical layer performance analysis,” Tech. Rep. 7986, NIST Interagency/Internal Report (NISTIR), May 2014.
- [3] “More than 50 billion connected devices,” tech. rep., Ericsson, 2011.
- [4] “Rising to meet the 1000x mobile data challenge,” tech. rep., Qualcomm, 2012.
- [5] A. Zappone and E. Jorswieck, “Energy efficiency in wireless networks via fractional programming theory,” *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.
- [6] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, “How much energy is needed to run a wireless network?,” *IEEE Wireless Communications*, vol. 18, pp. 40–49, October 2011.
- [7] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, “A parameterized base station power model,” *IEEE Communications Letters*, vol. 17, pp. 2033–2035, November 2013.
- [8] G. Li, S. Jin, F. Zheng, X. Gao, and X. Wang, “Energy efficient link adaptation for downlink transmission of LTE/LTE-A systems,” in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sept 2013.

- [9] Y. Zhang, H. M. Wang, T. X. Zheng, and Q. Yang, “Energy-efficient transmission design in non-orthogonal multiple access,” *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 2852–2857, March 2017.
- [10] Y. Niu, C. Gao, Y. Li, L. Su, D. Jin, Y. Zhu, and D. O. Wu, “Energy-efficient scheduling for mmwave backhauling of small cells in heterogeneous cellular networks,” *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 2674–2687, March 2017.
- [11] M. Zafer and E. Modiano, “Optimal rate control for delay-constrained data transmission over a wireless channel,” *IEEE Transactions on Information Theory*, vol. 54, pp. 4020–4039, Sept 2008.
- [12] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, “Energy-efficient packet transmission over a wireless link,” *IEEE/ACM Transactions on Networking*, vol. 10, pp. 487–499, Aug 2002.
- [13] W. Chen, M. J. Neely, and U. Mitra, “Energy-efficient transmissions with individual packet delay constraints,” *IEEE Transactions on Information Theory*, vol. 54, pp. 2090–2109, May 2008.
- [14] X. Li, X. Dong, and D. Wu, “On optimal power control for delay-constrained communication over fading channels,” *IEEE Transactions on Information Theory*, vol. 57, pp. 3371–3389, June 2011.
- [15] C. Gunaratne, K. Christensen, B. Nordman, and S. Suen, “Reducing the Energy Consumption of Ethernet with Adaptive Link Rate (ALR),” *Computers, IEEE Transactions on*, vol. 57, pp. 448–461, April 2008.
- [16] H. E. Kankaya and N. Akar, “Solving multi-regime feedback fluid queues,” *Stochastic Models*, vol. 24, pp. 425–450, July 2008.
- [17] C. Tunc and N. Akar, “Performance modeling of delay-based dynamic speed scaling,” in *The Ninth International Conference on Matrix-Analytic Methods in Stochastic Models (MAM9), Budapest, Hungary*, June 2016.

- [18] A. Wierman, L. L. H. Andrew, and M. Lin, “Speed scaling: An algorithmic perspective,” in *Handbook of Energy-Aware and Green Computing*, pp. 385–406, CRC Press, 2 Jan 2012.
- [19] L. Chen and N. Li, “On the interaction between load balancing and speed scaling,” *Selected Areas in Communications, IEEE Journal on*, vol. 33, pp. 2567–2578, Dec 2015.
- [20] J. M. George and J. M. Harrison, “Dynamic control of a queue with adjustable service rate,” *Operations Research*, vol. 49, no. 5, pp. 720–731, 2001.
- [21] J. Tang and X. Zhang, “Quality-of-service driven power and rate adaptation over wireless links,” *IEEE Transactions on Wireless Communications*, vol. 6, pp. 3058–3068, August 2007.
- [22] B. Tomasi and J. C. Preisig, “Energy-efficient transmission strategies for delay constrained traffic with limited feedback,” *IEEE Transactions on Wireless Communications*, vol. 14, pp. 1369–1379, March 2015.
- [23] M. Sinaie, A. Zappone, E. A. Jorswieck, and P. Azmi, “A novel power consumption model for effective energy efficiency in wireless networks,” *IEEE Wireless Communications Letters*, vol. 5, pp. 152–155, April 2016.
- [24] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, “Energy-efficient power control: A look at 5G wireless technologies,” *IEEE Transactions on Signal Processing*, vol. 64, pp. 1668–1683, April 2016.
- [25] J. Wu, E. W. M. Wong, Y. C. Chan, and M. Zukerman, “Energy efficiency-QoS tradeoff in cellular networks with base-station sleeping,” in *GLOBE-COM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–7, Dec 2017.
- [26] A. Zappone, L. Sanguinetti, and M. Debbah, “Energy-delay efficient power control in wireless networks,” *IEEE Transactions on Communications*, vol. 66, pp. 418–431, Jan 2018.

- [27] J. G. Kim and M. M. Krunz, “Bandwidth allocation in wireless networks with guaranteed packet-loss performance,” *IEEE/ACM Transactions on Networking*, vol. 8, pp. 337–349, Jun 2000.
- [28] M. M. Krunz and J. G. Kim, “Fluid analysis of delay and packet discard performance for QoS support in wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 384–395, Feb 2001.
- [29] L. Liu, P. Parag, J. Tang, W. Y. Chen, and J. F. Chamberland, “Resource allocation and quality of service evaluation for wireless communication systems using fluid models,” *IEEE Transactions on Information Theory*, vol. 53, pp. 1767–1777, May 2007.
- [30] Y. Li, G. Ozcan, M. C. Gursoy, and S. Velipasalar, “Energy efficiency of hybrid-ARQ under statistical queuing constraints,” *IEEE Transactions on Communications*, vol. 64, pp. 4253–4267, Oct 2016.
- [31] 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, pp. 1375–1395, March 2016.
- [32] M. Ozmen and M. C. Gursoy, “Energy-efficient power control in fading channels with Markovian sources and QoS constraints,” *IEEE Transactions on Communications*, vol. 64, pp. 5349–5364, Dec 2016.
- [33] G. Ozcan, M. Ozmen, and M. C. Gursoy, “QoS-driven energy-efficient power control with random arrivals and arbitrary input distributions,” *IEEE Transactions on Wireless Communications*, vol. 16, pp. 376–388, Jan 2017.
- [34] A. Badescu, S. Drekić, and D. Landriault, “On the analysis of a multi-threshold Markovian risk model,” *Scandinavian Actuarial Journal*, vol. 2007, no. 4, pp. 248–260, 2007.
- [35] A. da Silva Soares and G. Latouche, “Fluid queues with level dependent evolution,” *European Journal of Operational Research*, vol. 196, no. 3, pp. 1041 – 1048, 2009.

- [36] M. Mandjes, D. Mitra, and W. Scheinhardt, “Models of network access using feedback fluid queues,” *Queueing Systems*, vol. 44, no. 4, pp. 2989–3002, 2003.
- [37] M. A. Yazici and N. Akar, “The finite/infinite horizon ruin problem with multi-threshold premiums: a Markov fluid queue approach,” *Annals of Operations Research*, vol. 252, pp. 85–99, 2017.
- [38] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory (2nd Ed.)*. New York, NY, USA: John Wiley & Sons, Inc., 1985.
- [39] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures,” Technical Specification (TS) 36.213, 3rd Generation Partnership Project (3GPP), March 2012. Version 10.5.0.