ABSTRACT

Title of dissertation: AGE OF INFORMATION AND ENERGY EFFICIENCY IN COMMUNICATION NETWORKS

Maice Dutra da Costa, Doctor of Philosophy, 2015

Dissertation directed by: Professor Anthony Ephremides
Department of Electrical and Computer Engineering

This dissertation focuses on two important aspects of communication systems, namely energy efficiency and age of information. Both aspects have received much less attention than traditional performance metrics, such as throughput and delay.

The need to improve the energy efficiency in communication networks is apparent, given the high demand for power consuming applications to be implemented in devices with limited energy supplies. Additionally, improvements in energy efficiency are encouraged by possible reductions in network operation costs, and by the increasing awareness of the environmental impact caused by the information and communication technologies.

In this dissertation, energy efficiency is studied in the context of a cognitive wireless network, in which users have different priorities to access the network resources, possibly interfering and cooperating among themselves. A new parametrization is proposed to characterize performance trade-offs associated with energy efficiency for non-cooperative and cooperative network models. Additionally, a game theoretic model is proposed to study resource allocation in a cooperative cognitive

network, accounting for energy efficiency in the utility functions.

Age of information is a relatively new concept, which aims to characterize the timeliness of information. It is relevant to any system concerned with timeliness of information, and particularly relevant when information is used to make decisions, but the value of the information is degraded with time. This is the case in many applications of communications and control systems.

In this dissertation, the age of information is first investigated for status update communication systems. The status updates are samples of a random process under observation, transmitted as packets, which also contain the time stamp to identify when the sample was generated. The age of information at the destination node is the time elapsed since the last received update was generated. The status update systems are modeled using queuing theory. We propose models for status update systems capable of managing the packets before transmission, aiming to avoid wasting network resources with the transmission of stale information. In addition to characterizing the average age, we propose a new metric, called peak age, which provides information about the maximum value of the age, achieved immediately before receiving an update.

We also propose a new framework, based on the concept of age of information, to analyze the effect of outdated Channel State Information (CSI) on the performance of a communication link in which the source node acquires the CSI through periodic feedback from the destination node. The proposed framework is suitable to analyze the trade-off between performance and timeliness of the CSI, which is a fundamental step to design efficient adaptation functions and feedback protocols.

AGE OF INFORMATION AND ENERGY EFFICIENCY IN COMMUNICATION NETWORKS

by

Maice Dutra da Costa

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Anthony Ephremides, Chair/Advisor
Professor Sennur Ulukus
Professor Richard La
Professor Di Yuan
Professor Louiqa Raschid, Dean's representative

"In this moment, recognize the value of being alive."

(Prem Rawat)

Dedication

To my parents, Julio and Lena.

Acknowledgments

I would like to thank everyone that has helped, inspired, or encouraged me to pursue this degree, including Professors, collaborators, colleagues, staff members, my family and friends.

Special thanks to my advisor, Professor Anthony Ephremides, for sharing his valuable knowledge and experience, encouraging me to face the challenges, and supporting me during the past five years. It was my honor to be his student.

Special thanks to my co-authors, Dr. Marian Codreanu and Dr. Stefan Valentin, for their contributions to this work and to my professional development, including the enriching opportunities to visit other institutions abroad.

Thanks to Dr. Sennur Ulukus, Dr. Richard La, Dr. Di Yuan, and Dr. Louiqa Raschid for accepting to be part of my PhD dissertation committee. Your time and your comments are very much appreciated.

Thanks to the Professors I had along the way, including those from my undergraduate and graduate courses, and the friends who are now Professors. Special thanks to Dr. Paulo Cardieri, Dr. Michel Yacoub, Dr. Kasso Okoudjou, Dr. André Tits, Dr. P.S. Krishnaprasad and Dr. Prakash Narayan. Special thanks to my friends Dr. José Cândido Santos Filho and Dr. Ugo Dias.

I would also like to acknowledge the support from staff members from the University of Maryland, including Tracy Chung, Melanie Prange, Bill Churma, Maria Hoo, Carrie Hilmer, Vivian Lu, Vicky Berry, Wei Shi, Jeff Coriale and Regina King.

Thanks to all colleagues from the University of Maryland. Special thanks to

Marcos Vasconcelos, Michael Kuhlman, George Zoto, Khurram Shahzad, and Nadia Adem, for the endless hours of study, discussions and encouragement. Thanks to my academic family members Nof Abuzainab, Jeogho Jeon, Anthony Fanous and Mohamed Kaschef. Thanks to Jonathan Jesuraj for having me on board of the ECEGSA, and for the social events. Thanks to the WECE ladies, with special thanks to Dr. Min Wu and former presidents Filiz Yezilkoy, Nazre Batool, Sowmya Subramanian, Maya Kabkab, and Tian Li.

Thanks to all friends outside the University, who made my experience here more joyful. Special thanks to my friend Richard, for cheering and supporting me through good and bad moments. Words cannot express the gratitude I owe to my brother Daniel, who was the first one to encourage me to come to Maryland, and to his wife Natasha, for their support and for bringing the Brazilian vibe to my life in the US.

Last, but not least, special thanks to my family and friends back home, who encouraged and celebrated this achievement. Special thanks to my mother Lena, to my father Julio, and to my sister Letícia, for inspiring and supporting me, and for keeping strong besides the distance.

Contents

List of Figures

Acronyms

BER Bit Error Rate

ACK Acknowledgment Message

AWGN Additive White Gaussian Noise

BS Base Station

CSI Channel State Information

DF Decode and Forward

FCFS First-Come-First-Served

FDD Frequency Division Duplexing

FSMC Finite State Markov Channel

GHG Green-House Gases

HPU High-Priority User

LCFS Last-Come-First-Served

LTE Long Term Evolution

LPU Low-Priority User(s)

MDP Markov Decision Processes

MIMO Multiple-Input Multiple-Output

MMSE Minimum Mean Square Error

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

PASTA Poisson Arrivals See Time Averages

QoS Quality of Service

SNR Signal-to-Noise Ratio

SINR Signal-to-Interference-plus-Noise Ratio

Chapter 1: Introduction

1.1 Motivation

The large number of applications and the increasing demand for communication systems requires these systems to be constantly adapted, or reinvented, to become more efficient. In past years, much effort has been made to improve spectral efficiency, in order to deliver higher network throughput, and to reduce delay in communication networks. More recently, the need to focus on different performance metrics has become apparent, and we identify two important aspects to be considered in the deployment of communication systems, namely the *energy efficiency* and the *age of information*, which are the focus of this dissertation.

The energy efficiency of wireless networks has recently received much attention, both in industry and academia, given the urgency to address environmental and economical concerns, and also to handle the increasing demand for high-rate, power consuming applications implemented in battery powered devices with limiting energy constraints.

In addition to the increasing concern with energy efficiency, wireless networks are experiencing a paradigm shift regarding their architecture, departing from the centralized model to a user-centric model, with more autonomous nodes, and dis-

tributed solutions for resource allocation, based on local information and cooperation among users. For this scenario of a user-centric network, it is necessary to develop techniques and protocols that address the heterogeneity of the users, who may have different needs and priorities to use the resources. We identify two additional concepts to be considered when developing energy efficient solutions for future wireless networks, namely *cognition*, and *cooperation* among users.

In this dissertation, energy efficiency is investigated in the context of cognitive wireless networks, in which users with different priorities share the spectrum, accounting for the possibility of interference and cooperation among those users. A new parametrization is proposed to characterize the intricate relationship between the transmission parameters of high-priority and low-priority users, and the energy efficiency is analyzed for these two classes of users. Multiple spectrum sharing schemes are considered, including cases with and without network level cooperation. Performance trade-offs involving the energy efficiency are discussed, making apparent the difficulties in selecting transmission parameters in cognitive wireless networks. The cooperation among users with low- and high-priority is investigated using two models. In the first proposed model, the low-priority users are allowed to cause some interference to the high-priority user, as long as they act as relays when necessary. In the second proposed model, the spectrum is allocated during specified time intervals for individual transmissions and cooperation among the low- and high-priority user. In the second case, we proposed a game theoretic model to determine the time allocation of the spectrum while encouraging cooperation among the users.

The second part of this dissertation focuses on the concept of age of information. This relatively new concept is being developed with the objective to characterize the timeliness of information. Its scope is not confined to wireless communication networks. Several systems are sensitive to the timeliness of information, particularly when the information loses its value with time. This is the case, for example, in systems that take actions based on the status of an observed process, as in sensor networks, or control systems.

In this dissertation, the age of information is investigated for communication systems that transmit status update messages. These messages contain the status of a random process under observation and the time stamp indicating when the message was generated. The status updates are transmitted to a destination node, which is interested in timely information about the observed process. The age of information is defined as the difference between the current time instant and the time stamp of the most recent update message received at the destination node. The transmission medium is modeled using queuing theory. In addition to the analysis of the average age, a new metric called peak age is proposed to characterize the age of information for multiple system models. This dissertation also presents a new framework, based on the concept of age of information, to study the effect of outdated Channel State Information (CSI) on the performance of communication networks that make use of feedback.

In the remainder of this chapter, we briefly introduce the most relevant topics considered in this dissertation, and the related work in the literature. Section 1.2 provides an introduction to the topic of energy efficiency in wireless communication

networks. The following sections, 1.3 and 1.4, discuss the use of cognition and cooperation, focusing primarily on aspects of energy efficiency. Section 1.5 brings a brief description of the use of game theoretic tools in resource allocation for cognitive cooperative networks. Then we proceed to give a brief introduction on age of information, in section 1.6, and on the use of channel state feedback, in section 1.7. Finally, the outline of this dissertation is presented in section 1.8.

1.2 Energy Efficiency in Wireless Networks

As previously mentioned, the interest in developing energy efficient communication systems has increased in recent years. The reasons to support enhancements in the energy efficiency of wireless networks include:

(*i*) Environmental concerns – The awareness of the environmental problems related to the emission of Green-House Gases (GHG) has increased in the past few years, and action is needed to identify and improve the main sources of GHG in order to reduce the emissions. It is estimated that the information and communication technologies are responsible for approximately three percent of the worldwide consumption of electric energy, and two percent of the total carbon dioxide (CO_2) emissions [1], numbers which are rapidly increasing. Wireless networks are partially responsible for those numbers, and should be object of study to improve energy efficiency. The effort to reduce the emission of GHG will require not only improvements on the energy efficiency of the networks, but also the use of renewable energy supplies in large networks. For example, renewable energy supplies can be used in the

base stations, which are responsible for more than half of the energy consumption in mobile networks, particularly due to the power amplification devices [2].

(*ii*) Economical incentives – Energy efficiency is economically appealing, since the energy represents a significant share of the operation expenditure of the service providers [3]. Clearly, it is of interest of the operator that the network is as efficient as possible, reducing the costs of operation.

(*iii*) Technological challenges – Many terminals in wireless networks are expected to operate autonomously, with limited source of energy (battery supplied). While the power consumption is increasing due to new multimedia applications, the battery technology is evolving very slowly, increasing the gap between the energy demand and the energy available in the batteries [4]. Additionally, the size of the devices also imposes constraints in the available battery capacity.

The deployment of energy efficient (green) communication networks will require multidisciplinary efforts, to develop new materials, energy efficient circuits, new communication protocols and network architectures. Including the energy efficiency as an objective to design communication systems inevitably impacts on other performance metrics [4, 5].

A fundamental framework for research of energy efficient (green) wireless communications was proposed in [6], based on four fundamental trade-offs related to the energy efficiency, namely (1) the trade-off between energy efficiency and the deployment efficiency, given by the throughput per unit of deployment cost, (2) the trade-off between energy efficiency and spectrum efficiency, (3) the trade-off between bandwidth and power needed to achieve a target rate, and (4) the trade-off

between power consumption and delay.

The recognition of a trade-off between energy efficiency and spectral efficiency dates back to Shannon's theory. Using the capacity formulation for an Additive White Gaussian Noise (AWGN) channel, it can be shown that the energy efficiency is monotonically decreasing with the spectral efficiency [7]. The analysis of this trade-off in more realistic scenarios results in complicated multi-objective optimization problems, that account for fading conditions, multi-user systems, and the total power consumption. Therefore, further investigation is needed to develop tractable models to analyze this and other trade-offs related to energy efficiency.

Numerous previous works have addressed the trade-off between energy efficiency and spectral efficiency in more specific scenarios. For example, in [8], the authors aimed to solve a multi-objective optimization problem to maximize both energy efficiency and spectral efficiency in the case of a cellular network using Orthogonal Frequency Division Multiple Access (OFDMA), and a new metric named resource efficiency has been proposed to capture the trade-off between the two objectives. The case of a multi-hop wireless network was considered in [9] to characterize the trade-off between the total energy consumed and the communication rate, accounting for transmission energy and circuit processing energy.

In this dissertation, we also focus on aspects related to energy efficiency and spectral efficiency, considering the common definition of energy efficiency given in bits of information per unit of transmit energy, with unit of *bits per Joule*. The proposed parametrization to analyze the energy efficiency in cognitive networks adds to the much needed simple frameworks needed to study the trade-offs between

energy efficiency and other performance metrics in wireless networks.

1.3 Cognitive Networks

The static spectrum allocation policies implemented by regulatory agencies have created an apparent scarcity of this natural resource. While there seems to be no available spectrum bands to meet the increasing demand for high data rates in communication networks, many allocated frequency bands are underutilized by the licensee. A more dynamic model for spectrum allocation is enabled when the radio terminals have the capability to perceive the environment, taking advantage of transmission opportunities while avoiding to disrupt those transmissions with higher priority. This new paradigm is based on the concept of cognitive radios with the capability to adapt and learn, following a cognitive cycle, and significantly increasing the spectral efficiency in communication systems [10].

The adaptability of cognitive radios has been identified as a valuable tool to develop energy efficient (green) communication systems, even though cognitive networks were primarily envisioned to overcome spectrum underutilization. The most common assumption is that one network, the Primary User, has the license to use the frequency resources, while the other network, the Secondary User, has constrained access to network resources. This scenario is one of the cases considered in this dissertation, but we go beyond to study the interaction of users with different priorities, accounting for other scenarios with interference and cooperation. We will refer to Primary and Secondary Users as High-Priority User (HPU) and Low-Priority

User(s) (LPU), respectively, not to allude to the traditional model of cognitive networks. In our work, the HPU and the LPU could even be part of a single network, but they have different priorities to use the resources.

Overviews on the use of cognitive radios to improve energy efficiency of wireless networks have been presented in [11–14]. In [11], the authors described two steps to deploy energy efficient cognitive radio network. The first step would be to exploit the flexibility enabled by cognitive radios, but only when the penalty in power consumption is small or can be compensated by the other functionalities added to the network with the use of cognitive radios. The second step is to exploit the increased flexibility to actively save energy by adapting the reconfigurable radios to the conditions in real time. The work in [12] defended the benefits of the use of cognitive radios to address the trade-offs involving energy efficiency in wireless networks, given the increased energy efficiency awareness enabled by cognitive capabilities. The design of energy efficient spectrum sensing, the coordination of multiple users or networks with secondary (low-priority) access to the spectrum, and the management of the secondary networks have all been discussed in [13]. The work in [14] has pointed to interesting challenges to be addressed in the deployment of energy efficient cognitive networks, identifying the following four fundamental trade-offs concerning energy efficiency:

(*i*) Energy efficiency versus Quality of Service (QoS) – this trade-off exists in communication networks even without the use of cognitive radios, since attending QoS requirements becomes more challenging with energy efficiency constraints. Nonetheless, the existence of different levels of priority among the users brings an-

other dimension to the problem, since the QoS requirements may be focused on the HPU, on the LPU, or on both. The QoS may also be differentiated among the LPU, creating additional classes of users.

(*ii*) Energy efficiency versus fairness – similarly to what is observed for QoS, fairness in cognitive networks also has additional dimensions, since there may be bias to serve the HPU, while also being fair to the LPU. Fairness issues also need further investigation in scenarios with multiple service providers sharing spectrum and also network infrastructure.

(*iii*) Energy efficiency versus interference caused to the HPU – in the paper, the authors mention two cases of interference to be addressed: the misdetection HPU's activity, and the reappearance of the HPU which would require the LPU to vacate the spectrum. Reducing the interference caused to the HPU in both cases may increase energy consumption. The parametrization proposed in this work is useful in the analysis of this trade-off, as it characterizes the interdependence between the transmission parameters of the HPU and the LPU, considering different spectrum sharing schemes.

(*iv*) Energy efficiency versus network architecture – this trade-off is related to the energy spent when modifying the network architecture, for example by adding small cells, relays, and multiple hops in a network. In this dissertation, we address the trade-off of energy efficiency versus throughput when the LPU acts as a relay for the HPU. Our game theoretic formulation for such cooperation also considered the energy spent by the LPU. The energy cost of cooperation, as well as the cost of other modifications in architecture, deserve further investigation, in order to

create mechanisms to reduce operation costs while still benefiting from the new architecture.

In [15], the power efficiency of the LPU is maximized, subject to a constraint on the interference caused to the HPU, and considering spectrum sensing time and accuracy. Resource allocation for the network of low-priority users has been addressed in [16], where a heuristic scheduling algorithm is proposed to improve energy efficiency of the LPU. In [17], the optimal transmission duration and power allocation is found for a LPU sensing multiple channels. The optimal transmission power to maximize energy efficiency for the LPU under a particular spectrum sharing scheme is presented in [18]. Finally, the work in [19] presents sensing-access strategies for the LPU so that energy efficiency is maximized. All these are important contributions, but the development of energy efficient cognitive networks and the use of cognition to improve energy efficiency still deserve further investigation.

The need for spectrum sensing is a major concern when considering the energy efficiency for cognitive networks. The work in [20] proposes a convex optimization problem formulation to minimize the energy consumption subject to constraints on detection and false alarm probabilities. The sensing duration under energy constraints has been considered in [15]. In [21], the authors identify when the spectrum sensing energy consumption is overweighted by the energy savings due to reduced contention among multiple low-priority users.

Our work aims to contribute to the understanding of trade-offs related to energy efficiency in scenarios with cognitive networks, studying the impact of different cognitive capabilities on the energy efficiency of the network. In contrast with the

aforementioned works, our analysis is not restricted to a a single spectrum sharing scheme, nor to a single optimization problem, and we consider both the HPU and the LPU. For the schemes that require spectrum sensing we do not focus on possible improvements of the spectrum sensing energy consumption. Instead, we analyze the trade-off between energy efficiency and the sensing accuracy, regardless of the sensing technology.

1.4 Cooperative Networks

The use of cooperation among nodes in wireless networks has received a great deal of attention in the literature. It is well known that wireless networks can benefit from spatial diversity to mitigate the effects of time varying channels, and to increase reliability in communication, thereby improving spectrum efficiency and reducing the time used to deliver a message. Recently, it has been shown that cooperation at the network protocol level can improve performance in terms of stable throughput, and delay, benefiting all the users [22].

The benefits of cooperation have been previously verified in the context of cognitive radio networks. Relay assisted transmission in the low-priority network was considered in [23] and [24]. Models with the LPU acting as relay for the HPU have been investigated in [25–28], where the stable throughput regions have been analyzed, assuming that cooperative relays are activated during idle periods of the HPU. It has been shown that the use of cooperative relays would help to empty the queue of the HPU, thus creating better transmitting opportunities for the LPU, and

increasing the stable throughput of both users as compared to the non-cooperative case. Routing mechanisms in multi-hop networks that use several LPU as relays have been investigated in [29].

Cooperative networks with full-duplex relays have also been previously investigated. In [30], the authors considered a multi-access problem, assisted by a relay with capability of receiving and transmitting simultaneously. The model considered in [30] is not a cognitive cooperative model, and the relay does not have traffic of its own. In cognitive cooperative networks the use of full-duplex nodes is of interest because the LPU has traffic of its own to transmit, so the nodes acting as relays are not exclusively dedicated to this activity. In [31] the authors analyzed the effect of different degrees of accuracy in self interference cancellation on the average sum throughput and the congestion of a cooperative network. The self-interference is modeled as a variable power gain between transmitter and receiver in the relay node, and this approach will be used in our work presented in chapter 2.

In addition to the increased stable throughput and reduced delays, cooperative techniques potentially increase the energy efficiency in the network. The work in [7] has shown that cooperation may reduce the energy to deliver a message. In [32], the authors considered a cognitive network, with a cooperation scheme in which part of the bandwidth is allocated to the LPU, and the remaining bandwidth is used for HPU transmission and relay transmission, with time sharing. The authors reported up to 80% power savings for the HPU in comparison to non-cooperative schemes.

Although some of the benefits of cognitive cooperation have been indicated in the previous literature, much remains to be done to deploy energy efficient cognitive

cooperative networks. The contributions in this dissertation include two different cooperative schemes between users with different levels of priority. Our observation is that the LPU can be encouraged to cooperate in exchange for transmission opportunities. Our cooperative models were presented in [33] and [34], and the LPU can be either allowed to transmit simultaneously with the HPU subject to a constraint on the caused interference, or it can be allowed to access the spectrum exclusively for a period of time.

1.5 Game Theory in Cognitive Cooperative Networks

The study of cognitive cooperation has been further enhanced by using game-theoretic tools to assist with the resource allocation among users [35–38]. A scenario with multiple high-priority networks is considered in [35], where the authors compare three pricing models for the spectrum to be sold to the LPU. A game theoretic model for spectrum access is presented in [36], with low-priority users searching for transmission opportunities in portions of the spectrum unused by high-priority users. A spectrum leasing scheme was proposed in [37], where a HPU allocates the channel to a low-priority ad hoc network for a fraction of the time, and the low-priority users help forward the HPU's packets using distributed space-time coding technique. The proposed scheme uses a hierarchical Stackelberg game model, where the HPU is the game leader and selects the fractions of time to be used for cooperative and individual transmissions, aiming to maximize its own rate. Following the decision made by the leader user, the group of follower users optimizes their own power to

obtain higher transmission rates. A priced-based game model for spectrum leasing is proposed in [38], where the time allocation and also the price of spectrum are set by the primary, while the selected secondary user may increase its transmission rate by optimizing its transmission power.

Cooperation is not an inherent characteristic of multi-user communication networks, as users contend for resources, and often present selfish and rational behavior. Therefore, different incentive-based approaches have been studied in the literature to encourage cooperative packet forwarding. These schemes can be categorized into pricing-based and reputation-based schemes [39]. In pricing-based schemes, the relay node earns credits when it forwards other users' packets. The credit is usually in the form of virtual currency. Thus, a central controller is required to ensure the payment among the users [40–42]. In reputation-based schemes, the nodes adjust their strategies based on the reputation of other nodes. Hence, each user tries to maintain a good reputation to benefit from cooperation in following interactions [43–45].

The study of game theoretic tools is not the main objective of our work, but we have proposed a game theoretic model to study the interaction of two users with different priorities in a cognitive cooperative scenario. The proposed model is a Stackelberg game [46], in which the HPU is the leader, while the LPU is the follower player in the game. The objective of the proposed game model is to obtain the time allocation for transmissions of each user, and an interval for cooperation, in which the user with lower priority acts as a relay. The proposed model uses spectrum as the real currency to be exchanged between users, instead of the virtual currency often used in pricing mechanisms. Our model also includes a reputation mechanism

based on cooperation credits. The players sequentially make their decisions, observing the reputation of their opponent, and taking the best actions considering how cooperative the other player was in the previous rounds. As a result, the HPU is encouraged to allow the LPU to transmit data of its own, while the LPU is encouraged to act as a relay for the HPU in exchange for the exclusive use of the spectrum.

Our model for cognitive cooperative networks was presented in [34]. In comparison to other models presented in the literature, in particular in [37] and [38], the main contributions of our model include (i) the use of a reputation-based game to perform spectrum allocation, providing the framework to analyze the response of the users in the long run, (ii) the use of utility functions that account for fairness and energy efficiency in resource allocation, (iii) the use of a more complete formulation for the achievable rates, which considers not only the information exchange through a decode and forward multi-hop channel, but also the information flow through the direct link from source to destination, and (iv) the definition of a Stackelberg game in which the resource allocation is defined not only by the leader, but by the follower as well.

1.6 Age of Information

The need to define metrics to describe the age of information was first identified in the context of vehicular ad-hoc networks [47, 48]. In this particular application, information about the status of a node (vehicle) is broadcast to neighbor nodes for navigation and safety purposes, for example. Clearly, it is desirable that neighbor

nodes have access to up-to-date information about each other. However, as multiple nodes contend for limited network resources, the update messages cannot be transmitted very often. Due to this resource limitation, the information about the status of a neighbor node can become outdated. Metrics associated with the age of information can be used in this context to determine when the status update messages should be transmitted under resource constraints, so that the information available at neighbor nodes is as timely as possible. In [47], the authors addressed the problem of congestion control in large vehicular networks, proposing a rate control algorithm to minimize the average age of information throughout the system. The effect of piggybacking messages throughout the vehicular network was investigated in [48], and shown to be effective in reducing the average age of information.

Aiming to better understand the effect of information aging in communication networks, a new line of work was initiated by the same research group, modeling the network using simple queuing models. The initial research in [49], [50], and [51] consider a general model for a system transmitting status update messages, assuming that these messages become available to the source node at random instants and wait in a buffer to be transmitted to a destination node. The average age was characterized in [49], considering that status update messages arrive at the source node according to a Poisson process, and are transmitted using a First-Come-First-Served (FCFS) discipline. With the simplifying assumption that the time it takes to successfully transmit the message is exponentially distributed, the system is modeled as a M/M/1 queue. The average age for the M/D/1 and D/M/1 models was also calculated in [49], illustrating the cases with deterministic service time and pe-

riodic sampling, respectively. The case with multiple sources transmitting through a M/M/1 queue was studied in [50], where the authors derived a region of feasible values of average age. In [51] the authors analyzed the case with Last-Come-First-Served (LCFS) queue discipline, and showed that, for very large arrival rates, this transmission discipline yields the same average age as a system capable of transmitting *just in time* updates. The transmission of *just in time* updates assumes that a packet is generated as soon as the server becomes idle, eliminating queuing delays, and transmitting the updates as fresh as possible. Later in chapter 3, we show that the same result can be achieved with FCFS discipline and packet management policies. Although it has been mentioned in the literature that the transmission of *just in time* updates would be a lower bound on the average age, new results indicate that it may be desirable to let the server idle, and then transmit updates that cause further reduction in the value of age [52]. Clearly, further investigation is necessary to understand this fundamental trade-off of age of information.

Other authors have also investigated the age of information using queuing models. The average age was calculated in [53] for the case that all packets are transmitted immediately after generation through a network cloud, and some packets are rendered obsolete, due to the random service times in the network. This model corresponds to the extreme case of an infinite number of servers, and it calls the attention to the fact that network resources may be wasted transmitting all the available samples of the observed process. The trade-off between the average age of information and the waste of network resources was also investigated in [54], considering the case with only two servers.

For applications in which the destination further processes the samples (e.g. for estimation purposes), it may be necessary to transmit all samples, even at the cost of a larger age. However, when the destination node is interested solely on the current value of the process under observation, it would be acceptable to discard some packets before transmission, with the objective to reduce congestion in the network, and to avoid using network resources in the transmission of outdated information. In this thesis, we propose models for status update systems in which the source node is able to manage the packets before transmitting to the destination, possibly discarding packets, as opposed to storing them in a buffer. We have presented some of the results in [55].

As an alternative to the average age, we proposed a new metric, named peak age. The peak age is the value of age achieved immediately before a status update message is received. The peak age is more adequate than the average age when the application requires that the age is kept below a threshold level. In that case, we can characterize the probability that the peak age is below the required threshold. The probability distribution of a peak has been characterized for the three proposed schemes with packet management. We also observed that the average peak age presents similar behavior to the average age, hence it can be used as an alternative metric to characterize the age of information, with the advantage that the average peak age can be more easily calculated in many interesting cases. The analysis of the peak age was considered in [56], assuming a system with packet management and a single place in the buffer and no packet replacement, extending our results to the case with general distribution for the service time.

1.7 Channel State Feedback

A feedback system can be understood as an example of a status update system. In feedback systems, one or more variables of interest are observed and information about the state of these variables may be reported periodically or aperiodically to a controller, or a central node, where decisions are made based on the received feedback. Consequently, the concept of age of information plays a key role in the performance of any application using feedback.

In communication networks, a feedback scheme is used to report CSI back to the transmitter node, whenever the communication channel between two nodes in a link is non-reciprocal. Non-reciprocal channels are common, since the propagation environment differs between transmitter and receiver, e.g., due to different antenna positions or different frequency bands at the two ends of the link, as with Frequency Division Duplexing (FDD). Without reciprocity, the transmitter cannot directly observe the channel state but has to rely on CSI as reported from the receiver to adapt transmission parameters to the instantaneous state of a time-variant channel.

The CSI reported using a feedback scheme is not available immediately to the transmitter. It experiences delays due to the time it takes to measure, process and transmit the information. Upon reception, the use of CSI may require additional processing time to run the adaptation functions.

The effect of delayed CSI has been considered in the literature mainly under the assumption that past samples are used to estimate the current channel state. Using this approach, Goeckel has shown in [57] the need to consider the delay of

the CSI in the design of adaptive codes, in order to promote throughput gains in comparison with nonadaptive coding. The Orthogonal Frequency Division Multiplexing (OFDM) case was presented in [58], where the authors show that the gain of adaptive OFDM schemes is severely reduced due to the delay of CSI, and propose the use of multiple channel estimates to increase the tolerance to the delay of CSI. The achievable rates for multi-user Multiple-Input Multiple-Output (MIMO) were investigated in [59], where the authors show that a fraction of the multiplexing gain is achievable with delayed feedback, under a constraint in the maximum Doppler shift. A different approach was adopted in [60], where the authors assume a MIMO broadcast channel model and show that outdated channel information can be used to learn side information, improving the achievable rates in wireless networks with multiple flows.

We note that besides the delay experienced by the information reported using feedback channels, the frequency at which the reports are sent also has an impact on the timeliness of the CSI available at the transmitter for adaptation functions. If the CSI is reported infrequently, the information available to the transmitter may be outdated before it receives the CSI again. For this reason, we envision that the concept of age of information is more appropriate than the concept of delay to characterize the effects of outdated CSI on the performance of communication networks.

In this dissertation we propose a new framework, based on the age of information, to study the effect of outdated CSI. Our focus is not on the use of delayed CSI to obtain channel estimation or side information. We call attention to the

fact that the delay is only one of the components of the age of CSI, and focus our analysis on the use of *periodic feedback*. This feedback mode is common in cellular communication systems [61, Chapter 10].

Our initial results have been presented in [62] and [63]. The proposed framework is expected to lead to further work on the development of resource allocation and feedback schemes for multiple users, taking into account the age of information used in adaptation functions.

1.8 Outline of the Dissertation

In chapter 2, we present the results related to energy efficiency. We begin with an overview on the topic of energy efficiency in cognitive networks, presented in section 2.1. Section 2.2 presents the system models to be considered in the analysis of performance trade-offs in cognitive networks for non-cooperative cognitive networks and for the Cooperative Underlay model. The performance for both the non-cooperative and the cooperative cases is investigated in section 2.3, where the throughput and the energy efficiency are calculated for the LPU and the HPU. The performance trade-offs are discussed in section 2.4. The game theoretic model for a cooperative cognitive network is presented in section 2.5. Possible directions for future work are discussed in section 2.6. The contributions of this chapter are then summarized in section 2.7.

In chapter 3, we discuss the concept of age of information. In section 3.2, we formally define the age of information in status update systems and two metrics

that characterize the age. In section 3.3, we characterize the age of information for status update systems, cosidering five different models. At first, we present the analysis for one source-destination link, modeled as a M/M/1/ queue, and discussed in [49]. Then we present the results for multiple servers, and discuss the possibility of resource waste, as presented in [53] and [54]. Finally, we discuss the age of information for three proposed packet management schemes. Section 3.4 presents our numerical results, and section 3.5 presents a discussion about possible future work directions. We summarize the results in this chapter in section 3.6.

In chapter 4, we present the results for the application of age of information to the case of channel state feedback. We describe the system model in section 4.2. In section 4.3 we present the proposed feedback system, while the performance metric which is a utility function is defined in section 4.4, where we describe possible models for the utility using rate rewards and accounting for the cost of feedback. In section 4.5, we analyze the effect of the age of CSI on the probability of error in the channel estimation for a simple channel model. An application of our framework in the case of Rayleigh fading channel is discussed in section 4.6. The application of our framework to multiplexing systems is presented in section 4.7. The discussion about future lines of investigation is presented in section 4.8, and the contributions in this chapter are summarized in section 4.9.

The conclusion and summary of the contributions of this dissertation are presented in chapter 5.

Chapter 2: Energy Efficiency

2.1 Overview

Energy efficiency is an important objective in the analysis and design of wireless networks, in addition to the conventional interest in higher rates and quality of service. Its importance is due to the environmental concerns about energy consumption in communication technologies, and due to the possible reduction in operation expenditures in the networks. Additionally, the need to improve the energy efficiency becomes apparent as the number of power consuming functionalities required from mobile devices increases. Unfortunately, improvements in energy efficiency often require some compromise with respect to other performance metrics, such as throughput.

This chapter presents a contribution to a better understanding of the intricate relations between the energy efficiency and different system parameters in wireless networks. A new parametrization is proposed to investigate the trade-offs that should be addressed while designing energy efficient networks. In particular, we consider the case of a cognitive radio network, accounting for the possibility of interference and cooperation among users with different levels of priority, referred to as High-Priority User (HPU) and Low-Priority User(s) (LPU). The throughput and

energy efficiency are calculated first considering non-cooperative spectrum sharing schemes. We emphasize the possibility that the LPU causes interference to the HPU, and propose that the HPU establishes an interference tolerance, which is a new network design parameter. Given the level of interference tolerated by the HPU, a power constraint is imposed to the LPU, and it is shown to be effective in protecting the HPU from interference. In addition to the trade-off between energy efficiency and throughput, we discuss the trade-off between energy efficiency and the accuracy in spectrum sensing, when sensing is required by the spectrum sharing etiquette.

The energy-throughput trade-off discussed in our work can be understood as one of the trade-offs mentioned in [14], between the quality of service and the interference constraints. As opposed to qualitative discussions, our work describes a system parametrization to obtain analytic results, characterizing trade-offs related to energy efficiency for both HPU and LPU, under different spectrum sharing schemes.

Given the possibility of interference caused by the LPU to the HPU, we envision the use of cooperation as means of repayment from the LPU to the HPU for using the spectrum. In [33], we proposed a new cooperative spectrum sharing scheme, named Cooperative Underlay, which assumes network layer cooperation as proposed in [25, 64], using the LPU as a relay. Later in this chapter, we present the trade-off between energy efficiency and throughput for both the HPU and the LPU, and show the advantage of using cooperation in the case of severe interference.

The energy efficiency of our proposed Cooperative Underlay model is also analyzed in the case of full-duplex relays, since the LPU is assumed to have traffic

of its own to transmit. In our work, we model the self-interference as a variable power gain between transmitter and receiver in the relay node, as proposed in [31]. We investigate the effect of the accuracy in self-interference cancellation on the energy efficiency of the HPU, which benefits from the cooperation of a LPU with capability of simultaneous transmission and reception.

While the Cooperative Underlay model assumes that the LPU and the HPU make use of the spectrum simultaneously, it is also possible to define an alternative cooperative model, in which the LPU should transmit only when the HPU is not transmitting. In [34] we proposed a simple game model to study the interaction of a HPU and a LPU, aiming to obtain the optimum time allocation for the individual transmissions and the cooperative transmission. Given the different priorities in using the resources, a natural approach is to use a Stackelberg game model, in which the HPU is the leader, while the LPU is the follower player in the game [37]. Our game model considers concepts of pricing and reputation, and can be implemented in a distributed manner. The pricing mechanism is implemented using the spectrum as a real currency to be exchanged between users, instead of a virtual one. The reputation mechanism is based on cooperation credits, and is used to monitor the behavior of both users over the course of time. The players sequentially make their decisions, observing the reputation of their opponent, and take the best actions considering how cooperative the other player was in the previous rounds. As a result, the HPU is encouraged to allow the LPU to access the spectrum, and selfish misbehavior of the LPU in packet forwarding is discouraged.

The remaining of this chapter in organized as follows: Section 2.2 presents the

system models to be considered in the analysis of performance trade-offs in cognitive networks for non-cooperative cognitive networks and for the Cooperative Underlay model. The performance for both the non-cooperative and the cooperative cases is investigated in Section 2.3, where the throughput and the energy efficiency are calculated for the LPU and the HPU. The performance trade-offs are discussed in Section 2.4. The game theoretic model for a cooperative cognitive network is presented in Section 2.5. Possible directions for future work are discussed in Section 2.6. The contributions of this chapter are then summarized in Section 2.7.

2.2 System Model for Trade-Off Analysis

We assume that multiple source-destination pairs share a single channel of bandwidth W. Each pair, identified with $(S_i - D_i)$, $i \in \{0, 1, \ldots, n\}$, is referred to as a user. User $(S_0 - D_0)$ has higher priority to use the spectrum, and it is referred to as HPU. The remaining source-destination pairs $(S_i - D_i)$, $i = 1, \ldots, n$, have lower priority to access the channel, and are referred to as a LPU.

Time is slotted, with a slot of duration L used to transmit a single packet. Source nodes are assumed to be saturated, meaning that there is always a packet waiting for transmission. Random access to the channel is assumed for the HPU, so the source node S_0 transmits in a time slot with probability ρ_0.

Spectrum sharing schemes for cognitive radio networks have been extensively studied [65], and the terminology is not completely standard in the literature. Three spectrum sharing schemes will be considered in this chapter, and their definitions

are provided below, together with the assumptions on the transmission probabilities for low-priority users.

- *Underlay* - The HPU and the LPU are allowed to transmit simultaneously, but the LPU has a power constraint. In each time slot, the i-th LPU transmits with probability $\rho_{i,U} = \rho_U$, $i = 1, \ldots, n$.

- *Interweave* - In each time slot, the i-th LPU performs spectrum sensing and transmits with probability $\rho_{i,I} = \rho_I$ if the channel is identified to be idle, remaining silent otherwise. We assume that $0 < \rho_U < \rho_I \leq 1$.

- *Hybrid* - The LPU performs spectrum sensing, transmitting as in Underlay scheme if the channel is sensed occupied, and as in Interweave scheme if the channel is sensed idle.

When spectrum sensing is required, we assume that each low-priority source node senses the channel in the beginning of the time slot, and makes a decision independent from the other users. The signal from the HPU is correctly detected with probability Q_d, and a false alarm is emitted with probability Q_f. As a result, the LPU may cause inadvertent interference to the HPU in the case of Interweave or Hybrid spectrum sharing.

No scheduling scheme is assumed among the LPU, hence the links between a source node S_i and a destination node D_j with $i \neq j$ are interference links, $i, j \in \{0, 1, \ldots, n\}$.

Figure 2.1 illustrates a non-cooperative network model with $n = 2$ LPU. The main links ($S_i - D_i$) are shown with full lines, while the interference links ($S_i - D_j$, $i \neq$

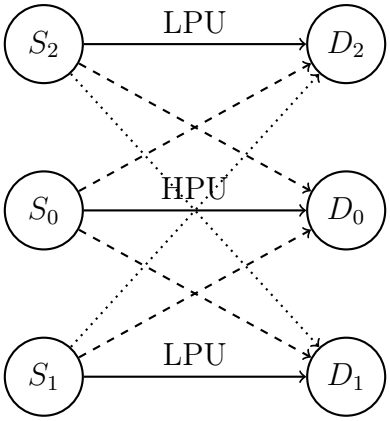


Figure 2.1: Non-cooperative network model. Example with $n = 2$. Main links illustrated with full lines, and interference links with dashed or dotted lines.

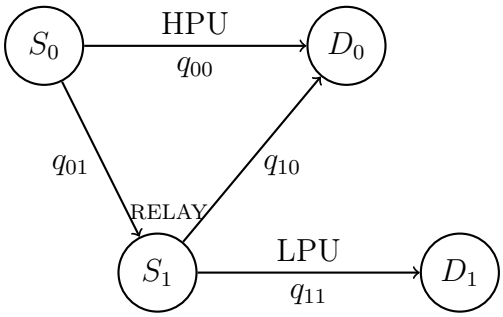


Figure 2.2: Simple cooperative network model, with success probabilities indicated in the links.

j) are shown in dashed or dotted lines. The packet transmission is always intended to the corresponding destination in the main link, and subject to interference from the other transmitting nodes.

In addition to the non-cooperative network model, we also propose the use of cooperation between one LPU and one HPU. Our analysis is restricted to a simple cooperative model, illustrated in Figure 2.2, which also shows the notation for success probability in each link. The Cooperative Underlay Scheme will be described in more detail and analyzed in Subsection 2.3.2.

A block Rayleigh fading channel is assumed in each link, with constant coefficients during one time slot. Let $h_{ij}(t)$ be the power gain of the link between a source S_i and a destination D_j during time slot t. Then

$$h_{ij}(t) = (d_{ij})^{-\alpha} c_{ij}(t), \tag{2.1}$$

where d_{ij} is the distance between source node S_i and destination node D_j, α is the path loss exponent assumed to be the same constant to all links, and $c_{ij}(t)$ are exponentially distributed random variables, which model the Rayleigh fading.

Denote with μ_{ij} the parameter that characterizes the probability distribution of c_{ij}, so that the expected value is $\mathbb{E}(c_{ij}) = \mu_{ij}$. The channel gains are assumed to be identically distributed, and independent between different links and different time slots, and defined as

$$g_{ij} = \mu_{ij} (d_{ij})^{-\alpha}. \tag{2.2}$$

We also assume that LPU transmitting with the same power have the same interference effect on other users, and are affected similarly by the transmissions of the HPU. This symmetry assumption considerably simplifies the presentation of the results, considering only the number of contending nodes, which is one of the most important parameters for the energy efficiency [21]. The asymmetric case can be handled in a similar manner, but it is necessary to keep track of the arbitrary channel coefficients, increasing the mathematical burden.

We restrict our analysis to the case of single user detectors, assuming that the receivers treat the interfering signals as noise. Let $\gamma_i(t)$ denote the Signal-to-

Interference-plus-Noise Ratio (SINR) for user i at time slot t, defined as

$$\gamma_i(t) = \frac{h_{ii}(t)P_i}{\sigma^2 + \sum_{k\neq i} h_{ki}(t)P_k}, \tag{2.3}$$

where σ^2 is the power of thermal noise at the receiver, P_i is the power used by the transmitter in link $(S_i - D_i)$, and P_k is the transmission power of the interfering nodes. We allow $P_k = 0$ if the interfering node S_k is silent during time slot t.

A generalized packet erasure channel is assumed, and we say that a packet is successfully received at the destination node D_i if $\gamma_i(t)$ exceeds a threshold denoted with β_i. This threshold is a constant depending on many system parameters, such as terminal sensitivity, modulation, coding, desired bit-error-rate (BER) and transmission rate. The complete characterization of the threshold β_i is out of the scope of this work, but it is always an increasing function of the rate, and decreasing function of the Bit Error Rate (BER). As an approximation for a feasible transmission rate, we use $r_i = \log_2(1 + \beta_i)$.

Instantaneous error-free Acknowledgment Message (ACK) is assumed to be available to the source nodes through a dedicated control channel of negligible bandwidth. The success probability for a packet transmitted through link $(S_i - D_i)$ is described by $\mathbb{P}(\gamma_i(t) > \beta_i)$, calculated as in (2.4), for any number of interfering nodes. The complete derivation can be found in [66].

$$\mathbb{P}(\gamma_i(t) > \beta_i) = \exp\left(-\frac{\beta_i\sigma^2}{g_{ii}P_i}\right)\prod_{k\neq i}\left(1 + \beta_i\frac{g_{ki}P_k}{g_{ii}P_i}\right)^{-1}. \tag{2.4}$$

Let N be the number of LPU that are active in a time slot. We denote with $q_0^{(N)}$ the probability that destination D_0 will successfully receive a packet sent by source node S_0, under the interference of N other LPU, $N \in \{0, 1, \ldots, n\}$. Given the

assumption of a random access channel, with a common transmission probability ρ for all LPU, the number of interfering nodes is a binomial random variable, $N \sim \mathcal{B}(n, \rho)$. Under the symmetry assumption, the success probabilities are the same for all LPU.

To define the transmission power, we assume that partial channel state information (statistics) is available at the transmitter. The transmission power of the HPU is then selected to achieve a target success probability $q_0^{(0)} = \theta$, $(0 < \theta < 1)$.

Using (2.4), it results that

$$P_0 = -\frac{\beta_0 \sigma^2}{g_{00} \log(\theta)}. \tag{2.5}$$

In the case of Underlay spectrum sharing, the LPU are allowed to transmit under some power constraint. To model the power constraint imposed to the LPU, we define an interference tolerance factor, denoted with τ, $(0 < \tau < 1)$, which describes the acceptable reduction in success probability of HPU, due to the presence of LPU. This new parameter leads to an alternative approach to design cognitive networks.

The total number of LPU was denoted with n. Let the HPU impose a constraint of the form

$$q_0^{(n)} \geq (1 - \tau) q_0^{(0)}.$$

That is, the worst case success probability for the HPU does not suffer a reduction greater than $\tau \times 100\%$, in comparison to the success probability without any interfering nodes.

Under the symmetry assumption, the power constraint imposed to the LPU is

$$P_i \leq \frac{P_0}{\beta_0}\frac{g_{00}}{g_{i0}}\left((1-\tau)^{-\frac{1}{n}}-1\right), \quad i=1,\ldots,n. \tag{2.6}$$

The right hand side in (2.6) is the maximum transmission power to be used by each LPU. Throughout the paper, the index $i = 1$ is used to identify parameters related to the LPU. The maximum transmission power will be denoted with P_1^{max}.

2.3 Throughput and Energy Efficiency

2.3.1 Non-cooperative cognitive network

We assume that n LPU share the channel with the HPU, without cooperation. Low-priority users transmit in a time slot following the etiquette defined by the spectrum sharing scheme, as described in Section 2.2.

The throughput is calculated in bits per second multiplying the feasible transmission rate r_i, the transmission probability ρ_i, and the success probability q_i. That is, the throughput is defined as $T_i = r_i\rho_i q_i$ *bits per second*.

The energy efficiency is calculated as the ratio between the number of bits successfully transmitted in a time slot of duration L and the energy spent in that time slot. The energy efficiency, in *bits per Joule*, is then defined as

$$\eta_i = \frac{T_i L}{P_i L} = \frac{T_i}{P_i}. \tag{2.7}$$

In order to describe the throughput of the HPU we define the auxiliary variable

$$a_0 = \beta_0 \frac{g_{10}P_1}{g_{00}P_0}. \tag{2.8}$$

Also, according to the power allocation for the HPU, defined in (2.5), we have

$$\theta = \exp\left(-\frac{\beta_0 \sigma^2}{g_{00} P_0}\right). \tag{2.9}$$

Theorem 2.1 (Throughput of High-Priority User)**.** *The throughput of the HPU, when it transmits with fixed feasible rate r_0, sharing the spectrum with n LPU, using the Interweave, Underlay, and Hybrid schemes, is given respectively by:*

$$T_{0,U} = \theta r_0 \rho_0 \left(1 - \frac{a_0}{1+a_0}\rho_U\right)^n, \tag{2.10}$$

$$T_{0,I} = \theta r_0 \rho_0 \left(1 - \frac{a_0}{1+a_0}\rho_I(1-Q_d)\right)^n, \tag{2.11}$$

$$T_{0,H} = \theta r_0 \rho_0 \left(1 - \frac{a_0}{1+a_0}\left(\rho_I(1-Q_d) + \rho_U Q_d\right)\right)^n. \tag{2.12}$$

Proof: The throughput of the HPU with the Underlay spectrum sharing scheme is calculated averaging the success probability with respect to the number of interfering nodes, as

$$T_{0,U} = r_0 \rho_0 \sum_{i=0}^{n} \binom{n}{i} \rho_U^i (1-\rho_U)^{n-i} q_0^{(i)}. \tag{2.13}$$

The result in (2.10) is obtained using the definition in (2.4) for the success probability $q_0^{(i)}$, as well as the definitions in (2.8) and (2.9) that simplify the final expression.

Next, consider the Interweave spectrum sharing scheme. If k out of n LPU fail to detect the signal from the HPU during spectrum sensing, and i out of k LPU transmit in that time slot, then the transmission of the HPU is successful with probability $q_0^{(i)}$. The throughput is then calculated as

$$T_{0,I} = r_0 \rho_0 \sum_{k=0}^{n} \binom{n}{k} (1-Q_d)^k Q_d^{n-k} \sum_{i=0}^{k} \binom{k}{i} \rho_I^i (1-\rho_I)^{k-i} q_0^{(i)}. \tag{2.14}$$

Using the success probability as defined in (2.4), and the definitions in (2.8) and (2.9), after some algebra, we obtain (2.11).

To calculate the throughput with the Hybrid spectrum sharing scheme, consider n LPU, and k out of n missed detections. Among these k LPU, each will transmit with probability ρ_I. But the group of LPU that correctly detected the transmission from the HPU may also transmit in the same time slot. Among these $n-k$ users, each will transmit with probability ρ_U. The throughput is calculated as

$$T_{0,H} = r_0\rho_0 \sum_{k=0}^{n} \binom{n}{k} (1-Q_d)^k Q_d^{n-k} \sum_{i=0}^{k} \binom{k}{i} \rho_I^i (1-\rho_I)^{k-i} \times \sum_{j=0}^{n-k} \binom{n-k}{j} \rho_U^j (1-\rho_U)^{n-k-j} q_0^{(i+j)} \tag{2.15}$$

After some algebra, we obtain (2.12), using the success probability in (2.4), and definitions (2.8) and (2.9). ■

The energy efficiency of the HPU is then calculated dividing the expressions (2.10)-(2.12) by the transmission power P_0, defined in (2.5).

Next we describe the value of transmission power that maximizes the energy efficiency of the HPU with Underlay spectrum sharing, if all the other parameters are held constant.

Proposition 2.1 (Optimal Transmission Power with Underlay)**.** *With the Underlay spectrum sharing scheme, the transmission power P_0 that maximizes the energy efficiency η_0 for the HPU is $P_0^\star = \frac{\beta_0\sigma^2}{g_{00}}$.*

Proof: Let each LPU transmit with power $P_i = mP_1^{max}$, $i = 1, 2, \ldots, n$, where m is a positive real constant value, common to all LPU. It is straightforward

to verify that the proposition holds for any value of m. Using the expression for P_0 in (2.5), and the P_1^{max} as in (2.6), the auxiliary variable defined in (2.8) assumes the value $a_0 = m\left[(1-\tau)^{-1/n} - 1\right]$.

From the definition in (2.7), and the expression for throughput of the HPU using Underlay spectrum sharing in (2.10), the energy efficiency of the HPU can be written as

$$\eta_{0,U} = -\rho_0 r_0 \frac{g_{00}}{\beta_0 \sigma^2} \left[1 - \rho_U \frac{a_0}{1+a_0}\right]^n \theta \log(\theta). \tag{2.16}$$

Therefore, the energy efficiency can be written in the form $\eta_{0,U} = K\theta\log(\theta)$, where K does not depend on θ. With all other parameters constant, an extremum of η_1 with respect to θ occurs in the open interval $(0,1)$, and satisfies $\partial\eta_{0,U}/\partial\theta = 0$. As a result, the extremum is obtained with $\theta^\star = e^{-1}$. Using (2.5), we obtain $P_0^\star = \frac{\beta_0\sigma^2}{g_{00}}$.

■

Note that $P_0^\star$ does not optimize the throughput for the HPU, as the throughput is an increasing function of power, but it optimizes energy efficiency for fixed transmission rate r_0, probability ρ_0, and tolerance factor τ.

The result in Proposition 2.1 still holds exactly for Interweave and Hybrid spectrum sharing schemes if the detection probability is held constant with respect to the power P_0, while other parameters in the detector are adjusted accordingly. In our numerical evaluations we observe that the result in Proposition 2.1 is a good approximation even if Q_d is varying with P_0

The throughput for one LPU (target LPU, $i = 1$), is calculated under the symmetry assumptions described in Section 2.2. To simplify the final expressions,

we define

$$a_1 = \beta_1 \frac{g_{01} P_0}{g_{11} P_1}, \tag{2.17}$$

$$\nu = \exp\left(-\frac{\beta_1 \sigma^2}{g_{11} P_1}\right). \tag{2.18}$$

Additionally, define the probabilities that a LPU will access the channel in the case of Hybrid spectrum sharing scheme, with a Busy or Idle channel, respectively, as

$$\rho_H^B = (1 - Q_d)\rho_I + Q_d \rho_U \tag{2.19}$$

$$\rho_H^I = (1 - Q_f)\rho_I + Q_f \rho_U \tag{2.20}$$

Theorem 2.2 (Throughput of a Low-Priority User)**.** *Under the symmetry assumption, the throughput of a LPU, when it transmits with fixed feasible rate r_1, sharing the spectrum with $(n-1)$ other LPU and the HPU, subject to the constraints imposed by the Underlay, Interweave, and Hybrid schemes, is given respectively by:*

$$T_{1,U} = \nu r_1 \rho_U \left(\frac{\rho_0}{(1 + a_1)} + (1 - \rho_0)\right)\left(1 - \rho_U \frac{\beta_1}{1 + \beta_1}\right)^{n-1} \tag{2.21}$$

$$\begin{aligned} T_{1,I} = \nu r_1 \rho_I \Bigg[& \frac{\rho_0 (1 - Q_d)}{(1 + a_1)} \left(1 - \rho_I (1 - Q_d) \frac{\beta_1}{1 + \beta_1}\right)^{n-1} \\ & + (1 - \rho_0)(1 - Q_f) \left(1 - \rho_I (1 - Q_f) \frac{\beta_1}{1 + \beta_1}\right)^{n-1} \Bigg] \end{aligned} \tag{2.22}$$

$$\begin{aligned} T_{1,H} = \nu r_1 \Bigg[& \frac{\rho_0}{(1 + a_1)} \rho_H^B \left(1 - \rho_H^B \frac{\beta_1}{1 + \beta_1}\right)^{n-1} \\ & + (1 - \rho_0) \rho_H^I \left(1 - \rho_H^I \frac{\beta_1}{1 + \beta_1}\right)^{n-1} \Bigg] \end{aligned} \tag{2.23}$$

Proof: The proof follows the same reasoning used in the case of the HPU, averaging the success probabilities with respect to the number of interfering nodes. When n LPU are active in a time slot, the transmission suffers interference of i LPU, $i \in \{0, 1, \ldots, n-1\}$. The definitions in (2.4) and (2.18) yield

$$q_1^{I,i} = \nu(1+\beta_1)^{-i}, \tag{2.24}$$
$$q_1^{B,i} = \nu\left(1+\beta_1\frac{g_{01}P_0}{g_{11}P_1}\right)^{-1}(1+\beta_1)^{-i}.$$

In the case of Underlay spectrum sharing, each interfering node is active with probability ρ_U, and there are $(n-1)$ LPU that may cause interference, in addition to one HPU, which is active with probability ρ_0. As a result, the throughput is calculated as

$$\begin{aligned} T_{1,U} = & r_1\rho_U\Bigg[\rho_0\sum_{i=0}^{n-1}\binom{n-1}{i}\rho_U^i(1-\rho_U)^{n-1-i}q_1^{B,i} \\ & +(1-\rho_0)\sum_{i=0}^{n-1}\binom{n-1}{i}\rho_U^i(1-\rho_U)^{n-1-i}q_1^{I,i}\Bigg]. \end{aligned} \tag{2.25}$$

We substitute the success probabilities as defined in (2.24), and utilize the auxiliary variables in (2.17) and (2.18) to simplify the expressions. After some algebra, we obtain the final formulation in (2.21).

With the Interweave scheme, the number of interfering nodes depends on the spectrum sensing results. First consider the case in which the HPU is active. If k out of $(n-1)$ LPU cannot detect the signal from the HPU, and i out of k LPU transmit in the time slot, the success probability of the target LPU is $q_1^{B,i}$. If the HPU is not active, the LPU that do not emit a false alarm may access the channel. If l out of $(n-1)$ LPU do not emit false alarm, and j out of l LPU decide to

transmit, the success probability of the target LPU is $q_1^{I,j}$. Following this reasoning, the throughput is calculated as

$$\begin{aligned} T_{1,I} = {} & r_1\rho_0(1-Q_d)\rho_I \sum_{k=0}^{n-1} \binom{n-1}{k} (1-Q_d)^k Q_d^{n-1-k} \\ & \times \sum_{i=0}^{k} \binom{k}{i} \rho_I^i (1-\rho_I)^{k-i} q_1^{B,i} \\ & + r_1(1-\rho_0)(1-Q_f)\rho_I \sum_{l=0}^{n-1} \binom{n-1}{l} (1-Q_f)^l Q_f^{n-1-l} \\ & \times \sum_{j=0}^{l} \binom{l}{j} \rho_I^j (1-\rho_I)^{l-j} q_1^{I,j}. \end{aligned} \tag{2.26}$$

Using the success probabilities as defined in (2.24), the auxiliary variables in (2.17) and (2.18), and routine algebra, we obtain the final expression in (2.22).

In the case of Hybrid scheme, we separate the calculations in the two cases: busy channel, when the target LPU transmits together with the HPU, and idle channel, when the target LPU may suffer interference only from other LPU. In the case of a busy channel, suppose that k out of the remaining $(n-1)$ LPU fail to detect the signal from the HPU. Each of these k will transmit with probability ρ_I. Say i out of k LPU will transmit in the time slot. The other LPU that correctly detected the presence of the HPU may also transmit in the same time slot, and each one does so with probability ρ_U. Following this reasoning, the throughput of the target LPU with Hybrid spectrum sharing is calculated as follows:

$$
\begin{aligned}
T_{1,H} = {} & r_1 \rho_0 \rho_H^B \sum_{k=0}^{n-1} \binom{n_2 - 1}{k} (1 - Q_d)^k Q_d^{n-1-k} \\
& \times \sum_{i=0}^{k} \binom{k}{i} \rho_I^i (1 - \rho_I)^{k-i} \\
& \times \sum_{j=0}^{n-1-k} \binom{n-1-k}{j} \rho_U^j (1 - \rho_U)^{n-1-k-j} q_1^{B,i+j} \\
& + r_1 (1 - \rho_0) \rho_H^I \sum_{k=0}^{n-1} \binom{n-1}{k} (1 - Q_f)^k Q_f^{n-1-k} \\
& \times \sum_{i=0}^{k} \binom{k}{i} \rho_I^i (1 - \rho_I)^{k-i} \\
& \times \sum_{j=0}^{n-1-k} \binom{n-1-k}{j} \rho_U^j (1 - \rho_U)^{n-1-k-j} q_1^{I,i+j}. \qquad (2.27)
\end{aligned}
$$

Once more, we substitute the success probabilities as defined in (2.24), and the auxiliary variables in (2.17) and (2.18), to obtain the final expression in (2.23).

■

The energy efficiency of the target LPU is obtained dividing the expressions in (2.21)-(2.23) by the transmission power, which has to satisfy the constraint in (2.6).

2.3.2 Cooperative Cognitive Network

In this Subsection, we describe the proposed spectrum sharing scheme, inspired by the non-cooperative Underlay scheme. The Cooperative Underlay scheme assumes the network topology illustrated in Figure 2.2.

In each time slot, the HPU and the LPU transmit simultaneously. The LPU relays packets from the HPU as a repayment for the caused interference. Error-free acknowledgment messages are assumed to be available to both users through a

control channel of negligible bandwidth. If the primary destination fails to receive a packet, then a cooperation frame is initiated. In a cooperation frame, the HPU will retransmit the packet until either the destination D_0 or the relay node S_1 receive it. If the relay node receives the packet before the destination, it will assume the responsibility to deliver the packet.

Let X denote the number of time slots in a cooperation frame. Let X_{00} be the number of time slots used for transmission between S_0 and D_0 using the direct link, X_{01} be the number of time slots used for transmission between S_0 and S_1, and X_{10} denote the number of time slots used by S_1 to relay the packet to D_0. The variable X is described by

$$X = \begin{cases} X_{01} + X_{10} & \text{if } \{X_{01} < X_{00}\}, \\ X_{00} & \text{else} \end{cases} \tag{2.28}$$

Following the notation for success probability under interference, we write $q_{ij}^{(1)}$ for the success probability of a transmission from source node i to destination node j under interference of the other user. The variable $X_{ij} \in \{1, 2 \ldots\}$ has Geometric distribution with parameter q_{ij} and the expected value is $\mathbb{E}(X_{ij}) = 1/q_{ij}$. We are interested in the probability $\mathbb{P}(\{X_{01} < X_{00}\})$, and the calculation steps are shown below.

We calculate the probability of the event $\{X_{00} \leq X_{01}\}$, conditioning on X_{01},

$$\begin{aligned} \mathbb{P}(\{X_{00} \leq X_{01}\}) &= \sum_{x=1}^{\infty} \left(1 - (1 - q_{00})^x\right) \mathbb{P}(X_{01} = x) \\ &= 1 - \sum_{x=1}^{\infty} (1 - q_{00})^x (1 - q_{01})^{x-1} q_{01} \\ &= \frac{q_{00}}{q_{00} + q_{01}(1 - q_{00})}. \end{aligned} \tag{2.29}$$

Then calculate $\mathbb{P}(\{X_{01} < X_{00}\}) = 1 - \mathbb{P}(\{X_{00} \leq X_{01}\})$, and we obtain

$$\mathbb{P}(\{X_{01} < X_{00}\}) = \frac{q_{01}(1-q_{00})}{q_{00} + q_{01}(1-q_{00})}. \tag{2.30}$$

The expected value of the random variable X defined in (2.28) will be denoted with $\bar{X}$, and calculated using the result in (2.30),

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X|X_{01} < X_{00}]\mathbb{P}(X_{01} < X_{00}) \\
&\quad + \mathbb{E}[X|X_{00} \leq X_{01}]\mathbb{P}(X_{00} \leq X_{01}) \\
&= \mathbb{E}[X_{01} + X_{10}|X_{01} < X_{00}]\mathbb{P}(X_{01} < X_{00}) \\
&\quad + \mathbb{E}[X_{00}|X_{00} \leq X_{01}]\mathbb{P}(X_{00} \leq X_{01}) \\
&= \mathbb{E}[\min\{X_{00}, X_{01}\}] + \mathbb{E}[X_{10}]\mathbb{P}(X_{01} < X_{00}) \\
&= \frac{q_{10} + q_{01}(1-q_{00})}{q_{10}(q_{00} + q_{01}(1-q_{00}))}.
\end{aligned} \tag{2.31}$$

Figure 2.3 illustrates an example with a few time slots in which the relay assumes the responsibility to deliver the packet. The cooperation frame is shaded and identified with (HPU+R). In the detail, the HPU retransmits the packet for a number X_{10} of time slots, until the relay successfully receives it. The relay then transmits for X_{10} consecutive slots, until the destination receives the packet successfully.

Theorem 2.3 (Throughput and Energy Efficiency of High-Priority User With Cooperative Underlay)**.** *Let the HPU transmit packets with k_0 bits, in time slots of duration L seconds, using power P_0. The throughput of the HPU, in bits per second, is calculated as*

$$T_0^{(C)} = \frac{k_0}{L}\left(q_{00}^{(1)} + \frac{1 - q_{00}^{(1)}}{\bar{X} + 1}\right). \tag{2.32}$$

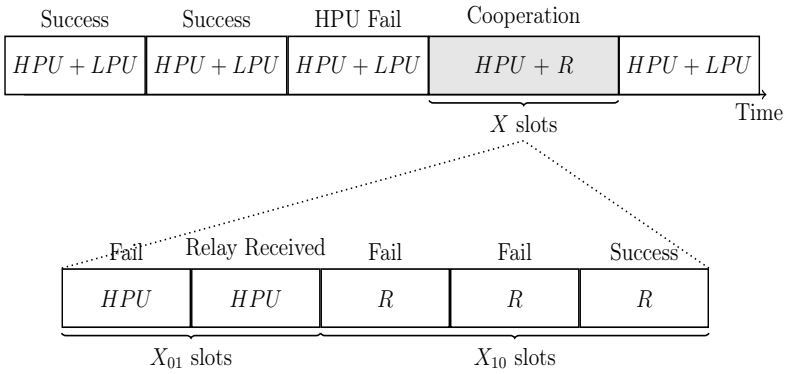


Figure 2.3: Cooperative Underlay: Cooperation frame example.

The energy efficiency of the HPU, in bits per Joule, is

$$\eta_0^{(C)} = \frac{k_0}{P_0 L}\left(1 + \frac{(1-q_{00}^{(1)})}{q_{00} + q_{01}(1-q_{00})}\right)^{-1}. \tag{2.33}$$

Proof:

The expression for the throughput follows from the cooperation scheme, as the HPU transmission will require one time slot with probability $q_{00}^{(1)}$, and it will require an average of $\bar{X} + 1$ if a cooperation frame is initiated.

The energy to transmit a single packet is described by the random variable

$$E_0^{(C)} = \begin{cases} P_0 L, & q_{00}^{(1)}, \\ P_0 L(1+X_{01}), & (1-q_{00}^{(1)})\mathbb{P}(\{X_{01} < X_{00}\}), \\ P_0 L(1+X_{00}), & (1-q_{00}^{(1)})(1-\mathbb{P}(\{X_{01} < X_{00}\})). \end{cases} \tag{2.34}$$

Let $\mathbb{E}(X_{ij}) = \bar{X}_{ij}$. The expected value is calculated assuming independence of

X_{00} and X_{01}, using the results in (2.30), which yields

$$\begin{aligned}\bar{E}_0^{(C)} &= P_0 L q_{00}^{(1)} \\ &\quad + P_0 L\, \mathbb{E}[1 + X_{01} | X_{01} < X_{00}](1 - q_{00}^{(1)})\mathbb{P}(X_{01} < X_{00}) \\ &\quad + P_0 L\, \mathbb{E}[1 + X_{00} | X_{00} \leq X_{01}](1 - q_{00}^{(1)})\mathbb{P}(X_{00} \leq X_{01}) \\ &= P_0 L\, \left(1 + (1 - q_{00}^{(1)})\mathbb{E}[\min\{X_{00}, X_{01}\}]\right) \\ &= P_0 L \left(1 + \frac{1 - q_{00}^{(1)}}{q_{00} + q_{01}(1 - q_{00})}\right). \end{aligned} \tag{2.35}$$

The energy efficiency, in *bits per Joule*, is

$$\eta_0^{(C)} = \frac{k_0}{\bar{E}_0^{(C)}}, \tag{2.36}$$

which, together with (2.35), yields (2.33). ■

Theorem 2.4 (Throughput and Energy Efficiency of Low-Priority User With Cooperative Underlay)**.** *Let the LPU transmit packets with k_1 bits, in time slots of duration L seconds, using power P_1.*

The throughput of the LPU, in bits per second, is

$$T_1^{(C)} = \frac{k_1 q_{11}^{(1)}}{L} \left(q_{00}^{(1)} + \frac{1 - q_{00}^{(1)}}{\bar{X} + 1}\right). \tag{2.37}$$

The energy efficiency of the LPU, in bits per Joule, is

$$\eta_1^{(C)} = \frac{k_1 q_{11}^{(1)}}{P_1 L} \left(1 + \frac{q_{01}(1 - q_{00})(1 - q_{00}^{(1)})}{q_{10}(q_{00} + q_{01}(1 - q_{00}))}\right)^{-1}. \tag{2.38}$$

Proof:

The throughput of the LPU is affected by the time slots spent relaying packets for the HPU. A transmission from the LPU is successful with probability $q_{11}^{(1)}$, and

it takes one time slot if the simultaneous transmission from HPU is also successful, which occurs with probability $q_{00}^{(1)}$. Otherwise, if a cooperation frame is initiated, the average interval is $1 + \bar{X}$ time slots.

To calculate the energy efficiency, we include the energy spent in relaying. The energy used by the LPU to transmit a single packet using the Cooperative Underlay scheme is described by the random variable

$$E_1^{(C)} = \begin{cases} P_1 L, & q_{00}^{(1)}, \\ P_1 L(1 + X_{10}), & (1 - q_{00}^{(1)})\mathbb{P}(\{X_{01} < X_{00}\}), \\ P_1 L, & (1 - q_{00}^{(1)})(1 - \mathbb{P}(\{X_{01} < X_{00}\})). \end{cases} \tag{2.39}$$

The expected value is calculated as in the case of HPU, and we obtain

$$\bar{E}_1^{(C)} = P_1 L \left(1 + \frac{q_{01}(1 - q_{00})(1 - q_{00}^{(1)})}{q_{10}(q_{00} + q_{01}(1 - q_{00}))} \right). \tag{2.40}$$

The energy efficiency, in *bits per Joule*, is then obtained calculating the ratio

$$\eta_1^{(C)} = \frac{k_0 q_{11}^{(1)}}{\bar{E}_0^{(C)}}, \tag{2.41}$$

which, together with (2.40), yields (2.38). ■

Consider now that the relay node has the ability to transmit and receive simultaneously (full-duplex). Note that the relay node S_1 is a common node between links $(S_0 - S_1)$ and $(S_1 - D_1)$, and a transmission in $(S_1 - D_1)$ would cause interference with the reception in $(S_0 - S_1)$. In this case, the self-interference in the relay node has to be taken into account.

Different methods have been proposed to address the self-interference problem, but the study of self-cancellation mechanisms is out of the scope of this work. It

is assumed that the node S_1 is capable of applying some interference cancellation mechanism, which is modeled as a (deterministic) variable power gain between the transmitter and receiver [31]. The self-interference is parametrized by a scalar $g \in [0, 1]$, with $g = 1$ if no interference cancellation mechanism is applied, and $g = 0$ if perfect cancellation is achieved.

The probability that the relay node S_1 will successfully receive the packet from S_0, while it was transmitting its own data in the same time slot is

$$\begin{aligned} q_{01}^{(1)} &= \mathbb{P}\left(\frac{h_{01}P_0}{\sigma^2 + gP_1} > \beta_1\right) \\ &= \mathbb{P}\left(h_{01}P_0 > \beta_1(\sigma^2 + gP_1)\right) \\ &= \exp\left(-\frac{\beta_1(\sigma^2 + gP_1)}{g_{01}P_0}\right). \end{aligned} \tag{2.42}$$

The superscript (SI) is used in what follows to identify the expressions as results that consider the self-interference.

Theorem 2.5 (Throughput and Energy Efficiency of High-Priority User with Full-Duplex Relay)**.** *Let the HPU transmit packets with k_0 bits, in time slots of duration L seconds, using power P_0.*

The throughput for the HPU, in bits per second, using Cooperative Underlay with full-duplex relay, is

$$T_0^{(SI)} = \frac{k_0}{L}\left(q_{00}^{(1)} + \frac{(1 - q_{00}^{(1)})(1 - q_{01}^{(1)})}{\bar{X} + 1} + \frac{(1 - q_{00}^{(1)})q_{01}^{(1)}}{\bar{X}_{10} + 1}\right).$$

The energy efficiency for the HPU, in bits per Joule, using Cooperative Underlay with full-duplex relay, is

$$\eta_0^{(SI)} = \frac{k_0}{P_0 L}\left(1 + \frac{(1 - q_{00}^{(1)})(1 - q_{01}^{(1)})}{q_{00} + q_{01}(1 - q_{00})}\right)^{-1}. \tag{2.43}$$

Proof:

A transmission from the HPU is successful in one time slot with probability $q_{00}^{(1)}$. Otherwise two cases unfold. If the relay received the packet in the first time, it takes an average of $1 + X_{10}$ slots until D_0 receives the packet. If the relay did not receive the packet, which occurs with probability $1 - q_{01}^{(1)}$, then a cooperation frame is initiated, and the packet transmission takes $1 + X$ time slots.

The energy spent by the HPU to transmit one packet is described by the random variable

$$E_0^{(SI)} = \begin{cases} P_0 L & q_{00}^{(1)} + (1 - q_{00}^{(1)}) q_{01}^{(1)}, \\ P_0 L(1 + X_{01}), & (1 - q_{00}^{(1)})(1 - q_{01}^{(1)}) \mathbb{P}(\{X_{01} < X_{00}\}), \\ P_0 L(1 + X_{00}), & (1 - q_{00}^{(1)})(1 - q_{01}^{(1)})(1 - \mathbb{P}(\{X_{01} < X_{00}\})). \end{cases}$$

The expected value is calculated similarly to the other cases, and as a result

$$\bar{E}_0^{(SI)} = P_0 L \left(1 + \frac{(1 - q_{00}^{(1)})(1 - q_{01}^{(1)})}{q_{00} + q_{01}(1 - q_{00})} \right). \tag{2.44}$$

Energy efficiency, in *bits per Joule*, is calculated as the ratio between bits per packet and average energy per packet,

$$\eta_0^{(SI)} = \frac{k_0}{\bar{E}_0^{(SI)}}, \tag{2.45}$$

which, together with (2.44), yields (2.43). ■

2.4 Performance Trade-Offs

In this Section we discuss important trade-offs and challenges in the design of energy efficient cognitive wireless networks, given the multitude of parameters to be selected.

For the generalized collision channel model, we assume that all users require a fixed SINR, and we set the threshold $\beta_i = 0.2$, $i = 0, 1, \ldots, n$. High-priority and low-priority source nodes are assumed to transmit with a fixed, feasible rate. We assume unit bandwidth, and use Shannon's formula to relate transmission rate and β_i, as an approximation, that is, $r_i = \log_2(1+\beta_i)$ *bits per second*. The thermal noise added at the receiver is assumed to have power 1 mW. Unless otherwise stated, we suppress the channel power gains, making $g_{ij} = 1$.

For the channel access, we assume that the PU transmits with probability $\rho_0 = 0.5$. The LPU transmits with probabilities $\rho_U = 0.5$ and $\rho_I = 1$, with the Underlay and Interweave schemes, respectively. The interference tolerance is set to $\tau = 0.1$ in the results for non-cooperative networks, and $\tau = 0.2$ in the results for cooperative networks, unless otherwise is stated.

In most figures, we change the transmission power of the HPU varying θ over the open interval $(0, 1)$. The transmission power P_0 is determined by (2.5). We assume that the LPU transmits with maximum power, as determined by the constraint in (2.6). Therefore, increasing θ increases the transmission power of both HPU and LPU, and we can observe the energy-throughput trade-off. The detection probability will also change when P_0 increases, producing interesting effects in the performance when spectrum sensing is required from the LPU.

2.4.1 Sensing Accuracy and Energy Efficiency

Reliability in spectrum sensing is desirable from the point of view of the HPU and the LPU, as it reduces interference and improves efficiency in spectrum usage. In this Subsection, we discuss the effect of the detection accuracy on energy efficiency, assuming that all the LPU can detect the presence of the HPU with probability Q_d, and emit a false alarm with probability Q_f.

A number of methods are presented in the literature to detect the presence of signals in the channel, including matched filtering, energy detection, cyclostationarity, and process models [67], [68]. To illustrate the trade-off between the accuracy and energy efficiency, we use the formulations for Q_d and Q_f in the case of energy detection under Rayleigh fading presented in closed form in [69]. Nonetheless, our observations are not restricted to the sensing technique, since only the probabilities of error for a given transmission power are taken into account.

Figure 2.4 illustrates the trade-off between energy efficiency the detection probability, increasing the Signal-to-Noise Ratio (SNR). The transmission power of both HPU and LPU is increasing, and that is the reason we observe a proper trade-off, with a single maximum for each curve. The value of Q_d that maximizes energy efficiency of LPU is lager than the one that maximizes the energy efficiency of the HPU, hence the LPU may have more interest in sensing accuracy than the HPU. Note that larger values of energy efficiency for the LPU in comparison to the HPU come at the expense of a smaller throughput for the LPU, accompanied by smaller values of transmission power.

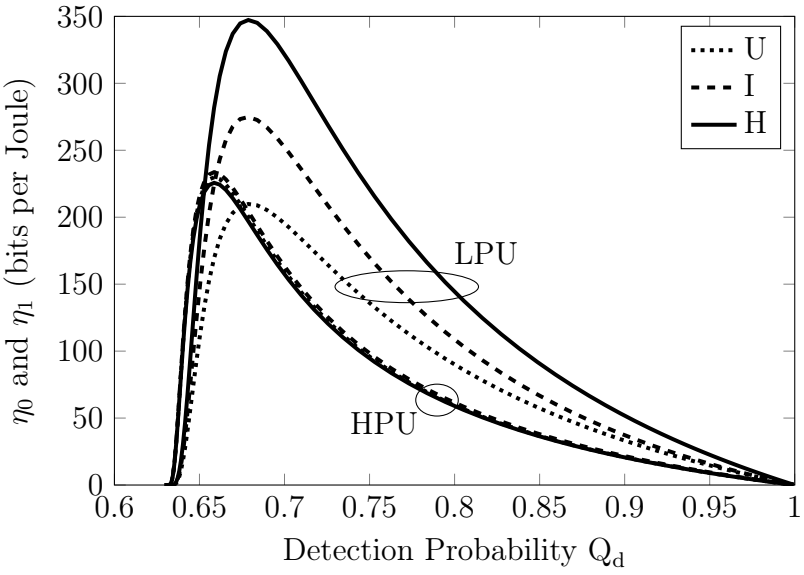


Figure 2.4: Energy efficiency versus detection probability with $Q_f = 0.1$. LPU benefits from Hybrid spectrum sharing. LPU is more energy efficient due to lower power, but has lower throughput.

Figure 2.5 shows the effect of false alarm probabilities, with $Q_f = 0.1$ and $Q_f = 0.2$. The figure is generated as Figure 2.4, increasing the SNR to change Q_d. We observe that Q_f can be selected so that the maximum energy efficiency is shifted to a different value of Q_d (and the corresponding transmission powers imposed by the PU). The reduction in the maximum energy efficiency when the false alarm rate increases is due to the missed transmission opportunities that result in a penalty in the throughput of the LPU.

Note that the curves for Underlay scheme are presented for comparison. This scheme does not utilize the spectrum sensing, but the detection probabilities are varying due to an increase in power transmission, which will also affect the performance of the Underlay scheme.

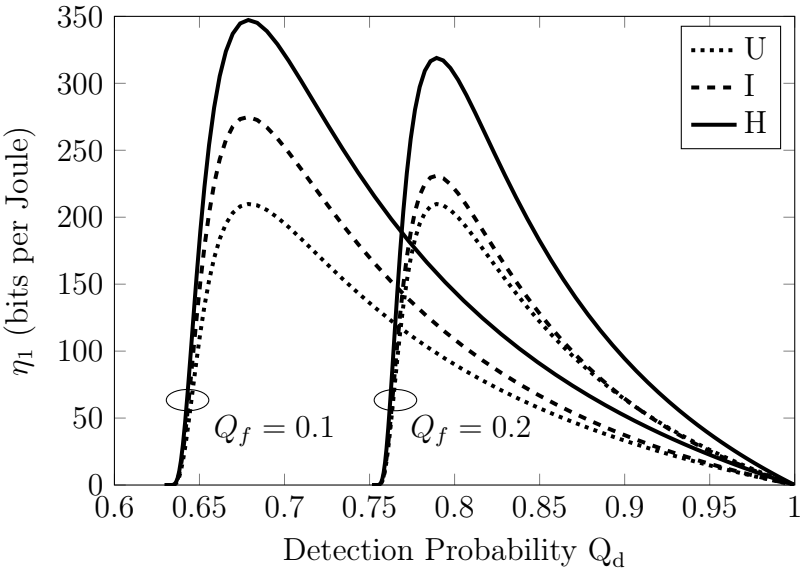


Figure 2.5: Energy efficiency of LPU versus detection probability. Effect of false alarm rate.

2.4.2 Performance Trade-Offs in Non-Cooperative Networks

Figure 2.6 illustrates the effect of the power constraint imposed on LPU. For a single user, the maximum transmission power may be larger than P_0, depending on the required SINR by the HPU. In this numerical example, $P_1^{max} = 1.25P_0$, and the energy efficiency of the LPU is slightly larger than that for the HPU. As the number of interfering nodes increases, the transmission power of the LPU is reduced to protect the HPU. Note that the HPU remains oblivious to the number of LPU, and its performance remains virtually unaltered. Figure 2.7 illustrates the energy-throughput trade-off for the LPU, increasing θ and the transmission powers. The throughput of the LPU is not strictly increasing with power if Interweave or Hybrid spectrum sharing is used, resulting in peculiar curves. This result indicates that even if some desired throughput value is achieved, the energy efficiency of the LPU

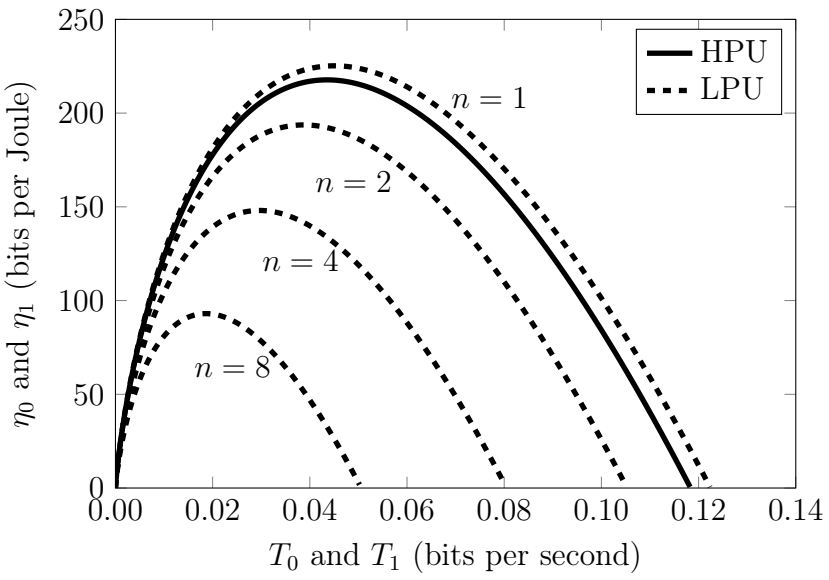


Figure 2.6: Energy-throughput trade-off with multiple LPU and Underlay spectrum sharing. Effect of number of interfering nodes with $\tau = 0.2$.

may be significantly reduced, depending on the parameters selected by the HPU, the conditions for spectrum sensing, and the constraints on channel access.

The effect of the channel power gains is shown in Figure 2.8, which depicts the energy-throughput trade-off for a single LPU. In contrast with the case analyzed in Figure 2.7, the present case considers that the transmission power P_0 remains fixed, and also the detection probability remains fixed. The distance between source node S_1 and destination node D_0 is increased, so that the interference caused by the LPU is attenuated. We assume an attenuation factor $\alpha = 3$. In this case, the maximum transmission power of the LPU is allowed to increase, even though the power of the HPU remains constant. Note that the effects caused by spectrum sensing are no longer observed.

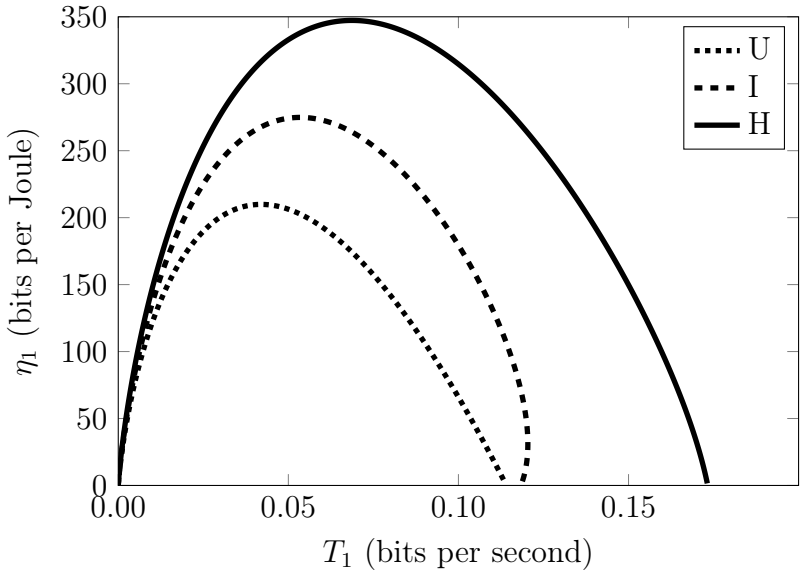


Figure 2.7: Energy-throughput trade-off for LPU changing θ. Power increases with θ for both HPU and LPU.

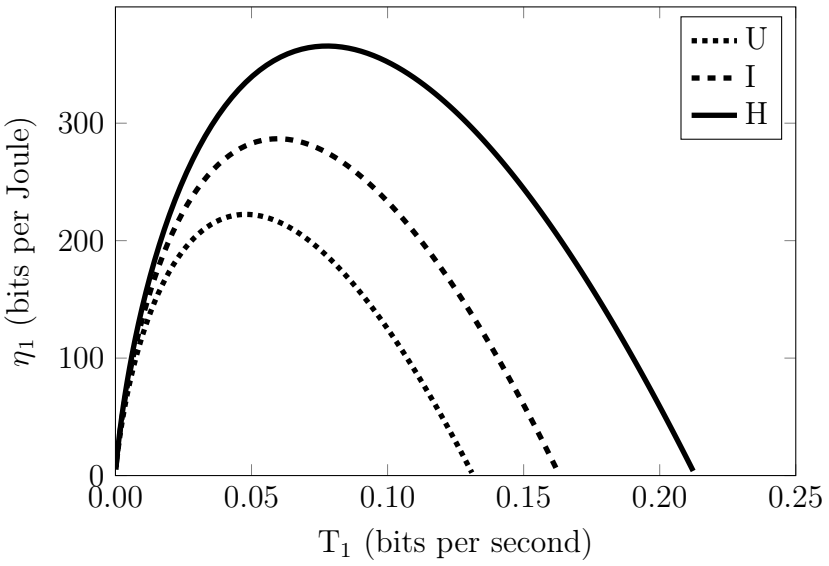


Figure 2.8: Energy-throughput trade-off changing channel quality. For fixed P_0, power constraint of the LPU is relaxed due to attenuation.

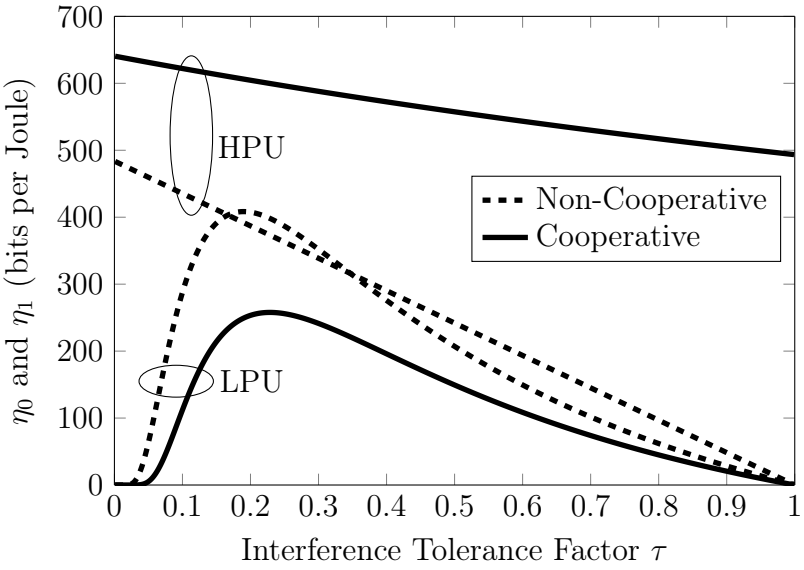


Figure 2.9: Energy efficiency versus interference tolerance factor τ, for non-cooperative and Cooperative Underlay spectrum sharing. Half-duplex relay. $\theta = \exp(-1)$.

2.4.3 Performance Trade-Offs in Cooperative Networks

The energy efficiency for both the HPU and the LPU is depicted in Figure 2.9 versus the interference tolerance factor τ. For the HPU, the advantage of using cooperation increases with the interference caused by the LPU. For the LPU, we observe a trade-off between energy efficiency and τ, since the maximum transmission power increases with τ. Consequently, it is not straightforward that the LPU should increase transmission power, even if the HPU accepts more severe interference.

Figure 2.10 presents the energy efficiency of the HPU versus τ. Reducing the parameter g is equivalent to a more efficient self-interference cancellation technique. The highest values of energy efficiency are achieved with perfect cancellation ($g = 0$), which can be regarded as an upper bound for the energy efficiency of this cooperative

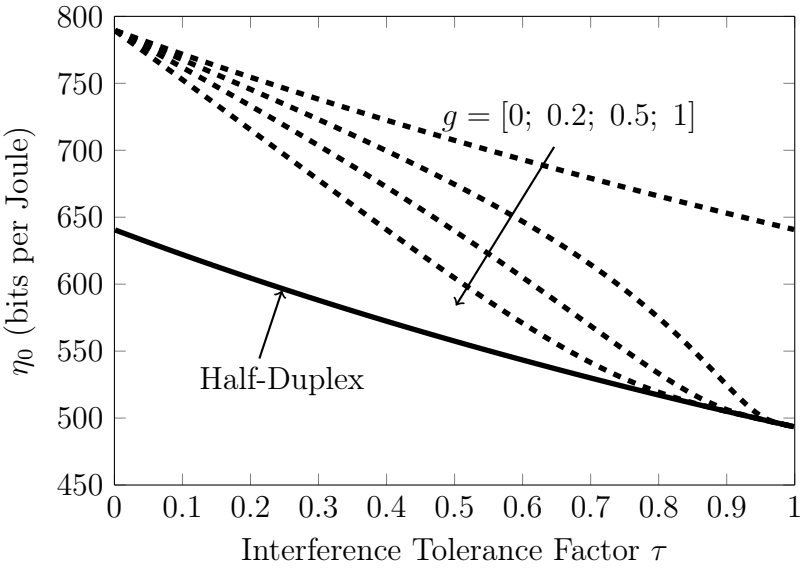


Figure 2.10: Energy efficiency of HPU versus interference tolerance factor τ with cooperative underlay spectrum sharing and $\theta = \exp(-1)$. Effect of self-interference cancellation: no cancellation when $g = 1$, and perfect cancellation when $g = 0$.

scheme. The corresponding energy-throughput trade-off is shown in Figure 2.11, where the advantage of a full-duplex relay node is validated once more.

2.5 A Game Model for Cooperative Cognitive Networks

In the present chapter, we have already investigated one scheme for cooperation between the HPU and the LPU. The scheme presented earlier was based on the Underlay spectrum sharing, and assumed that the LPU could transmit simultaneously with the HPU, as long as it was willing to cooperate when necessary.

In this Section we present a different approach for cooperation, assuming that the LPU will be allocated part of the resources for exclusive use, in exchange for a period of cooperation. The system model consists of two point-to-point wireless links sharing the same channel, one pair is the HPU, while the other pair is the

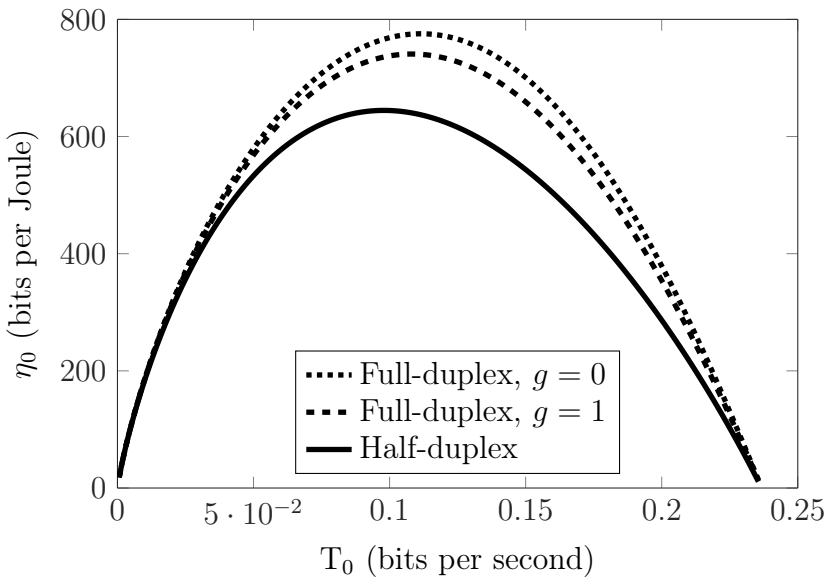


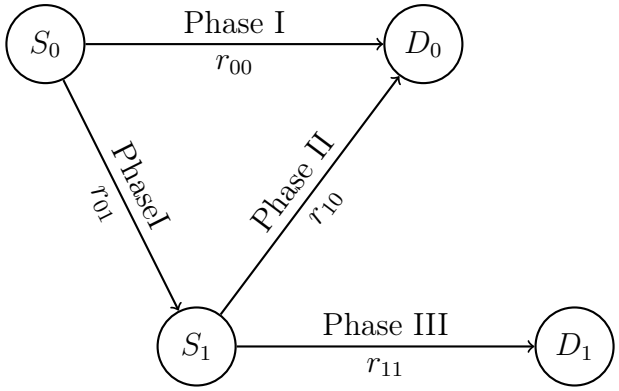
Figure 2.11: Energy-throughput trade-off for HPU with cooperation from LPU. Effect of self-interference cancellation: no cancellation when $g = 1$, and perfect cancellation when $g = 0$. Interference tolerance $\tau = 0.2$.

LPU. The cooperative spectrum sharing scenario analyzed in this Section assumes that the low priority source node relays packets from the high priority source node, in exchange for an interval to transmit its own packets.

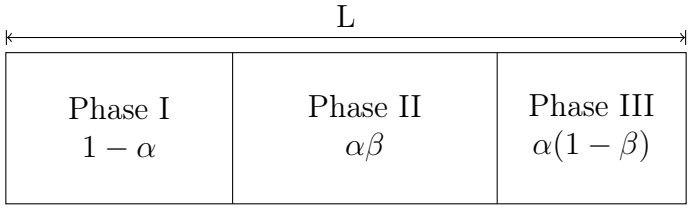
We consider a time frame of duration L, to be allocated between three activities: the high- and low-priority users' individual transmissions, and the cooperative relaying. Each time frame is associated with a cooperation cycle, characterized by two variables, α and β. In each frame, the channel is allocated as follows [38]:

- Phase I: only the HPU transmits its data for $(1 - \alpha)L$ seconds, $(0 \leq \alpha \leq 1)$;
- Phase II: the LPU relays HPU's data to D_0 for $\alpha\beta L$ seconds, $(0 \leq \beta \leq 1)$;
- Phase III: the LPU transmits its own data for $\alpha(1 - \beta)L$ seconds.

Figure 2.12(a) illustrates the network model with the four nodes. The HPU



(a) Network Model



(b) Time Frame Model

Figure 2.12: System model for game theoretic approach for cooperation. The network model in 2.12(a) shows the HPU and the LPU, which is also a relay. The phase of the resource allocation scheme and the transmission rate are shown in each link. The time allocated for each phase within a frame is illustrated in 2.12(b)

and LPU are identified with index $i = 0,\ 1$, respectively. The source nodes are denoted with S_0 and S_1, and the destination nodes with D_0 and D_1. The picture indicates the communication links that are active in each phase, and the transmission rates in each link. Figure 2.12(b) presents the channel allocation model for one time frame, to be determined by the game model.

We consider once more the block Rayleigh fading channel, assuming the channel gains to be constant during one time frame of duration L. The nodes transmit with power P, and White Gaussian Noise is added at the receiver. Denote with γ_{ij} the SNR of signal from source node i received at node j, with $i, j \in \{0, 1\}$. The transmission rate is approximated using the Shannon capacity formulation. That

is, the transmission rate between nodes i and j is given by $r_{ij} = \log_2(1 + \gamma_{ij})$.

A Decode-and-Forward (DF) relaying method is assumed at the LPU, hence at phase II the LPU forwards the fully decoded message received from the source of HPU.

2.5.1 Game Definition

A reputation-based Stackelberg game is proposed to model the interactions between the HPU and the LPU in a cooperative spectrum sharing scenario. A two-player game is defined, with the HPU as a leader and the LPU as a follower, to obtain the optimum time allocation of spectrum in a fair and energy-efficient manner.

In what follows, the round k of the game will be identified with the superscript k, $k \in \{1, 2, \ldots\}$. The users' parameters are identified by the index i, $i \in \{0, 1\}$. The game is run for each time slot. In each round of the game, the players interact to define the variables α^k and β^k, which determine the time allocation for HPU transmission, cooperation, and LPU transmission, as described in Figure 2.12. The two source nodes are assumed to communicate with each other using a low rate dedicated control channel.

The decisions of the users take into account the reputation of their opponent, represented by a *cooperation credit*. In order to keep track of the users' reputation through a cooperation credit, the users' strategies in each round are defined as intermediate variables, which have impact on both the cooperation credit, and the

values of α^k and β^k.

The remaining of this Subsection describes in more detail the four elements that define our game model, namely the strategies, the cooperation credits, the recursion rules for α and β, and the utility functions.

2.5.1.1 Users' Strategies

The strategy of user i in round k is defined as an intermediate variable s_i^k, taking values in a set $\mathcal{S}$. The set $\mathcal{S}$ is assumed to be a closed interval on the real line, $\mathcal{S} = [s_{min}, s_{max}]$, where $0 \leq s_{min} < s_{max} \leq 1$.

In general, the strategy s_i^k incurs variation of user i's cooperation during game round k. If the opponent has positive credit history, s_i^k will result in an increase of the i-th user cooperation time, while for an opponent with negative credit history, s_i^k results in a reduction of the i-th user cooperation time.

In a Stackelberg game, the strategies are presented sequentially. In each round, the leader HPU selects s_0^k which optimizes its own utility, denoted with $U_0^k(s_0^k, s_1^k)$, assuming that the LPU is a rational player and it will respond to the leader's action with its best strategy. In other words, the HPU maximizes its own utility anticipating the reaction of the LPU. The best response of the LPU is the strategy that maximizes its own utility, $U_1^k(s_0^k, s_1^k)$, and it depends not only on the action of the HPU, but also on the cooperation credits and other parameters. The detailed description of the optimization problems is provided in Subsection 2.5.2.

2.5.1.2 Cooperation Credits

The cooperation credits are defined as a mechanism to encourage cooperation by using the reputation of the users. The reputation of the HPU is based on its willingness to lease the spectrum to the LPU. For the LPU, the reputation is based on how reliable it is in forwarding the relayed packets from the HPU. Both users are encouraged to sustain a good reputation, so that they can benefit from cooperation in subsequent periods. In the common Stackelberg game models, the solution of each round of the game is obtained by the one-shot backward-induction process [70]. In our proposed Stackelberg game model, the reputation of the users is included in this process, encouraging the cooperation among the users.

The cooperation credit, denoted with C_i^k, encapsulates the history of the cooperative behavior of the user, reflecting its willingness to cooperate in consecutive game rounds up to, but not including, round k. To represent this concept of willingness, C_i^k assumes values on a symmetric interval $[-C, C]$, with negative values representing lack of cooperation, and positive values representing willingness to cooperate. The use of a single parameter avoids the need to keep track of all the actions of each user, saving memory space and simplifying the game model.

The cooperation credit is calculated using a recursion rule, with initial value C_i^0, updated based on the user's selected strategy and on the opponent's willingness to cooperate. The credit should be reduced if the opponent's credit is positive, but the selected strategy is not cooperative. The credit should be increased if the opponent's credit is positive and the selected strategy is cooperative. With this

reasoning, we define the credit change as

$$\Delta C_i^k = (2s_i^k - 1)sgn(C_{-i}^k),\ k \geq 0, \tag{2.46}$$

where C_{-i}^k is the credit of the opponent, and $sgn(x)$, $x \in \mathbb{R}$ is the Sign function, defined as $+1$ for non-negative values, and -1 for negative values of x.

To accumulate credit history, the recursion rule is defined with initial value C_i^0 and updated as

$$C_i^{k+1} = C_i^k + \Delta C_i^k,\ k \geq 0. \tag{2.47}$$

2.5.1.3 Recursion Rule for α and β

The ultimate goal in each round of the game is to define $\alpha^k \in [0, 1]$ and $\beta^k \in [0, 1]$. The game is initiated with values $\alpha^0 = 0.5$ and $\beta^0 = 0.5$. The recursion rules that update these variables are as follows:

$$\alpha^k = \max\left(0, \min(\alpha^{k-1} + \alpha_s s_0^k C_1^k, 1)\right),\ k \geq 1, \tag{2.48}$$

$$\beta^k = \max\left(0, \min(\beta^{k-1} + \beta_s s_1^k C_0^k, 1)\right),\ k \geq 1, \tag{2.49}$$

where the functions max and min are used to bound the values in the interval $[0, 1]$, and α_s and β_s are nonzero constant steps to modify α and β, respectively.

2.5.1.4 Utility Functions

To complete the game definition, we define the utility functions of the players. The proposed utility functions account for energy efficiency and fairness in cooperative spectrum sharing, and consist of two parts, namely the throughput utility

$U_{i,t}$ and the energy utility $U_{i,e}$, where subindex i identifies the HPU ($i = 0$) and the LPU ($i = 1$). In round k of the game the users select their strategies s_i^k with the objective of maximizing the throughput and minimizing energy. Therefore, the utility functions are of the form

$$U_0^k(s_0^k, s_1^k) = U_{0,t}^k - U_{0,e}^k, \tag{2.50}$$

$$U_1^k(s_0^k, s_1^k) = U_{1,t}^k - U_{1,e}^k. \tag{2.51}$$

The energy utilities $U_{i,e}$ are functions of the energy spent with transmission during the time slot k, and they introduce a cost of transmission to improve energy efficiency. The energy utilities are defined as

$$U_{0,e}^k = \delta_0(1 - \alpha^k)P, \tag{2.52}$$

$$U_{1,e}^k = \delta_1\alpha^k(1 - \beta^k)P + \delta_2\alpha^k\beta^k P, \tag{2.53}$$

where δ_0, δ_1 and δ_2 are normalizing coefficients, necessary to make rate and energy comparable. We will refer to the coefficients δ_0 and δ_1 as *transmission costs* for HPU and LPU, respectively. We will refer to δ_2 as the *cooperation cost coefficient*.

The throughput utility of each user is defined as the logarithm of its achievable rate to incorporate the fairness in spectrum sharing. Although we consider a non-cooperative game, the use of logarithmic utility functions still assigns higher priority to the user with lower effective rate, approximating the proportional fair resource allocation introduced in [71].

For the LPU, the transmission using a single link between S_1 and D_1, with rate r_{11}, results in the following expression for the throughput utility:

$$U_{1,t}^k = \log(1 + \alpha^k(1 - \beta^k)r_{11}). \tag{2.54}$$

Remark 1: For the LPU, we sum one unit to the effective rate, to avoid the utility function convergence problem at $\alpha^k = 0$ and assign a zero value to it for no cooperation between the users.

For the HPU, the transmission using both the direct path and the relay path requires a more elaborate expression for the throughput utility. A single-relay system with decode-and-forward (DF) and Time Division (TD) was considered in [72], where one cooperation cycle is divided into two phases: the relay is either in receive (RX) or in transmit (TX) mode.

In our proposed system model, we assumed that the HPU is silent in the second stage. Let the transmission powers be P. Considering the Decode and Forward (DF) relaying method at the LPU, the achievable rate of the HPU is [72]

$$\begin{aligned} R_0 = \min \Bigg\{ & \frac{1-\alpha}{2} \log_2(1 + |h_{01}|^2 P), \\ & \frac{1-\alpha}{2} \log_2(1 + |h_{00}|^2 P) + \frac{\alpha\beta}{2} \log_2(1 + |h_{10}|^2 P) \Bigg\} \end{aligned} \tag{2.55}$$

The final expressions for the utility functions are

$$U_0^k(s_0^k, s_1^k) = \log\left(\min\left\{\frac{1-\alpha^k}{2} r_{01}, \frac{1-\alpha^k}{2} r_{00} + \frac{\alpha^k\beta^k}{2} r_{10}\right\}\right) - \delta_0(1-\alpha^k)P, \tag{2.56}$$

$$U_1^k(s_0^k, s_1^k) = \log\left(1 + \alpha^k(1-\beta^k)r_{11}\right) - \delta_1\alpha^k(1-\beta^k)P - \delta_2\alpha^k\beta^k P. \tag{2.57}$$

2.5.2 Equilibrium Analysis

In this Subsection, the equilibrium of the proposed Stackelberg game model is analyzed. In each round of the game, the players take their actions sequentially, in order of priority, noticing the credit of each other. At the first stage, the HPU (leader) optimizes its utility, under the assumption that the LPU is rational and it selects its best strategy in response to the leader's strategy. At the second stage, the LPU (follower) selects its optimal strategy observing the strategy of the HPU. Hence, the equilibrium solution can be found by a backward-induction process.

Theorem 2.6 (Stackelberg Equilibrium)**.** *The Stackelberg equilibrium of the proposed model exists in each round of the game, and it is unique if* $\alpha^k \neq 0$, $\beta^k \neq 0, 1$ *and* $C_0^k \neq 0$.

Proof: The leader of game (HPU) maximizes its own utility with respect to its own strategy s_0^k, assuming the LPU's rational reaction s_1^k. This is a single parameter maximization problem, for which the solution always exists, and is unique if for any s_0^k, there exists only one possible LPU's reaction, s_1^k. For any given strategy of the HPU, the optimal response (rational reaction) of the LPU, denoted with $\bar{s}_1^k$, is obtained by solving the following optimization problem:

$$\begin{aligned} &\max_{s_1^k} \;\; U_1^k(s_0^k, s_1^k) \\ &\text{s.t.} \;\; s_{min} \leq s_1^k \leq s_{max}. \end{aligned} \tag{2.58}$$

The inequality constraints in optimization problem (2.58) are affine functions, and the objective function is strictly concave if $\alpha^k \neq 0$ and $C_1^k \neq 0$, since the second

derivative is

$$\frac{\partial^2 U_1^k}{\partial {s_1^k}^2} = -\Big(\frac{\alpha^k \beta_s C_0^k r_{11}}{1+\alpha^k(1-\beta^k) r_{11}}\Big)^2 \leq 0. \tag{2.59}$$

Hence, under these conditions the convex optimization problem (2.58) admits a unique solution $\bar{s}_1^k$. Following the KKT conditions, the solution $\bar{s}_1^k$ assumes one of the three values in (2.60), based on the HPU's strategy and system parameters.

$$\bar{s}_1^k = \begin{cases} s_{min} \\ \dfrac{a s_0^k + b}{c s_0^k + d} \\ s_{max} \end{cases} \tag{2.60}$$

where, a,b,c and d are constant values, defined as follows:

$$a = \alpha_s C_1^k (1-\beta^{k-1}),$$

$$b = \frac{1}{r_{11}} + \alpha^{k-1}(1-\beta^{k-1}) - \frac{1}{(\delta_1 - \delta_2)P},\ \delta_1 \neq \delta_2$$

$$c = \alpha_s \beta_s C_0^k C_1^k,$$

$$d = \alpha^{k-1} \beta_s C_0^k.$$

The best response of the HPU, $\bar{s}_0^k$ is then obtained as a solution to

$$\max_{s_0^k} \ U_0^k(s_0^k, \bar{s}_1^k), \tag{2.61}$$

subject to the constraint $s_{min} \leq s_0^k \leq s_{max}$.

The special cases of $\alpha^k = 0$, $\beta^k = 0$ and $\beta^k = 1$ result in trivial time allocations: (i) individual transmission for HPU during whole time frame if $\alpha = 0$, (ii) no cooperation if $\beta = 0$, and (iii) no individual transmission for LPU if $\beta = 1$. $C_0^k =$

0 refers to the case that the LPU has no ground to judge whether the HPU is cooperative or selfish, hence it takes the strategy of the previous round.

Consequently, under the conditions $\alpha^k \neq 0$ and $C_1^k \neq 0$, the existence of the Stackelberg equilibrium is ensured. Moreover, noting equation (2.59), under these conditions the utility of the LPU is strictly concave, hence the solution of the game is unique. ■

Theorem 2.7 (Nash Equilibrium)**.** *The solution of the proposed game in each round,* $(\bar{s}_0^k, \bar{s}_1^k)$*, is a Nash equilibrium.*

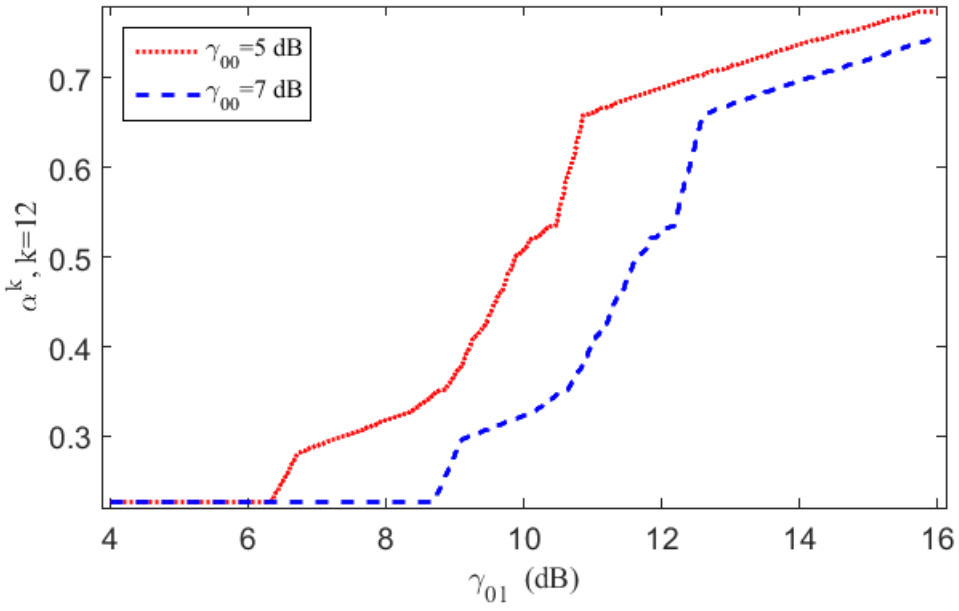
Proof: A strategy set $(s_0^{k\star}, s_1^{k\star})$ achieves Nash equilibrium if, and only if

$$\forall i \in \{0,1\},\ \forall s_i^k \in \mathcal{S},$$

$$U_i^k(s_i^{k\star}, s_{-i}^{k\star}) \geq U_i^k(s_i^k, s_{-i}^{k\star}), \tag{2.62}$$

where s_{-i}^k, denotes the strategy of the opponent of player i in round k of the game.

The Stackelberg equilibrium solution $(\bar{s}_0^k, \bar{s}_1^k)$ is obtained with the backward-induction process, as described in Theorem 1. First the leader (HPU) selects its strategy to maximize its utility $U_0^k(s_0^k, s_1^k(s_0^k))$ with respect to s_0^k, where $s_1^k(s_0^k)$ is the corresponding optimal response of the LPU. Once the strategy $\bar{s}_0^k$ of the HPU is announced, the LPU selects its best response $\bar{s}_1^k = s_1^k(\bar{s}_0^k)$, which maximizes the utility $U_1^k(\bar{s}_0^k, s_1^k)$ with respect to s_1^k. Under the conditions in Theorem 1, this optimal strategy is unique for any choice of HPU's strategy. The solution $(\bar{s}_0^k, \bar{s}_1^k)$ is achieved upon observing the best strategy of the HPU, $\bar{s}_0$ by the LPU. Therefore, at both stages of the backward-induction process, the players set their strategy as the best possible response to the other one, which follows the definition of Nash equilibrium.



;

Figure 2.13: α^k versus γ_{01} for different values of γ_{00}, $k = 12$.

In other words, $U_i^k(\bar{s}_0^k, \bar{s}_1^k)$ can not be improved by varying s_i^k, since it violates the aforementioned Stackelberg optimization procedure. Hence, it is concluded that Stackelberg solution of the game is a Nash equilibrium. ■

2.5.3 Numerical Results for Game Model

In Figure 2.13, we show the behavior of the parameter α when varying the quality of the relay channel, represented by the SNR in the link (S_0-S_1), denoted with γ_{01}. We set $\gamma_{10} = 20$ dB, $\gamma_{11} = 10$ dB. As shown in Fig. 2.13 for small values of γ_{01}, a small value is obtained for parameter α, meaning that the HPU prefers to use the direct link noting the bad quality of the relay channel. When γ_{01} becomes larger than γ_{00}, the HPU is encouraged to cooperate, since the achievable rate can be improved by using the relay. Therefore, we observe that α increases

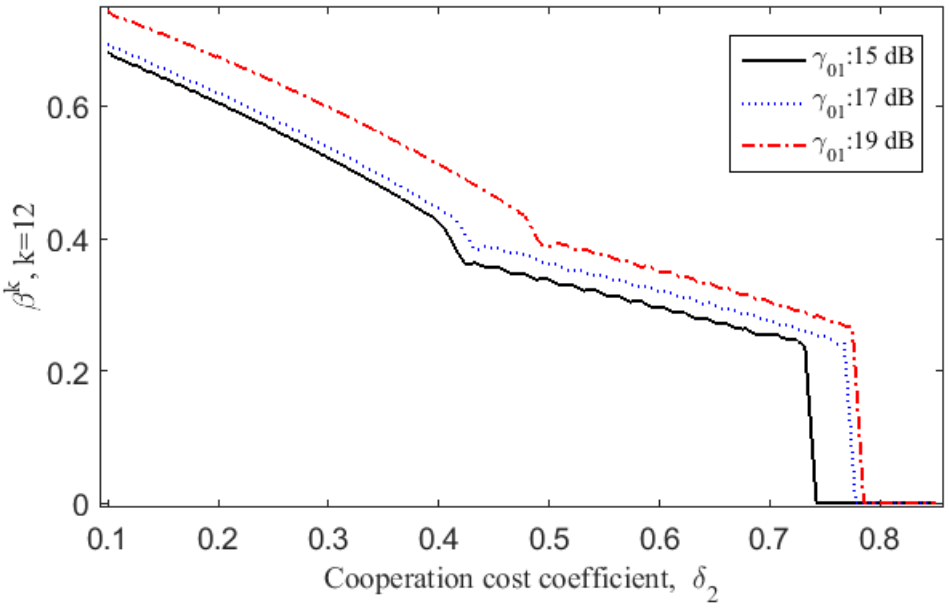
with increasing γ_{01}, resulting in a larger interval allocated for cooperation. We also observe that for a smaller value of γ_{00}, α assumes larger values. This is because a smaller γ_{00} represents worse conditions in the direct channel, which also encourages cooperation.

In Figure 2.14, we present the behavior of the parameter β while varying the cooperation cost coefficient δ_2. We set $(\delta_0, \delta_1) = (0.7, 1.8)$, $\gamma_{00} = 5$ dB, $\gamma_{10} = 20$ dB, $\gamma_{11} = 10$ dB. We observe that β decreases as the cooperation cost coefficient increases. This is because for large values of δ_2 the cooperation is discouraged, and the portion of time allocated to cooperation ($\alpha\beta$) is reduced. If δ_2 is significantly large, the LPU becomes more concerned with the energy cost, and may prefer not to cooperate with the HPU. Additionally, when the relay channel has better quality, represented by larger values of γ_{01}, the users are more encouraged to cooperate, increasing the resulting values of β.

2.6 Future Work Discussion

Even though the problem of energy efficiency in wireless networks has received considerable attention in the past few years, many challenges remain to be addressed in order to deploy truly green networks. Regarding the specific scenarios considered in this chapter, we mention the need to study more comprehensive network models and the overhead incurred by cooperation.

Additionally, the effects of spectrum sensing on the energy efficiency deserve further investigation, since the energy cost of sensing, and the penalty in throughput



;

Figure 2.14: β^k versus Cooperation Cost Coefficient δ_2, for different values of γ_{01}, $k = 12$.

due to sensing time were not included in our analysis.

Regarding the game theoretic model, the distributed mechanisms implemented as part of the game can be further investigated to obtain a fair resource allocation for heterogeneous networks. We note that it may not be interesting for the LPU to cooperate if the time allocated for its transmissions is very small, or if the amount of energy spent with cooperation is very large, not compensated by the transmission opportunity received as a reward. In this sense, incorporating fairness into the game definition is also a mechanism to encourage cooperation among users. Also, the concept of fairness in those heterogeneous scenarios is in itself an interesting topic for future research.

Game theory may also be further exploited to investigate energy efficient resource allocation for a larger number of users. That will require the use of different

game models, and it could also include other performance metrics, such as age of information, which will be described in the following chapters.

2.7 Chapter Summary

The concept of quality of service from the point of view of a user in a communication network is not a trivial one to define, and this concept should certainly account for energy-efficient solutions. We have observed that accounting for energy efficiency in the performance of wireless networks requires some compromise with respect to other metrics such as throughput or spectrum sensing accuracy, leading to interesting and complicated trade-offs.

For the particular case of cognitive wireless networks, with HPU and LPU sharing the spectrum, we proposed a new parametrization using the required success probability of the HPU, and defining a new design parameter to impose an interference tolerance. The parametrization models the interdependence of the transmission parameters for the HPU and LPU, accounting for possible interference and cooperation among users. The trade-offs between performance and energy efficiency have been described for three non-cooperative spectrum sharing schemes, and one cooperative scheme.

We have shown that the proposed cooperation scheme improves energy efficiency in comparison to the traditional non-cooperative scheme, and can be used as an alternative to compensate for caused interference. The use of low-priority nodes as full-duplex relays is shown to be beneficial for the HPU, even without

self-interference cancellation.

We have also proposed a Stackelberg game model to study spectrum sharing among the HPU and LPU in a cooperative scenario, with utility functions accounting for the energy efficiency. The solution of the game contains two parameters, α and β, which define the time allocation to individual transmissions and cooperative transmission. Our model considers the reputation of each user, represented by the cooperation credits, to encourage the cooperation and prevent misbehavior.

Chapter 3: Age of Information

3.1 Overview

Age of information is a new emerging concept, as one of the attributes of the information to be used in improving the efficiency of communication and control systems. The concept of age is relevant to systems which are sensitive to the timeliness of the information, which is the case when actions are taken based on the status of a process of interest, as in sensor networks, or feedback systems.

In this chapter, we consider communication systems in which status update messages are transmitted through a communication network as packets containing information about the status of a random process under observation, together with a time stamp to identify when the update was generated. Numerous applications of communication systems require the transmission of information about the state of a process of interest between a source and a destination. This is the case, for example, in sensor networks, where sensor nodes report the observations to a central processor, to monitor health or environment conditions [73, 74]. In these applications, the timeliness of the transmitted message is an important and often critical objective, since an outdated message may lose its value.

To characterize the timeliness of the transmitted message, we define the age of

the information available to the destination node as the time elapsed since the last received update was generated. When the application requires that fresh information about the observed process is delivered to the destination, it is not sufficient to minimize the delay associated to the transmission of a message. The frequency at which the status update messages are generated is also important. Note that update messages delivered with a small delay, but sent very infrequently, would result in outdated information available at the destination. On the other hand, given that network resources are limited, sending messages too frequently would congest the network, and the objective of having fresh information at the destination may be compromised. Once this trade-off is identified, we note that optimizing a system to deliver timely update messages is not equivalent to maximizing throughput, nor it is the same as minimizing delay. Hence, different metrics are needed to characterize the age of information.

The initial steps in this characterization of age of information use queuing systems to model the status update system. It is assumed that status update messages become available to a source node at different time instants, and could wait in a buffer before being transmitted to a destination node interested in the status of the process under observation. Previous works have analyzed the average age considering a queue model with a single server, First-Come-First-Served (FCFS) discipline, and infinite buffer capacity [49–51]. In those models with infinite buffer capacity, the packets may experience long waiting times in the buffer, becoming obsolete even before transmission.

The waste of network resources in the transmission of obsolete packets may also

occur when multiple status update messages are transmitted simultaneously with random service times. In this case, a packet containing more recent sample of the observed process may complete service before another packet that was generated before, rendering the last one obsolete. The average age in this case has been considered in [53, 54], where the authors modeled the status update system as a queue with multiple servers.

Depending on the application, it may be unnecessary to transmit all the samples obtained from the observed process. If the destination node has interest only in the current value of the process, we propose that the source node discards some of the arriving status update messages even before transmitting to the destination [33]. That is, the source node is able to manage the packets even before transmitting to the destination, possibly discarding packets, as opposed to storing them in a buffer, avoiding the waste of network resources with the transmission of outdated information. In this chapter we present three models for status update systems with packet management, modeled as queuing systems with a single server, FCFS discipline, but finite buffer. One of these models is a very peculiar queue, for which the classic result known as Little's theorem does not apply. For the systems with packet management, we characterize the average age, and discuss the limit of the average age when the packet arrival rate at the source node is very large, comparing the results with a system capable of producing *just in time* status updates which was first discussed in [49].

In addition to the average age, we also proposed a different metric, named peak age. The peak age characterizes the value of age achieved immediately before

a status update message is received. This new metric is particularly interesting when the application requires that the age of information is kept below a certain threshold. We have characterized the probability distribution of the peak age for the three models with packet management. The peak age also presents an expected value close to the expected value of age for the observed models, indicating that it is an appropriate metric to describe age, with the advantage that the average peak age can be easily obtained in many cases of interest.

The remaining of this chapter is organized as follows. In Section 3.2, we formally define the age of information in status update systems and two metrics that characterize the age. In Section 3.3, we characterize the age of information for status update systems, cosidering five different models. At first, we present the analysis for one source-destination link, modeled as a M/M/1/ queue, and discussed in [49]. Then we present the results for multiple servers, and discuss the possibility of resource waste, as presented in [53] and [54]. Finally, we discuss the age of information for three proposed packet management schemes. Section 3.4 presents our numerical results, and Section 3.5 presents a brief discussion about possible future work directions. We summarize the results in this chapter in Section 3.6.

3.2 Definitions

Consider a communication link with one source-destination pair. A random process $H(t)$ is observed, and the status of this process is available to the source node at random time instants. The destination node has interest in timely information

about the status of the process $H(t)$, but this information is not available instantaneously, since it has to be transmitted from the source through a link with limited resources. The source node obtains samples of the random process of interest $H(t)$, and transmits status updates in the form of packets, containing the information about the status of the process, and the time instant that the sample was generated. We denote the time stamps with t_k, and each status update message contains the information $\{H(t_k), t_k\}$, $k = 1, 2, \ldots$, to be transmitted in a source-destination link. Denote with t'_k the corresponding time instants at which the status update messages arrive at the destination node.

Definition 3.1. *Index of the most recently received update*

At a time instant s, we define the index of the most recently received update

$$N(s) \coloneqq \max\{k | t'_k \leq s\}, \tag{3.1}$$

Definition 3.2. *Time stamp of the most recently received update*

Once the index of the last received update is defined, we define its time stamp

$$U(s) \coloneqq t_{N(s)}. \tag{3.2}$$

Definition 3.3. *Age of Information*

Finally, we define the age of information as the random process

$$\Delta(s) \coloneqq s - U(s). \tag{3.3}$$

The age of information is given by the difference between the current time instant and the time at which the last received update was generated. The age

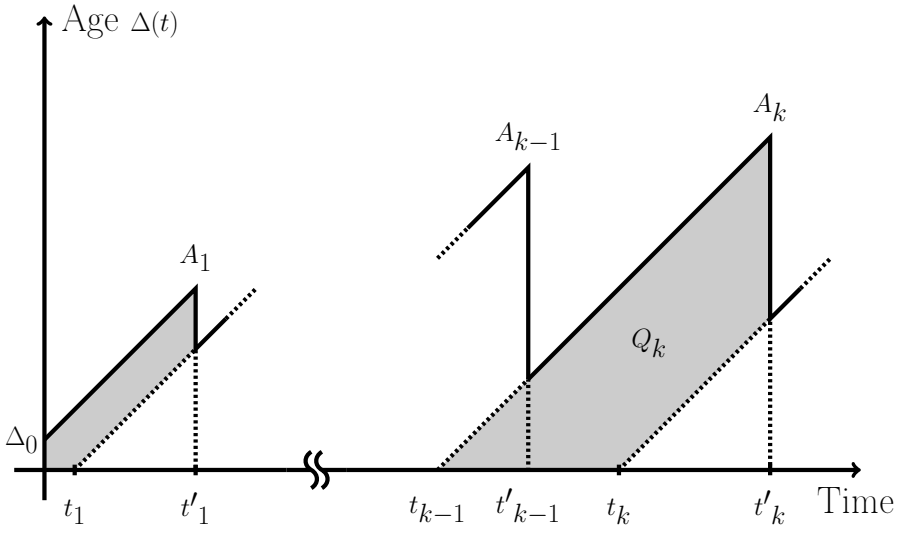


Figure 3.1: Sample path for age of information.

increases linearly with time and, upon reception of a status update message, it is reset to the difference $t'_k - t_k$. In order to better understand the behavior of the process $\Delta(t)$, we observe that a sample path would take the form of a sawtooth curve, as illustrated in Figure 3.1. Assuming an initial value of age Δ_0, the value of age will increase, until the status update message that was generated at time t_1 is received at time t'_1. The value achieved at the peak will be identified with A_1. The age increases with time, until a message generated at time t_k is received at time t'_k, and the value of age is reset to the difference $t'_k - t_k$. We identify the peak A_k, and the shaded area identified with Q_k. The properties of this sawtooth curve and the variables identified in Figure 3.1 will be exploited later in this chapter, aiming to characterize some of the properties of the process $\Delta(t)$.

3.2.1 Average Age

Assuming ergodicity of the process $\Delta(t)$, the average age can be calculated using a time average. Consider an observation interval $(0, \tau)$. The time average age is

$$\Delta_\tau = \frac{1}{\tau} \int_0^\tau \Delta(t) dt. \tag{3.4}$$

The integration in (3.4) can be interpreted as the area under the curve for $\Delta(t)$. The area can be calculated as the sum of disjoint geometric parts with areas identified by Q_k, as the one shadowed in Figure 3.1, with $k = 1, 2, \ldots N(\tau)$, and $N(\tau)$ as defined in (3.1). In the case that $\tau > t'_{N(\tau)}$, there will also be a partial area to be added in the end, which we will denote with $\tilde{Q}$. Summing all the areas, we write the time average age as

$$\begin{aligned} \Delta_\tau &= \frac{1}{\tau} \left(Q_1 + \tilde{Q} + \sum_{k=2}^{N(\tau)} Q_k \right) \\ &= \frac{Q_1 + \tilde{Q}}{\tau} + \frac{N(\tau) - 1}{\tau} \frac{1}{N(\tau) - 1} \sum_{k=2}^{N(\tau)} Q_k. \end{aligned} \tag{3.5}$$

The average age is calculated as we take the length of the observation interval to infinity

$$\Delta = \lim_{\tau \to \infty} \Delta_\tau. \tag{3.6}$$

The ratio of the number of transmitted packets by the length of the interval converges to the rate of transmitted packets, referred to as the effective arrival rate. We define

Definition 3.4. *Effective Arrival Rate*

Given an observation interval τ, and the number of packets transmitted during this interval as defined in (3.1)*, we define the effective packet arrival rate as the limit*

$$\lambda_e \coloneqq \lim_{\tau \to \infty} \frac{N(\tau)}{\tau}. \tag{3.7}$$

Also, as $\tau \to \infty$, the number of transmitted packets grows to infinity, $N(\tau) \to \infty$. Due to the ergodicity of Q_k, the average age can be calculated as

$$\Delta = \lambda_e \mathbb{E}[Q_k] \tag{3.8}$$

To calculate the area, Q_k, we take the area of the bigger isosceles triangle with sides $T_{k-1} + Y_k$, and subtract the area of the smaller triangle with sides T_k. The average area is

$$\begin{aligned} \mathbb{E}[Q_k] &= \frac{1}{2}\mathbb{E}[(T_{k-1} + Y_k)^2] - \frac{1}{2}\mathbb{E}[T_k^2] \\ &= \frac{1}{2}\mathbb{E}[Y_k^2] + \mathbb{E}[T_{k-1}Y_k], \end{aligned} \tag{3.9}$$

where the second equality follows from the fact that T_{k-1} and T_k are equally distributed.

3.2.2 Peak Age

Depending on the application, it may be necessary to characterize the maximum value of the age of information immediately before an update is received. It may also be desirable to optimize the system so that the age remains below a threshold with a certain probability.

With that motivation, we propose an alternative metric to the average age, obtained observing the peak values in the sawtooth curve. The peak age provides information about the worst case age, and its expected value can be easily calculated in many cases of interest.

Observe the example sample path shown in Figure 3.1. Consider the peak corresponding to the kth successfully received packet. The value of age in the peak is denoted with A_k. We present the definition of peak age below.

Definition 3.5. *Peak Age*

Let T_{k-1} be the time it takes to successfully transmit the $(k-1)$th packet, that is, $T_{k-1} = t'_{k-1} - t_{k-1}$. Let Y_k be the time elapsed between service completion of the $(k-1)$th packet and the service completion of the kth packet, that is, $Y_k = t'_k - t'_{k-1}$. The value of age achieved immediately before receiving the kth update is called peak age, and defined as

$$A_k \coloneqq T_{k-1} + Y_k. \tag{3.10}$$

3.3 Age of Information in Status Update Systems

In this section, we calculate the average age and the peak age for specific system models. We represent the random process of interest $H(t)$, which is observed in different time instants t_k, $k = 1, 2, \ldots$. Each observation becomes immediately available to the source node as a packet containing $\{H(t_k), t_k\}$. Hence, t_k is also regarded as the arrival time of the packet to the source node. For all the models discussed in this section, the arrival process is modeled as a Poisson process of rate λ,

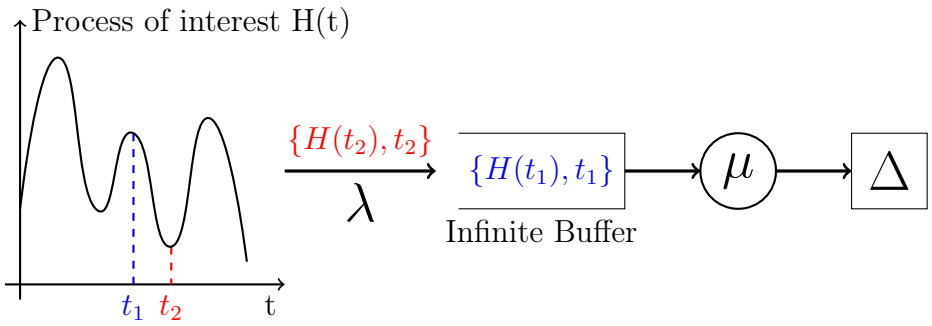


Figure 3.2: Model for status update system with a single link, and infinite capacity buffer. The system is modeled as a M/M/1 queuing system.

and packets are transmitted using a FCFS policy. The transmission of a packet takes a random amount of time, which depends on channel conditions such as fading, and network conditions such as congestion. As a simplifying assumption, we consider that the time for transmission of a packet is exponentially distributed, with mean $1/\mu$. At the destination, we are interested in characterizing the timeliness of the information available about the process $H(t)$.

3.3.1 Status Updates Through a Single Link

Consider a single source-destination link. The status update messages arrive at the source node according to a Poisson process of rate λ, and should be transmitted through the communication link to the destination. This link is modeled as a single server, and service time is assumed to be exponentially distributed with parameter μ. If the server is found busy upon arrival of a packet, this packet can wait for its turn in an infinite capacity buffer. This system is modeled as a M/M/1 queue, illustrated in Figure 3.2.

We illustrate a sample path for this model in Figure 3.3. Recall that t_k denotes

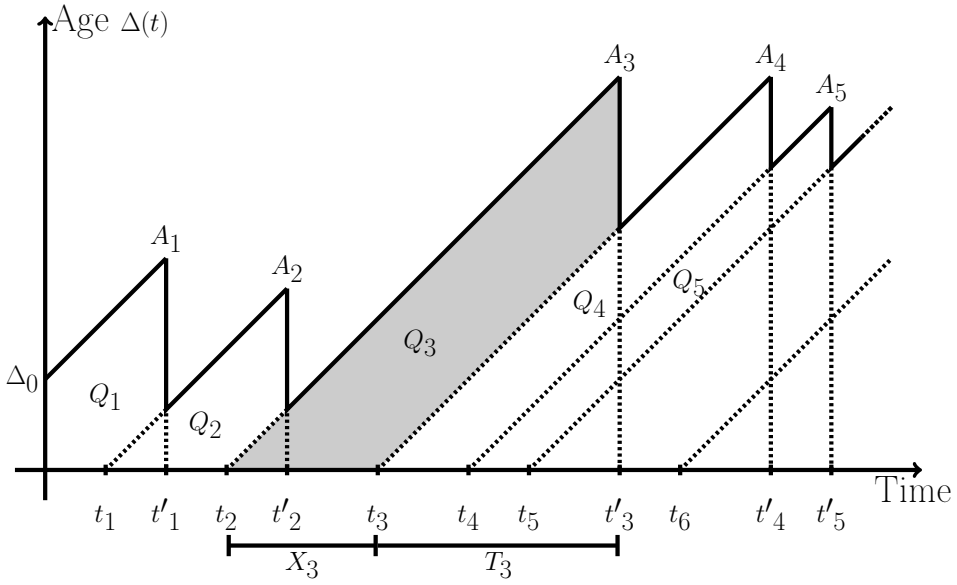


Figure 3.3: Sample path for age of information, illustrating how the age evolves in time in the case of a M/M/1 queuing model.

the time instant the kth packet was generated, and t'_k the time instant that this packet completes service. The kth interarrival time is defined as $X_k \coloneqq t_k - t_{k-1}$, and the time in the system is defined as $T_k \coloneqq t'_k - t_k$. We identify X_3 and T_3 in the figure. We also indicate the peak age A_k, and the areas of the geometric parts Q_k. The trapezoid with area Q_3 is shadowed in the figure.

The average age for the M/M/1 model has been presented in [49]. The calculation follows the steps presented in subsection 3.2.1, that is, the average age is the product of the effective arrival rate λ_e and the expected area of the trapezoids $\mathbb{E}[Q_k]$, as illustrated in the sawtooth curve in Figure 3.3. In the M/M/1 model, every packet is admitted into the queuing system. Hence, the effective arrival rate is equal to the arrival rate λ. Disregarding the area identified with Q_1, the areas of the trapezoids, Q_k, can be calculated as the difference of the areas of two triangles,

which yields

$$\begin{aligned}\mathbb{E}[Q_k] &= \mathbb{E}\left[\frac{(X_k+T_k)^2}{2} - \frac{T_k^2}{2}\right] \\ &= \mathbb{E}\left[\frac{X_k^2}{2} + X_k T_k\right]. \end{aligned} \tag{3.11}$$

In the M/M/1 model, the interarrival times X_k are exponentially distributed with parameter λ, with a second moment $\mathbb{E}[X_k^2] = 2/\lambda^2$. To calculate the second term in (3.11), we write the time in the system as the sum of a waiting time and a service time, $T_k = W_k + S_k$. The waiting time is zero if and only if the previously transmitted packet has already completed service. We can write $W_k = (T_{k-1} - X_k)^+$, where $(\cdot)^+$ represents $\max\{\cdot, 0\}$. Clearly, the time in the system T_k and the interarrival time X_k are not independent. To calculate the second term in (3.11), we proceed as follows:

$$\begin{aligned}\mathbb{E}\left[X_k T_k\right] &= \mathbb{E}\left[X_k\left((T_{k-1} - X_k)^+ + S_k\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[X_k\left((T_{k-1} - X_k)^+\right)|X_k\right]\right] + \mathbb{E}[X_k]\mathbb{E}[S_k], \end{aligned} \tag{3.12}$$

where we have used the independence of the service times and interarrival times in the second equality.

To calculate the conditional expectation we use the probability distribution of the time in the system for a M/M/1 queue, denoted with $f_T(t)$ [75]. Let $\rho \coloneqq \lambda/\mu$, and write

$$\begin{aligned}\mathbb{E}\left[\left((T_{k-1} - x)^+\right)|X_k = x\right] &= \int_x^\infty (t-x) f_T(t) dt \\ &= \int_x^\infty (t-x)\mu(1-\rho)e^{-\mu(1-\rho)t} dt \\ &= \frac{e^{-\mu(1-\rho)x}}{\mu(1-\rho)}. \end{aligned} \tag{3.13}$$

Using the result in (3.13), we calculate the first term in (3.12) using the distribution of the interarrival times, denoted with $f_X(x)$, to obtain

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[X_k\left((T_{k-1}-X_k)^+\right)|X_k\right]\right] &= \int_0^\infty x\mathbb{E}\left[\left((T_{k-1}-x)^+\right)|X_k=x\right]f_X(x)dx \\ &= \int_0^\infty x\frac{e^{-\mu(1-\rho)x}}{\mu(1-\rho)}\lambda e^{-\lambda x}dx \\ &= \frac{\rho}{\mu^2(1-\rho)} \end{aligned} \tag{3.14}$$

Using (3.14), and the expected values for interarrival and service times, we calculate the expected value in (3.12) as

$$\mathbb{E}\left[X_kT_k\right] = \frac{\rho}{\mu^2(1-\rho)} + \frac{1}{\lambda}\frac{1}{\mu}. \tag{3.15}$$

Finally, using (3.15) in (3.11), we calculate the average age $\Delta = \lambda\mathbb{E}[Q_k]$, which in the case modeled as M/M/1 queue yields [49]

$$\Delta_{M/M/1} = \frac{1}{\lambda} + \frac{1}{\mu} + \frac{\lambda^2}{\mu-\lambda}. \tag{3.16}$$

3.3.2 Status Updates Through Multiple Paths

Consider a communication system designed to deliver status update messages from a source node to a destination node using a network, with the status update messages arriving at the source node according to a Poisson process of rate λ. In [53] and [54], the authors investigated the case that multiple paths are available to transmit information between source and destination. The results in [53] show the average age in the extreme case that a status update message is transmitted immediately upon arrival at the source node, assuming there is always a link available. The time to deliver the message was assumed to be exponentially distributed,

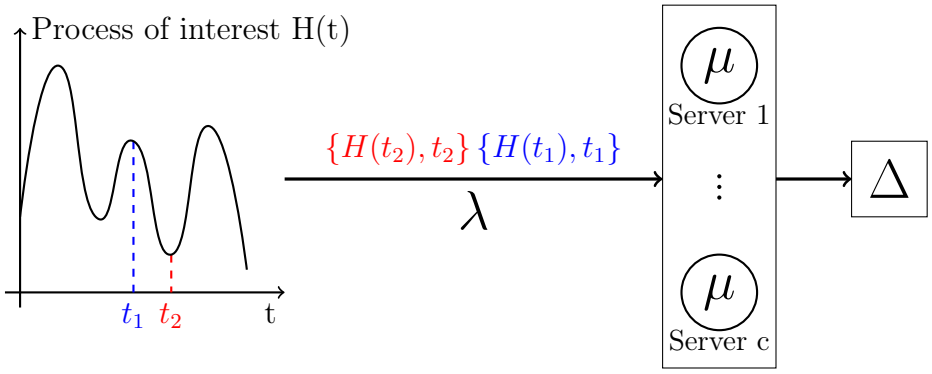


Figure 3.4: Model for status update system with multiple (possibly infinite) parallel servers. The system is modeled as a M/M/c queuing system.

hence this system is equivalent to a M/M/∞ queue, with infinite number of servers. An approximation to the average age in the case with two servers was presented in [54]. We will refer to the cases with multiple paths as the M/M/c queue model, where c is the number of servers. In these cases, given that multiple status update messages can be transmitted concurrently, it is possible that those messages arrive out of order. Therefore, newer status update message may render other messages obsolete, and network resources are wasted in the transmission of obsolete information. The queuing model with c servers is illustrated in Figure 3.4. In what follows, we present the main results in [53], namely the average age for the M/M/∞ model, and the probability that a status update message is non-informative due to the fact that it was rendered obsolete by a newer message that completed service before.

Figure 3.5 illustrates a sample path for the age of information in the case that multiple paths exist between source and destination. Recall that t_k denotes the time instant the kth packet was generated, and t'_k the time instant that this packet completes service. It may be the case that packets arrive out of order. In this example, the packet generated at time t_3 was rendered obsolete by the packet

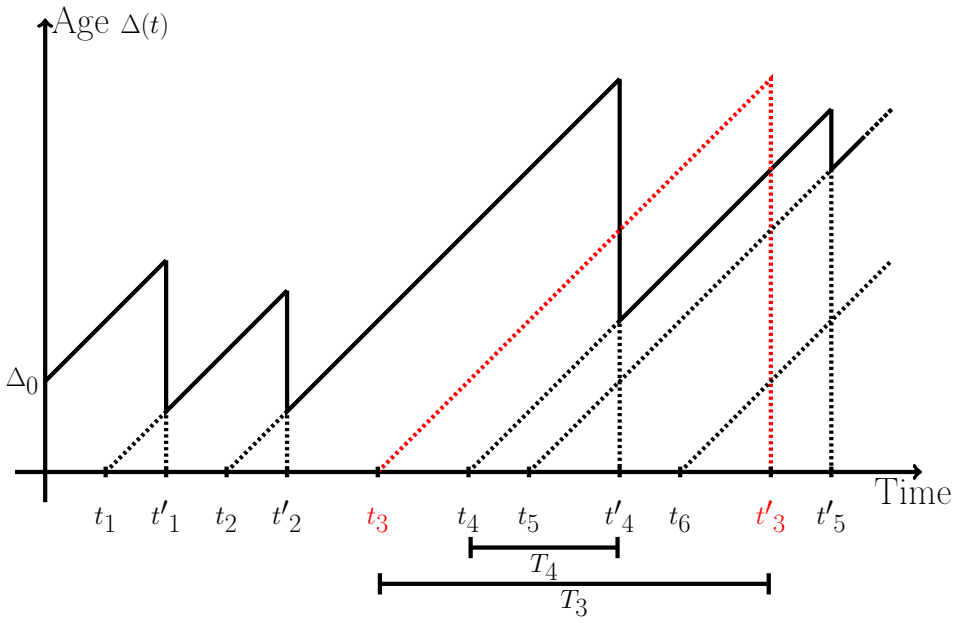


Figure 3.5: Sample path for age of information, illustrating how the age evolves in time in the case of a M/M/c queuing system.

generated at time t_4, since $t'_4 < t'_3$, and the obsolete packet has no impact on the age value, having wasted the network resources allocated for its transmission.

In [53], the authors have defined an event $E_1(k)$ that the kth transmitted packet is informative, i.e. not rendered obsolete. Assuming that there has been at least $n+1$ transmissions, define $E_2(n)$ as the event that the current packet has rendered exactly n of the previous packets obsolete. The intersection of these two events is defined as $\mathcal{E}(n) \coloneqq E_1(k) \cap E_2(n)$, and its probability is shown to be [53]

$$\mathbb{P}(\mathcal{E}(n)) = \frac{\lambda^n \mu}{\prod_{j=1}^{n+1}(\lambda + j\mu)}. \tag{3.17}$$

The average age for the M/M/∞ model is [53]

$$\Delta_{M/M/\infty} = \lambda \sum_{n=0}^{\infty} \mathbb{P}(\mathcal{E}(n)) \times \left[\sum_{j=0}^{n} \left(\frac{1}{\lambda + j\mu} \left(\frac{(n+2)\mu}{(\lambda + (n+1)\mu)(\lambda + (n+2)\mu)} + \sum_{i=j}^{n} \frac{1}{\lambda + i\mu} \right) \right) + \frac{(n+1)\sigma(n)}{\lambda} \left(\frac{(n+2)\mu}{(\lambda + (n+1)\mu)(\lambda + (n+2)\mu)} + \sum_{i=0}^{n+1} \frac{1}{\lambda + i\mu} \right) \right], \quad (3.18)$$

where

$$\sigma(n) = \sum_{r=1}^{\infty} \left[\frac{\lambda^r}{(n+r+1) \prod_{l=1}^{r} (\lambda + (n+l)\mu)} \left(1 - \frac{\lambda}{\lambda + (n+r+1)\mu} \right) \right]. \quad (3.19)$$

The probability that a packet is rendered obsolete can be used as an indicator of the amount of network resources wasted with the transmission of non-informative packets [53]. This probability is a function of the channel utilization $\rho = \lambda/\mu$, and is calculated as

$$1 - \mathbb{P}(E_1(k)) = \frac{\rho}{1+\rho} - \sum_{r=1}^{\infty} \frac{\rho^r}{(r+1) \prod_{l=1}^{r} (\rho + l)} \left(1 - \frac{\rho}{\rho + r + 1} \right). \quad (3.20)$$

3.3.3 Status Updates With Packet Transmission Management

So far, we have discussed the average age for two different models of status update systems. One model assumed a single source-destination link, and was studied as a M/M/1 queue. The other model assumed that multiple paths are available through a network to transmit the status update messages between a source and a destination, and was studied as a M/M/c queue, possibly with infinite number of servers. We have noted that these models share one common feature: every status update message that arrives at the source node will be transmitted to the

destination. As a result, network resources are sometimes wasted transmitting outdated information, or non-informative packets that have been rendered obsolete by a newer one. In the M/M/1 model, the transmission of obsolete messages is due to the fact that the packets wait in queue, being subject to the aging process even before the transmission. In the M/M/∞ model, there was no queue, but transmitting many packets simultaneously increases the percentage of packets that are rendered obsolete and waste the resources used for its transmission.

To address these issues, we proposed that the source discard some of the arriving packets, in a process that is referred to as *packet management*. In this subsection we compare three policies of packet management:

(*i*) assume that samples which arrive while a packet is being transmitted are discarded, and the ones that find the source node idle are immediately transmitted to the destination. In this case, no packets are kept in a queue waiting for transmission. Assuming a Poisson arrival process, FCFS policy, and exponentially distributed service time, this scheme is modeled as an M/M/1/1 queue, where the last entry in the Kendall notation refers to the total capacity of the queuing system, which is a single packet in service;

(*ii*) assume that a single packet may be kept in queue, waiting for transmission if the server is busy transmitting another packet. If the server is idle, the service starts immediately. Under the assumptions of Poisson arrivals, and exponential service times, this second model of packet management can be studied as an M/M/1/2 queue. Here the total capacity of the queuing system is one packet in service and one packet in queue;

(*iii*) The third packet management policy also assumes that a single packet may be kept in the buffer waiting for transmission. We propose that packets waiting for transmission are replaced upon arrival of a more up-to-date packet. Keeping the other assumptions, we can model this proposed system as a modified M/M/1/2 queue, identified as a M/M/1/2* queue. This is a peculiar queuing model, for which some of the classic results from queuing theory, such as Little's result [76, Chapter 2], fail to apply. The analysis of the M/M/1/2* queue is provided in the appendix at the end of this chapter.

Examples of sample paths for each model with packet management are presented in Figure 3.6. The age of information at the destination node increases linearly with time. Upon reception of a new status update, the age is reset to the difference of the current time instant and the time stamp of the received update. Recall that t_k denotes the time instant the kth packet was generated, and t'_k the time instant that this packet completes service. We identify the time spent in the system T_k, defined as

$$T_k \coloneqq t'_k - t_k,$$

and the interdeparture time Y_k, defined as

$$Y_k \coloneqq t'_k - t'_{k-1}.$$

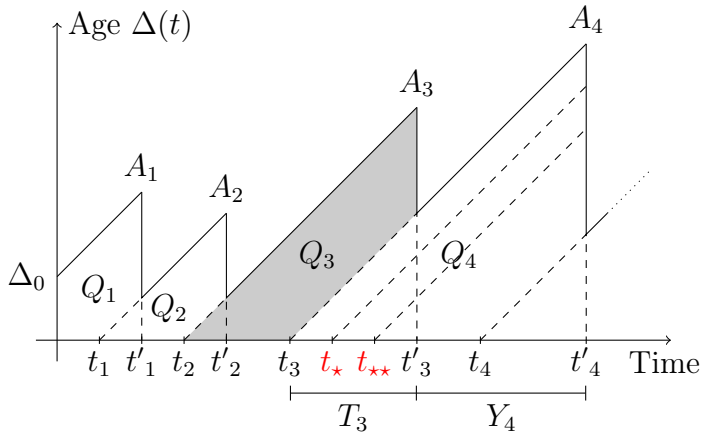
We also identify the peaks A_k, and the areas of geometric forms Q_k, which will be used in the characterization of the age. Figure 3.6(a) illustrates the case **(*i*)** of a M/M/1/1 queue, and the packets arriving at times $t_\star$ and $t_{\star\star}$ are discarded, since the server was found busy upon arrival. A sample path example for the case **(*ii*)**

is shown in Figure 3.6(b), which considers a M/M/1/2 queue model. In this model, there is a single buffer space, and packets that find the server busy and the buffer occupied are discarded, as illustrated for a packet arriving at time $t_\star$. Figure 3.6(c) presents an example of the evolution of age with the policy described in item ***(iii)***, modeled as a M/M/1/2* queue. No packets are blocked from entering the queue. If a new packet arrives while the system is full, the packet waiting is discarded. In the illustration, the packet that arrived at time $t_\star$ spends some time in the buffer, but it is substituted when a new packet arrives at time t_4. The transmission of this last packet begins when the server becomes available, at time t'_3.

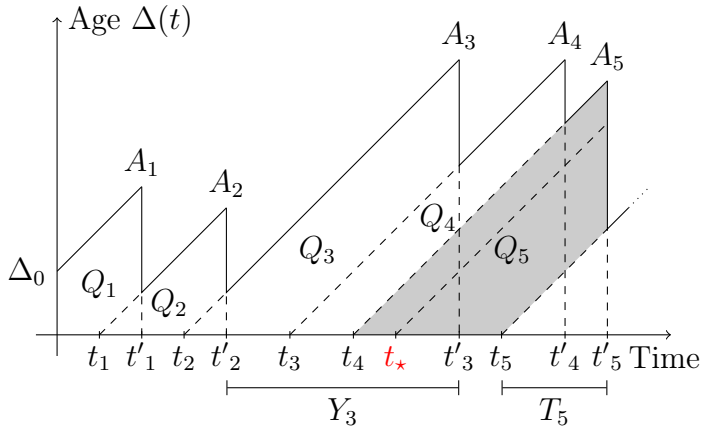
Notation: the subindex k will be used to refer to successfully transmitted packets only. We will not identify the time of arrival of discarded packets. As an example, consider the sample path illustrated in Figure 3.6(a). The fourth and fifth packets to arrive are discarded. In this case, the fourth transmitted packet was the sixth packet to arrive, but we use index $k = 4$, arrival time t_4, and departure time t'_4. We identify the fourth peak with A_4, and the area of the fourth trapezoid with Q_4. In general we will write A_k, Q_k, Y_k, T_k to refer to quantities associated with the kth transmitted packet.

Using queuing system models, we are able to characterize both the time in the system and the interdeparture times but, in general, these two random variables are not independent and we do not have information about the joint distribution. Nonetheless, it is possible to describe an event such that T_{k-1} and Y_k are conditionally independent.

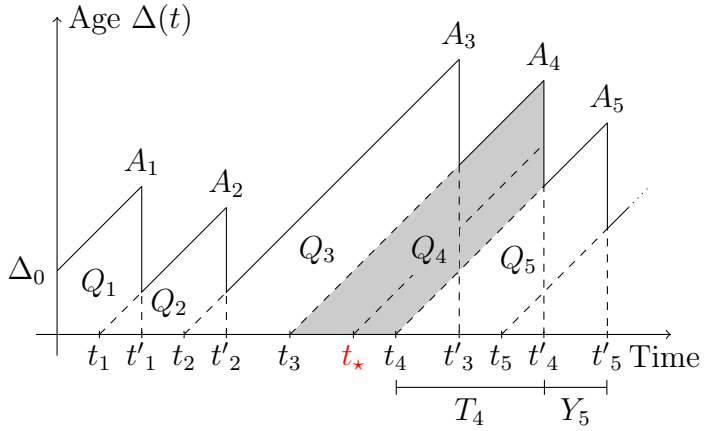
Let ψ be the event that a packet left behind an empty system upon depar-



(a) Sample path with M/M/1/1 model



(b) Sample path with M/M/1/2 model



(c) Sample path with M/M/1/2* model

Figure 3.6: Sawtooth curve - Examples of sample path for the age of information for the proposed models with packet management.

ture. Under the assumption of Poisson arrivals, the time until the next arrival is exponentially distributed with parameter λ, due to the memoryless property of interarrival times. The arriving packet will be served immediately, and the service time is assumed to be exponentially distributed with parameter μ. The interarrival and service times are independent, and they are also independent from the time in the system of the packet that just left.

We denote with $\bar{\psi}$ the complement event that the system is not empty upon departure, i.e. there was at least one packet waiting in queue. In this case, the packet waiting in queue starts service immediately, and the time until the next departure is simply a service time, which is independent of the time in the system for the packet previously served.

The variables T_{k-1} and Y_k are conditionally independent, given the event ψ that the $(k-1)$th packet leaves behind an empty system upon departure. While the conditional distribution of the time in the system depends on the selected queuing model, the conditional distribution of the interdeparture time can already be stated here. It is given by the convolution of the distributions of two exponential random variables, with parameters λ and μ, which yields

$$f(y|\psi) = \frac{\lambda\mu}{\mu - \lambda}\left[e^{-\lambda y} - e^{-\mu y}\right]. \tag{3.21}$$

$$\mathbb{E}[Y_k|\psi] = \frac{1}{\lambda} + \frac{1}{\mu}, \tag{3.22}$$

$$\mathbb{E}[Y_k^2|\psi] = \frac{2(\lambda^2 + \lambda\mu + \mu^2)}{\lambda^2\mu^2}. \tag{3.23}$$

We also have that

$$f(y|\bar{\psi}) = \mu \exp(-\mu y). \tag{3.24}$$

$$\mathbb{E}[Y_k|\bar{\psi}] = \frac{1}{\mu}, \tag{3.25}$$

$$\mathbb{E}[Y_k^2|\bar{\psi}] = \frac{2}{\mu^2}. \tag{3.26}$$

For queuing systems with single, as opposed to batch, arrivals, the probability of leaving the system empty upon departure is equal to the steady state probability that the system is empty [77, Chapter 5]. The steady state distribution is particular to each queuing model, and will be presented later in this section.

3.3.3.1 M/M/1/1 Model

A M/M/1/1 queue can be described using a two-state Markov chain, with each state representing the server as idle or busy. We denote with p_0 the probability of an empty system, and p_1 is the probability of one packet in the system. Analyzing the two-state Markov chain, it is straightforward to obtain [76, Chapter 3]

$$p_0 = \frac{\mu}{\lambda + \mu}; \;\; p_1 = \frac{\lambda}{\lambda + \mu}. \tag{3.27}$$

In the case of an M/M/1/1 queue, a packet is accepted in the system only if the server is idle. The effective arrival rate is denoted with λ_e, and we have $\lambda_e = \lambda(1 - p_1)$. The time in the system for transmitted packets is equal to the service time, which is assumed to be exponentially distributed with mean $1/\mu$. The

interdeparture time is distributed as described in (3.21), since in the M/M/1/1 model the event ψ that the system is left empty upon departure of a transmitted packet is a certain event. In this case, the time in the system and the interdeparture time are independent.

Theorem 3.1. *Average Age for M/M/1/1 Queue*

The average age for the M/M/1/1 model can be writen as

$$\Delta_{M/M/1/1} = \frac{1}{\lambda} + \frac{2}{\mu} - \frac{1}{\lambda+\mu}. \tag{3.28}$$

Proof: We use the steady state distribution in (3.27) to calculate the effective arrival rate, the results in (3.22) and (3.23) for expectations involving the interdeparture time, and the fact that the time in the system for a transmitted packet is a service time, exponentially distributed with expected value $1/\mu$. The average age can be calculated as

$$\begin{aligned}
\Delta_{M/M/1/1} &= \lambda_e \mathbb{E}[Q_k] \\
&= \lambda_e \left(\frac{1}{2}\mathbb{E}[(Y_k)^2] + \mathbb{E}[T_{k-1}]\mathbb{E}[Y_k] \right) \\
&= \frac{\lambda\mu}{\lambda+\mu} \left[\left(\frac{\lambda+\mu}{\lambda\mu} \right)^2 - \frac{1}{\lambda\mu} + \frac{1}{\mu}\frac{\lambda+\mu}{\lambda\mu} \right] \\
&= \frac{1}{\lambda} + \frac{2}{\mu} - \frac{1}{\lambda+\mu}.
\end{aligned} \tag{3.29}$$

■

Theorem 3.2. *Peak Age for M/M/1/1 Queue*

For the M/M/1/1 system, the complementary cumulative distribution function, which describes the probability that the peak age surpasses a threshold $\bar{a}$, is given by

$$\mathbb{P}(A_k > \bar{a})_{M/M/1/1} = \left(\frac{\mu}{\lambda-\mu} \right)^2 e^{-\lambda\bar{a}} + \left[1 - \left(\frac{\mu}{\lambda-\mu} \right)^2 \right] e^{-\mu\bar{a}} + \frac{\lambda\mu}{\lambda-\mu}\bar{a}e^{-\mu\bar{a}} \tag{3.30}$$

Proof: Recall that the peak age is $A_k = T_{k-1} + Y_k$. In the M/M/1/1 model, the time in the system T_{k-1} and the interdeparture time Y_k are independent. Hence, the probability density function of the peak age is given by the convolution of an exponential distribution for the service time with the distribution of the interdeparture times shown in (3.21). As a result, we have

$$\begin{aligned} f(a)_{M/M/1/1} &= f(a|\psi) \\ &= f(t|\psi) * f(y|\psi) \\ &= \left(\frac{\mu}{\lambda-\mu}\right)^2 \left(\lambda e^{-\lambda a} - \lambda e^{-\mu a} + \lambda(\lambda-\mu) a e^{-\mu a}\right) \end{aligned} \tag{3.31}$$

The result in (3.30) is obtained integrating the probability density function in (3.31) over an interval of the form $(\bar{a}, \infty)$, where $\bar{a}$ is the threshold to be imposed on the age of information. ■

3.3.3.2 M/M/1/2 Model

The M/M/1/2 queue can be described by a Markov chain with three states, that represent an empty system, a single packet being served, or a packet in service with a packet waiting in the buffer. Let $\rho \coloneqq \lambda/\mu$. The analysis of the three-state Markov chain yields the steady state probabilities [76, Chapter 3]

$$p_j = \frac{\rho^j}{1+\rho+\rho^2}, \quad j \in \{0,1,2\}. \tag{3.32}$$

The effective arrival rate is calculated as the product of the arrival rate and the probability that the system is not found full. Due to the PASTA property (Poisson Arrivals See Time Averages) [77, Chapter 3], the state of the system found upon

arrival is distributed according to the steady state distribution in (3.32), and we can write $\lambda_e = \lambda(1 - p_2)$.

In order to characterize the age of information, we need the expected values $\mathbb{E}[Y_k^2]$ and $\mathbb{E}[T_{k-1}Y_k]$. We first obtain the probability distribution of the time in the system for transmitted packets. The time in the system can be written as $T_{k-1} = W_{k-1} + S_{k-1}$, where W_{k-1} is a random variable representing the waiting time, and S_{k-1} represents the service time. We are interested in the conditional distributions given the events ψ and $\bar{\psi}$.

The waiting time is zero if the system is found idle upon arrival, otherwise it is a remaining service time, exponentially distributed with parameter μ, due to the memoryless property of service times. The probabilities that the system is found idle or busy upon arrival can be calculated using the steady state distribution in (3.32), normalizing due to the fact that a packet is only accepted into the queue if the system is not full. The waiting time of a transmitted packet is independent of future arrivals, hence independent of the event ψ. Its complementary cumulative distribution is

$$\mathbb{P}(W_{k-1} > w) = \frac{p_1}{p_0 + p_1} e^{-\mu w}, \ w > 0. \tag{3.33}$$

The conditional distribution of service time given the event ψ is calculated noting that the $(k-1)$th transmitted packet leaves the system idle upon departure if and only if zero arrivals occur while the $(k-1)$th transmitted packet is being served. We define the probability $\mathbb{P}(\psi | S_{k-1} = s)$ by requiring that for every measurable set

$A \subset [0, \infty)$

$$\mathbb{P}(\psi, S_{k-1} \in A) = \int_A f(s)\mathbb{P}(\psi|S_{k-1} = s)ds, \tag{3.34}$$

where $f(s)$ is the probability density function of the service time. The conditional distribution of service time given ψ is then calculated as

$$\begin{aligned} f(s|\psi) &= \frac{\mathbb{P}(\psi|S_{k-1} = s)f(s)}{\int_0^\infty \mathbb{P}(\psi|S_{k-1} = s)f(s)ds} \\ &= \frac{\frac{(\lambda s)^0}{0!}e^{-\lambda s}\mu e^{-\mu s}}{\int_0^\infty \frac{(\lambda s)^0}{0!}e^{-\lambda s}\mu e^{-\mu s}ds} \\ &= (\lambda + \mu)e^{-(\lambda+\mu)s}, \end{aligned} \tag{3.35}$$

which yields the conditional expectation

$$\mathbb{E}[S_{k-1}|\psi] = \frac{1}{\lambda + \mu}. \tag{3.36}$$

Using the results in (3.33) and (3.35) we can obtain the conditional probability distribution of the time in the system, given the event ψ,

$$\mathbb{P}(T_{k-1} > t|\psi) = \mathbb{P}(W_{k-1} + S_{k-1} > t|\psi) \tag{3.37}$$

$$= \int_0^\infty \mathbb{P}(W_{k-1} > t - s)f(s|\psi)ds \tag{3.38}$$

$$= e^{-\mu t}, \tag{3.39}$$

with conditional expectation

$$\mathbb{E}[T_{k-1}|\psi] = \frac{1}{\mu}. \tag{3.40}$$

The conditional distribution of the time in the system given the event $\bar{\psi}$ is obtained similarly. To obtain the conditional distribution of the service time, we note that the system is left behind with another packet waiting for transmission

if and only if at least one arrival occurs during the service time of the $(k-1)$th transmitted packet. Then calculate

$$\begin{aligned} f(s|\bar{\psi}) &= \frac{\mathbb{P}(\bar{\psi}|S_{k-1}=s)f(s)}{\int_0^\infty \mathbb{P}(\bar{\psi}|S_{k-1}=s)f(s)ds} \\ &= \frac{\left(1-\frac{(\lambda s)^0}{0!}e^{-\lambda s}\right)\mu e^{-\mu s}}{\int_0^\infty \left(1-\frac{(\lambda s)^0}{0!}e^{-\lambda s}\right)\mu e^{-\mu s}ds} \\ &= \frac{\lambda+\mu}{\lambda}\left(1-e^{-\lambda s}\right)\mu e^{-\mu s}, \end{aligned} \tag{3.41}$$

with conditional expectation given by

$$\mathbb{E}[S_{k-1}|\bar{\psi}] = \frac{1}{\mu} + \frac{1}{(\lambda+\mu)}. \tag{3.42}$$

The conditional distribution of the time in the system is obtained as follows:

$$\begin{aligned} \mathbb{P}(T_{k-1}>t|\bar{\psi}) &= \mathbb{P}(W_{k-1}+S_{k-1}>t|\bar{\psi}) \\ &= \int_0^\infty \mathbb{P}(W_{k-1}>t-s)f(s|\bar{\psi})ds \\ &= e^{-\mu t}(1+\mu t), \end{aligned} \tag{3.43}$$

which corresponds to a probability density function given by the convolution of two exponential distributions with parameter μ. The conditional expectation is

$$\mathbb{E}[T_{k-1}|\bar{\psi}] = \frac{2}{\mu}. \tag{3.44}$$

The probability of the event ψ that the system is left idle upon departure is obtained using the steady state probability that the system is empty [77, Chapter 5]. Normalizing the probabilities to account only for transmitted packets, we write

$$\mathbb{P}(\psi) = \frac{p_0}{p_0+p_1} = \frac{\mu}{\lambda+\mu}, \tag{3.45}$$

$$\mathbb{P}(\bar{\psi}) = \frac{p_1}{p_0+p_1} = \frac{\lambda}{\lambda+\mu}. \tag{3.46}$$

At this point, we have all the elements needed to prove the following theorem.

Theorem 3.3. *Average Age for M/M/1/2 Queue*

For the M/M/1/2 model, the average age is

$$\Delta_{M/M/1/2} = \frac{1}{\lambda} + \frac{3}{\mu} - \frac{2(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2} \tag{3.47}$$

Proof: Recall that the average age is calculated as

$$\begin{aligned}\Delta_{M/M/1/2} &= \lambda_e \mathbb{E}[Q_k] \\ &= \lambda_e \left(\frac{1}{2}\mathbb{E}[(Y_k)^2] + \mathbb{E}[T_{k-1}Y_k] \right).\end{aligned} \tag{3.48}$$

The effective arrival rate is

$$\lambda_e = \lambda(1-p_2) = \frac{\lambda\mu(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2}. \tag{3.49}$$

Using (3.23) and (3.26), together with (3.45) and (3.46), we calculate

$$\begin{aligned}\frac{1}{2}\mathbb{E}[Y_k^2] &= \frac{1}{2}\left(\mathbb{E}[Y_k^2|\psi]\mathbb{P}(\psi) + \mathbb{E}[Y_k^2|\bar{\psi}]\mathbb{P}(\bar{\psi})\right) \\ &= \frac{1}{2}\left(\frac{2(\lambda^2+\lambda\mu+\mu^2)}{(\lambda\mu)^2}\frac{\mu}{\lambda+\mu} + \frac{2}{\mu^2}\frac{\lambda}{\lambda+\mu}\right) \\ &= \frac{1}{\lambda^2} + \frac{1}{\mu^2}.\end{aligned} \tag{3.50}$$

The expected value of the product $T_{k-1}Y_k$ is calculated using the probabilities in (3.45) and (3.46), together with the conditional expectations calculated in (3.22), and (3.25) for Y_k, and in (3.40) and (3.44) for T_{k-1}. Given that T_{k-1} and Y_k are

conditionally independent, we may write

$$\begin{aligned}\mathbb{E}[T_{k-1}Y_k] &= \mathbb{E}[T_{k-1}Y_k|\psi]\mathbb{P}(\psi) + \mathbb{E}[T_{k-1}Y_k|\bar{\psi}]\mathbb{P}(\bar{\psi}) \\ &= \mathbb{E}[Y_k|\psi]\mathbb{E}[T_{k-1}|\psi]\frac{\mu}{\lambda+\mu} + \mathbb{E}[Y_k|\bar{\psi}]\mathbb{E}[T_{k-1}|\bar{\psi}]\frac{\lambda}{\lambda+\mu} \\ &= \left(\frac{1}{\lambda}+\frac{1}{\mu}\right)\frac{1}{\mu}\frac{\mu}{\lambda+\mu} + \frac{1}{\mu}\frac{2}{\mu}\frac{\lambda}{\lambda+\mu} \\ &= \frac{2\lambda^2+\lambda\mu+\mu^2}{\lambda\mu^2(\lambda+\mu)}. \end{aligned} \tag{3.51}$$

The average age is obtained using (3.49), (3.50), and (3.51) in (3.48)

$$\begin{aligned}\Delta_{M/M/1/2} &= \frac{\lambda\mu(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2}\left(\frac{1}{\lambda^2}+\frac{1}{\mu^2}+\frac{2\lambda^2+\lambda\mu+\mu^2}{\lambda\mu^2(\lambda+\mu)}\right) \\ &= \frac{1}{\lambda}+\frac{3}{\mu}-\frac{2(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2}, \end{aligned} \tag{3.52}$$

which is the result stated in the theorem. ■

Theorem 3.4. *Peak Age for M/M/1/2 Queue*

The complementary cumulative distribution of the peak age for the M/M/1/2 model, which gives the probability that the peak age is larger than a threshold $\bar{a}$, is

$$\begin{aligned}\mathbb{P}(A>\bar{a})_{M/M/1/2} &= \frac{\mu^3}{(\lambda-\mu)^2(\lambda+\mu)}e^{-\lambda\bar{a}} \\ &+\frac{\lambda}{2(\lambda-\mu)^2(\lambda+\mu)}e^{-\mu\bar{a}}\left[\mu^2\bar{a}^2(\lambda-\mu)^2+2\lambda\mu(\lambda-\mu)\bar{a}+2(\lambda^2-\lambda\mu-\mu^2)\right] \end{aligned} \tag{3.53}$$

Proof: The peak age is given by $A_k = T_{k-1} + Y_k$. These variables are conditionally independent given the event ψ or the event $\bar{\psi}$ and, in this case, the distribution of their sum is calculated as the convolution of their individual conditional probability density functions. Using the probability distributions described

in (3.21), (3.24), (3.39), and (3.43), we obtain

$$f(a|\psi)_{M/M/1/2} = \left(\frac{\mu}{\lambda-\mu}\right)^2 \left(\lambda e^{-\lambda a} - \lambda e^{-\mu a} + \lambda(\lambda-\mu)ae^{-\mu a}\right) \tag{3.54}$$

$$f(a|\bar{\psi})_{M/M/1/2} = \frac{1}{2}a^2\mu^3 e^{-\mu a}, \tag{3.55}$$

and the distribution of the peak age is the mixture, using the probabilities in (3.45) and (3.46), which yields

$$f(a) = \left(\frac{\mu}{\lambda-\mu}\right)^2 \frac{\lambda\mu}{2(\lambda+\mu)} \left(2e^{-\lambda a} + e^{-\mu a}(a^2(\lambda-\mu) + 2a - 2)\right) \tag{3.56}$$

Integrating (3.56) over the interval $(\bar{a}, \infty)$ we obtain (3.53). ■

3.3.3.3 M/M/1/2* Model

To model the packet management scheme in which the packet waiting in queue is replaced if a new packet arrives, we have proposed a new queue model, named M/M/1/2*. The characterization of the proposed queue model is peculiar because some of the classic results in queuing theory do not apply, and it is shown in Appendix 3.7.

The M/M/1/2* model behaves as the M/M/1/2 queue regarding the number of packets in the system. In both queuing systems, a packet is discarded when the system is found full upon arrival. The difference is that in the M/M/1/2 system the discarded packet is the one that just arrived, while in the M/M/1/2* system the discarded packet is the one found in the buffer. The steady state probabilities in (3.32) still hold, and the effective arrival rate is $\lambda_e = (1-p_2)\lambda$, as in the M/M/1/2 model.

We proceed to characterize the conditional distributions of the time in the system for a transmitted packet, given the event ψ that the system is left empty by a departing packet, and the event $\bar{\psi}$ that the system is left busy upon departure. We write the time in the system as the sum of the waiting time and the service time, $T_{k-1} = W_{k-1} + S_{k-1}$ and proceed as in the M/M/1/2 model, following the steps described in (3.39) or (3.43) to obtain the conditional distribution of the time in the system given the events ψ or $\bar{\psi}$, respectively. The conditional distribution of the service time is the same as in the case of a M/M/1/2 queue, as given in (3.35) or (3.41). The next step is to describe the probability distribution of the waiting time for a transmitted packet.

If the arriving packet finds the server idle, the packet is served immediately, and the waiting time is equal to zero. This event occurs with probability $\mathbb{P}(\text{idle}) = p_0$, due to the Poisson Arrivals See Time Averages (PASTA) property [77, Chapter 3].

Due to the same property, the server is found busy upon arrival with probability $\mathbb{P}(\text{busy}) = 1 - p_0$. The arriving packet will be admitted in the system, but it will be transmitted if and only if no other arrival occurs while the service in progress is not completed. Let R represent a remaining service time, with probability density function $f(r)$, and let ϕ be the event that no other packet arrives during the remaining service. We define the probability $\mathbb{P}(\phi|R = r)$ by requiring that for every measurable set $A \subset [0, \infty)$, we have

$$\mathbb{P}(\phi, R \in A) = \int_A f(r)\mathbb{P}(\phi|R = r)dr.$$

The probability of the packet being transmitted given that the server was

found busy upon arrival is calculated as

$$\begin{aligned}\mathbb{P}(\text{tx}|\text{busy}) &= \int_0^\infty \mathbb{P}(\phi|R=r)f(r)dr \\ &= \int_0^\infty \frac{(\lambda r)^0}{0!}e^{-\lambda r}\mu e^{-\mu r}dr \\ &= \frac{\mu}{\lambda+\mu}. \end{aligned} \tag{3.57}$$

As a result,

$$\mathbb{P}(\text{busy,tx}) = (1-p_0)\frac{\mu}{\lambda+\mu} \tag{3.58}$$

$$\mathbb{P}(\text{busy,drop}) = (1-p_0)\frac{\lambda}{\lambda+\mu} \tag{3.59}$$

The conditional distribution of the waiting time given the event {busy, tx} is equal to the conditional distribution of the remaining service given the event ϕ, and it can be calculated as

$$\begin{aligned} f(w|\text{busy, tx}) &= f(r|\phi) \\ &= \frac{\mathbb{P}(\phi|R=r)f(r)}{\int_0^\infty \mathbb{P}(\phi|R=r)f(r)dr} \\ &= \frac{\frac{(\lambda r)^0}{0!}e^{-\lambda r}\mu e^{-\mu r}}{\int_0^\infty \frac{(\lambda r)^0}{0!}e^{-\lambda r}\mu e^{-\mu r}dr} \\ &= (\lambda+\mu)e^{-(\lambda+\mu)r}, \end{aligned} \tag{3.60}$$

hence we have for the transmitted packets in the M/M/1/2* model

$$\begin{aligned}\mathbb{P}(W_{k-1} > w) &= \frac{(1-p_0)\frac{\mu}{\lambda+\mu}}{p_0+(1-p_0)\frac{\mu}{\lambda+\mu}}e^{-(\lambda+\mu)w} \\ &= \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)w}. \end{aligned} \tag{3.61}$$

Using the results in (3.61) and (3.35) we can obtain the conditional probability

distribution of the time in the system, given the event ψ,

$$\begin{aligned}\mathbb{P}(T_{k-1} > t|\psi) &= \mathbb{P}(W_{k-1} + S_{k-1} > t|\psi) \\ &= \int_0^\infty \mathbb{P}(W_{k-1} > t - s) f(s|\psi) ds \\ &= e^{-(\lambda+\mu)t}(1 + \lambda t). \end{aligned} \tag{3.62}$$

The conditional expectation can be obtained integrating (3.62), or adding the conditional expectations for the waiting time and the service time,

$$\begin{aligned}\mathbb{E}[T_{k-1}|\psi] &= \mathbb{E}[W_{k-1}|\psi] + \mathbb{E}[S_{k-1}|\psi] \\ &= \mathbb{E}[W_{k-1}] + \mathbb{E}[S_{k-1}|\psi] \\ &= \frac{\lambda}{(\lambda+\mu)^2} + \frac{1}{\lambda+\mu} \\ &= \frac{2\lambda+\mu}{(\lambda+\mu)^2}, \end{aligned} \tag{3.63}$$

Similarly, conditioning on the event $\bar{\psi}$, we use (3.61) and (3.41) to calculate

$$\begin{aligned}\mathbb{P}(T_{k-1} > t|\bar{\psi}) &= \mathbb{P}(W_{k-1} + S_{k-1} > t|\bar{\psi}) \\ &= \int_0^\infty \mathbb{P}(W_{k-1} > t - s) f(s|\bar{\psi}) ds \\ &= \frac{\lambda+2\mu}{\lambda} e^{-\mu t} - \frac{\mu}{\lambda} e^{-(\lambda+\mu)t}(2 + \lambda t). \end{aligned} \tag{3.64}$$

The conditional expectation of the time in the system for transmitted packets can be calculated as

$$\begin{aligned}\mathbb{E}[T_{k-1}|\bar{\psi}] &= \mathbb{E}[W_{k-1}|\bar{\psi}] + \mathbb{E}[S_{k-1}|\bar{\psi}] \\ &= \mathbb{E}[W_{k-1}] + \mathbb{E}[S_{k-1}|\bar{\phi}] \\ &= \frac{\lambda}{(\lambda+\mu)^2} + \frac{1}{\mu} + \frac{1}{(\lambda+\mu)} \\ &= \frac{1}{\mu} + \frac{2\lambda+\mu}{(\lambda+\mu)^2}. \end{aligned} \tag{3.65}$$

The probabilities of the events ψ and $\bar{\psi}$ are obtained using the steady state distribution for the number of packets in the system, therefore these probabilities are the same as in the M/M/1/2 model, as described in (3.45) and (3.46). The expected values related to the interdeparture times Y_k are also the same as in the M/M/1/2 model. At this point, we have all the elements to prove the following theorem:

Theorem 3.5. *Average Age for M/M/1/2* Queue*

The average age for the M/M/1/2 model is*

$$\Delta_{M/M/1/2^*} = \frac{1}{\lambda} + \frac{2}{\mu} + \frac{\lambda}{(\lambda+\mu)^2} + \frac{1}{\lambda+\mu} - \frac{2(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2} \tag{3.66}$$

Proof: The average age is calculated as

$$\begin{aligned}\Delta_{M/M/1/2^*} &= \lambda_e \mathbb{E}[Q_k] \\ &= \lambda_e \left(\frac{1}{2}\mathbb{E}[(Y_k)^2] + \mathbb{E}[T_{k-1}Y_k] \right)\end{aligned} \tag{3.67}$$

The effective arrival rate is calculated as in the M/M/1/2 case,

$$\lambda_e = \lambda(1-p_2) = \frac{\lambda\mu(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2}. \tag{3.68}$$

Using (3.23) and (3.26), together with (3.45) and (3.46), we also have the same second moment for the interdeparture time

$$\frac{1}{2}\mathbb{E}[Y_k^2] = \frac{1}{\lambda^2} + \frac{1}{\mu^2}. \tag{3.69}$$

The expected value of the product $T_{k-1}Y_k$ is calculated using the probabilities in (3.45) and (3.46), together with the conditional expectations calculated in (3.22),

and (3.25) for Y_k, and in (3.63) and (3.65) for T_{k-1}. Given that T_{k-1} and Y_k are conditionally independent, we may write

$$\begin{aligned}
\mathbb{E}[T_{k-1}Y_k] &= \mathbb{E}[T_{k-1}Y_k|\psi]\mathbb{P}(\psi) + \mathbb{E}[T_{k-1}Y_k|\bar{\psi}]\mathbb{P}(\bar{\psi}) \\
&= \mathbb{E}[Y_k|\psi]\mathbb{E}[T_{k-1}|\psi]\frac{\mu}{\lambda+\mu} + \mathbb{E}[Y_k|\bar{\psi}]\mathbb{E}[T_{k-1}|\bar{\psi}]\frac{\lambda}{\lambda+\mu} \\
&= \left(\frac{1}{\lambda}+\frac{1}{\mu}\right)\frac{2\lambda+\mu}{(\lambda+\mu)^2}\frac{\mu}{\lambda+\mu} + \frac{1}{\mu}\left(\frac{1}{\mu}+\frac{2\lambda+\mu}{(\lambda+\mu)^2}\right)\frac{\lambda}{\lambda+\mu} \\
&= \frac{1}{\mu^2}+\frac{1}{\lambda\mu}-\frac{2\lambda+\mu}{(\lambda+\mu)^3}. \qquad (3.70)
\end{aligned}$$

As a result, the average age for the M/M/1/2* model is calculated using (3.68), (3.69), and (3.70) in (3.66), which yields the result in the theorem

$$\begin{aligned}
\Delta_{M/M/1/2^*} &= \frac{\lambda\mu(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2}\left(\frac{1}{\lambda^2}+\frac{1}{\mu^2}+\frac{1}{\mu^2}-\frac{1}{\lambda\mu}-\frac{2\lambda+\mu}{(\lambda+\mu)^3}\right) \\
&= \frac{1}{\lambda}+\frac{2}{\mu}+\frac{\lambda}{(\lambda+\mu)^2}+\frac{1}{\lambda+\mu}-\frac{2(\lambda+\mu)}{\lambda^2+\lambda\mu+\mu^2} \qquad (3.71)
\end{aligned}$$

■

Theorem 3.6. *Peak Age for M/M/1/2* Queue*

The complementary cumulative distribution of the peak age for the M/M/1/2 model, which gives the probability that the peak age is larger than a threshold $\bar{a}$, is*

$$\begin{aligned}
&\mathbb{P}(A_k > \bar{a})_{M/M/1/2^*} = \frac{e^{-(\lambda+\mu)a}}{\lambda(\lambda+\mu)(\lambda-\mu)}(\lambda^3-3\mu^3+\lambda\mu(\lambda+\mu)(1+(\lambda-\mu))) \\
&+\frac{e^{-\mu a}}{\lambda(\lambda+\mu)(\lambda-\mu)}(3\mu^3+\lambda(\lambda+\mu)(\lambda-\mu)+\lambda\mu a(\lambda^2+\lambda\mu-2\mu^2)) \\
&-\frac{e^{-\lambda a}}{(\lambda+\mu)(\lambda-\mu)}(\lambda^2+\lambda\mu+\mu^2) \qquad (3.72)
\end{aligned}$$

Proof: The conditional distribution of the *peak age*, given the events ψ and $\bar{\psi}$ is obtained as the convolution of the conditional distributions for T_{k-1} and

Y_k, since these variables are conditionally independent.

$$f(a|\psi)_{M/M/1/2^*} = \left(\lambda(\lambda+\mu)a - \frac{(2\mu^3-\lambda^3-\lambda^2\mu)}{\mu(\lambda-\mu)}\right)e^{-(\lambda+\mu)a} + \frac{\lambda\mu+2\mu^2}{\lambda-\mu}e^{-\mu a} - \frac{\lambda\mu(\lambda+\mu)+\lambda^3}{\mu(\lambda-\mu)}e^{-\lambda a} \tag{3.73}$$

$$f(a|\bar{\psi})_{M/M/1/2^*} = \frac{\mu^2}{\lambda^2}e^{-(\lambda+\mu)a}(3\mu+2\lambda+\lambda(\lambda+\mu)a) - \frac{\mu^2}{\lambda^2}e^{-\mu a}(3\mu+2\lambda-\lambda(\lambda+2\mu)a) \tag{3.74}$$

The distribution of the *peak age* is the mixture, using the probabilities in (3.45) and (3.46), and integrating it over the interval $(\bar{a},\infty)$ gives the result in (3.72). ■

3.3.4 Just in Time Updates

In this section we discuss the limiting behavior of the proposed packet management schemes when the arrival rate becomes very large. We compare the results with a system capable of generating status update messages as soon as the server is idle. This *just in time* status update system was first presented in [49], and it eliminates any queuing delays, since the packets are transmitted immediately after generated.

Consider the aforementioned *just in time* status update system, in which a message is generated as the server becomes available. That is, the arrival time is equal to the departure time of the previous packet, $t_k = t'_{k-1}$. In this case, no packet waits in queue for transmission, and every status update is as fresh as possible when transmitted. The time in the system for each transmitted packet is simply its service time, and the arrival rate is the inverse of the expected service

time. The interdeparture time is also the service time, which is independent of the time in the system for a previously transmitted packet.

For the *just in time* status update system, the following equalities hold:

$$\mathbb{E}[T_{k-1}] = \mathbb{E}[S_{k-1}] = \mathbb{E}[S_k] \tag{3.75}$$

$$\mathbb{E}[Y_k] = \mathbb{E}[S_k] \tag{3.76}$$

$$\mathbb{E}[Y_k^2] = \mathbb{E}[S_k^2] \tag{3.77}$$

$$\lambda_e = \lambda = \frac{1}{\mathbb{E}[S_k]} \tag{3.78}$$

The average age for the *just in time* status update system is calculated using the same procedure described before, that is

$$\begin{aligned}\Delta_{JIT} &= \lambda_e \mathbb{E}[Q_k] \\ &= \lambda \left(\mathbb{E}\left[\frac{Y_k^2}{2}\right] + \mathbb{E}[T_{k-1}Y_k] \right) \\ &= \lambda \left(\mathbb{E}\left[\frac{Y_k^2}{2}\right] + \mathbb{E}[T_{k-1}]\mathbb{E}[Y_k] \right),\end{aligned} \tag{3.79}$$

where the last equality follows from the independence between the interdeparture time and the time in the system for the previously transmitted packet, since these two correspond to different service times.

Using the equalities in (3.75)-(3.78) in (3.79), considering FCFS queuing systems with expected service time $\mathbb{E}[S_k] = \mathbb{E}[S]$, yields [49]

$$\Delta_{JIT} = \frac{1}{\mathbb{E}[S]} \left(\mathbb{E}\left[\frac{S^2}{2}\right] + (\mathbb{E}[S])^2 \right). \tag{3.80}$$

In the case of exponentially distributed service time with rate μ, we have $\mathbb{E}[S] = 1/\mu$, and

$$\Delta_{JIT} = \mu \left(\frac{1}{\mu^2} + \frac{1}{\mu^2} \right) = \frac{2}{\mu}. \tag{3.81}$$

Next, we investigate the limit as the arrival rate goes to infinity for the queuing systems considered in this thesis. With large arrival rates, we expect to approximate the desirable feature of the *just in time* system described above, that is, the availability of a fresh status update message as soon as the server becomes available.

Consider the system modeled as a M/M/1/1 queue. In this case, no packets are kept in the buffer. When the arrival rate is finite, the performance of this system differs from the *just in time* system because it may be the case that the server is idle, and no packet arrives. As the arrival rate increases, the expected interarrival time will be reduced, and we expect the behavior of the M/M/1/1 system to be closer to the *just in time* system. In fact, using the average age presented in Theorem 3.1, the limit as the arrival rate goes to infinity is

$$\lim_{\lambda\to\infty} \Delta_{M/M/1/1} = \frac{2}{\mu}, \tag{3.82}$$

which is equal to the average age Δ_{JIT} for just in time updates.

The average age in for the M/M/1/2 model was presented in Theorem 3.3. The limit as the arrival rate goes to infinity is

$$\lim_{\lambda\to\infty} \Delta_{M/M/1/2} = \frac{3}{\mu}, \tag{3.83}$$

which is much larger than the case modeled as M/M/1/1. Intuitively, for very large arrival rates, keeping a packet in the buffer is not advantageous, because the information ages while waiting in queue, and it is better to wait for a new packet to arrive instead of keeping one in the buffer.

The packet management scheme modeled as a M/M/1/2* queue considered the case in which a packet is kept in the buffer, but it is replaced if a new status update

message is available. The average age for this model was presented in Theorem 3.5. The limit as the arrival rate goes to infinity is

$$\lim_{\lambda \to \infty} \Delta_{M/M/1/2^*} = \frac{2}{\mu}, \tag{3.84}$$

indicating that the model with packet replacement behaves as the model without buffer (M/M/1/1) in the limit. Intuitively, as the arrival rate goes to infinity, a fresh packet will always be available for transmission in both models. The M/M/1/2* behaves as the M/M/1/1 as the arrival rate goes to infinity but, for finite arrival rates and a fixed service rate, the packet replacement provides better performance with respect to the average age, as will be illustrated in section 3.4.

3.4 Numerical Results

This section presents numerical results to illustrate the average age and peak age for the queuing models investigated in this chapter. The service rate is assumed to be $\mu = 1$, unless otherwise stated. In this case, the arrival rate λ in the x axis can be interpreted as the channel utilization $\rho = \lambda/\mu$. In addition to the analytical formulations, some figures present simulation points, obtained from a simple discrete-event simulator built in MATLAB. The simulation results are presented as circular markers in the curves, and corroborate the analytical results.

The average age stated in Theorems 3.1, 3.3, and 3.5 is illustrated in Figure 3.7. The average age is decreasing with the channel utilization inside the observed range $0 < \rho \leq 1.5$. For very small arrival rates, keeping a packet in the buffer is preferable than waiting for a new arrival, which may take a long time, hence

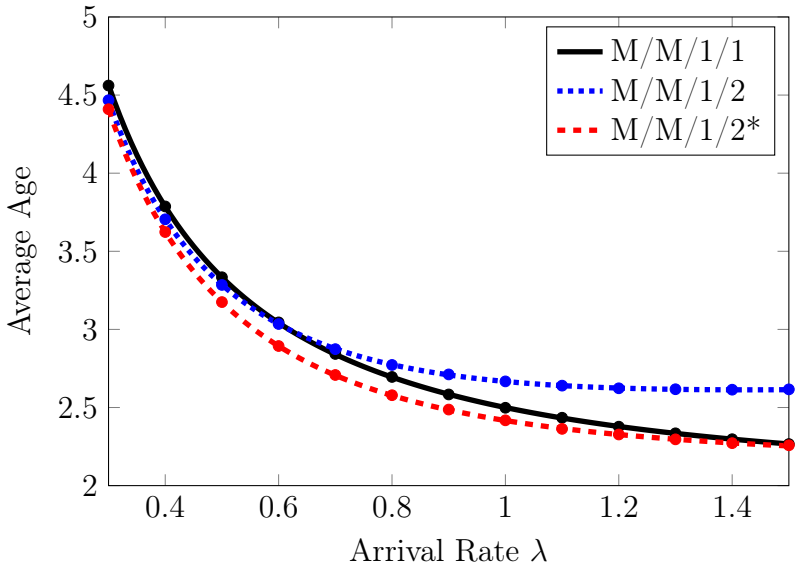


Figure 3.7: Average age versus arrival rate (λ) for the queuing models with packet management.

the M/M/1/2 model yields smaller average age than the M/M/1/1 model. As the arrival rate increases, the average interarrival times become smaller, on average, and the performance of the M/M/1/1 model gets closer to the M/M/1/2* model, while the M/M/1/2 model yields much larger average age. The M/M/1/2* model yields the smallest average age. For example, when $\lambda = 0.6$, packet replacement promotes a reduction of approximately 5% in the average age, in comparison to the other models.

In Figure 3.8, the average age for the M/M/1/2* model is compared to that of a M/M/1 model, which corresponds to a transmission scheme with no packet management, where all packets that find the server busy are kept in the buffer waiting for transmission in the order of arrival. The average age for the M/M/1 model was presented in subsection 3.3.1, with the final result given in (3.16), which is due to Kaul et al. [49]. When $\lambda \ll \mu$, the dominant effect is the large interarrival

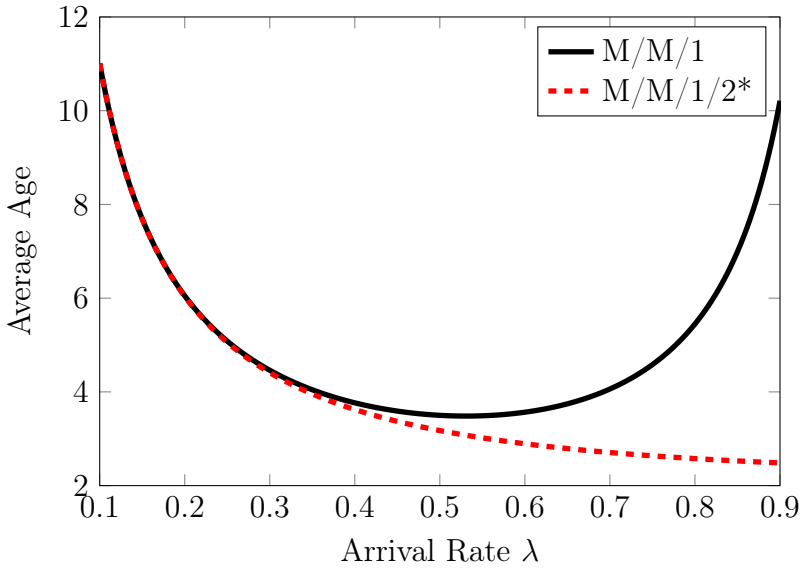


Figure 3.8: Average age versus arrival rate (λ) comparing the results with the M/M/1/2* model for packet management with the M/M/1 model which assumes an infinite capacity buffer and does not apply any packet management.

time and the two models have similar behavior because they are idle for a large fraction of time. As the arrival rate λ increases, the dominant effect in the M/M/1 system is the queuing time which approaches infinity as the arrival rate approaches the service rate. As a result, the average age increases to infinity as the arrival rate approaches the service rate. This effect is eliminated with the packet management scheme, which promotes significant improvement in the average age, particularly for large values of channel utilization $\rho > 0.5$.

The average age for the M/M/1/2* model is compared in Figure 3.9 to the average age when the status update messages can be transmitted as soon as they are generated, using a network cloud, as proposed in [53]. Recall that our model assumes a single server and the packet management scheme keeps a single packet in the buffer, replaced upon arrival of a new packet. The comparison scheme is a

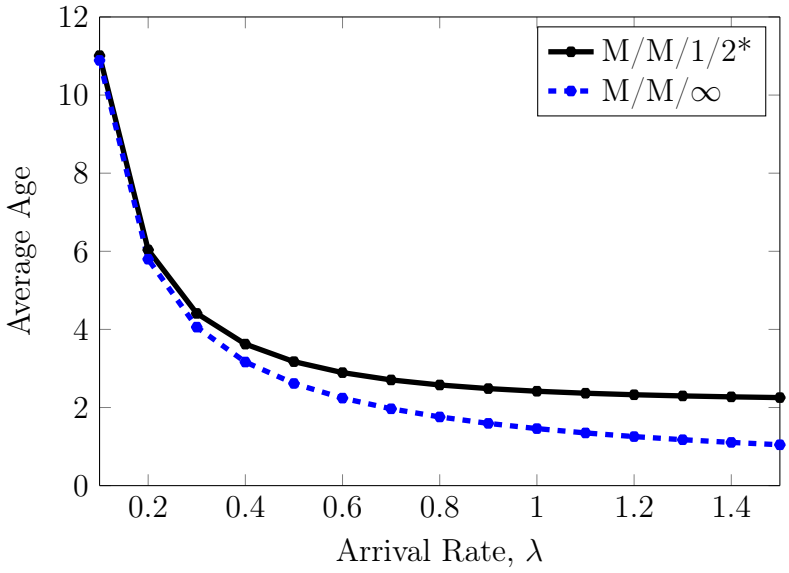


Figure 3.9: Average age versus arrival rate (λ) comparing the results with the M/M/1/2* model for packet management with the M/M/∞ model which assumes an infinite number of servers.

system with infinite number of servers, modeled as a M/M/∞ queue, which does not keep any packet in a buffer, but it may transmit packets out of order, wasting network resources with the transmission of non-informative packets. The M/M/∞ model assumes independent service times among the servers, and it provides a very optimistic result. We use the service rate $\mu = 1$. In the observed range of arrival rates, the gap in the average age between the two schemes is increasing, and the scheme with infinite servers reaches an average age value which is approximately 53% smaller when $\lambda = 1.5$. We note that this gain on the average age comes at the cost of packets being rendered obsolete after transmission, which occurs with probability 0.351378, obtained using the expression in (3.20) with $\lambda = 1.5$ and $\mu = 1$.

Figure 3.10 shows the effect of the service rate on the average age. The arrival

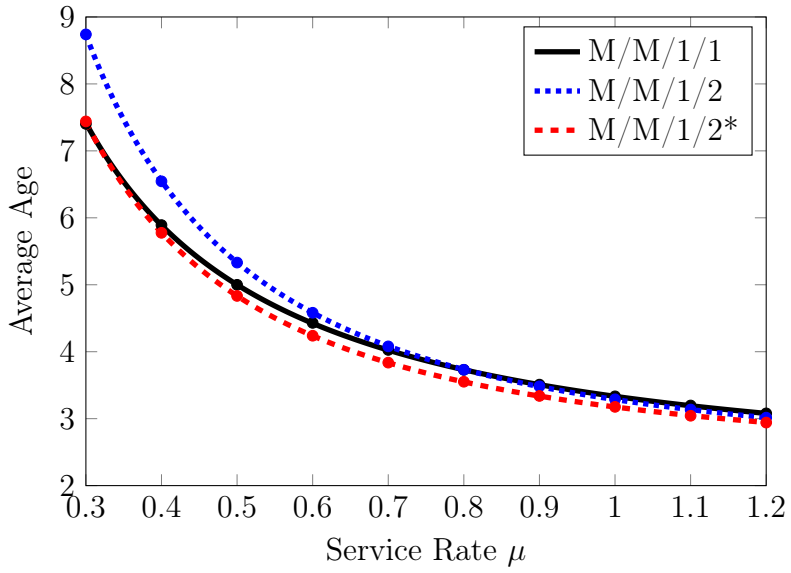


Figure 3.10: Average age versus service rate (μ) for the queuing models with packet management.

rate is fixed to $\lambda = 0.5$. The M/M/1/2* model promotes the smallest average age, but the schemes have a common limit value as μ goes to infinity, which is $1/\lambda$. That is, for very small service times, the average age is limited by the interarrival times.

Figure 3.11 presents the complementary cumulative distribution function for the peak age. The service rate is $\mu = 1$, while the arrival rate takes the values $\lambda = 0.5$ and $\lambda = 1.3$. The probability that the *peak age* is larger than a given threshold $\bar{a}$ can be reduced with larger arrival rates for all three schemes. Comparing the two groups of curves, we also note that the M/M/1/2* model produces the best results regarding the peak age for arrival rates below and above the service rate. The M/M/1/2 model presents the second best results for small arrival rates, but for large arrival rates the M/M/1/1 model performs better than the M/M/1/2 with respect to the peak age. This is so because for larger arrival rates it is advantageous to wait for the arrival of a new packet, instead of keeping one in the buffer.

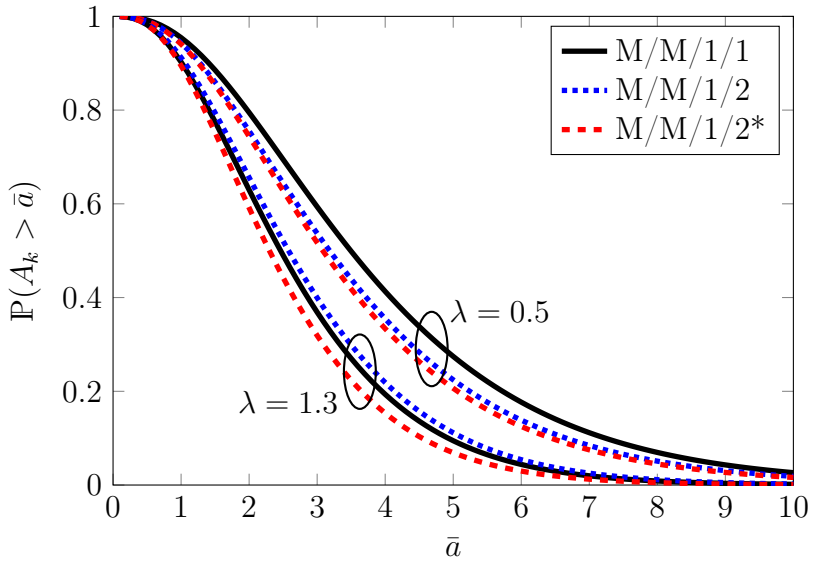


Figure 3.11: Peak age distribution: $\mathbb{P}(A_k > \bar{a})$ versus threshold a.

The average peak age for the three queuing models is illustrated in Figure 3.12. The M/M/1/2* model with packet replacement is shown to be the most adequate model for applications that require the age of information available to the receiver to be below a certain threshold. That is the case when the outdated information looses its value due to small correlation with the current state of the process under observation. We also observe that keeping a packet in the buffer is preferable in the case of small arrival rates, while discarding all the packets that find the server busy could be adopted for very large arrival rates ($\lambda > \mu$).

3.5 Future Work Discussion

The characterization of age of information is still incipient, and the investigation of simple models and new metrics remains important to understand its impact on the performance of communication systems. The fundamental limits and trade-

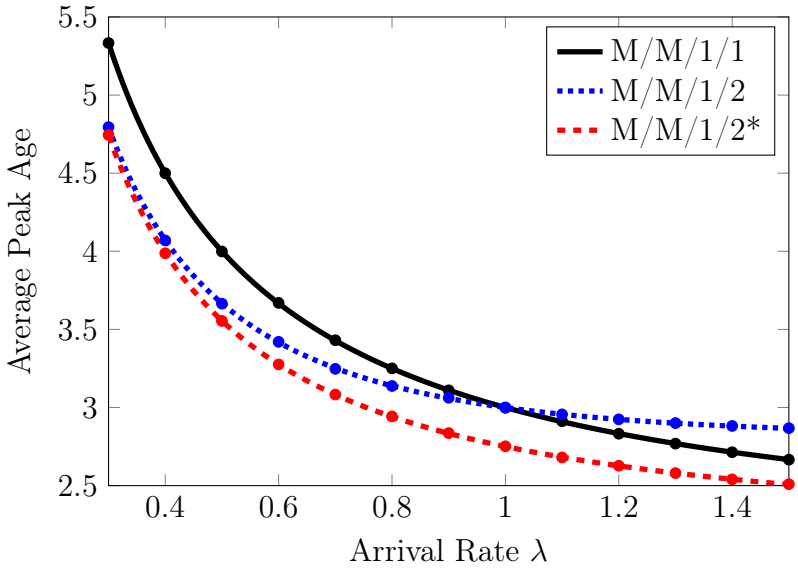


Figure 3.12: Average peak age versus arrival rate (λ) for the queuing models with packet management.

offs involving age of information deserve further investigation. In particular, the line of work initiated in [52] indicates that sending *just in time* updates does not provide a lower bound on the average age. Intuitively, it may be more effective to let the server idle and then transmit status update messages which cause larger reduction in the value of age, as opposed to transmitting status updates as soon as the server becomes available. This trade-off between the network resource allocated to deliver the message and the effect of that message on metrics of age needs to be understood.

The performance of multi-user networks with respect to new metrics which describe the timeliness of the transmitted messages still needs to be addressed. In this context, problems such as scheduling and resource allocation need to be investigated under the light of age of information. One possible approach to study the scenario with multiple users competing for network resources is to use game theory, formulating utility functions accounting for the age of information. In addition to

minimizing the age of the information transmitted by the users in the network, it is also relevant to consider the age of the information used for network control, such as the channel state information, and the network topology information used for routing protocols.

The information theoretic aspects of age of information also remain to be investigated. We envision that a new notion of capacity for point-to-point communication channels could be developed by associating the value of a bit of information with the age of that bit. If the value of a bit of information decreases with its age, then finite blocklength coding could be optimal, under this new notion of capacity. The problem of data compression can also be revisited under the light of age of information, with the definition of a distortion function which accounts for different values of a bit of information, depending on its age.

The development of a framework that accounts for the value of information as a function of its age would also impact other areas, such as control systems. Control systems are developed to manage the behavior of another system in order to improve performance, and the action of a controller can be determined by one or more observed variables in the system. Under resource constraints, the state of these variables is not available to the controller instantaneously. Instead, information about the state of these variables may be reported periodically or aperiodically to the controller. As a result, the controller may have to choose its action based on outdated information. In general, the effect of age of information is not incorporated in the decision process, although it is intuitive that timely information would lead to more accurate decision and control, so this is another promising area of investigation.

Finally, we identify the need to characterize the intuitive notion of *effective age*. A definition of the effective age should describe the value of the information as a function of its age. Depending on the application, and on the characteristics of the random process under observation, the value of the information available at a destination node may be degraded in a different fashion, other than linearly with time. For example, a definition of the effective age could account for the autocorrelation of the random process. When the process is highly correlated in time, a received message containing the status of the process can be valuable, despite a large age. In the extreme case that the observed process takes one constant value, a received status update message would age, but its effective age would be zero. The effective age of a received status update message could be described, for example, using the error in estimating the current value of the process with a Minimum Mean Square Error (MMSE) estimator. These ideas require further investigation, but the effective age will certainly play a key role in optimizing systems to deliver timely information.

3.6 Chapter Summary

In this chapter we have defined the process called age of information, with the objective to characterize the timeliness of the information transmitted in a communication system. This process is of interest particularly when an action is taken based on available information, but the information loses its value with time.

The contributions of this work to the initial steps in the characterization of

age in communication systems that transmit status update messages are twofold:

(*i*) We have characterized the average age of information for three models of status update systems, assuming that the source node can manage the samples of a process of interest, deciding to discard or to transmit status updates to the destination. We proposed a new queue model, named M/M/1/2*, to represent a packet management scheme in which packets are replaced in the buffer upon arrival of a newer packet. As a result of the packet replacement, the proposed queuing model cannot be analyzed using the classic approach in queuing theory, as Little's theorem does not apply to this peculiar queue. We have shown that package management with packet replacement in the buffer promotes smaller average age, when compared to schemes without replacement. For very large arrival rates, the proposed scheme also delivers *just in time* update messages, while avoiding the waste of network resources in the transmission of stale information.

(*ii*) In addition to the characterization of the average age, we proposed a new metric, named peak age. The peak age is a suitable metric to characterize the age of information in applications that impose a threshold on the value of age. The peak age has the advantage of a more simple mathematical formulation, and we have described its probability distribution for the three queuing models investigated in this chapter. This new metric will certainly benefit future investigations regarding the optimization of a network with respect to the timeliness of the transmitted information.

3.7 Appendix: Characterization of the M/M/1/2* Queue

In this thesis, we have proposed a modified queuing model, in which the packet waiting in queue is replaced if a new packet arrives. The new model was called M/M/1/2* queue. This appendix presents the characterization of this peculiar queuing model, in order to obtain the time in the system and waiting time of transmitted packets in a status update system with packet management.

Consider a queuing system with a single server, and a single space in a buffer. Let the packets arrive according to a Poisson process, so that the interarrival times are independent exponentially distributed random variables. The service time is assumed to be exponentially distributed, independent of the state of the queue. Packets are admitted if there are less than two packets in the system, that is, a maximum of two packets can be in the system at a given time, one in service and one waiting in the buffer. Packets that find the system full are discarded. Packets admitted to the system are served following a First-Come-First-Served (FCFS) discipline. A queuing system with these characteristics is identified with the notation M/M/1/2.

Modifying the M/M/1/2 model so that the packet waiting in queue is replaced if a new packet arrives yields the proposed model, named M/M/1/2* queue. Regarding the number of packets in the system, the M/M/1/2* model behaves exactly as a M/M/1/2 queue. That is because a replacement only occurs when an arriving packet finds the system full, so the replacement results in a packet being discarded. Intuitively, discarding the packet that just arrived or the packet that was already in

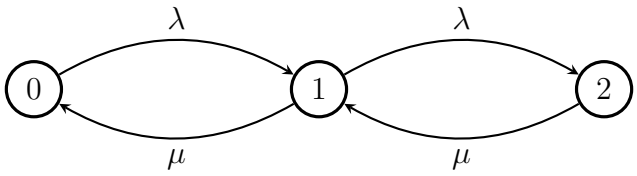


Figure 3.13: Three State Markov Chain for M/M/1/2* Model.

the buffer does not change the number of packets in the system.

To characterize the M/M/1/2* queue, we consider the stochastic process describing the number of packets in the system at a given time $\{N(t), t \geq 0\}$. The variable $N(t)$ takes values in the set $\{0, 1, 2\}$, and it evolves as in a three-state Markov chain, illustrated in Figure 3.13.

Denote with p_j the steady state probability of j packets in the system, with $j \in \{0, 1, 2\}$. The steady state probability distribution is calculated solving the set of equations

$$p_j \lambda = p_{j+1} \mu, \quad j \in \{0, 1\} \tag{3.85}$$

$$p_0 + p_1 + p_2 = 1. \tag{3.86}$$

We define the offered load $\rho \coloneqq \lambda/\mu$, and write the result of (3.85)-(3.86) as [76, Chapter 3]

$$p_j = \frac{\rho^j}{1 + \rho + \rho^2}, \quad j \in \{0, 1, 2\}. \tag{3.87}$$

The average total number of packets in the system is calculated as

$$\mathbb{E}[N] = 0p_0 + 1p_1 + 2p_2 \tag{3.88}$$

The effective arrival rate is given by the product of the arrival rate and the probability that the packet is admitted into the system. For Poisson arrivals, the

probability distribution of the state seen upon arrival is equal to the steady state probability distribution. This result is known as the Poisson Arrivals See Time Averages (PASTA) property [77, Chapter 3]. As a result, we have

$$\lambda_e = (1 - p_2)\lambda. \tag{3.89}$$

Next, we would like to characterize the time in the system. The classic result in queuing theory known as Little's theorem establishes the relationship between the expected time in the system, the expected number in the system, and the arrival rate [77, Chapter 5]. In the M/M/1/2* model, some packets leave the system after spending some time waiting in the buffer, while other packets end up being transmitted. For this reason, Little's theorem fails to apply to the transmitted packets in the M/M/1/2* model.

We will denote with T (without any index) the time in the system for any packet, which may or may not be discarded. Let $\mathbb{P}(T > t)$ describe the probability that a packet stays in the system for a time longer than t. Conditioning on the state of the server upon arrival (idle or busy), and on the event that the packet is transmitted (tx) or dropped (drop), we calculate

$$\begin{aligned}\mathbb{P}(T > t) = {} & \mathbb{P}(\text{idle})\mathbb{P}(T > t|\text{idle}) \\ & + \mathbb{P}(\text{busy, tx})\mathbb{P}(T > t|\text{busy, tx}) \\ & + \mathbb{P}(\text{busy, drop})\mathbb{P}(T > t|\text{busy, drop}).\end{aligned} \tag{3.90}$$

In what follows, we present the main arguments to calculate each term in (3.90).

An arriving packet finds the server idle with probability $\mathbb{P}(\text{idle}) = p_0$, due to

the PASTA property. In this case, the packet is served immediately, and

$$\mathbb{P}(T > t|\text{idle}) = e^{-\mu t}. \tag{3.91}$$

A packet finds the server busy with probability $\mathbb{P}(\text{busy}) = 1 - p_0$. According to the packet management scheme, it will be admitted in the system, but it will be transmitted if and only if no other arrival occurs while the service in progress is not completed. Let R represent a remaining service time, with probability density function $f(r)$, and let ϕ be the event that no other packet arrives during the remaining service. We define the probability $\mathbb{P}(\phi|R = r)$ by requiring that for every measurable set $A \subset [0, \infty)$, we have

$$\mathbb{P}(\phi, R \in A) = \int_A f(r)\mathbb{P}(\phi|R = r)dr.$$

The probability of transmission, given that the server was busy upon arrival is calculated as

$$\begin{aligned}\mathbb{P}(\text{tx}|\text{busy}) &= \int_0^\infty \mathbb{P}(\phi|R = r)f(r)dr \\ &= \int_0^\infty \frac{(\lambda r)^0}{0!}e^{-\lambda r}\mu e^{-\mu r}dr \\ &= \frac{\mu}{\lambda + \mu}.\end{aligned} \tag{3.92}$$

As a result,

$$\mathbb{P}(\text{busy,tx}) = (1 - p_0)\frac{\mu}{\lambda + \mu} \tag{3.93}$$

$$\mathbb{P}(\text{busy,drop}) = (1 - p_0)\frac{\lambda}{\lambda + \mu} \tag{3.94}$$

The time in the system for a transmitted packet can be written as the sum of waiting and service times, $T = W + S$. Conditioned on the event {busy, tx}, the

waiting time is distributed as the conditional distribution of the remaining service, given the event ϕ, calculated as

$$\begin{aligned} f(w|\text{busy, tx}) &= f(r|\phi) \\ &= \frac{\mathbb{P}(\phi|R=r)f(r)}{\int_0^\infty \mathbb{P}(\phi|R=r)f(r)dr} \\ &= \frac{\frac{(\lambda r)^0}{0!}e^{-\lambda r}\mu e^{-\mu r}}{\int_0^\infty \frac{(\lambda r)^0}{0!}e^{-\lambda r}\mu e^{-\mu r}dr} \\ &= (\lambda+\mu)e^{-(\lambda+\mu)r}. \end{aligned} \tag{3.95}$$

Hence, conditioned on the event {busy, tx}, W is exponentially distributed with parameter $(\lambda+\mu)$. The conditional probability density function of the time in the system is given by the convolution of the individual densities, since W and S are independent. The corresponding complementary cumulative distribution function we are looking for is

$$\mathbb{P}(T>t|\text{busy, tx}) = \frac{\lambda+\mu}{\lambda}e^{-\mu t} - \frac{\mu}{\lambda}e^{-(\lambda+\mu)t} \tag{3.96}$$

If the packet finds the server busy and is dropped while waiting, the time it spends in the system is exactly the time until the next arrival, to be denoted with X, which is exponentially distributed with parameter λ, conditioned on the event that the next arrival occurs before the end of the service in progress R. As a result, the conditional distribution can be calculated as

$$\begin{aligned} \mathbb{P}(T>t|\text{busy, drop}) &= \mathbb{P}(X>t|X<R) \\ &= \frac{\mathbb{P}(X>t, X<R)}{\mathbb{P}(X<R)} \\ &= \frac{\int_t^\infty (e^{-\lambda t}-e^{-\lambda r})\mu e^{-\mu r}dr}{\int_0^\infty (1-e^{-\lambda r})\mu e^{-\mu r}dr} \\ &= e^{-(\lambda+\mu)t} \end{aligned} \tag{3.97}$$

Finally, using the arguments above, we rewrite (3.90) as

$$\begin{aligned}\mathbb{P}(T > t) = p_0 \exp(-\mu t) \\ &+(1-p_0)\frac{\mu}{\lambda+\mu}\left[\frac{\lambda+\mu}{\lambda}e^{-\mu t} - \frac{\mu}{\lambda}e^{-(\lambda+\mu)t}\right] \\ &+(1-p_0)\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}. \end{aligned} \tag{3.98}$$

From (3.98), we also have

$$\begin{aligned}\mathbb{P}(T > t|\text{tx}) = \frac{p_0}{p_0 + (1-p_0)\frac{\mu}{\lambda+\mu}}e^{-\mu t} \\ &+\frac{(1-p_0)\frac{\mu}{\lambda+\mu}}{p_0 + (1-p_0)\frac{\mu}{\lambda+\mu}}\left[\frac{\lambda+\mu}{\lambda}e^{-\mu t} - \frac{\mu}{\lambda}e^{-(\lambda+\mu)t}\right], \end{aligned} \tag{3.99}$$

and the expected value

$$\mathbb{E}[T|\text{tx}] = \frac{1}{\mu} + \frac{\lambda}{(\lambda+\mu)^2}. \tag{3.100}$$

The expected value of the waiting time for a transmitted packet is calculated using (3.95) to be

$$\begin{aligned}\mathbb{E}[W|\text{tx}] &= \frac{(1-p_0)\frac{\mu}{\lambda+\mu}}{p_0 + (1-p_0)\frac{\mu}{\lambda+\mu}}\left(\frac{1}{\lambda+\mu}\right) \\ &= \frac{\lambda}{(\lambda+\mu)^2}. \end{aligned} \tag{3.101}$$

Chapter 4: Age of Channel State Information

4.1 Overview

In this chapter, we discuss a specific application of the concept of age of information, proposing a new framework to analyze the effect of outdated channel state information on the performance of communication networks. We consider communication systems which rely on the feedback of Channel State Information (CSI) to perform adaptation functions based on the channel conditions between the source and the destination nodes. The CSI feedback is needed when the channel between source and destination is non-reciprocal with the reverse channel, hence the source node is not able to learn about the channel conditions directly, and requires the destination node to report this information. The CSI may be reported periodically or aperiodically.

As a result of the intervals between consecutive reports, in addition to the delays due to framing, transmission, and processing times, the information available to the transmitter can be an outdated CSI – an effect called *aging* of the CSI. The aging of the CSI may lead to inefficient or even erroneous communication, thus it should be taken into account when designing feedback channels and protocols for use of this feedback. Among the factors causing the *aging* of the CSI we list

(i) measurement times, (ii) transmission delay, (iii) processing time required to decode and estimate the channel quality, (iv) processing time to run adaptation functions, (v) frame times, and (vi) the interval between consecutive feedbacks.

Our proposed framework based on the age of information is focused mainly on the effect of periodic feedback on the performance of communication networks. We consider a Finite State Markov Channel (FSMC) model, assuming that the channel state takes values in a discrete, finite set. This simple channel model provides mathematical tractability, and yields reasonably useful results and insights.

Finite state channel models have also been used before to analyze the effect of delayed CSI. Viswanathan presented the capacity of a FSMC for a single user as a function of the feedback delay in [78]. The capacity region of the finite state multiple access channel with delayed CSI at the transmitter was presented in [79]. As opposed to [78, 79], we do not focus on the characterization of the achievable data rates. Instead, we formulate utility functions as a general performance metric that can even account for the cost of feedback. The cost of training and feedback has also been addressed in [80], but assuming a Gaussian broadcast channel, a one step Minimum Mean Square Error (MMSE) estimator of the channel, and the sum rate as performance metric.

The performance degradation caused by delayed CSI was also verified by simulation. In [81], the authors have studied the degradation in the sum capacity caused by the feedback delay, and concluded that the delay is a dominant factor in defining the upper bound for achievable data rates in multi-user diversity systems. In [82], the authors also analyzed through simulation the effect of channel information delays

on the sum-rate, showing that multi-user Multiple-Input Multiple-Output (MIMO) retains the advantage over the single-user case for typical operating conditions regarding Doppler shift and channel information reporting intervals. Our work complements these observations with an analytic characterization of the effect of the age of CSI on the performance of a wireless link and, thus, provides valuable insight for design and parametrization of channel adaptation functions.

The remaining of this chapter is organized as follows: section 4.2 presents the system model, including the FSMC model adopted in this work. The periodic feedback scheme is described in section 4.3, where we exploit the Markovian property of the channel, using the CSI to estimate the current state of the channel. We associate rewards (or penalties) with the decision about the channel state and choose the decision rule that maximizes the expected reward with respect to the distribution of the current channel state. Section 4.4 presents possible choices for the rewards, and the definition of a general utility function as a performance metric for the network, also accounting for the cost of feedback. In section 4.5 we use a simple channel model with two states to investigate the effect of the age of CSI on the probability of error in the channel estimation. Section 4.6 presents an application of our proposed framework for the case of Rayleigh fading channels, and we discuss the effect of the age of CSI on the performance, considering the Signal-to-Noise Ratio (SNR) and maximum Doppler shift in the channel. The application of our framework to multiplexing systems is presented in section 4.7, where we consider that multiple orthogonal resource blocks are allocated to one user which should report a vector with the CSI of all channels, and we analyze an alternating feedback

scheme. The discussion about future lines of investigation is presented in section 4.8, and the contributions in this chapter are summarized in section 4.9.

4.2 System Model

We consider a communication link between a Base Station (BS) and a mobile user. Information is transmitted in time slots. The mobile user obtains samples of the channel conditions in the downlink and reports the CSI to the BS in the beginning of each frame of duration τ slots. Let n denote the index of each time slot in a frame, $n \in \{0, 1, \ldots, \tau - 1\}$.

We adopt a FSMC model, assuming that the channel conditions evolve as in a Markov chain χ. The channel is assumed to be stationary, so that the probability transition matrix for this Markov chain is independent of the time instant. We also assume that the channel conditions remain constant for the duration of one time slot.

Figure 4.1 illustrates the channel model. The Markov chain has M states in the set $\mathcal{S} = \{1, 2 \ldots, M\}$ which represent different fading conditions. Let S_n denote the channel state at time slot n, $S_n \in \mathcal{S}$, $n \in \{0, 1, \ldots \tau - 1\}$. The transition probability from $S_n = s$ to $S_{n+1} = m$ is denoted with $\mathrm{P}_{s,m}$, where $s, m \in \mathcal{S}$. In this Figure, we illustrate solely the self-transitions and the transitions to neighbor states, for simplicity.

The use of FSMC to model time-varying fading channels is well established in the literature [83]. The FSMC is adequate to model channels with memory

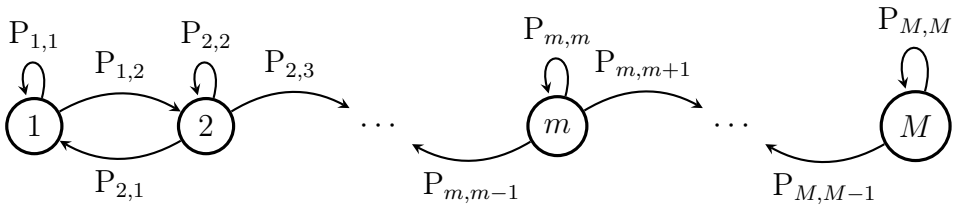


Figure 4.1: Finite State Markov Channel Model.

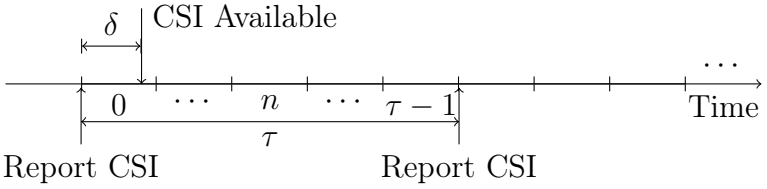


Figure 4.2: Periodic reports of Channel State Information (CSI), with period τ. The CSI may be available only after an interval δ.

and to exploit the advantages of the use of feedback. The FSMC is also suitable to model channel error bursts, to design coding and decoding schemes for channels with memory, and to devise efficient adaptive schemes with reduced feedback bandwidth. It has mathematical tractability, at the same time it provides useful insights.

We have mentioned several factors that cause the *aging* of the CSI. We represent with a random variable δ the time elapsed since the user obtained the channel knowledge, until the instant it will be used by the BS, which includes delay and processing times, mentioned in section 4.1 as factors $(\boldsymbol{i})$-$(\boldsymbol{v})$.

The interval between consecutive reports mentioned in factor $(\boldsymbol{vi})$ will be denoted with τ, which is deterministic in the case of periodic feedback, and random in the case of aperiodic feedback. In current standards, such as Long Term Evolution (LTE), both periodic and aperiodic feedback modes are supported [61, Chapter 10]. The case of periodic feedback, with constant interval between consecutive reports, is illustrated in Figure 4.2. The interval τ should be selected such that $\mathbb{P}(\delta > \tau) = 0$.

The work presented in this thesis investigated the effect of the interval between consecutive reports, τ, under the light of the age of information concept. Small values of τ correspond to frequent updates of the CSI sent to the BS through the feedback channel, but this parameter can not be made arbitrarily small, due to signaling overhead and energy consumption at the user equipment. On the other hand, very large values of τ may compromise the performance of the system, since the adaptation functions performed by the BS make use of outdated CSI. For this reason, the effect of the age of CSI should be well understood in order to implement efficient adaptation.

4.3 Channel Feedback and Channel Estimation

In this section we describe the feedback mechanism and the channel state estimate to be used by the BS, considering the periodic feedback scheme illustrated in Figure 4.2.

The CSI is sent from the mobile to the BS at the beginning of a frame, and each frame has τ time slots. At first, we assume that the channel state is available to the BS without errors or delays in the first time slot of the frame. That is, to study the effect of τ on the performance, we disregard the variable δ by assuming $\mathbb{P}(\delta = 0) = 1$.

The channel state in a time slot takes one value in the set $\mathcal{S}$. Let P represent the probability transition matrix of the Markov chain illustrated in Figure 4.1. The matrix P is a $M \times M$ matrix with entry $\mathrm{P}_{s,m}$ in the sth row and mth column, where

$\mathrm{P}_{s,m} = \mathbb{P}[S_{j+1} = m | S_j = s]$, with $s, m \in \mathcal{S}$. Under the assumption of a stationary channel, $\mathrm{P}_{s,m}$ is independent of j. The n-step transition probability matrix $\mathrm{P}^{(n)}$ is the $M \times M$ matrix with entries $\mathrm{P}^{(n)}_{s,m} = \mathbb{P}[S_{j+n} = m | S_j = s]$, for $s, m \in \mathcal{S}$, and it is also obtained as the nth power of the transition probability matrix P.

The probability distribution of the channel state S_n in time slot n can be determined using the n-step transition probability matrix, and the initial channel state in the frame, denoted with S_0, $S_0 \in \mathcal{S}$, which is learned through the feedback channel. We define the vector

$$\mathbb{P}(S_n | S_0 = s) = \mathbb{P}(S_0 = s)\mathrm{P}^n, \quad n = 1, \ldots, \tau - 1, \tag{4.1}$$

where $\mathbb{P}(S_0 = s) = e_s$, a row vector in $\mathbb{R}^M$ with all entries equal to zero, except for the entry in the sth position which is equal to one, since we assume that S_0 is known to the BS in the beginning of the frame. In other words, given the initial observation in the frame, $S_0 = s$, the probability distribution of the channel state in time slot n of the same frame is equal to the sth row in the n-step transition probability matrix P^n, for $n = 1, \ldots, \tau - 1$.

At the BS a decision is made about the channel state at time slot n, $n \in \{0, 1, \ldots, \tau - 1\}$. We define a decision variable, to be denoted with D_n, which could represent the channel state estimated at the BS, or it could be a map from the channel state to an action. We will assume the first, hence $D_n \in \mathcal{S}$. For a pair $(S_n, D_n) = (m, d)$, which indicates the channel state and the decision at time slot n, we associate a reward, to be denoted with $\mathrm{R}_{m,d}$. The reward takes values in the real numbers, $\mathrm{R}_{m,d} \in \mathbb{R}$ for all $m, d \in \mathcal{S}$. The reward matrix is the $M \times M$ matrix

R, with entries $\mathrm{R}_{m,d}$, $m, d \in \mathcal{S}$.

Denote with $r(n)$ the received reward at time slot n. At time slot n, only the distribution of the channel state S_n is known. Hence, for a given decision variable $\{D_n = d\}$, we calculate the conditional expected received reward, conditioning on the initial observation $\{S_0 = s\}$, as

$$\mathbb{E}[r(n)|S_0 = s, D_n = d] = \sum_{m \in \mathcal{S}} \mathrm{R}_{m,d} \mathbb{P}(S_n = m|S_0 = s), \quad n = 1, \ldots, \tau - 1, \tag{4.2}$$

where the probability in the right-hand side does not depend on the decision variable D_n because, given the initial channel state S_0, the decision made at the BS does not bring any additional information about the actual state of the channel at the time slot n.

Let $\bar{\mathrm{R}}$ be the matrix containing the conditional expected rewards at time slot n ($n = 0, 1, \ldots, \tau - 1$), given the initial observation $\{S_0 = s\}$, and the decision variable $\{D_n = d\}$. Note that we can write this matrix using the n-step transition probability and the reward matrix, as $\bar{\mathrm{R}} = \mathrm{P}^n \, \mathrm{R}$. Each entry in row s and column d will be

$$\bar{\mathrm{R}}_{s,d} = \mathbb{E}[r(n)|S_0 = s, D_n = d], \quad s \in \mathcal{S}, \; d \in \mathcal{S}. \tag{4.3}$$

We define the decision rule to be the determination of the value of D_n that maximizes the expected reward in a time slot, given that $\{S_0 = s\}$. In this case, the decision variable takes the value

$$d^* \coloneqq \arg\max_{d \in \mathcal{S}} \mathbb{E}[r(n)|S_0 = s, D_n = d]. \tag{4.4}$$

When the decision variable takes the value in (4.4), we write the reward in the time slot as $r^*(n)$. The expected received reward is obtained averaging over

the probability distribution of S_0. We assume that S_0 is distributed according to the steady state distribution, denoted with $\pi = [\pi_1, \pi_2, \ldots, \pi_M]$. For each time slot $n = 1, \ldots, \tau - 1$, we have

$$\mathbb{E}[r^*(n)] = \sum_{s \in \mathcal{S}} \mathbb{E}[r(n)|S_0 = s, D_n = d^*]\pi_s. \tag{4.5}$$

4.4 Rewards and Utility Definition

In this section, we describe the selection of the rewards $\mathrm{R}_{m,d}$, $m, d \in \mathcal{S}$, for many cases of interest. Recall that $\mathrm{R}_{m,d}$ is the reward associated with channel state $S_n = m$ and decision about the channel state $D_n = d$. First, we define the rewards in order to obtain the probability of error in the decision regarding the channel state. Later, we present alternative definitions of rewards, as functions of the transmission rate achievable under certain channel conditions.

A general utility function is defined taking into account the rewards accumulated in a frame of duration τ, as well as the cost of feedback.

4.4.1 Error Probability

In this subsection we define the rewards so that the decision rule defined in (4.4) minimizes the probability of error in selecting the channel state. One possibility is to treat all error events equally, by attributing penalties $\mathrm{R}_{m,d} = -1$ if $m \neq d$, and $\mathrm{R}_{m,m} = 0$ for all $m \in \mathcal{S}$. In this case, all the entries in the matrix $\bar{\mathrm{R}}$ of expected

rewards are of the form

$$\bar{\mathrm{R}}_{s,d} = -\sum_{\substack{m\in\mathcal{S}\\ m\neq d}} \mathbb{P}(S_n = m|S_0 = s), \quad s \in \mathcal{S},\ d \in \mathcal{S}, \tag{4.6}$$

hence the decision rule defined in (4.4) selects the most likely state according to the conditional distribution $\mathbb{P}(S_n|S_0 = s)$, in order to minimize the error defined as

$$E_n(s) \coloneqq \sum_{\substack{m\in\mathcal{S}\\ m\neq d}} \mathbb{P}(S_n = m|S_0 = s), \quad s \in \mathcal{S},\ d \in \mathcal{S}. \tag{4.7}$$

It may be desirable to identify different types of errors, and attribute different penalties to them. For example, consider that $\mathcal{S}$ is an ordered set, representing better channel conditions with a larger number $m \in \{1, \ldots, M\}$, so the order of the states is important. When overestimating has a different effect than underestimating, it may be desired to penalize only the overestimation as an error, which could be associated with an outage, as the channel conditions are assumed to be better than they actually are, and the selected transmission rate is not supported by the channel. A trivial result is obtained if we set $\mathrm{R}_{m,d} = -1$ when $m < d$, and $\mathrm{R}_{m,d} = 0$ when $m \geq d$, for all $m \in \mathcal{S}$. In this case, the decision rule in (4.4) yields $d^* = 1$, for all $s \in \mathcal{S}$. In other words, if there is no reward to risk, the optimal decision is very conservative, and always selects the smallest channel state. In the next subsection, we define the rewards proportional to the transmission rates supported by the channel conditions, and obtain non-trivial decision rules.

4.4.2 Rate Rewards

Let the states of the Markov chain χ be associated with different transmission modes over the communication link. For each mode of operation identified by m, $m \in \{1, 2, \ldots, M\}$, the SNR takes values in the interval (Γ_m, Γ_{m+1}). This work is not concerned with the modulation and coding for each transmission mode. Instead, we simply describe the spectral efficiency associated with mode m, using the SNR value $\gamma_m \in [\Gamma_m, \Gamma_{m+1})$, and the approximation $\log(1 + \gamma_m)$ for the rate.

The reward associated with the pair $(S_n = m, D_n = d)$ can be defined to penalize overestimation errors as follows:

$$\mathrm{R}_{m,d} = \begin{cases} -\log(1+\gamma_d) & \text{if } m < d, \\ \log(1+\gamma_d) & \text{if } m \geq d. \end{cases} \tag{4.8}$$

It is also possible to penalize both types of errors, but with different penalties. One interesting case is to set the penalties as the difference between the rate obtained and the rate supported by the channel. When the channel is overestimated, the rate obtained is zero, and this decision is penalized with the negative of the rate that could have been transmitted. When the channel is underestimated, the communication with a lower rate is assumed to be successful, but the transmission rate could have been larger, so this larger rate is subtracted as a penalty. The rate rewards would be defined as follows:

$$\mathrm{R}_{m,d} = \begin{cases} -\log(1+\gamma_m) & \text{if } m < d, \\ \log(1+\gamma_d) & \text{if } m = d, \\ \log(1+\gamma_d) - \log(1+\gamma_m) & \text{if } m > d. \end{cases} \tag{4.9}$$

4.4.3 Utility function and the age-performance trade-off

We have defined a reward $r(n)$, received in each time slot, as a function of the real channel state S_n and the decision made about the channel state D_n. Because the current channel state is unknown, we calculate its expected value, assuming that the channel state has the Markov property, and that the channel state is sent through a feedback channel once per frame.

Reporting the CSI once per frame has a cost. One can model, for example, the energy spent on feedback, or the time and bandwidth dedicated to this activity. In any of these cases, as the frame lasts longer, the cost per time slot becomes smaller. On the other hand, the channel state estimate becomes less accurate if the channel feedback is too sparse. Clearly, when we consider the cost of feedback, we identify the trade-off between the age of the CSI and the performance of the system.

To define a general utility function, we denote with C_τ the average cost per time slot to report the CSI. We require that C_τ is inversely proportional to τ. For example, C_τ can be simply the total feedback cost, expressed in appropriate units, divided by τ. We also define the average expected reward in a frame, denoted with R_τ, and obtained by dividing the cumulative expected reward by the length of the frame. Finally, the general utility function is defined as

$$U_\tau = \frac{1}{\tau} \sum_{n=0}^{\tau-1} \mathbb{E}[r^*(n)] - C_\tau, \tag{4.10}$$

where $\mathbb{E}[r^*(n)]$ was defined in (4.5) as the expected reward under the decision rule in (4.4).

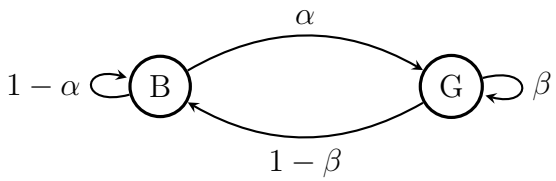


Figure 4.3: Gilbert-Elliot Channel Model.

4.5 Analysis for a Gilbert-Elliot Model

In this section, we consider a particular case of FSMC, namely the Gilbert-Elliot channel model, assuming that the number of states is $M = 2$. We use this simple model to illustrate the effect of the age of CSI on the probability of error in estimating the channel state. We exploit the simplicity of the channel model to understand the effect of age under different channel conditions.

Consider the channel model presented in Figure 4.3. The two-state Markov chain represents the fading conditions as *bad* or *good*, described by the set $S \coloneqq \{B, G\}$. The channel state at time slot n is denoted with S_n, $S_n \in S$. The transition probability from (B) to (G) is denoted with α, and transition probability from (G) to (B) is given by $1 - \beta$.

For $n \geq 0$, the n-step transition matrix is given by

$$P^{(n)} = \begin{bmatrix} \frac{(1-\beta)+\alpha(\beta-\alpha)^n}{\alpha+1-\beta} & \frac{\alpha-\alpha(\beta-\alpha)^n}{\alpha+1-\beta} \\ \frac{(1-\beta)-(1-\beta)(\beta-\alpha)^n}{\alpha+1-\beta} & \frac{\alpha+(1-\beta)(\beta-\alpha)^n}{\alpha+1-\beta} \end{bmatrix}, \tag{4.11}$$

and the steady state distribution of the states is

$$\pi = [\pi_B, \ \ \pi_G] = \left[\frac{1-\beta}{\alpha+1-\beta}, \ \ \frac{\alpha}{\alpha+1-\beta}\right], \tag{4.12}$$

which is obtained from the simple linear system $\pi P^{(1)} = \pi$, under the condition that

$\pi_B + \pi_G = 1$.

Recall that $\mathbb{P}(S_n|S_0 = s)$ is the conditional distribution of S_n, given that the channel state in the first time slot of the frame is $S_0 = s$. Then

$$\begin{aligned}\mathbb{P}(S_n|S_0 = B) &= \left[\frac{(1-\beta)+\alpha(\beta-\alpha)^n}{\alpha+1-\beta},\ \frac{\alpha[1-(\beta-\alpha)^n]}{\alpha+1-\beta}\right], \\ \mathbb{P}(S_n|S_0 = G) &= \left[\frac{(1-\beta)[1-(\beta-\alpha)^n]}{\alpha+1-\beta},\ \frac{\alpha+(1-\beta)(\beta-\alpha)^n}{\alpha+1-\beta}\right].\end{aligned} \tag{4.13}$$

For each time slot n, let $\mathbb{P}(S_n|S_0 = B)(i)$ or $\mathbb{P}(S_n|S_0 = G)(i)$ be the ith entry in the vectors defined in (4.13), with $i \in \{1, 2\}$. The conditional probability of error at time slot n, given the initial observation S_0, will be denoted with $E_n(S_0)$, and is given by

$$\begin{aligned}E_n(B) &= \min\left\{\mathbb{P}(S_n|S_0 = B)(1),\ \mathbb{P}(S_n|S_0 = B)(2)\right\}, \\ E_n(G) &= \min\left\{\mathbb{P}(S_n|S_0 = G)(1),\ \mathbb{P}(S_n|S_0 = G)(2)\right\}.\end{aligned} \tag{4.14}$$

The probability of error in the nth time slot of a frame is obtained averaging over the distribution for the initial state, which is assumed to be equal to the steady state distribution, yielding

$$E_n = \pi_B E_n(B) + \pi_G E_n(G). \tag{4.15}$$

Note that $|\beta - \alpha| < 1$, since $\alpha \in (0, 1)$ and $\beta \in (0, 1)$. Hence, in the distributions in (4.13), the terms depending on n are decreasing as n increases. Since the channel is modeled by a finite-state Markov chain (which is irreducible, aperiodic, and positive recurrent), each line of the n-step transition matrix converges to the invariant distribution π [84, Chapter 4]. As a result, in the limit, the error tends to

the smallest value in the steady state distribution:

$$\lim_{n\to\infty} E_n = \min\{\pi_B, \pi_G\}. \tag{4.16}$$

We will say that the channel is unbiased if both states have equal steady state probability, or $\pi = [0.5,\ 0.5]$, and the transition probabilities are such that $\beta = 1-\alpha$. In the particular case of unbiased channels, we can write the probability of error as

$$E_n = \frac{1-(2\alpha-1)^n}{2}, \quad \alpha \in (0.5, 1), \tag{4.17}$$

which tends to 0.5 as n goes to infinity. In the case of $\alpha = 0.5$, the dependence on n is eliminated, and the probability of error is equal to that using the steady state distribution, which reflects the case in which no channel knowledge is reported.

Figure 4.4 illustrates the probability of error for unbiased channels, $\pi = [0.5, 0.5]$. With the assumption of $\delta = 0$, the age is zero in time slot $n = 0$, when the BS is assumed to know the current channel state. Then, the channel knowledge ages, since no information is reported. As α gets closer to unity, the probabilities of transitioning between the two channel states becomes large, and the probability of error increases slowly with the channel age. When α decreases towards 0.5, the probability of error is more severely affected by increasing the age.

To understand the effect of the channel information age under different channel conditions, we plot the probability of error versus the parameters α and β in Figure 4.5. Figure 4.5(a) shows the probability of error versus the transition probability α. We consider $\beta \in \{0.1, 0.5, 0.8\}$. The probability of error plotted for $n = 1$ is shown in dotted curves with circles, and maximized for $\alpha = 0.5$. The limiting cases with $n \to \infty$ are shown as full lines, and maximized when $\alpha = 1 - \beta$. Clearly, when

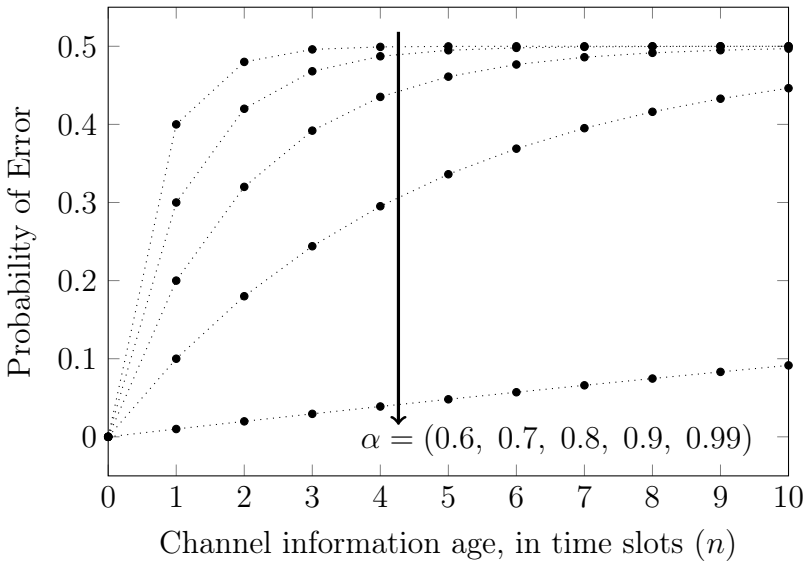
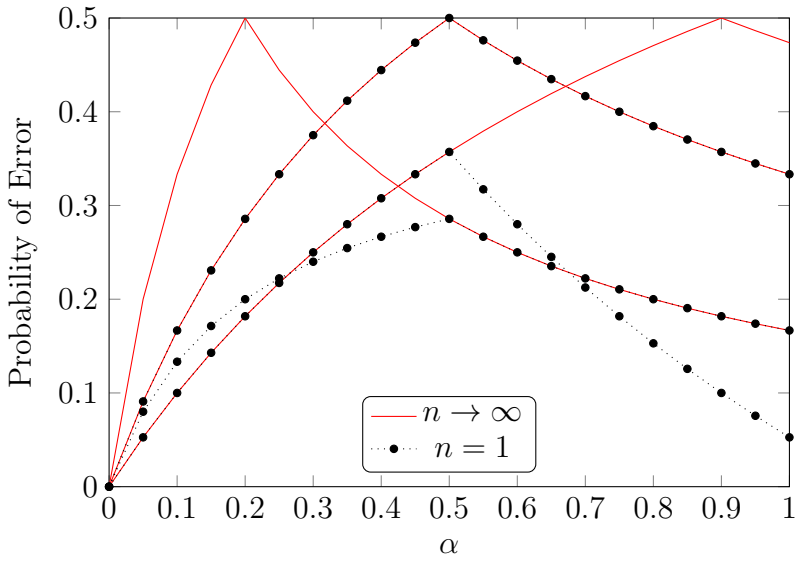


Figure 4.4: Probability of error E_n versus frame duration for unbiased channel model.

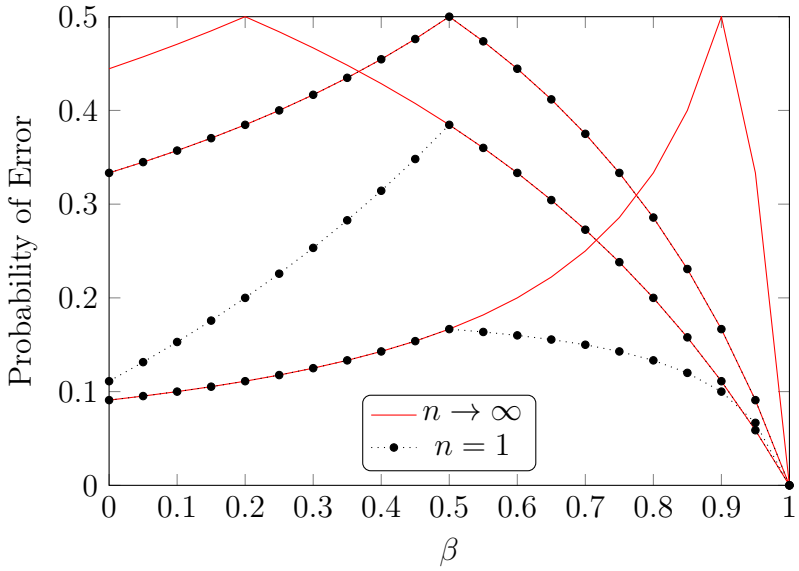
$\alpha = 1 - \beta = 0.5$, the effect of channel information age is irrelevant, and the two curves overlap for all α. When $1 - \beta < 0.5$, the probability of error is affected by the age of channel information only if $\alpha < 0.5$. On the other hand, if $1 - \beta > 0.5$, then the probability of error is affected by n only if $\alpha > 0.5$.

Figure 4.5(b) corroborates the previous observations, showing the probability of error versus the transition probability β. Again, dotted lines with circles identify the case $n = 1$, while full lines identify the limiting case as $n \to \infty$. The full lines achieve the maximum value of probability of error for $\beta = 1 - \alpha$. The three values shown in Figure 4.5(b) are for $\alpha \in \{0.1, 0.5, 0.8\}$. For $n = 1$, the probability of error is maximized with $\beta = 0.5$, independent of α. When $1 - \alpha < 0.5$, the probability of error increases with n only if $\beta < 0.5$. In the cases with $1 - \alpha > 0.5$, the probability of error is affected by n only if $\beta > 0.5$.

The observations from Figure 4.5 are summarized in Figure 4.6. The values



(a) Probability of error E_n versus α.



(b) Probability of error E_n versus β.

Figure 4.5: Probability of error versus the channel state distribution parameters α and β.

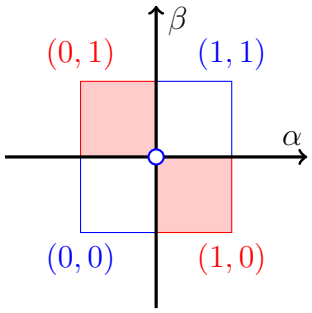


Figure 4.6: Sensitivity of probability of error to the age of information.

for the pair (α, β), which characterize the channel, are divided in four regions, and the point at the center corresponds to the values $(0.5, 0.5)$. When the channel conditions are characterized by parameters α and β within the shadowed regions, then the probability of error is affected by the age of the information, represented in number of time slots by the parameter n.

4.6 Application to Rayleigh Fading

In this section, we consider the general FSMC model as described in Figure 4.1. Each channel state is used to represent a different level of SNR under the assumption of a Rayleigh fading channel between the mobile user and the BS.

We define M intervals for the values of SNR, which are associated to each of the M states in the Markov chain χ. Let γ represent the instantaneous SNR value. The channel state S_n belongs to a set $\mathcal{S} = \{1, 2, \ldots, m, \ldots, M\}$, with $\{S_n = m\}$ if and only if $\gamma \in [\Gamma_m, \Gamma_{m+1})$ during time slot n. To characterize the FSMC, we define the thresholds Γ_m, and then describe the transition probability matrix.

First recall that, for a Rayleigh fading channel, the received SNR is exponen-

tially distributed as

$$f(\gamma) = \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right), \tag{4.18}$$

where γ_0 is the average SNR.

The steady state probability that the channel is in state $\{S_n = m\}$ is calculated as

$$\pi_m = \int_{\Gamma_m}^{\Gamma_{m+1}} \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right) d\gamma, \quad m \in \mathcal{S}. \tag{4.19}$$

To determine the thresholds Γ_m, we require that the steady state distribution is uniform. That is, all the states have equal probabilities $\pi_m = 1/M$, for all $m \in \mathcal{S}$ [85]. Other partitions are of course possible, as shown in [86], and [87].

Define $\Gamma_1 = 0$, and $\Gamma_{M+1} = \infty$. Then we obtain the remaining thresholds by solving

$$\exp\left(-\frac{\Gamma_m}{\gamma_0}\right) - \exp\left(-\frac{\Gamma_{m+1}}{\gamma_0}\right) = \frac{1}{M}, \quad m \in \mathcal{S}. \tag{4.20}$$

To calculate the transition probabilities, we assume that the transitions occur only between neighboring states or are self-transitions. The transition probabilities are approximated as the product of the level crossing rate $N(\Gamma)$ by the duration of the time slot (T), divided by the probability of being in the first state [88, Equation 3.48], as follows:

$$\mathrm{P}_{m,m+1} = \frac{N(\Gamma_{m+1})T}{\pi_m}, \quad m \in \{1, \ldots, M-1\}, \tag{4.21}$$

$$\mathrm{P}_{m,m-1} = \frac{N(\Gamma_m)T}{\pi_m}, \quad m \in \{2, \ldots, M\}, \tag{4.22}$$

where, in the case of a Rayleigh fading, the level crossing rate N for the received signal power, for a given level Γ_m, average SNR γ_0, and maximum Doppler shift f_m,

can be written as [88, Equation 3.44]

$$N(\Gamma_m) = \sqrt{2\pi \frac{\Gamma_m}{\gamma_0}} f_m \exp\left(-\frac{\Gamma_0}{\gamma_0}\right). \tag{4.23}$$

To illustrate the effect of the age of CSI under the FSMC model for the Rayleigh application, we consider the rate rewards, as defined in (4.9). For each channel state, the SNR is such that $\gamma \in [\Gamma_m, \Gamma_{m+1})$, and we let

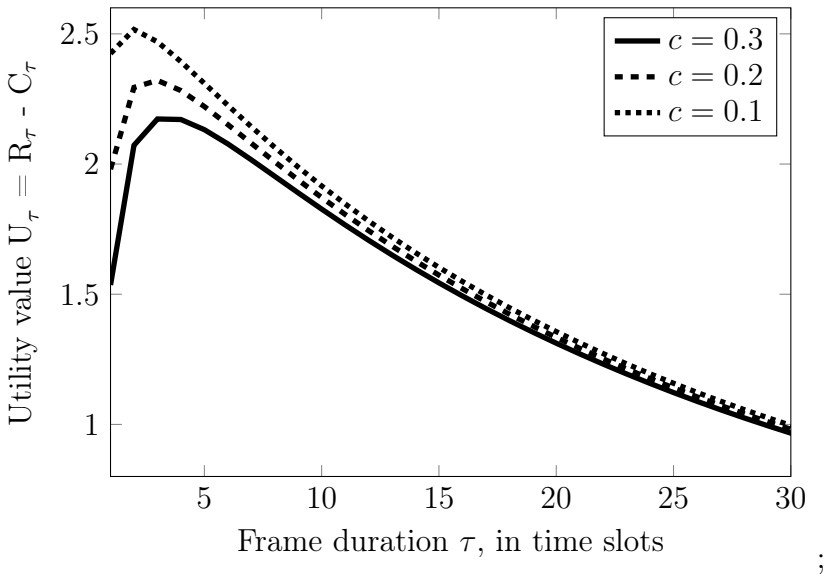
$$\gamma_m \coloneqq \frac{\Gamma_m + \Gamma_{m+1}}{2}, \quad m \in \{1, \ldots, M-1\}, \tag{4.24}$$

$$\gamma_M \coloneqq \Gamma_M, \tag{4.25}$$

$$\mathrm{R}_{m,m} = \log_2(1+\gamma_m) \quad m \in \{1, \ldots, M\}. \tag{4.26}$$

Figure 4.7 illustrates the utility value U_τ as described in (4.10), assuming a Rayleigh fading channel model with $M = 8$ states, $\gamma_0 = 10$ dB, and $f_m = 5$ Hz. The feedback cost is modeled as a fraction of the maximum rate, that is, $C_\tau = c \log_2(1+\gamma_M)/\tau$, with $c \in (0,1)$, modeling in-band feedback cost. We observe a trade-off between the utility value and the frame duration, since both the average reward and the cost are reduced when a longer frame is considered. As expected, larger values of c yield smaller utility values, but we also note that this difference is significantly reduced as the frame duration increases. That indicates the importance of addressing the age of the CSI available in the first place, and then promoting a reduction of the feedback cost.

Figure 4.8 illustrates the utility value U_τ in (4.10) for a Rayleigh fading channel model with $M = 4$ states, assuming $\gamma_0 = 10$ dB. The in-band feedback is assumed to be fixed $C_\tau = \log_2(1+\gamma_1)/\tau$, while we vary the maximum Doppler shift f_m. Clearly,



;

Figure 4.7: Utility value versus frame duration for different feedback costs.

the utility values are smaller for larger values of f_m, which is expected, since the channel conditions change faster in this scenario. For larger values of f_m, the utility value also decays faster with the age of CSI. We conclude that reducing the age of CSI is particularly important in scenarios with large values of f_m, which is the case, for example, in high-mobility applications.

Figure 4.9 presents the utility value U_τ in (4.10) for a Rayleigh fading channel model, varying the number of states M, assuming $\gamma_0 = 10$ dB, and $f_m = 5$ Hz. The in-band feedback is assumed to be proportional to the smallest rate reward, $C_\tau = \log_2(1 + \gamma_1)/\tau$. Larger values of M yield more SNR levels, reducing the value of γ_1. Consequently, the feedback cost as we modeled here is decreasing with M. For small feedback cost, the utility function may be strictly decreasing with the frame duration, since the predominant effect is the decreasing rate reward, as observed in the model with $M = 32$. Modeling the channel with larger M increases the sensitivity to the age of CSI, but it may also yield larger utility values for small

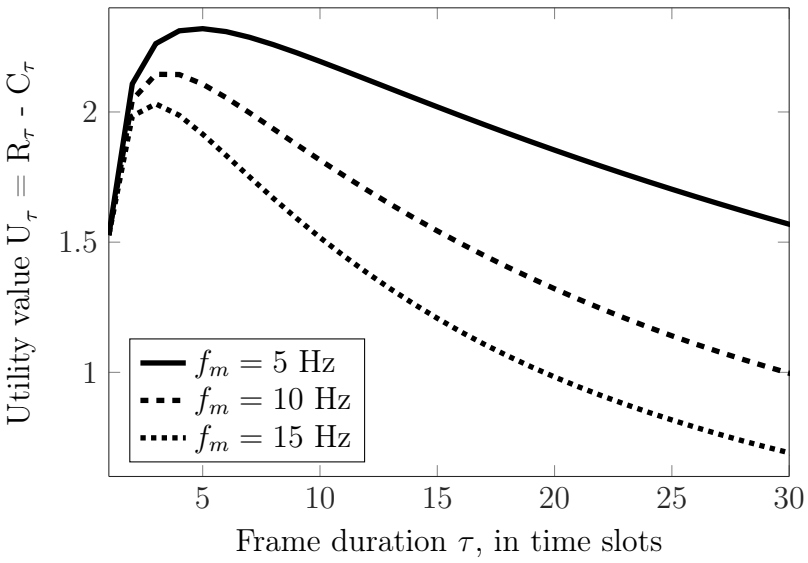


Figure 4.8: Utility value versus frame duration for different values of maximum Doppler shift.

frames, when the cost of feedback is decreasing with M.

4.7 Application to Multiplexing

Let us now investigate the effect of the age of channel state information when a mobile user is assigned multiple orthogonal resource blocks for communication. This application is motivated by the use of multiple access techniques as Orthogonal Frequency Division Multiple Access (OFDMA) in current standards for mobile networks. The ultimate goal would be to develop resource allocation and feedback schemes when a large number of resource blocks are allocated to multiple users, making decisions based on aged CSI. The results in this section give the initial steps in that direction, considering a single user and a fixed number of assigned resource blocks. The results have been partially presented in [62].

For each resource block, let the channel states be time-variant according to an

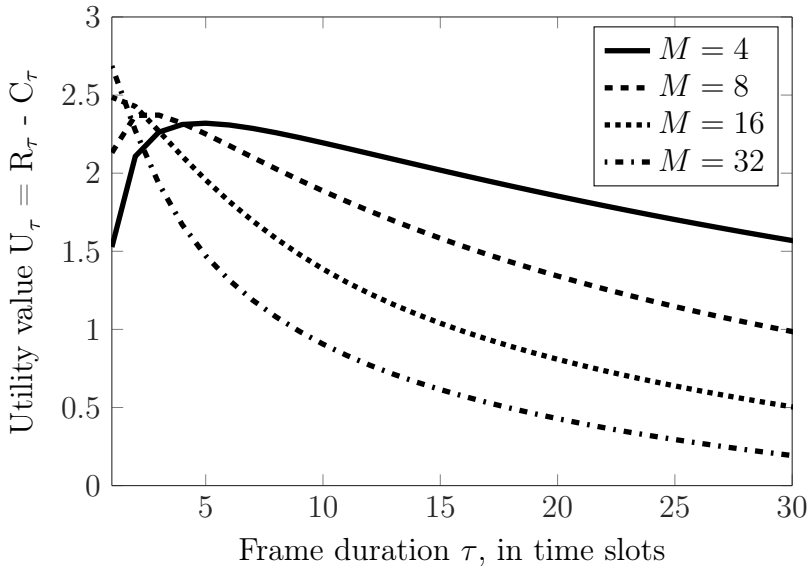


Figure 4.9: Utility value versus frame duration for different channel models. We consider Rayleigh fading with states associated to levels of SNR, and vary the number of states M.

M-state Markov chain χ, as described in section 4.2. For simplicity we assume the channel conditions to be independent and identically distributed among the resource blocks. This assumption is realistic if the resource blocks are sufficiently far apart in the appropriate dimension, such as space or frequency. For example, in frequency multiplexing, this assumption holds if the carrier spacing exceeds the coherence bandwidth or if interleaving is applied in the frequency domain [89, Chapter 3].

Let N be the number of orthogonal resource blocks allocated to one user. The channel state per resource block belongs to the set $\mathcal{S}$. Consequently, the overall channel state can be expressed as a M-ary vector $\mathbf{k}$ with N elements and the Markov chain models M^N states in a set denoted with $\mathcal{S}_N$.

Based on this concept of a channel state vector, let us now use $P_{\mathbf{s},\mathbf{k}}$ to denote the transition probability between the states $\mathbf{s}$ and $\mathbf{k}$, with $\mathbf{s}, \mathbf{k} \in \mathcal{S}_N$. We represent

the entries in these N-dimensional vectors as $\mathbf{s} = (s_1, \ldots, s_N)$. Under the assumption that the channel state of each resource block is independent of the others, we have

$$\mathrm{P}_{\mathbf{s},\mathbf{k}} = \prod_{i=1}^{N} \mathrm{P}_{s_i,k_i}, \tag{4.27}$$

with $s_i,\ k_i \in \mathcal{S}$, and P_{s_i,k_i} as the transition probability of the FSMC model for each resource block $i \in \{1, \ldots, N\}$.

Using the transition matrix P which characterizes the channel per resource block, the Kronecker product $A \otimes B$ between two matrices A and B, and the Kronecker power [90, Chapter 12] we define

$$\mathrm{P}^{[0]} := 1 \tag{4.28}$$

$$\mathrm{P}^{[1]} := \mathrm{P} \tag{4.29}$$

$$\mathrm{P}^{[2]} := \mathrm{P} \otimes \mathrm{P}, \tag{4.30}$$

$$\mathrm{P}^{[k]} := \mathrm{P} \otimes \mathrm{P}^{[k-1]}, \quad k = 3, 4, \ldots \tag{4.31}$$

With the definition of the Kronecker power in (4.31), we can write the transition matrix P_N, which characterizes the Markov chain for the M^N channel states in the multiplexing application, under the aforementioned assumption of independent and identically distributed channel conditions.

Proposition 4.1. *The transition matrix P_N, which describes the Markov chain for the M^N channel states in the case with multiplexing, is given by $P_N = P^{[N]}$, with entries $P_{\mathbf{s},\mathbf{k}}$, $\mathbf{s}, \mathbf{k} \in \mathcal{S}_N$.*

Proof: The proposition follows immediately from the definition of the Kronecker power in (4.31), which yields the formulation described in (4.27). ■

Expressing the Markov chain through the transition matrix P_N allows to apply the theory from section 4.3 to study the effect of the age of CSI on the utility of the user. In order to estimate the channel conditions in the subsequent time slot, we characterize the n-step transition matrix of the channel.

Let the n-step transition matrix of the channel be denoted with P_N^n. Using the power property of the Kronecker product, this matrix can be calculated as the Nth Kronecker power of the n-step transition matrix for a single channel.

Proposition 4.2. *The n-step transition matrix of the channel, denoted with P_N^n, can be calculated as the Nth Kronecker power of the n-step transition matrix for a single resource block. That is, the equalities*

$$P_N^n = \left(P^{[N]}\right)^n = (P^n)^{[N]} \tag{4.32}$$

hold.

Proof: The statement in the proposition follows from the property for powers of Kronecker products. If M_1 and M_2 are square matrices, then [91]

$$(M_1 \otimes M_2)^k = M_1^k \otimes M_2^k, \quad k = 1, 2, \ldots \tag{4.33}$$

Using a simple induction argument, we can show that the same property holds for a general Kronecker product. That is, assuming

$$(A_1 \otimes A_2 \cdots \otimes A_m)^k = A_1^k \otimes A_2^k \cdots A_m^k, \tag{4.34}$$

we have

$$\begin{aligned}(A_0 \otimes A_1 \otimes A_2 \cdots \otimes A_m)^k &= [A_0 \otimes (A_1 \otimes A_2 \cdots \otimes A_m)]^k \\ &= A_0^k \otimes (A_1 \otimes A_2 \cdots \otimes A_m)^k \\ &= A_0^k \otimes A_1^k \otimes A_2^k \cdots A_m^k. \end{aligned} \tag{4.35}$$

From Proposition 4.1, the transition matrix for the multiplexing application with N resource blocks allocated to one mobile user is $\mathrm{P}_N = \mathrm{P}^{[N]}$, where P is the transition matrix for a M-state Markov chain, which describes the channel conditions for a single resource block. The n-step transition matrix is obtained as the nth matrix power of the transition matrix.

In general, when N resource blocks are allocated to a mobile user, and the probability transition matrix is $\mathrm{P}_N = \mathrm{P}^{[N]}$, the n-step probability transition matrix is obtained using (4.35), with all matrices equal to P, and we have

$$\left(\mathrm{P}^{[N]}\right)^n = (\mathrm{P}^n)^{[N]}, \quad n \in \{1, 2, \ldots\},\ N \in \{1, 2, \ldots\}, \tag{4.36}$$

as stated in Proposition 4.2. ■

The next step is to define the reward matrix. For the multiplexing application, the reward matrix will be denoted with R_N, and it has entries $\mathrm{R}_{\mathbf{s},\mathbf{d}}$, with $\mathbf{s}, \mathbf{d} \in \mathcal{S}_N$.

Let $\mathrm{R}_1 = \mathrm{R}$, as described in section 4.4. This matrix will define the reward associated to a single resource block, and it could associate a rate reward, or a negative penalty to identify estimation errors, for example. We assume that the channel conditions in each resource block can be estimated separately, and we add

the rewards. That is, we define

$$\mathrm{R}_{\mathbf{s},\mathbf{d}} = \sum_{i=1}^{N} \mathrm{R}_{s_i,d_i}, \tag{4.37}$$

where $s_i,\ d_i \in \mathcal{S}$, and R_{s_i,d_i} is an entry in R which defined the reward when the channel state of the ith resource block is s_i, and the decision is d_i, $i \in \{1,\ldots,N\}$.

Let $\mathbb{1}_{M^z}$ be a $M^z \times M^z$ matrix with all entries equal to one, for $z \in \{0,1,2\ldots\}$. The reward matrix R_N, with entries $\mathrm{R}_{\mathbf{s},\mathbf{d}}$ satisfying (4.37), can be constructed from the individual reward matrix R using the following recursion rule:

$$\mathrm{R}_1 = \mathrm{R} \tag{4.38}$$

$$\mathrm{R}_2 = \mathrm{R} \otimes \mathbb{1}_{M^1} + \mathbb{1}_{M^1} \otimes \mathrm{R} \tag{4.39}$$

$$\mathrm{R}_N = \mathrm{R} \otimes \mathbb{1}_{M^{N-1}} + \mathbb{1}_{M^1} \otimes \mathrm{R}_{N-1}, \quad N \in \{3,4,\ldots\}. \tag{4.40}$$

Proposition 4.3. *Reward Matrix with Multiplexing*

The reward matrix R_N, with entries $R_{\mathbf{s},\mathbf{d}}$ satisfying (4.37)-(4.40) *can be described by*

$$R_N = \sum_{i=1}^{N} (\mathbb{1}_M)^{[i-1]} \otimes R \otimes \mathbb{1}_{M^{N-i}}, \tag{4.41}$$

where $(\mathbb{1}_M)^{[i-1]}$ is the $(i-1)$th Kronecker power, as defined in (4.31)*, and we make the definition of $(\mathbb{1}_M)^{[0]} \coloneqq 1$.*

Proof: In the multiplexing link, when N resource blocks are allocated to a mobile user, the channel conditions are represented in the state vector $\mathbf{s} = (s_1,\ldots,s_N)$. Let the decision about the channel states be described by the vector $\mathbf{d}$. The reward associated with the pair of vectors $(\mathbf{s},\mathbf{d})$ is the sum of the rewards

associated to each resource block, as stated in (4.37). The rewards associated to each resource block are described by the $M \times M$ rewards matrix R.

In the case of $N = 2$ resource blocks, the rewards matrix can be written as

$$\mathrm{R}_2 = \mathrm{R} \otimes \mathbb{1}_M + \mathbb{1}_M \otimes \mathrm{R}, \tag{4.42}$$

with $\mathbb{1}_M$ representing the $M \times M$ matrix of ones.

When the dimension of the vector $\mathbf{s}$ is increased by one unit, the rewards will have one additional term added. This additional term can assume values in the set of entries of R. As a result, the rewards matrix can be constructed, in general, using the recursion presented in (4.38). That is, once we have R_2 in (4.42), we construct the reward matrices as

$$\mathrm{R}_N = \mathrm{R} \otimes \mathbb{1}_{M^{N-1}} + \mathbb{1}_{M^1} \otimes \mathrm{R}_{N-1}, \quad N \in \{3, 4, \ldots\}. \tag{4.43}$$

With the definitions of $\mathbb{1}_M^{[0]} \coloneqq 1$ and $\mathbb{1}_M = 1$, we may also write

$$\mathrm{R}_1 = \mathbb{1}_M^{[0]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^0} \tag{4.44}$$

$$\mathrm{R}_2 = \mathbb{1}_M^{[0]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^1} + \mathbb{1}_M^{[1]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^0}. \tag{4.45}$$

We proceed by induction to show that, for $N \in \{3, 4 \ldots\}$,

$$\mathrm{R}_N = \sum_{i=1}^{N} (\mathbb{1}_M)^{[i-1]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^{N-i}}. \tag{4.46}$$

Assume that

$$\mathrm{R}_{N-1} = \sum_{i=1}^{N-1} (\mathbb{1}_M)^{[i-1]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^{(N-1)-i}}. \tag{4.47}$$

By construction, the rewards matrix in the case of N resource blocks is obtained

using (4.43). Using (4.47) in (4.43) yields the desired formulation

$$\begin{aligned}
\mathrm{R}_N &= \mathrm{R} \otimes \mathbb{1}_{M^{N-1}} + \mathbb{1}_{M^1} \otimes \left[\sum_{j=1}^{N-1} \left(\mathbb{1}_M\right)^{[j-1]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^{N-1-j}} \right] \\
&= \mathbb{1}_M^{[0]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^{N-1}} + \sum_{j=1}^{N-1} \left(\mathbb{1}_M\right)^{[j]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^{N-1-j}} \\
&= \sum_{i=1}^{N} \left(\mathbb{1}_M\right)^{[i-1]} \otimes \mathrm{R} \otimes \mathbb{1}_{M^{N-i}}. \qquad (4.48)
\end{aligned}$$

■

We illustrate how the age of channel state information affects a multiplexed link by a numerical example, considering a simple channel model with $M = 2$ states, as the one considered in section 4.5 and illustrated in Figure 4.3. The transition probabilities are arbitrarily set to $\alpha = 0.03$ and $\beta = 0.97$. We assume that the rewards per resource block are $(\mathrm{R}_{1,1}, \mathrm{R}_{1,2}, \mathrm{R}_{2,1}, \mathrm{R}_{2,2}) = (0, -r, -r, r)$, with $r = 1$ bits per second. This choice of rewards would be adequate, for example, if the user only transmits when the channel is identified to be in good state. The two possible wrong decisions about the channel state are penalized by a negative rate. This penalty is attributed to $R_{1,2}$ when a too optimistic decision leads to an outage event. The penalty assigned to $R_{2,1}$ reflects a too pessimistic decision, resulting in a missed opportunity to transmit.

The cost of feedback is modeled proportional to the number of resource blocks N and to the rate reward as $C_\tau = Nc\mathrm{R}_{2,2}$. Here, $c \in (0, 1)$ and $\mathrm{R}_{2,2}$ is the rate reward attributed to a single resource block when the channel is in good condition and it is correctly identified to be so. This representation is suitable for in-band feedback, where a part of the channel resources available for data transmission is

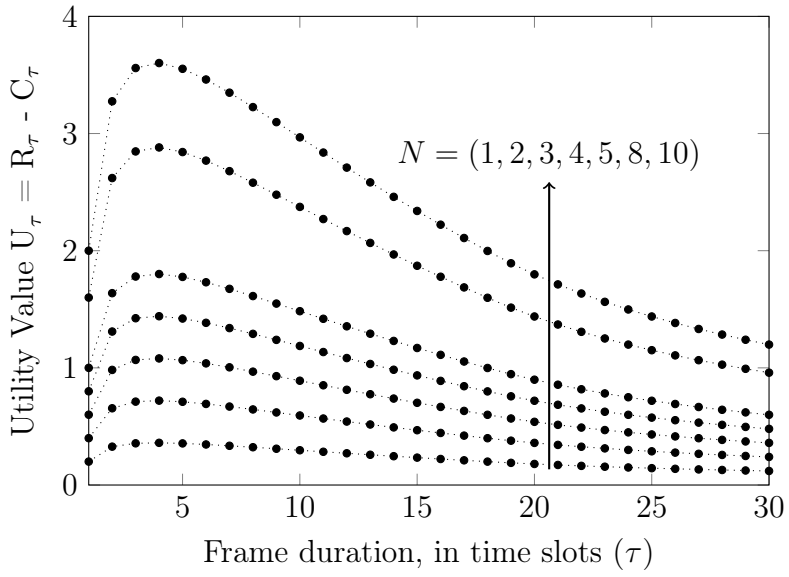


Figure 4.10: Utility value versus frame duration with multiplexing, with N resource blocks allocated to one user. Gilbert-Elliot channel model with $M = 2$ states assumed, with transition probabilities $\alpha = 0.03$ and $\beta = 0.97$.

allocated to the feedback of channel states. Consequently, the cost for occupying the channel to transmit the feedback messages can be modeled as a fraction of the achievable data rate.

The results using the aforementioned numerical values are presented in Figure 4.10. We observe that, if the frame duration is optimized, any increase of the number of resource blocks results in a proportional increase of the utility. The gain is expected, since we add the rewards for each resource block, but to fully benefit from the multiple resource blocks, the age of channel state information clearly has to be taken into account, since the cost to obtain channel state information also increases with the number of allocated resource blocks.

In Figure 4.11 we remove the assumption that S_0 is perfectly known in the first time slot of a frame, to illustrate the effect of delay δ on the channel state

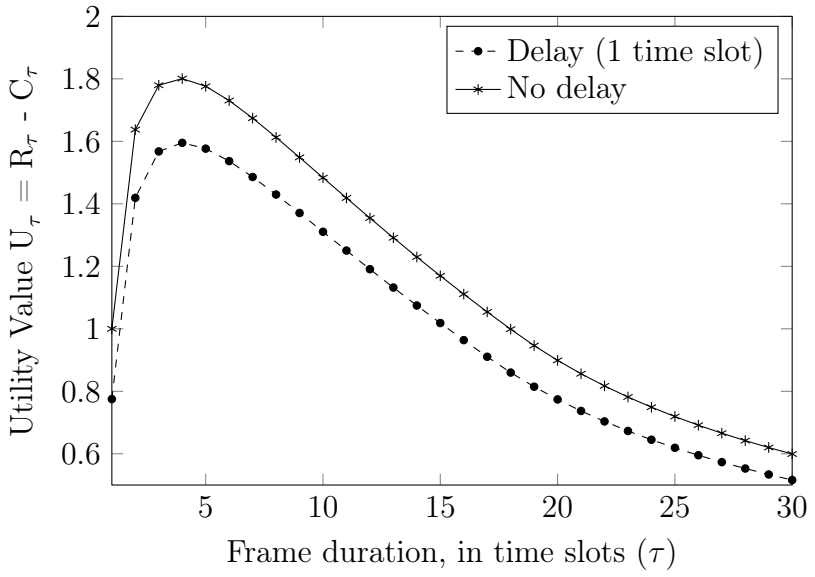


Figure 4.11: Utility value versus frame duration with multiplexing, with $N = 5$ resource blocks allocated to one user. Gilbert-Elliot channel model with $M = 2$ states assumed, with transition probabilities $\alpha = 0.03$ and $\beta = 0.97$.

information. Assuming $\delta < \tau$ we illustrate the case in which the channel state is known without errors at the mobile user, but it is only available to the BS in the subsequent time slot. In this case, the state S_0 is already estimated using a one step transition probability matrix. We compare with the previous result without delay, in which the state S_0 is known in the first time slot of the frame. In this numerical example, when $\tau = 4$ time slots, the delay causes a significant reduction of 11.4% in the utility value. For $\tau = 20$, the reduction is of 13.9%. This observation indicates the importance of addressing the delay δ, even though the interval τ has the major impact on the utility value.

In the multiplexing application with N resource blocks, each taking values in a set $\mathcal{S}$ of cardinality M, the problem dimension grows exponentially with N. To mitigate this problem we propose that the feedback sent in the beginning of a frame

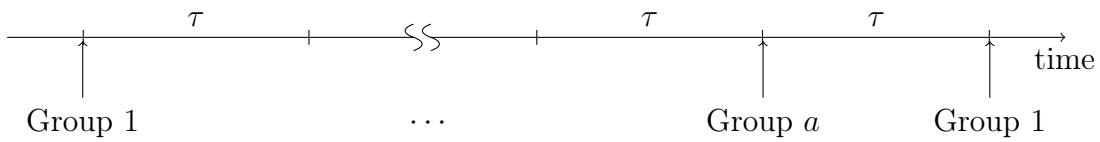


Figure 4.12: Alternating feedback scheme for $N = ab$ resource blocks. Feedback is sent in a groups, each containing the CSI of b resource blocks.

contains the CSI only of part of the resource blocks, alternating the resource blocks reported in each frame. We call this an alternating feedback scheme.

Let the number of allocated resource blocks be $N = ab$, with $a, b \in \mathbb{Z}_+$, where $\mathbb{Z}_+$ represents positive integer numbers. Such representation is always possible, since we can write $N = 1N$ in case N is a prime number.

We divide the resource blocks in a groups of b elements. In the beginning of each frame, the mobile user will report the channel states for one of the groups. We assume that the states of all b elements in the group are made available to the BS without errors or delays. In the beginning of the next frame, the states for a different group are reported, following a Round Robin scheme, as illustrated in Figure 4.12.

The expected rewards for all resource blocks are summed in each time slot. Without loss of generality, consider a frame in which the CSI for group 1 is reported. For all resource blocks belonging to group 1, the age of CSI in the nth time slot is equal to n, $n \in \{0, 1, \ldots, \tau - 1\}$. The CSI for resource blocks in group a was reported in the previous frame, as a result, the age of CSI at time slot n is equal to $\tau + n$, and so on. Also, resource blocks within the same group yield the same expected reward, under the assumption of identically distributed channels. As a

result, the average reward in a frame is denoted with R_τ and can be described as

$$R_\tau = \frac{1}{\tau} \sum_{n=0}^{\tau-1} \sum_{j=1}^{a} b\mathbb{E}[r^*((a-j)\tau + n)]. \tag{4.49}$$

We model the feedback cost proportional to the number of resource blocks within a group b and to the rate reward for a single resource block, denoted with r. Let the feedback cost per time slot be $C_\tau = bcr/\tau$, with $c \in (0, 1)$. This representation is suitable for in-band feedback, where a part of the channel resources available for data transmission is allocated to the feedback of channel states. Consequently, the cost for occupying the channel to transmit the feedback messages can be modeled as a fraction of the achievable data rate.

We illustrate the alternating feedback scheme for the multiplexing application in Figure 4.13. For this numerical example, we assume that $N = 60$ orthogonal resource blocks are allocated to the user, with independent and identically distributed channel states. The CSI is reported in the beginning of each frame of duration τ time slots, as before. We write $N = ab$, and change the values of a and b as indicated to study the effect of reporting the CSI of a few resource blocks each time. When $a = 1$, then $b = N$ and the CSI of every resource block is reported in the begining of every frame. When $a = 2$, the resource blocks are divided in two groups, and the CSI of all the resource blocks in one group is reported in the beginning of a frame, alternating between the groups. Similarly with the other values of a and b. We observe that subdividing the resource blocks in groups may yield larger utility values, as long as the duration of the frame is small enough. In this numerical example, we observe that for a frame duration larger than $\tau = 10$ time slots, the larger utility

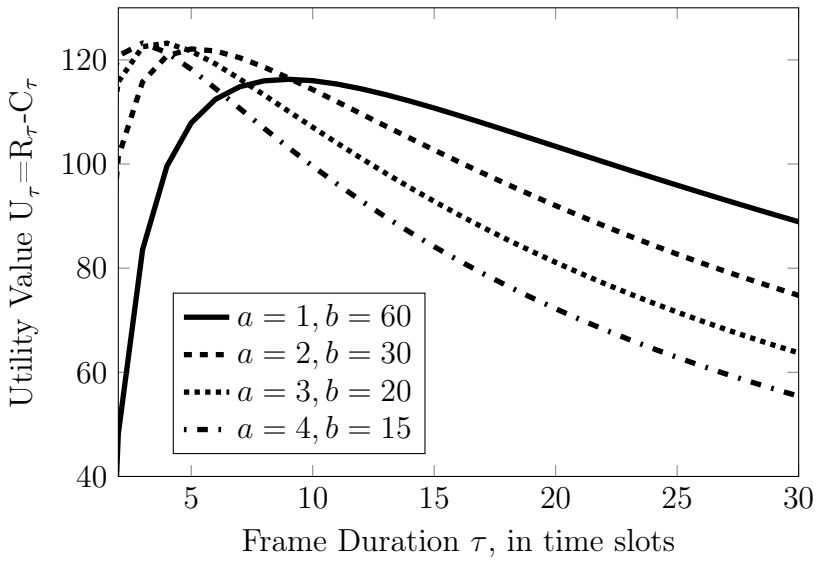


Figure 4.13: Utility value versus frame duration using the alternating feedback scheme. $N = 60$ resource blocks with CSI reported in groups of 60, 30, 20 or 15.

values are achieved reporting the CSI for all resource blocks.

4.8 Future Work Discussion

In general, the analytical framework proposed in this work can be used to derive limits and to choose the optimal operation parameters of channel adaptation functions and feedback protocols. Alternatively, our models can be used to study under which maximum Doppler shift and SNR a channel adaptation system with a given feedback delay performs best. By enabling both perspectives, our analytical tools provide new insight for wireless link design.

In the future, we intend to consider other channel models, extending the results to other forms of autocorrelation in the channel model.

Future work using the proposed framework should also extend the results regarding the allocation of multiple resource blocks to a user. The main goal would

be to define feedback schemes so that a number N of resource blocks can be selected among K, with $N \ll K$, while accounting for the age of Channel State Information (CSI) available to make resource allocation decisions. A scenario with multiple users should be investigated, and tools as Markov Decision Processes (MDP) can be useful for this purpose.

4.9 Chapter Summary

Assuming a non-reciprocal Finite State Markov Channel (FSMC) and a single channel-adaptive communication link with periodic feedback, we provided an analytical characterization of the effect of the age of Channel State Information (CSI) on the performance of wireless links. The resulting model allowed us to express the effect of the age of CSI on the utility of a wireless link, and to study the trade-offs between performance, feedback period, and feedback cost. We applied this framework to a Rayleigh fading channel, as an example, and studied the effect of feedback overhead, maximum Doppler shift, and Signal-to-Noise Ratio (SNR), analyzing the performance through a general utility function that accounts for rate rewards and feedback cost.

We have also investigated the effect of the age of CSI when multiple orthogonal resource blocks are allocated to one mobile user, and the channel conditions are represented by a vector. We proposed the use of an alternating feedback scheme, and verified its efficiency in improving the utility value, when compared to the periodic feedback of the CSI for a large number of resource blocks.

Chapter 5: Conclusion

In this dissertation we have considered two important aspects of communication systems, namely the energy efficiency and the age of information.

The energy efficiency has been studied in the context of cognitive networks, in which users may have high-priority or low-priority to access the spectrum. We proposed a new parametrization consisting of two parameters: the required quality of service of the high-priority user, and the interference tolerance. This parametrization is useful to characterize the intricate relationship between the transmission parameters for high- and low-priority users, allowing the investigation of performance trade-offs for both classes of users. We have considered three spectrum sharing schemes for non-cooperative cognitive networks, and studied the trade-offs between energy efficiency and throughput, and between energy efficiency and spectrum sensing accuracy. Further investigation is encouraged to account for the energy spent in spectrum sensing.

As an alternative to non-cooperative schemes, we envision that a low-priority user may be requested to act as a relay for the high-priority user, in exchange for an opportunity to use the network resources. We have considered two cooperative schemes for cognitive networks. The first cooperative scheme is inspired by the non-

cooperative scheme named Underlay, in which both classes of users may transmit simultaneously, given an interference tolerance imposed by the high-priority user. For this Cooperative Underlay scheme, we have analyzed the trade-off between energy efficiency and throughput, and verified the advantage of using full-duplex relay nodes, for different levels of self-interference cancellation. Further investigation is needed to account for the overhead incurred by cooperation and its impact on energy efficiency.

We have also proposed a second cooperative model, inspired by the non-cooperative model named Interweave, in which the user with lower priority may only access the spectrum when the high-priority user is silent. In our cooperative model, two users share the spectrum during a time frame, which is subdivided in three portions, allocated for individual transmissions and for cooperation. We proposed a game theoretic model to obtain the optimal allocation for these three activities. The proposed game model is a Stackelberg game, with the high-priority user acting as a leader, maximizing a utility function that accounts for transmission rates and energy spent in transmission. The low-priority user is the follower, and it responds to the leader's action with the rational response that maximizes the utility function, accounting for transmission rates, energy spent in transmission, and energy spent as a relay. The game model implements cooperation incentives in the form of credits, that are taken into account by the users when selecting their strategies. As a result of the game, we obtain two parameters that define how the frame will be allocated between the two users and the relaying activity.

We then proceeded to discuss the concept of age of information. We have

formally defined the age of information in status update systems and two metrics that characterize the age, namely the average age and the peak age. The peak age is a new proposed metric, which presents great potential in the characterization of age, given that it is more simple to calculate, and its average values are shown to be close to the average age. The peak age is also important for applications that require the age of information to stay below a threshold. We have discussed five different models for status update systems, using queuing theory. Three of those models are new, and consider the use of packet management before transmission, in order to avoid wasting network resources with the transmission of outdated information. Our proposed scheme with packet management and packet replacement in the buffer is modeled with a new queue model, which requires careful analysis, since classic results such as Little's theorem fail to apply. We have shown that this packet management scheme is efficient in reducing the average age and the peak age, when compared to the other analyzed schemes. Future investigations on age of information should characterize the information theoretic aspects of age, define the notion of effective age, and utilize age metrics as design criteria to deploy communication systems optimized to deliver timely information.

Finally, we have considered an application of the concept of age of information to the transmission of Channel State Information (CSI) using feedback. Assuming a non-reciprocal Finite State Markov Channel (FSMC) and a single channel-adaptive communication link with periodic feedback, we provided an analytical characterization of the effect of the age of CSI on the performance of wireless links. The resulting model allowed us to express the effect of the age on the utility of a wireless

link, and to study the trade-offs between performance, feedback period, and feedback cost. We applied this framework to a Rayleigh fading channel, as an example, and studied the effect of feedback overhead, maximum Doppler shift, and required Signal-to-Noise Ratio (SNR). We have also investigated the effect of the age of CSI when multiple orthogonal resource blocks are allocated to one mobile user, and the channel conditions are represented by a vector. We proposed the use of an alternating feedback scheme, and verified its efficiency in improving the utility value, when compared to the periodic feedback of the CSI for all resource blocks. In general, the proposed analytical framework can be used to derive limits and to choose the optimal operation parameters of channel adaptation functions and feedback protocols. Alternatively, our models can be used to study under which maximum Doppler shift and SNR a channel adaptation system with a given feedback delay performs best. By enabling both perspectives, our analytical tools provide new insight for wireless link design. Further extensions of this work should consider the decision process in adaptation functions, combining our framework with Partially Observable Markov Decision Processes to determine the best decision strategy for utility maximization, including the cases with multiple users.

Bibliography

[1] Gerhard Fettweis and Ernesto Zimmermann. ICT energy consumption - trends and challenges. In *Proc. International Symposium on Wireless Personal Multimedia Communications, WPMC*, 2008.

[2] Tao Chen, Honggang Zhang, Zhifeng Zhao, and Xianfu Chen. Towards green wireless access networks (invited paper). In *Proc. ChinaCom Beijing*, 2010.

[3] David Lister. An operator's view on green radio. In *Proc. IEEE International Workshop on Green Communications*, 2009.

[4] Guowang Miao, Nageen Himayat, Geoffrey Ye Li, and Ananthram Swami. Cross-layer optimization for energy-efficient wireless communications: A survey. *International Journal of Wireless Communications and Mobile Computing*, 9:529–542, 2009.

[5] Farhad Meshkati, H. Vincent Poor, and Stuart C. Schwartz. Energy-efficient resource allocation in wireless networks - an overview of game-theoretic approaches. *IEEE Signal Processing Magazine*, 24:58–68, 2007.

[6] Yan Chen, Shunqing Zhang, Shugong Xu, and Geoffrey Ye Li. Fundamental trade-offs on green wireless networks. *IEEE Communications Magazine*, 49:30–37, 2011.

[7] Geoffrey Li, Zhikun Xu, Cong Xiong, Chen yang Yang, Shunqing Zhang, Yan Chen, and Shugong Xu. Energy-efficient wireless communications: tutorial, survey, and open issues. *IEEE Wireless Communications*, 18:28–35, 2011.

[8] Jie Tang, Daniel K. C. So, Emad Alsusa, and Khairi Ashour Hamdi. Resource efficiency: A new paradigm on energy efficiency and spectral efficiency tradeoff. *IEEE Trans. on Wireless Communications*, 13:4656–4669, 2014.

[9] Changhun Bae and Wayne E. Stark. End-to-end energy-bandwidth tradeoff in multihop wireless networks. *IEEE Trans. on Information Theory*, 55:4051–4066, 2009.

[10] Hussein Arslan. *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*. Springer, 2007.

[11] Antoine Dejonghe, Bruno Bougard, Sofie Pollin, Jan Craninckx, André Bourdoux, Liesbet Van der Perre, and Francky Catthoor. Green reconfigurable radio systems. *IEEE Signal Processing Magazine*, 24:90–101, 2007.

[12] Gürkan Gür and Fatih Alagöz. Green wireless communications via cognitive dimension: an overview. *IEEE Network*, 25:50–56, 2011.

[13] Chunxiao Jiang, Haijun Zhang, Yong Ren, and Hsiao-Hwa Chen. Energy-efficient non-cooperative cognitive radio networks: Micro, meso, and macro views. *IEEE Communications Magazine*, 52:14–20, 2014.

[14] Salim Eryigit, Gürkan Gür, Suzan Bayhan, and Tuna Tugcu. Energy efficiency is a subtle concept: Fundamental trade-offs for cognitive radio networks. *IEEE Communications Magazine*, 52:30–36, 2014.

[15] Deah J. Kadhim, Shimin Gong, Wenfang Xia, Wei Liu, and Wenqing Cheng. Power efficiency maximization in cognitive radio networks. In *Proc. IEEE Wireless Communications and Networking Conference, WCNC*, 2009.

[16] Suzan Bayhan and Fatih Alagöz. Scheduling in centralized cognitive radio networks for energy efficiency. *IEEE Trans. on Vehicular Technology*, 62:582–595, 2013.

[17] Liying Li, Xiangwei Zhou, Hongbing Xu, Geoffrey Ye Li, Dandan Wang, and Anthony Soong. Energy-efficient transmission in cognitive radio networks. In *Proc. IEEE Consumer Communications and Networking Conference, CCNC*, 2010.

[18] Tao Qiu, Wenjun Xu, Tao Song, Zhiqiang He, and Baoyu Tian. Energy-efficient transmission for hybrid spectrum sharing in cognitive radio networks. In *Proc. IEEE 73rd Vehicular Technology Conference, VTC Spring*, 2011.

[19] Yiyang Pei, Ying-Chang Liang, Kah Chan Teh, and Kwok Hung Li. Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order. *IEEE Journal on Selected Areas in Communications*, 29:1648–1659, 2011.

[20] Sina Maleki, Ashish Pandharipande, and Geert Leus. Energy-efficient distributed spectrum sensing with convex optimization. In *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2009.

[21] Anm Badruddoza, Vinod Namboodiri, and Neeraj Jaggi. Does cognition come at a net energy cost in ad hoc wireless lans? *Computer Communications*, 43:43–54, 2014.

[22] Beiyu Rong and Anthony Ephremides. Cooperative access in wireless networks: Stable throughput and delay. *IEEE Trans. on Information Theory*, 58:5890–5907, 2012.

[23] Juncheng Jia, Jin Zhang, and Qian Zhang. Cooperative relay for cognitive radio networks. In *Proc. IEEE International Conference on Computer Communications, INFOCOM*, 2009.

[24] Khaled Ben Lataief and Wei Zhang. Cooperative communications for cognitive radio networks. *Proceedings of the IEEE*, 97:878–893, 2009.

[25] Ahmed K. Sadek, K.J. Ray Liu, and Anthony Ephremides. Cognitive multiple access via cooperation: Protocol design and performance analysis. *IEEE Trans. on Information Theory*, 53:3677–3696, 2007.

[26] Osvaldo Simeone, Yeheskel Bar-Ness, and Umberto Spagnolini. Stable throughput of cognitive radios with and without relaying capability. *IEEE Trans. on Communications*, 55:2351–2360, 2007.

[27] Sastry Kompella, Gam D. Nguyen, Jeffrey E. Wieselthier, and Anthony Ephremides. Stable throughput tradeoffs in cognitive shared channels with cooperative relaying. In *Proc. IEEE International Conference on Computer Communications, INFOCOM*, 2011.

[28] Anthony Fanous and Anthony Ephremides. Stable throughput in a cognitive wireless network. *IEEE Journal on Selected Areas in Communications*, 31:523–533, 2013.

[29] Davide Chiarotto, Osvaldo Simeone, and Michele Zorzi. Spectrum leasing via cooperative opportunistic routing techniques. *IEEE Trans. on Wireless Communications*, 10:2960–2970, 2011.

[30] Nikolaos Pappas, Anthony Ephremides, and Apostolos Traganitis. Relay-assisted multiple access with multi-packet reception capability and simultaneous transmission and reception. In *Proc. IEEE Information Theory Workshop, ITW*, 2011.

[31] Pradeep C. Weeraddana, Marian Codreanu, Matti Latva-aho, and Anthony Ephremides. The benefits from simultaneous transmission and reception in wireless networks. In *Proc. IEEE Information Theory Workshop, ITW*, 2010.

[32] Weifeng Su, John D. Matyjas, and Stella Batalama. Active cooperation between primary users and cognitive radio users in heterogeneous ad-hoc networks. *IEEE Trans. on Signal Processing*, 60:1796–1805, 2012.

[33] Maice Costa and Anthony Ephremides. Energy efficiency in cooperative cognitive wireless networks. In *Proc. 48th annual Conference on Information Sciences and Systems, CISS*, 2014.

[34] Fatemeh Afghah, Maice Costa, Abolfazl Razi, Ali Abedi, and Anthony Ephremides. A reputation-based stackelberg game approach for spectrum sharing with cognitive cooperation. In *Proc. 52nd Annual Conference on Decision and Control, CDC*, 2013.

[35] Dusit Niyato and Ekram Hossain. Market-equilibrium, competitive, and cooperative pricing for spectrum sharing in cognitive radio networks: Analysis and comparison. *IEEE Trans. Wireless Commun.*, 7:4273–4283, 2008.

[36] Michael Maskery, Vikram Khrishnamurthy, and Qing Zhao. Decentralized dynamic spectrum access for cognitive radios: Cooperative design of a non-cooperative game. *IEEE Trans. on Communications*, 57:459–469, 2009.

[37] Osvaldo Simeone, Igor Stanojev, Stefano Savazzi, Yeheskel Bar-Ness, Umberto Spagnolini, and Raymond Pickholtz. Spectrum leasing to cooperating secondary ad-hoc networks. *IEEE Journal on Selected Areas in Communications*, 26:203–213, 2008.

[38] X. Wang, K. Ma, Q. Han, Z. Liu, and X. Guan. Pricing-based spectrum leasing in cognitive radio networks. *Networks, IET*, 1(3):116 –125, sept. 2012.

[39] Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, and Dusit Niyato. A survey of trust and reputation management systems in wireless communications. *Proceedings of the IEEE*, 98:1755 –1772, 2010.

[40] Scheng Zhong, Jiang Chen, and Yang Richard Yang. Sprite: a simple, cheat-proof, credit-based system for mobile ad-hoc networks. In *Proc. IEEE International Conference on Computer Communications, INFOCOM*, 2003.

[41] Omer Ileri, Siun-Chuon Mau, and Narayan B. Mandayam. Pricing for enabling forwarding in self-configuring ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 23:151–162, 2005.

[42] Fatemeh Afghah, Abolfazl Razi, and Ali Abedi. Stochastic game theoretical model for packet forwarding in relay networks. *Springer Telecommunication Systems Journal, Special Issue on Mobile Computing and Networking Technologies*, 52:1877–1893, Jun. 2011.

[43] Qi He, Dapeng Wu, and Pradeep Khosla. SORI: A secure and objective reputation-based incentive scheme for ad-hoc networks. In *Proc. IEEE Wireless Communications and Networking Conference, WCNC*, 2004.

[44] Pietro Michiardi and Refik Molva. CORE: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In *Sixth Joint Working Conference on Communications and Multimedia Security*, 2002.

[45] Hyukjoon Kwon, HyungJune Lee, and John M. Cioffi. Cooperative strategy by stackelberg games under energy constraint in multi-hop relay networks. In *Proc. IEEE Global Telecommunications Conference, GLOBECOM*, 2009.

[46] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory.* MIT Press, 1994.

[47] Sanjit Kaul, Marco Gruteser, Vinuth Rai, and John Kenney. Minimizing age of information in vehicular networks. In *Proc. 8th Annual Communications Society Conference on Sensor, Mesh, and Ad-Hoc Communications and Networks, SECOM*, 2011.

[48] Sanjit Kaul, Roy Yates, and Marco Gruteser. On piggybacking in vehicular networks. In *Proc. IEEE Global Telecommunications Conference, GLOBECOM*, 2011.

[49] Sanjit Kaul, Roy Yates, and Marco Gruteser. Real-time status: How often should one update? In *Proc. IEEE International Conference on Computer Communications, INFOCOM*, 2012.

[50] Roy Yates and Sanjit Kaul. Real-time status updating: Multiple sources. In *Proc. IEEE International Symposium on Information Theory, ISIT*, 2012.

[51] Sanjit Kaul, Roy Yates, and Marco Gruteser. Status updates through queues. In *Proc. 46th Annual Conference on Information Sciences ans Systems, CISS*, 2012.

[52] Roy Yates. Lazy is timely: Status updates by an energy harvesting source. In *Proc IEEE International Symposium on Information Theory, ISIT*, 2015.

[53] Clement Kam, Sastry Kompella, and Anthony Ephremides. Age of information under random updates. In *Proc. IEEE International Symposium on Information Theory, ISIT*, 2013.

[54] Clement Kam, Sastry Kompella, and Anthony Ephremides. Effect of message transmission diversity on status age. In *Proc. IEEE International Symposium on Information Theory, ISIT*, 2014.

[55] Maice Costa, Marian Codreanu, and Anthony Ephremides. Age of information with packet management. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2014.

[56] Longbo Huang and Eytan Modiano. Optimizing age of information in a multiclass queuing system. In *in Proc. IEEE International Symposium on Information Theory, ISIT*, 2015.

[57] Dennis L. Goeckel. Adaptive coding for time-varying channels using outdated fading estimates. *IEEE Trans. on Communications*, 47:844–855, 1999.

[58] Sigen Ye, Rick S. Blum, and Leonard J. Cimini Jr. Adaptive OFDM systems with imperfect channel state information. *IEEE Trans. on Wireless Communications*, 5:3255–3265, 2006.

[59] Giuseppe Caire, Nihar Jindal, and Mari Kobayashi. Multiuser MIMO achievable rates with downlink training and channel state feedback. *IEEE Trans. on Information Theory*, 56:2845–2866, 2010.

[60] Mohammad Ali Maddah-Ali and David Tse. Completely stale transmitter channel state information is still very useful. *IEEE Trans. on Information Theory*, 58:4418–4431, 2012.

[61] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE - The UMTS Long Term Evolution: From Theory to Practice*. New York: Wiley, 2011.

[62] Maice Costa, Stefan Valentin, and Anthony Ephremides. On the age of channel state information for non-reciprocal wireless links. In *Proc. IEEE International Symposium on Information Theory, ISIT*, 2015.

[63] Maice Costa, Stefan Valentin, and Anthony Ephremides. On the age of channel information for a Finite State Markov model. In *Proc. IEEE International Conference on Coomunications, ICC*, 2015.

[64] Beiyu Rong and Anthony Ephremides. Cooperation above the physical layer: the case of a simple network. In *Proc. IEEE International Symposium on Information Theory, ISIT*, 2009.

[65] Antonio De Domenico, Emilio C. Strinati, and M. Di Benedetto. A survey on MAC strategies for cognitive radio networks. *IEEE Commun. Surveys & Tutorials*, 14:21–44, 2012.

[66] Gam D. Nguyen, Sastry Kompella, Jeffrey E. Wieselthier, and Anthony Ephremides. Optimization of transmission schedules in capture-based wireless networks. In *Proc. IEEE Military Communications Conference, MILCOM*, 2008.

[67] Tevfik Yücek and Hüseyin Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Commun. Surveys & Tutorials*, 11:116–130, 2009.

[68] Erik Axell, Geert Leus, Erik G. Larsson, and H. Vincent Poor. Spectrum sensing for cognitive radio : State-of-the-art and recent advances. *IEEE Signal Processing Magazine*, 29:101–116, 2012.

[69] Fadel F. Digham, Mohamed-Slim Alouini, and Marvin K. Simon. On the energy detection of unkown signals over fading channels. In *Proc. IEEE International Conference on Communications, ICC*, 2003.

[70] Martin J. Osborne. *An Introduction to Game Theory*. Oxford University Press, 2000.

[71] Frank P. Kelly, Aman Maulloo, and David Tan. Rate control in communication networks: Shaddow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[72] Anders Host-Madsen and Junshan Zhang. Capacity bounds and power allocation for wireless relay channels. *IEEE Trans. on Information Theory*, 51:2020 –2040, 2005.

[73] JeongGil Ko, Chenyang Lu, Mani B. Srivastava, John A. Stankovic, Andreas Terzis, and Matt Welsh. Wireless sensor networks for healthcare. *Proceedings of the IEEE*, 98:1947–1960, 2010.

[74] Peter Corke, Tim Wark, Raja Jurdak, Wen Hu, Philip Valencia, and Darren Moore. Environmental wireless sensor networks. *Proceedings of the IEEE*, 98:1903–1917, 2010.

[75] Athanasios Papoullis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. Mc Graw Hill, 2001.

[76] Leonard Kleinrock. *Queuing Systems: Theory Volume I*. Wiley, 1976.

[77] Robert B Cooper. *Introduction to Queuing Theory*. Elsevier North Holland Inc., 1981.

[78] Harish Viswanathan. Capacity of Markov channels with receiver CSI and delayed feedback. *IEEE Trans. on Information Theory*, 45:761–771, 1999.

[79] Uria Basher, Avihay Shirazi, and Haim H. Permuter. Capacity region of finite state multiple-access channels with delayed state information at the transmitters. *IEEE Trans. on Information Theory*, 58:3430–3452, 2012.

[80] Mari Kobayashi, Nihar Jindal, and Giuseppe Caire. Training and feedback optimization for multiuser MIMO downlink. *IEEE Trans. on Communications*, 59:2228–2240, 2011.

[81] Stefan Valentin and Thorsten Wild. Studying the sum capacity of mobile multiuser diversity systems with feedback errors and delay. In *Proc. IEEE Vehicular Technology Conference, VTC*, 2010.

[82] Qi Wang, Larry Greenstein, Leonard Cimini, Douglas Chan, and Ahmadreza Hedayat. Multi-user and single-user throughputs for downlink MIMO channels with outdated channel state information. *IEEE Wireless Communicatins Letters*, 3:321–324, 2014.

[83] Parastoo Sadeghi, Rodney A. Kennedy, Predrag B. Rapajic, and Ramtin Shams. Finite-state Markov modeling of fading channels - a survey of principles and applications. *IEEE Signal Processing Magazine*, 25:57–80, 2008.

[84] Sheldom M. Ross. *Stochastic Processes*. Wiley, 1996.

[85] Hong Shen Wang and Nader Moayeri. Finite-State Markov Channel - a useful model for radio communication channels. *IEEE Trans. on Vehicular Technology*, 44:163–171, 1995.

[86] Hong Shen Wang and Nader Moayeri. Modeling, capacity, and joint source/channel coding for rayleigh fading channels. In *Proc. IEEE Vehicular Technology Conference, VTC*, 1993.

[87] Qinqing Zhang and Saleem A. Kassam. Finite-state Markov model for Rayleigh fading channels. *IEEE Trans. on Communications*, 47:1688–1692, 1999.

[88] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[89] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005.

[90] Richard Bellman. *Introduction to Matrix Analysis*. RAND Corporation, 1997.

[91] Amy N. Langville and William J. Stewart. The kronecker product and stochastic automata networks. *Journal of Computational and Applied Mathematics*, 167:429–447, 2004.