TOWARD COGNITIVE COEXISTENCE: OPTIMAL SENSING-BASED RESOURCE MANAGEMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Stefan Geirhofer

January 2009

TOWARD COGNITIVE COEXISTENCE: OPTIMAL SENSING-BASED RESOURCE MANAGEMENT

Stefan Geirhofer, Ph.D.

Cornell University 2009

The rapid growth of wireless communication systems makes the coexistence of heterogeneous technologies more and more important. This dissertation studies how cognitive radio concepts may provide an efficient framework for accessing and sharing spectrum by sensing and predicting temporal activity patterns. In this way, the spectrum access of interfering devices can be coordinated implicitly and coexistence be improved.

The efficacy of this approach is studied for two common coexistence scenarios. First, we address the coexistence of a frequency hopping cognitive radio with a set of parallel ad-hoc bands, a setup that has conceptual similarities with interfering local and personal area networks. Temporal idle periods that remain between ad-hoc transmissions are reused efficiently by the cognitive radio through predicting ad-hoc radio activity and dynamically adapting the hopping pattern. Second, we address the coexistence of infrastructure and ad-hoc networks. Motivated by the superior resources of the infrastructure system, we study how its centralized resource allocation may accommodate the ad-hoc links based on adjusting the power and transmission time allocation. Despite adapting its behavior to coexist with the ad-hoc links, the infrastructure system maintains a specified quality of service level for its users by imposing rate constraints.

Both of the above formulations are based on a two-state continuous-time Markov chain model for the ad-hoc system's temporal behavior which approxi-

mates the carrier-sense multiple access typically employed in such systems. The model is discussed in detail and corroborated through empirical analysis of a practical system. Our analyses are based on the mathematical tools of constrained Markov decision processes and convex optimization and are validated by system-level simulations. Further, a real-time test bed has been developed for the cognitive frequency hopping protocol which enables us to corroborate model assumptions experimentally and gain further insight into fundamental tradeoffs.

The results presented in this dissertation demonstrate a significant performance gain compared to reference schemes without sensing capabilities. While various implementation details remain to be addressed in future work, our study clearly shows the conceptual merits of this framework and the importance it might play in future wireless systems.

BIOGRAPHICAL SKETCH

Stefan Geirhofer was born in Vienna, Austria. He received the "Dipl.-Ing." degree in Electrical Engineering from the Vienna University of Technology in March 2005 after having spent one year at the University of Illinois at Urbana-Champaign in the scope of an academic exchange program. Stefan joined Professor Lang Tong's research group at Cornell University in May 2005, where he has since been pursuing his doctoral studies. Stefan's research interests focus on cognitive radio networks and the optimization of wireless communication systems based on signal processing techniques and efficient medium access control methods.

To my parents, Helmut and Dorothea.

And to Karin.

In necessariis unitas, in dubiis libertas, in omnibus caritas.

ACKNOWLEDGEMENTS

This dissertation could not have been written without the support of many faculty and staff members as well as my colleagues, family, and friends. First and foremost, I would like to thank my advisor, Professor Lang Tong, for his guidance and support during the last years. I am grateful for his inspiring enthusiasm for research and cherish the many things I have learned from him. I have no doubt that they will continue to serve as important assets throughout my career.

I am indebted to Professor Zygmunt J. Haas and Professor Anna Scaglione for serving on my academic committee and many fruitful suggestions and comments. I value their support throughout my time at Cornell.

I would further like to thank Dr. Brian M. Sadler at the U.S. Army Research Laboratory for being a mentor and vital source of advice. He has contributed to every single research topic I have worked on at Cornell and his comments and suggestions have helped to shape and direct my thesis research considerably.

I am also grateful to Dr. Özgür Oyman who supervised my internship at Intel Research in the summer of 2007 and has become a valued mentor. Through him I have gained many new perspectives on cooperative wireless communications and practical systems design.

During my studies, I have had several opportunities to work with other students, particularly when implementing and testing some of the methods that I developed in my thesis research. I would like to thank John Sun, Jon San Ng, Bruce Lei, and Hemanth Mullur for their good work and commitment.

I have benefited tremendously from interacting with my colleagues and groupmates, many of whom have become good friends: Animashree Anandkumar, Matthew Ezovski, Ting He, Brandon Jones, Jin Sub Kim, Oliver Kosut, Amine Laourine, Chin-Chen Lee, Saswat Misra, Abhishek Nadamani, Youngchul Sung,

Parvathinathan Venkitasubramaniam, Meng Wang, and Zhiyu Yang.

Finally, I would like to thank my parents, Helmut and Dorothea, for raising me the way they did, and being a never-ending source of encouragement throughout my life. And I am grateful to Karin for her love and support during the last years.

This work is supported in part by the U.S. Army Research Laboratory under the Collaborative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Wireless communications have experienced a tremendous growth in recent years and are becoming ubiquitous. The rapid proliferation of wireless devices makes interference an important performance impediment and requires that coexistence aspects be addressed. Interference management is especially important in unlicensed bands, where heterogeneous technologies have equal spectrum access rights despite a lack of coordination among them.

In order to alleviate interference, it is necessary to find efficient ways for accessing and sharing spectrum among heterogeneous technologies. This thesis explores how cognitive radio concepts may provide a novel framework for deriving interference management techniques without requiring explicit communication among interfering systems. This leads to a new *cognitive coexistence* paradigm in which sensing and prediction provide the implicit coordination that is required to mitigate interference.

1.1 Cognitive Radio: Sensing and Adaptation

The term "cognitive radio" was coined by Joseph Mitola in 1999 [1, 2] and has evolved into an active research area during recent years. Despite some ongoing discussion on a precise definition and standardization of the term [3], the fundamental proposition of cognitive radio is enabling *sensing* and *adaptation* by leveraging advances in software-defined radio and machine intelligence. This concept is illustrated in Figure 1.1. A cognitive radio observes its radio environment by means of spectrum sensing and gathers a (statistical) representation of its surroundings.

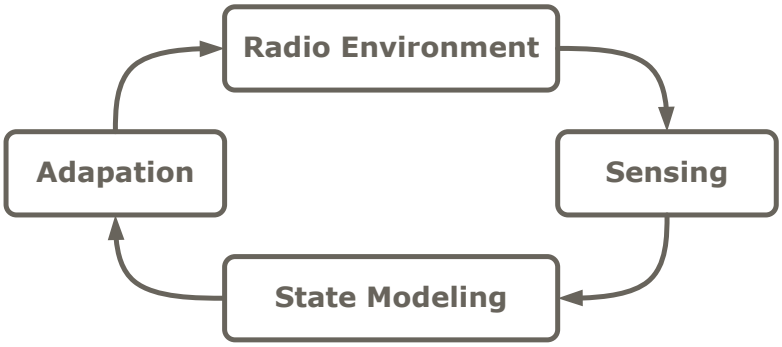


Figure 1.1: Illustration of the cognition cycle (adapted from [4]). A cognitive radio develops statistical models of its environment through spectrum sensing. Based on this information, transmission parameters are adapted according to some underlying objective.

This knowledge is then used to adapt transmission parameters according to some underlying objective such as, for example, minimizing interference to adjacent transmission links.

Cognitive radio enables wireless devices to access spectrum in a dynamic fashion based on the usage and interference patterns in its surroundings. This has led to the research area of dynamic spectrum access (DSA), which is discussed in more detail below and compared to cognitive coexistence.

1.1.1 Dynamic Spectrum Access

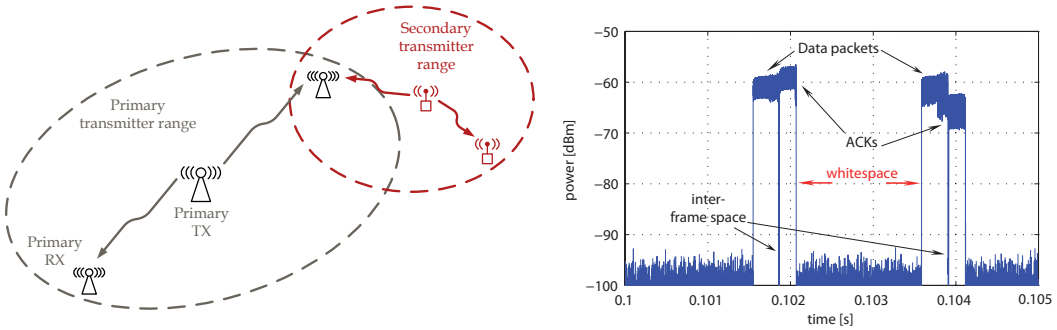
DSA employs cognitive radio techniques for reusing licensed but underutilized frequency bands; see [5, 6] for surveys of this research area. Spectrum property rights in licensed bands generally mandate a hierarchical approach, in which spectrum can only be accessed by a cognitive radio when this does not generate significant interference to license owners. In such setups, licensees are therefore typically referred to as primary users, whereas cognitive devices make up a secondary system,

which reuses spectrum opportunistically. Based on this hierarchical structure, the resource management of the secondary system is designed to maximize its throughput subject to maintaining interference constraints.

To avoid interference, knowledge of the radio environment (gathered through spectrum sensing) is used to orthogonalize secondary transmissions. DSA techniques can be classified according to which degrees-of-freedom are employed. Logically, the spatial and temporal domain are favorable from an implementation viewpoint. Orthogonalizing heterogeneous systems in other domains (such as the code domain) is more difficult in practice.

In the spatial domain, as illustrated in Figure 1.2(a), orthogonalizing heterogeneous systems is based on separating transmitters by a sufficient distance, or equivalently, controlling the transmission powers such that their transmission radii do not overlap. Consequently, the primary challenge in spatial DSA is guaranteeing accurate sensing performance. Since orthogonality is achieved through spatial separation, weak primary signals need to be detected reliably which may be difficult to ensure in practice [7]. Furthermore, the fact that only primary transmitters but not receivers can be detected leads to challenges similar to the hidden terminal problem [8].

Naturally, the spatial domain cannot be used to orthogonalize radio systems that are either collocated (for example, integrated into the same portable device) or that operate in close proximity of each other. In such scenarios the systems need to be separated in the temporal domain by scheduling the secondary system's transmissions during the idle periods of the primary's as shown in Figure 1.2(b). In addition to spectrum sensing, this approach requires a means for predicting the primary user's access behavior. In this thesis we will study this case from the



(a) Spatial Domain DSA

(b) Temporal Domain DSA

Figure 1.2: Illustration of spatial and temporal dynamic spectrum access (DSA). In spatial DSA (left), interfering systems are orthogonalized by controlling the cognitive radio's transmit power such that transmission radii do not overlap. In contrast, temporal DSA (right) separates transmissions in the time domain by detecting and predicting activity patterns.

viewpoint of cognitive coexistence.

1.1.2 Cognitive Coexistence

Similar to DSA, cognitive radio concepts can be used for reducing interference in coexistence setups. In unlicensed bands, sensing and prediction can be employed to implicitly coordinate the spectrum access of heterogeneous systems and thereby reduce interference. However, in contrast to DSA setups, where spectrum property rights mandate that secondary systems adapt to primaries, equal access rights in unlicensed bands allow for more flexibility. This provides more freedom in adapting transmission parameters and allows to better exploit a system's ability to accommodate adaption. Based on this paradigm, cognitive radio concepts may ultimately be used to realize different quality-of-service levels, for example by making service guarantees to some users (who pay for better service) while serving others with "best effort."

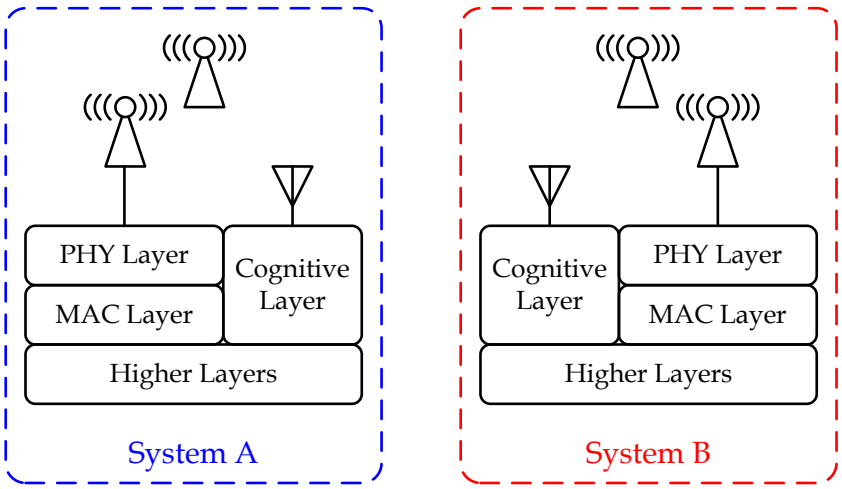


Figure 1.3: Cognitive cross-layer design. A cognitive layer gathers sensing results about interference in its surroundings. This information can be used to perform PHY/MAC cross-layer design dynamically.

In practical systems, cognitive radio functionality can lead to a novel, dynamic form of cross-layer design. In conventional communication systems, the layers of a communication system are optimized jointly for better performance. As illustrated in Figure 1.3, cognitive radio functionality can be viewed as adding a cognitive layer, which gathers knowledge of the radio-frequency environment and enables dynamic cross-layer optimization based on sensing results.

1.2 Summary of Contributions

Motivated by the fact that in many coexistence setups enough spatial separation cannot be achieved, this dissertation studies cognitive coexistence methods that achieve orthogonality in the temporal domain. Further, we focus on scenarios in which one of the systems is an ad-hoc network that utilizes carrier-sense multiple access (CSMA) similar to that in wireless local area networks (WLANs). Such systems have found widespread usage today, partly due to the rapid proliferation

of the IEEE 802.11 family of standards [9].

One reason for focusing on ad-hoc networks is the promising amount of temporal white space that generally exists in such setups. Due to the burstiness of traffic that is supported by such systems and the inefficiency of the random medium access employed, spectrum opportunities remain that can be used for orthogonalizing systems. The main obstacle for taking advantage of these opportunities, however, lies in the difficultly of predicting the ad-hoc system's medium access. To find analytical solutions, we establish a simple prediction model based on empirical data, which approximates the fundamental characteristics of the ad-hoc network's temporal activity patterns.

Based on this framework, we first consider a scenario in which the ad-hoc system takes the role of the primary user and is being interfered by a frequency hopping cognitive radio. Interference is reduced by adapting the cognitive radio's hopping behavior through decision-theoretic analysis. The conceptual similarities to Bluetooth/WLAN coexistence make this problem interesting from a practical perspective.

Subsequently, we analyze the case in which the ad-hoc network serves as the secondary user and is being interfered by an infrastructure system with higher priority. In this scenario, we show that sensing and prediction can be used successfully to alter the infrastructure system's medium access such that it accommodates the ad-hoc network as well as possible.

Finally, this dissertation presents a real-time test bed, which has been developed for demonstrating proof-of-concept and validating model assumptions associated with the cognitive frequency hopping protocol proposed in this thesis.

In the following, the main contributions within each area are discussed in more detail.

1.2.1 Sensing and Prediction Framework

We first propose a statistical prediction model which approximates the temporal activity patterns of ad-hoc communication systems. In subsequent sections of this dissertation we will build on this model to derive sensing-based medium access protocols. With this ultimate goal in mind, it is important to keep the prediction model simple enough to retain analytical tractability, therefore striking a balance between prediction accuracy and model complexity.

Finding such a tradeoff is difficult because ad-hoc networks may consist of many nodes that operate in a decentralized manner. While the behavior of each individual node may be standardized, we are interested in characterizing the overall behavior of the ad-hoc network. Deriving such a model may be complicated as the overall behavior depends on many parameters (such as the number of nodes, traffic characteristics, and many others) which increase model complexity and are difficult to estimate in practice.

The difficulty of deriving analytical prediction models motivates an empirical approach. By setting up a practical ad-hoc system based on WLAN devices, we record idle and busy periods for different traffic and model setups. Based on these empirical data, we identify common characteristics and compare classes of statistical models that can approximate the observed behavior. Of course, these selected models depend on some parameters that allow us to tailor them to the specific traffic and usage scenarios (for example, the traffic intensity). In practice,

these parameters will be estimated on-the-fly from previous observation of busy and idle periods.

This approach has conceptual similarity with tracking a fading channel in wireless communications systems. In many applications, it is reasonable to assume that the fading state remains constant over short periods of time; long-term variations are taken into account by periodically re-estimating the channel. Similarly, we approximate long-term variations in traffic intensity by continuously re-estimating model parameters.

For deriving cognitive medium access schemes, it is important to characterize the medium's idle and busy periods statistically. This leads to a simple two-state ON/OFF model, in which we can characterize the sojourn time (the time spent in either ON or OFF state) statistically. For arbitrary distributions, such a model is referred to as a semi-Markov model (for an arbitrary, finite number of states) or an alternating renewal process (in the special case of two states). Based on empirical data we found that a mixture of uniform and generalized Pareto distributions provides an excellent fit (a motivation for this mixture distribution will be discussed in detail). Similarly, phase-type distributions result in a concise and accurate description of the empirical data. While the above distributions lead to accurate models, this is at the cost of model complexity. To improve tractability we also consider fitting an exponential distribution whose memoryless property greatly simplifies the derivation of medium access schemes.

In order to verify the goodness-of-fit of these models quantitatively, we use the Kolmogorov-Smirnov test [10]. The semi-Markov model (for both mixture and phase-type distributions) provides an excellent fit with empirical data. The exponential distribution shows a larger statistical aberration; nevertheless the sim-

plicity of the model makes the Markovian framework attractive from an analytical viewpoint.

1.2.2 Cognitive Frequency Hopping Protocol

Based on the stochastic model for the ad-hoc system's behavior, we derive cognitive medium access protocols with improved coexistence. We first address a scenario in which multiple ad-hoc bands evolve independently of each other in separate bands. The temporal idle periods of these channels are used for transmissions of a frequency hopping cognitive radio, and we design the optimal hopping behavior based on sensing and predicting the activity patterns of the ad-hoc network.

The cognitive radio operates in a slotted fashion and detects the activity of the ad-hoc bands at the beginning of every slot, as shown in Figure 1.4. Based on these sensing results, the optimal medium access aims at maximizing the throughput of the cognitive radio subject to meeting interference constraints for the ad-hoc bands. These constraints result in a hierarchial setup in which the ad-hoc system has priority over the cognitive radio. The extent to which interference is tolerable is specified by interference constraints that quantify the impact on the ad-hoc bands. In this chapter, we explore two different metrics, that model interference either from the cognitive radio's or from the ad-hoc systems' perspective. In this way the performance and impact of different metrics can be understood as a function of the primary user's traffic intensity.

The optimal medium access is derived through a decision-theoretic analysis. The mathematical framework of constrained Markov decision processes is used to maximize throughput while adhering to interference constraints. Linear program-

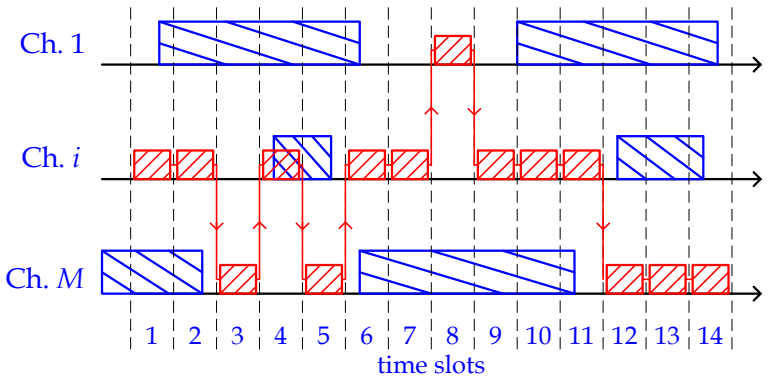


Figure 1.4: Cognitive frequency hopping setup. The time-slotted, frequency-hopping cognitive radio hops optimally among the ad-hoc bands such as to maximize throughput while adhering to interference constraints.

ming techniques are employed to find the optimal hopping pattern based on this formulation. Furthermore, it is shown that the optimal hopping pattern possesses a certain structure, which allows us to gain further insight into the problem.

The medium access scheme described above is based on the assumption that we observe the state of each channel at the beginning of every slot. In practice, however, the bandwidth of the sensing front-end may be limited and it may only be possible to observe a single ad-hoc band at a time. Finding the optimal hopping pattern under *partial observability* is much more challenging. A balance between exploring spectrum opportunities (by frequently switching among ad-hoc bands) and exploiting them (by staying in the band and transmitting) needs to be struck.

1.2.3 Cognitive Coexistence between Infrastructure and Ad-Hoc Networks

In the cognitive frequency hopping protocol, the ad-hoc system maintains priority over the cognitive radio, which adheres to interference constraints. In contrast, this section considers a setup in which the ad-hoc system coexists with an infrastructure system and has lower priority. Such setups are attractive for reducing interference between local and wide area networks and accommodating peer-to-peer functionality in future wireless networks [11]. Furthermore, this concept can be used to address coexistence aspects between a cellular network and femtocell base stations that are deployed in subscribers' homes [12].

While the ad-hoc system has lower priority, we assume that it does not adapt its behavior; instead the infrastructure system uses its superior communication resources to accommodate the ad-hoc links by predicting their activity patterns and adapting its resource allocation accordingly. In order to ensure priority, this adaption is performed subject to rate constraints that guarantee that a desired quality-of-service level remains met for infrastructure terminals.

We assume that the infrastructure network is a multicarrier system, evolves in frames of fixed duration, and allocates resources based on sensing results at the beginning of each frame. We consider two different scenarios for accommodating the ad-hoc system: either the power allocation is adapted based on knowledge of the interference channel, or power and transmission time are assigned based on the temporal prediction model proposed earlier.

We show that in both scenarios, the optimal interference-aware resource allocation can be formulated as a convex program. The problems are analyzed through

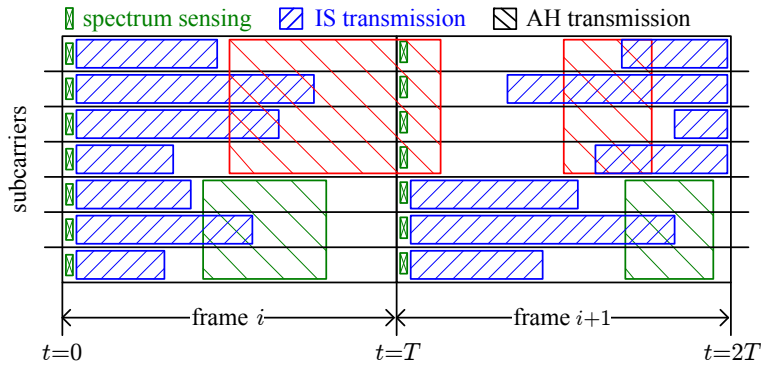


Figure 1.5: Cognitive coexistence between infrastructure and ad-hoc networks. The infrastructure network allocates power and transmission time such as to minimize the temporal transmission overlap between both systems.

Lagrangian techniques which leads to closed-form solutions for the optimal resource allocation as a function of Lagrange multipliers. Algorithms are presented for finding the optimal values of these Lagrange multipliers with guaranteed convergence.

The previous formulations require that rate constraints are met on a per-frame basis. In practice, however, it usually suffices to meet these constraints on the long-term average. This relaxation can significantly improve performance because less rate may be allocated in frames with poor interference conditions, provided that this is compensated during more advantageous frames. Ultimately, the average-rate formulation leads to a resource allocation across both frequency (the subchannels of the multicarrier system) and time (consecutive frames of the infrastructure system).

The above allocation techniques operate at the terminals of the infrastructure system and rely on local sensing results. The multi-terminal case is addressed by having clients operate on disjoint sets of subchannels, which are allocated by the infrastructure base station. The problem of optimal subchannel allocation therefore

arises naturally. While a comprehensive analysis of this problem is difficult, our numerical results suggest that conventional subchannel allocation methods, which minimize total transmit power subject to rate constraints, are almost optimal for the interference-aware formulation as well.

1.2.4 Cognitive Frequency Hopping Test Bed

The fifth chapter of this dissertation presents a real-time test bed for the cognitive frequency hopping protocol proposed in Chapter 3. By comparing theory and experiment we are able to demonstrate proof-of-concept, validate model assumptions, and gain a better understanding of the interaction between the cognitive radio and the ad-hoc bands. The measurements show a good fit with our analytical results.

The implementation of the test bed consists of the spectrum sensor, the cognitive frequency hopping controller, and the transmitter of the cognitive radio (see Figure 1.6). These components are discussed in detail and practical limitations are addressed. The measurement methodology also forms an important part of this work, as the performance of the cognitive radio needs to be quantified with respect to the behavior of the ad-hoc bands.

The experimental results gathered from this test bed validate the statistical model introduced in Chapter 2. Furthermore, they enable us to better understand the dynamic interaction between the ad-hoc system and the cognitive radio by measuring its open-loop and closed-loop behavior. The open-loop setup, in which transmissions of the test bed are not fed back to the ad-hoc system, enables us to draw a direct comparison with our analytical contribution. The closed-loop setup, on the other hand, allows us to better understand the impact that retransmissions

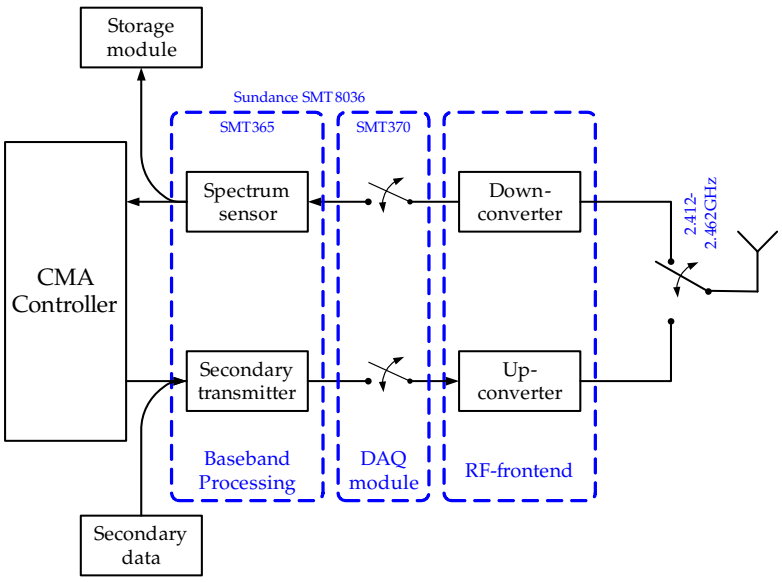


Figure 1.6: Block diagram of the cognitive frequency hopping test bed.

may have on the temporal activity of the ad-hoc system.

1.3 Related Publications

Some of the research results presented in this thesis have previously been published. These works are referenced below.

Portions of Chapter 2 have previously been published in [13, 14, 15] and [16]. In [13] and [14], the prediction model for the ad-hoc network was presented for the first time and statistical validation of the model was performed. An extension to tracking non-stationary traffic was presented in [15].

Some of the results in Chapter 3 were presented at [17, 18] and have appeared in [19, 20]. In [17, 19] we presented results for the fully and partially observable case. A periodic sensing scenario was addressed in [18, 20].

Portions of Chapter 4 have been presented at [21, 22] and a journal contribution is currently in submission [23]. The case of optimal power allocation based on knowledge of the interference channel has been considered in [21]. The case of optimal power and transmission time allocation based on temporal prediction of the ad-hoc system was considered in [22].

The development of the cognitive frequency hopping test bed was presented in [24, 25] and a journal contribution is currently in submission [26].

CHAPTER 2

SENSING AND PREDICTION FRAMEWORK

2.1 Summary of Contributions and Related Work

Cognitive radio enables wireless systems to exploit spectrum opportunities which result from the spectral underutilization of a primary system. The main contribution of this thesis is the development of medium access protocols that can take advantage of temporal idle periods while adhering to interference constraints with respect to the primary system.

The challenges associated with using temporal white space are quite different from those in other domains. The secondary system needs to have a means of predicting the behavior of the primary system's medium access in order to separate transmissions in time. In this chapter, we develop a prediction framework for the temporal activity of the ad-hoc system which will form the foundation for the protocols derived in subsequent chapters.

Using temporal white space is motivated by the significant amount of idle periods that typically exist between ad-hoc transmissions. This underutilization is due to the burstiness of the traffic that is usually supported by such systems, and the inefficiency of the decentralized random medium access protocols predominantly used. Moreover, in many coexistence setups the temporal domain is the only one that can be exploited from a practical viewpoint. Wireless ad-hoc systems often operate in close proximity of other devices and a separation in space is therefore not possible. In the extreme case where multiple radios are integrated into the same device (which becomes more and more typical in portable devices)

separating transmissions in the temporal domain may be the only way of reducing cross-talk.

Many ad-hoc systems use carrier-sense multiple access (CSMA) to avoid collisions and provide terminals with fair access to the medium (consider, for example, IEEE 802.11). Although the behavior of these protocols is standardized, predicting the activity of the overall ad-hoc system may be difficult due to its decentralized nature and the impact of many different parameters of the radio environment. This complexity motivates an empirical approach toward proposing a prediction model. Instead of modeling terminals explicitly and deriving a system model based on their individual behavior, we gather empirical observations by experiment and propose statistical models that approximate the empirical data. In this way, we can also strike a better balance between accurately representing the statistical behavior and maintaining the analytical tractability that will be needed in subsequent chapters.

2.1.1 Main Contribution

The main contribution of this chapter is the proposition of two statistical models for the idle and busy periods of an ad-hoc WLAN system. Both of the proposed models are simple two-state models in continuous time but differ in their statistical characterization of the holding time in each state. In particular, we consider a semi-Markov model and a continuous-time Markov chain:

Semi-Markov Model. For a semi-Markov model (SMM), it is possible to specify the holding times through arbitrary probability distributions. This flexibility is used to find an accurate fit with empirical data but leads to higher model complexity.

Continuous-Time Markov Chain. For a continuous-time Markov chain (CTMC), the holding time in any state needs to be exponentially distributed. Naturally, this results in a worse approximation of the empirical data but leads to a simple and tractable model. We will see in subsequent chapters that the memoryless property of the exponential distribution greatly simplifies the derivation of cognitive medium access protocols.

To solidify the proposition of the above models, we discuss the empirical setup in detail and describe the sensing methods that are used to find idle and busy periods accurately. The model is validated through extensive measurements for different traffic scenarios and the statistical fit is verified through the Kolmogorov-Smirnov test.

2.1.2 Related Work and Organization

Traffic analysis and modeling have become increasingly important to better understand and design wireless networks. Much attention has focused on demonstrating that Poisson models are inadequate for capturing some fundamental characteristics of internet traffic [27]. Instead, the focus has shifted to models that can incorporate the heavy-tailed behavior and long-range dependencies that are observed in practical systems. There have also been studies on the small time scale behavior of internet traffic [28].

Wireless networks may exhibit very different behavior due to the variability of the wireless medium as well as interference and contention between nodes. Even though some of the end-to-end applications that are typically supported by WLANs may be the same as the ones that are supported by wired networks (for exam-

ple, when WLAN serves as a means to access the internet), the medium access and contention behavior may have a profound impact. Many empirical studies have been conducted to gain a better understanding of ad-hoc network behavior [29, 30, 31, 32, 33]. These contributions study typical usage and performance patterns for some realistic scenarios and also discuss the higher complexity associated with reliably capturing packets on the air interface. For example, in order to draw meaningful conclusions from a trace it is necessary to capture almost all of the packets in a certain area. However, given propagation characteristics and differences in the sensitivity of adapter cards, this may be challenging to guarantee in practice [30]. Further, hidden and exposed nodes pose additional challenges.

While some experimental studies have focused on establishing models for WLAN system parameters (such as throughput or delay) finding a characterization of the idle and busy periods of the channel has not received attention. To the best of the author's knowledge our contribution is among the first to study such metrics.

Some contributions in temporal DSA have assumed simple Markovian models for characterizing the temporal behavior of the primary user. For example Zhao *et al.* [34] focus on a discrete-time Markov chain model for the occupancy of a slotted primary system. Continuous-time Markov chains are therefore natural extensions for modeling primary users that evolve in continuous time [19, 34, 35].

Finally, research in DSA has focused on analyzing the performance of spectrum sensing schemes. This is especially important in setups, where sufficient spatial separation is used to orthogonalize users. By the nature of such an approach, very weak primary signals need to be detected reliably, which requires sophisticated sensing methods [7]. Due to uncertainties about the propagation environment

and the statistics of the noise, it may even be impossible to guarantee a certain detection accuracy [7]. However, in contrast to spatial DSA networks, the propagation conditions are very different in our setup. The proximity of devices leads to moderate to high signal-to-noise ratios and consequently simple methods such as energy detection perform well. Ultimately, in our framework, orthogonality is not achieved by sufficient separation but by accurately predicting the temporal activity of the primary system.

2.2 Empirical Data Gathering

The complexity associated with modeling the medium access of ad-hoc systems motivates an empirical approach in which idle and busy periods are gathered from an experimental setup and then analyzed. Throughout this chapter, the ad-hoc system is represented by an IEEE 802.11b WLAN system, which operates in the 2.4 GHz band. While the measurement-based approach requires us to focus on a specific type of ad-hoc network, our results should extend to related networks that are also based on carrier-sense multiple access with collision avoidance (CSMA/CA).

The measurement setup consists of a vector signal analyzer which captures the raw complex baseband signal of the WLAN system. These raw data are stored and subsequently processed to find the exact start and end times of the packets. This guarantees an accurate and verifiable identification of the idle and busy periods and enables us to test different spectrum sensing methods. Our approach differs from many related publications, which gather traces of WLAN traffic by using commercial adapter cards that operate in promiscuous mode. While such

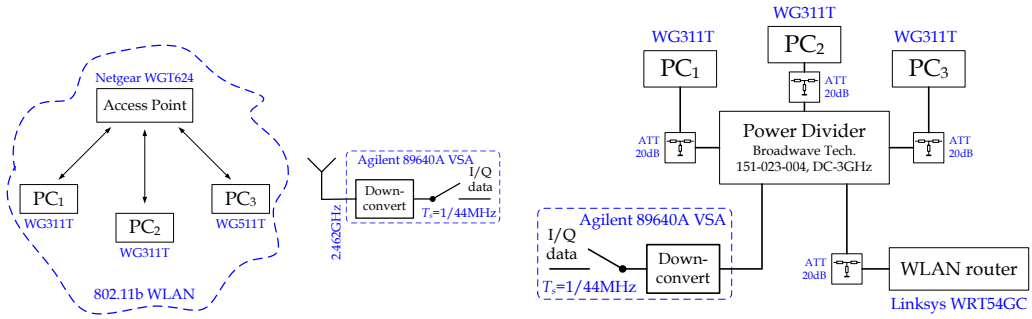
approaches are well suited for finding long-term statistics and coarse metrics of the WLAN system they may not be accurate enough to measure idle and busy periods accurately.

2.2.1 Experiment Setup

The experiment setup consists of an actual WLAN system, whose transmissions are recorded with an Agilent 89640A vector signal analyzer; see Figure 2.1. The device internally down-converts the WLAN signal from radio frequency and is configured to sample the baseband signal at a sampling frequency of 44 MHz. We consider both an antenna-based and RF-isolated setup. Since measurements are conducted in an unlicensed band, the latter configuration helps to reduce interference from unrelated sources operating in the vicinity.

Antenna-Based Setup. The antenna-based propagation setup is shown in Figure 2.1(a) and consists of a Netgear WGT624 wireless router and three computers with wireless adapter cards (two use a Netgear WG311T and one uses a WG511T adapter). The WLAN system operates in channel 6, which corresponds to a 22 MHz frequency band centered at 2.437 GHz [9]. The devices are all located in the same room, leading to a setup with moderate to high signal-to-noise ratio (SNR) and no hidden terminals. Even though a completely interference-free setup cannot be guaranteed we use the vector signal analyzer to select a channel with minimum interference.

RF-Isolated Setup. In addition to the antenna-based setup, we also consider the RF-isolated setup shown in Figure 2.1(b). It consists of a Linksys WRT54GC



(a) Antenna-based setup (b) RF-isolated setup

Figure 2.1: Experimental setup for gathering idle and busy periods of a WLAN system. Both antenna-based and RF-isolated setup are based on capturing transmissions using a vector signal analyzer.

wireless router and three workstations with Netgear WG311T wireless adapter cards. All devices are connected to a Broadwave Technologies resistive power divider via RG174U coaxial cables and SMA connectors. Since all devices are connected through cables, interference from unrelated sources is eliminated almost entirely and can be neglected.

The Netgear router used for the antenna-based setup could not be used for the isolated measurements as well since its built-in antenna was non-detachable. The use of two different routers caused our setup to differ in terms of the used synchronization preamble. While the Netgear router could be configured to use only long synchronization preambles, the Linksys router did not allow for specifying this option. As a consequence most of the time a short preamble was transmitted (due to the high SNR setup). While this leads to slightly different packet durations it does not impact the qualitative behavior of our results.

Traffic Generation. To analyze idle and busy periods for varying traffic types and intensities, each of the workstations generates traffic using the Distributed

Internet Traffic Generator (D-ITG) [36]. This software application allows for a flexible statistical specification of traffic characteristics, including varying packet lengths and inter-departure times.

In addition, we also consider practical traffic scenarios by using the popular "Skype" voice-over-IP (VoIP) client to set up a conference call within the WLAN, using the traffic generator to simulate G.711-codec-based voice communication, and using an SFTP client to download files from a central server.

2.2.2 Spectrum Sensing and Opportunity Detection

The measurement setups described in the last section yield time captures of the complex baseband signal, which we then process to find the exact start and end time of each packet. Clearly, this fully determines the channel's idle and busy periods. We consider two sensing strategies that differ in terms of their assumptions and detection performance: energy detection and feature-based detection.

Energy-Based Detection. If the primary user's transmission standard is unknown, a natural approach for detecting the start and end times of packets is based on the received energy. In order to achieve satisfactory performance we consider blocks of N samples whose length is much shorter than the smallest packet length. The detection problem can then be formulated as

$$\mathcal{H}_0 \quad : \quad Y_i = V_i, \ i = 1, \ldots, N \tag{2.1}$$

$$\mathcal{H}_1 \quad : \quad Y_i = S_i + V_i, \ i = 1, \ldots, N, \tag{2.2}$$

where Y_i denotes the complex baseband samples, V_i are noise samples, $V_i \sim \mathcal{CN}(0, \sigma_0^2)$, and S_i denotes the signal samples drawn independently from a complex

Gaussian distribution, $S_i \sim \mathcal{CN}(0, \sigma_1^2)$. Lacking any information on the transmission standard of the primary user, the Gaussian assumption for S_i appears reasonable.

The hypothesis testing problem defined above is standard [37] and the optimal Neyman-Pearson detector is given by

$$T(\mathbf{y}) = \sum_{i=1}^{N} |y_i|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{2.3}$$

where the threshold γ is determined according to the probability of false alarm, which amounts to

$$\alpha = \Pr(T(\mathbf{y}) > \gamma | \mathcal{H}_0) = 1 - \tilde{\Gamma}_r(N, \frac{\gamma}{\sigma_0^2}), \tag{2.4}$$

where

$$\tilde{\Gamma}_r(N, \xi) = \frac{1}{\Gamma(N)} \int_0^{\xi} t^{N-1} e^{-t} dt \tag{2.5}$$

is the regularized gamma function and $\Gamma(N)$ is the complete gamma function [38]. Similarly, the power of the detector is given by

$$\beta = \Pr(T(\mathbf{y}) > \gamma | \mathcal{H}_1) = 1 - \tilde{\Gamma}_r(N, \frac{\gamma}{\sigma_0^2 + \sigma_1^2}). \tag{2.6}$$

The above expressions show that the detection performance depends on the signal-to-noise ratio, which is defined as $\mathrm{SNR} = \sigma_1^2/\sigma_0^2$, as well as on the block length N. For our setup we choose $N = 44$ samples, which corresponds to 1 µs long blocks. If we demand $\alpha = 1 - \beta < 10^{-5}$ then we can see that we have to guarantee that the SNR is above 4.29 dB which is easily met in our setup.

The fact that energy detection works well in our setup is not surprising given the moderate to high SNR conditions. As a matter of fact, spectrum sensing in the context of cognitive coexistence is quite similar to carrier-sensing in multiple access systems. The main difference is that spectrum sensing detects the activity

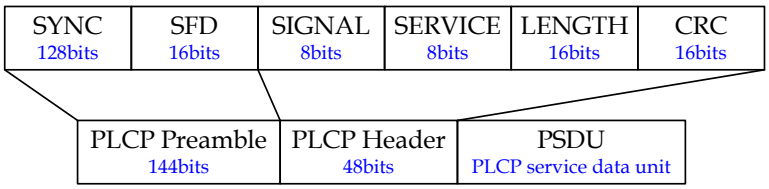


Figure 2.2: Physical layer preamble in IEEE 802.11b. The start-of-frame delimiter (SFD) and LENGTH field can be used to find the start and end times of WLAN packets.

patterns of interfering wireless systems, while carrier-sensing detects transmissions of other terminals within a homogenous wireless network. Consequently, despite the similarities in the physical-layer sensing, the detection outcomes need to be used in very different ways.

Feature-Based Detection. Energy detection does not assume knowledge of any information, any *features*, of the WLAN transmit signal. By tailoring the sensing process to these signals, the detection performance can be improved considerably.

To find the precise start and end times of WLAN packets of an IEEE 802.11b system, we can exploit its physical-layer preamble, which is shown in Figure 2.2. It consists of a physical-layer convergence procedure (PLCP) preamble, split into a block of scrambled '1's ('0's for the short-preamble) and the start-of-frame delimiter (SFD) indicating the beginning of the PLCP header. Clearly, the SFD can be used to precisely find the start time of the packet. The information provided in the header consists of a "SIGNAL", "SERVICE," and "LENGTH" field as well as a cyclic-redundancy check (CRC) which protects these three fields. Together, the SFD and the LENGTH field can naturally be used to determine the start and the end time of WLAN transmissions.

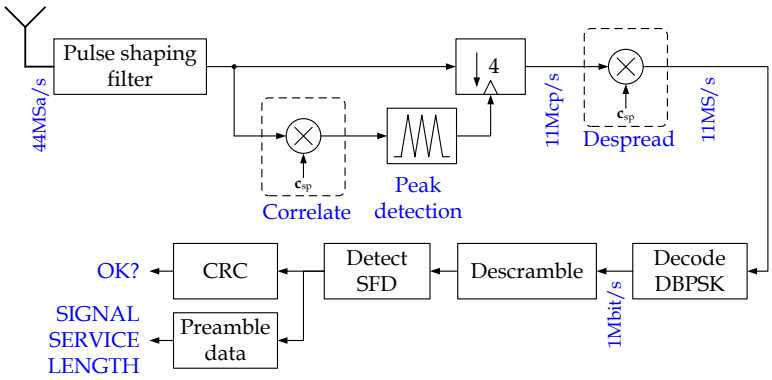


Figure 2.3: Receive processing for feature-based detection.

The receive processing for the feature-based detection scheme is depicted in Figure 2.3. The complex baseband data collected at a rate of 44 MHz is first passed through a Gaussian pulse shaping filter with a bandwidth/symbol-time product of $BT_s = 1/2$. In order to obtain chip synchronization the filtered signal is correlated with the 11-sample Barker sequence specified by the standard [39]. The resulting signal shows periodic peaks whenever the spreading sequence lines up with the input signal. We detect these peaks and downsample the signal to the symbol rate of 11 Mbps. Subsequently, we despread and demodulate the DBPSK/DQPSK encoded preamble. The frequency offset at the receiver is noticeable but can be neglected since the signals are differentially encoded. After successful decoding, the resulting bit stream is descrambled and the start-of-frame delimiter (SFD) is detected. In the same way, the SIGNAL, SERVICE, and LENGTH field are extracted and a cyclic redundancy check is performed to ensure that the extracted information is accurate.

In our empirical study we use both energy and feature-based detection to detect the idle and busy periods of the channel. By comparing the results from both methods, we are able to validate the accuracy of both schemes and indeed observe a good match. Due to the moderate to high SNR conditions of our setup, both

schemes work very well.

In a practical system, other implementation considerations may impact the choice of the sensing method. For example, if the cognitive radio system operates in a slotted fashion (as it will be the case in Chapter 3), energy detection is better suited because it does not require that the channel be continuously monitored for detecting the start of a synchronization preamble. Instead, it suffices to capture a set of a samples at the beginning of the slot and make a decision based on the energy contained in them.

2.3 Modeling Spectrum Opportunities

The setups described in the last section are used for gathering experimental data. In this section we present some of the empirical results and develop two statistical models that approximate their behavior.

2.3.1 Empirical Results

The empirical results presented in this section focus on a scenario, in which each of the WLAN terminals supports stationary traffic with a specified level of intensity. This enables us to measure the distribution of the idle and busy periods accurately by capturing long traces and assuming that time and ensemble averages are identical.

In addition to stationary traffic scenarios, we also consider more practical types of traffic such as voice-over-IP (VoIP) and file-transmission-protocol (FTP) traf-

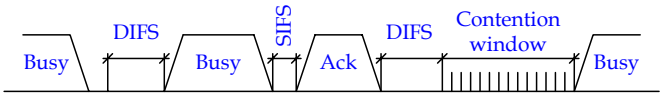


Figure 2.4: Illustration of carrier-sense multiple access used in IEEE 802.11b WLAN systems. ©2006 IEEE. Reprinted, with permission, from [14].

fic. We observe similar qualitative trends as compared to the stationary traffic scenario. After proposing the statistical model, we will demonstrate in detail how non-stationary traffic can be tracked by adjusting model parameters. For a better understanding of the results we briefly review some characteristics of WLAN medium access.

WLAN Medium Access Protocol. The IEEE 802.11 WLAN standard [9, 39] uses CSMA/CA to control stations' access to the medium. As depicted in Figure 2.4, this requires that before transmitting a packet, a station first senses the medium. If the channel is free, the station continues sensing for a duration equal to the distributed coordination function inter-frame space (DIFS). If the channel remains idle throughout the entire period, the station may initiate a transmission.

After a packet transmission has been completed, the receiver needs to confirm reception by responding with an acknowledgement. Only a short inter-frame space (SIFS) is necessary as to give priority over new packet transmissions.

If the channel is busy in the first place the station has to defer access until the medium becomes idle again. Then, after a DIFS, a contention window is used to avoid collisions between multiple stations trying to access the medium. Specifically, each station generates a uniform random number $i \in \{0, \ldots, 31\}$ and defers transmission for $iT_{\text{slot}} = i \cdot 20\,\mu\text{s}$ before accessing the channel (provided that no other station has already started to access the channel before).

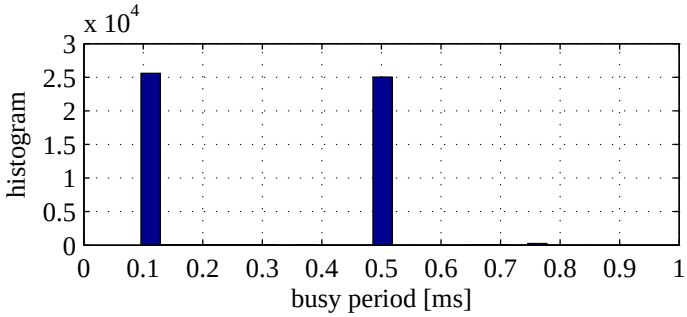
The standard provides more technical details that go beyond the basic features discussed above. Since these specifics do not manifest themselves in our study, we refer to the standard for further details [9].

Validation Experiment. We first perform a simple experiment to validate the measurement setup. In particular, we consider the isolated measurement setup shown in Figure 2.1(b), but with only the wireless router and one PC being active. The D-ITG traffic generator is used to generate UDP packets with a constant length of 512 Bytes, constant inter-arrival times, and a traffic intensity of 10^5pkts/s. This rate exceeds the capacity of the WLAN system and therefore ensures that all workstations always have packets to transmit.

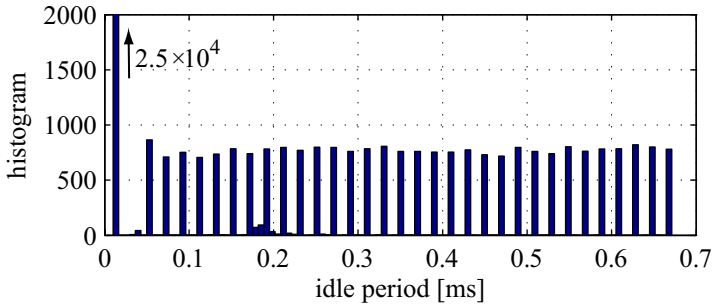
Based on this setup, we use a vector signal analyzer to capture 100 blocks of complex baseband data with a duration of 0.25 s each. The blocks are processed to find idle and busy periods. A histogram of the empirical data thus obtained is shown in Figure 2.5.

As can be expected, the histograms reflect the standard characteristics. First, the histogram of the busy periods depicted in Figure 2.5(a) shows only three components, corresponding to the transmission of acknowledgement packets ($t \approx 0.11$ ms), data packets ($t \approx 0.51$ ms) and beacon frames ($t \approx 0.76$ ms), respectively. Given that we deal with constant-length packets, this result is in accordance with our expectations.

The histogram of the idle durations reflects the standard as well. We observe a discrete component with a length of approximately 10 µs, corresponding to the SIFS. The 32 discrete components spaced about 20 µs apart correspond to the contention window described previously.



(a) Histogram for the busy periods.



(b) Histogram for the idle periods.

Figure 2.5: Validation experiment for a fully loaded WLAN system. The busy periods show two components, which correspond to data and acknowledgement packets, respectively. The idle periods show the contention window of the WLAN terminals.

Stationary Traffic Scenarios. The last section has validated the measurement setup based on a simple traffic scenario. In this section, we now use all three workstations together with the wireless router as discussed in Section 2.2. We continue to focus on stationary UDP traffic with constant-length packets but specify exponentially distributed inter-departure times and varying traffic intensity. Specifically, the WLAN traffic load is represented by σ, where $\sigma = 0$ corresponds to a completely inactive WLAN system and $\sigma = 1$ denotes a system that is active to the maximum extent. Experimentally, the traffic load σ is determined by measuring the maximum WLAN capacity (in packets per second) similar to the last section, and normalizing the traffic load accordingly. For example, if a maximum

traffic load of 450 packets per second can be supported by the WLAN then configuring each of the traffic generators to transmit at a rate of 50 packets per second corresponds to $\sigma = 1/3$.

The histograms for busy and idle periods are shown for $\sigma = 0.4$ in Figure 2.6. The busy periods are again discrete with the components corresponding to acknowledgement, data, and beacon packets, respectively. The histogram of the idle periods, on the other hand, allows us to make two important observations. First, there is a significant component around 0.7 ms which corresponds to the effect of the contention window and the DIFS. Second, the tail of the distribution appears to decay approximately exponentially, suggesting that besides some impact from the standard, the idle periods are modeled approximately by an exponential distribution.

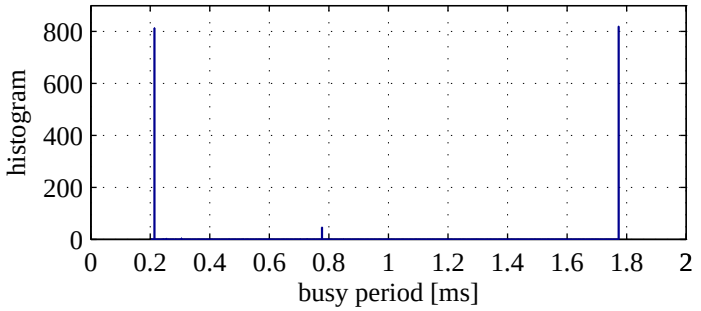
Based on the above observations it is natural to define the following channel states:

DATA The channel is busy due to the transmission of a data packet. The sojourn time in this state is deterministic and amounts to the time required to transmit the 1024 B size packet (namely 1.77 ms).

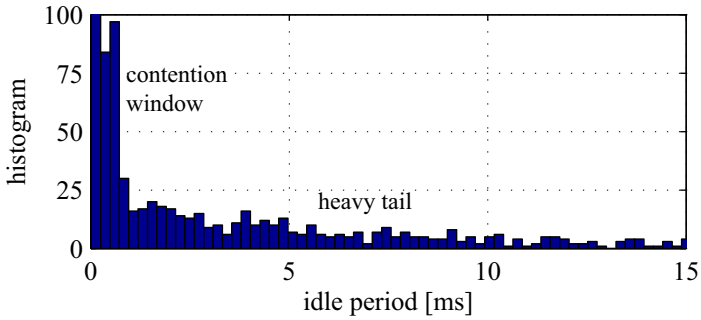
SIFS The channel is idle due to the short inter-frame space required between a data packet and its subsequent acknowledgement. The sojourn time in this state is deterministic and amounts to 10 µs.

ACK The channel is busy due to the transmission of an acknowledgement packet. The sojourn time is deterministic and amounts to 0.21 ms.

CW The channel is idle but there are primary users contending for the medium. The sojourn time in this state can be approximately derived from the stan-



(a) Histogram for the busy durations.



(b) Histogram for the idle durations.

Figure 2.6: Histograms of idle and busy periods for a stationary traffic scenario. As in Figure 2.5, the busy periods show two components, corresponding to data and acknowledgement packets, respectively. The idle periods, show the impact of contention among nodes some heavy-tailed behavior corresponding to the random packet arrival.

dard. Because all stations defer transmission for uniformly distributed random time periods, the sojourn time can be assumed to have finite support as well. Further, even though the exact distribution depends on the number of stations contenting for the medium, a uniform distribution leads to a good approximation as long as few stations are contenting for the medium a the same time.

FREE The channel is idle because none of the primary users has packets to transmit. From the viewpoint of cognitive coexistence the time spent in this state essentially defines to what extent the channel can be reused. A generalized Pareto distribution will turn out to be a good fit for the sojourn time.

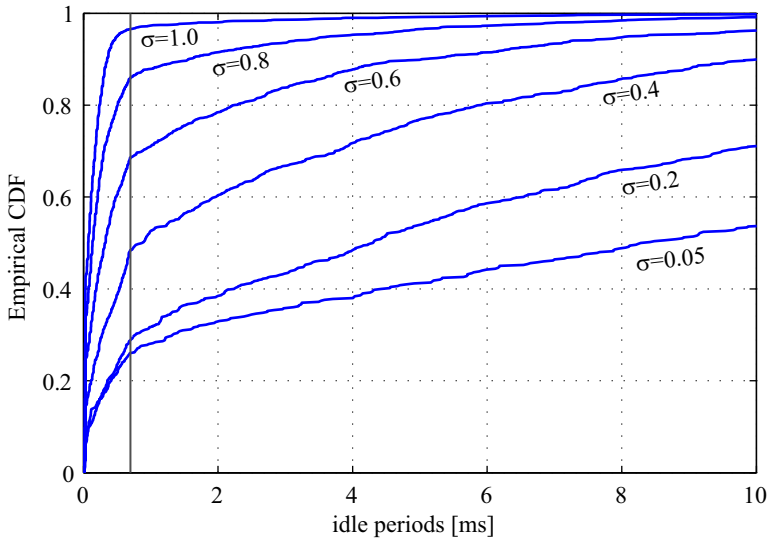


Figure 2.7: Empirical cumulative distribution functions for the idle periods at several WLAN traffic intensities. Due to contention among nodes, the curves exhibit a bend at approximately 0.7 ms.

While the histograms depicted in Figure 2.6 give a first impression of the distribution of the idle durations, more insights can be gained from the empirical distribution function, which is defined as the fraction of observations smaller than some t [40]

$$F_e(t) = \frac{\#i : y_i \leq t}{n}, \tag{2.7}$$

where $y_i, i = 1, \ldots, n$ correspond to n independent samples. The empirical distribution function is shown in Figure 2.7 for different values of the traffic intensity σ. We can make two important observations. First, idle periods decrease stochastically with σ, that is, if $\sigma_1 < \sigma_2$ then $F_e(t;\sigma_1) < F_e(t;\sigma_2)$ for all t. Second, for low to medium traffic load we can clearly observe that the distribution consists of two mixture components, corresponding to the contention window and the random packet arrival, respectively (note the bend at approximately $T_c = 0.7$ ms). The approximately constant slope of $F_e(t)$ within the contention window suggests that a uniform distribution is indeed a good fit. The tail distribution corresponding to a free channel will be analyzed in more detail.

Non-Stationary Traffic Scenarios. In addition to stationary traffic scenarios, we also consider two non-stationary traffic setups consisting of FTP and VoIP traffic, respectively. The empirical CDFs of the idle periods are shown in Figure 2.8 for both scenarios.

First, consider file transfer via secure-FTP from a remote server. In order to collect enough baseband data a text file of approximately 100 kB was transferred 1000 times using a secure-FTP client. The resulting curve shows that there is little remaining whitespace. The effect of the contention window is well-visible by the bend in the empirical CDF at 0.7 ms.

Second, we used D-ITG to generate traffic according to the G.711 codec (used by some VoIP clients). We consider the case of one and three codecs running simultaneously on each of the workstations. The resulting curves show an almost idle channel in the case of one active codec, while the channel appears quite busy in the case of three.

Finally, we used the popular "Skype" client to set up a conference call within the WLAN. A prerecorded audio sample was used to simulate a speech conversation on each of the workstations. The resulting empirical CDF shows that the channel is lightly used.

In summary, the empirical CDFs for the above traffic scenarios show similar qualitative behavior compared to the stationary traffic scenarios considered before. The tail of the distribution and the effect of the contention window are again similar.

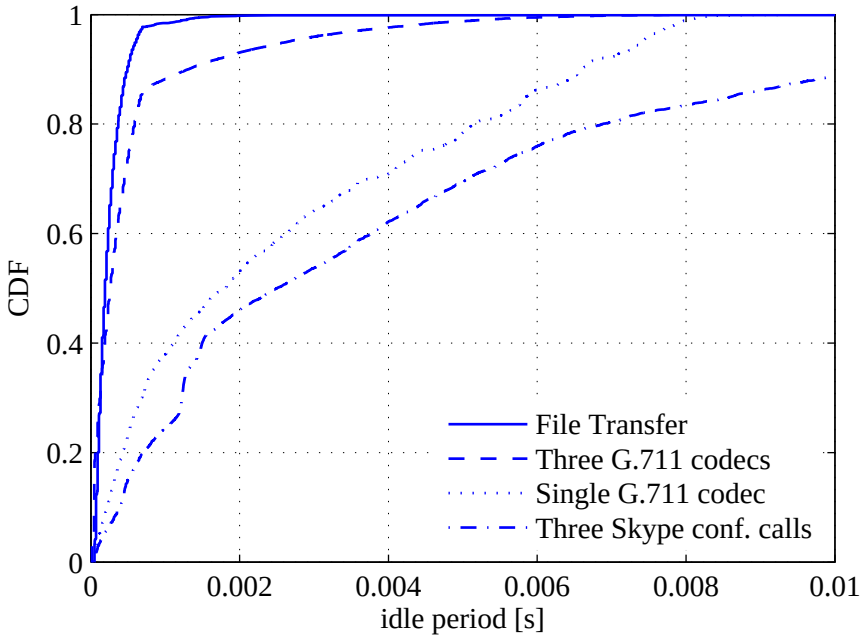


Figure 2.8: Empirical cumulative distribution functions of file-transfer protocol (FTP) and voice-over-internet protocol (VoIP) traffic.

2.3.2 Semi-Markov Model

Based on the classification of states, we can analyze the transition behavior from the empirical data. This analysis shows that the transition DATA→SIFS→ACK occurs with almost certainty, which is not surprising as this state sequence simply corresponds to a successful packet transmissions followed by the mandatory inter-frame space and an acknowledgement packet. Further, we note that the states CW and FREE are not observable because, by observing the medium only, it is impossible to differentiate whether a terminal is contending for the medium or simple does not have any packets to transmit.

The above observation motivates lumping DATA, SIFS, and ACK state into a TRANSMIT state, and introducing an IDLE state which consists of the states CW and FREE; see Figure 2.9. This leads to a simple two-state ON/OFF model

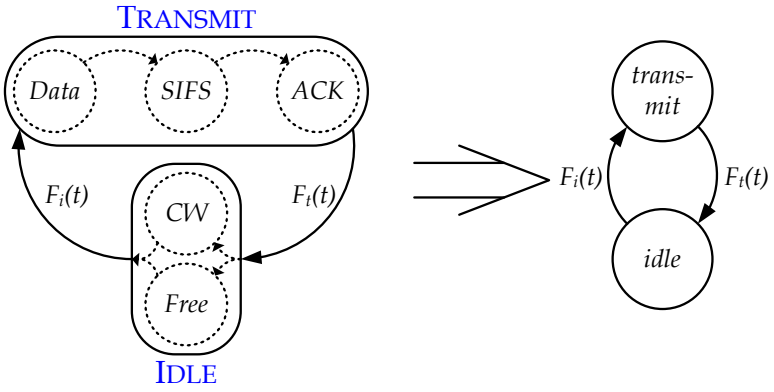


Figure 2.9: State transition diagram of the temporal activity model. By lumping states, a simple two-state model can be obtained which facilitates the derivation of cognitive coexistence protocols.

for the WLAN medium access and allows us to concentrate on modeling the idle and busy periods statistically by specifying the holding times in TRANSMIT and IDLE state.

To obtain an accurate fit, we first consider a semi-Markov model which enables us to approximate the holding times by arbitrary probability distributions (because the model only has two states this is also referred to as an alternating renewal process). In the remainder of this section we discuss two fitting approaches, namely a mixture and a hyper-Erlang fit, which approximate the empirical data while depending on few model parameters. In the next section, we will compare this approach to simply fitting an exponential distribution.

Fitting a Mixture Distribution. We first focus on fitting a mixture distribution to the empirical distributions of the idle periods based on the observation that the impact of the contention behavior is well approximated by a uniform distribution, while the tail of the distributions decays approximately exponentially. By fitting different distributions to the data we found that a generalized Pareto distri-

bution provides an accurate fit for the distribution's tail. This distribution includes the special case of the exponential distribution but can model some heavy-tailed behavior through appropriate choice of a shape parameter.

Mathematically, the mixture distribution is given by

$$F_m(t) = p_c F_c(t) + (1 - p_c) F_f(t; k, \omega), \tag{2.8}$$

where p_c denotes the mixture coefficient, $F_c(t)$ is the CDF of a uniform distribution between $t = 0$ and $t = T_c = 0.7\,\text{ms}$, and $F_f(t; k, \omega)$ denotes the CDF of a generalized Pareto distribution

$$F_f(t; k, \omega) = 1 - \left(1 + k\frac{t}{\omega}\right)^{-1/k}. \tag{2.9}$$

with shape parameter k (quantifying the deviation from an exponential distribution) and scale parameter ω (corresponding to the decay rate). There are several standard approaches for estimating the model parameters of the mixture distribution. The Expectation-Maximization (EM) algorithm [41] is a well-known technique, but computationally expensive to apply in this case. Therefore, we consider a simpler approach, which is based on the specific structure of the mixture distribution.

In particular, we can exploit that the uniform component of the mixture distribution has finite support $[0, T_c]$. Therefore, by removing all observations that fall into this interval, we can eliminate the impact of the uniform component entirely and estimate the parameters of the generalized Pareto distribution directly from the remaining samples.

Following this approach, let the truncated data obtained by discarding idle periods smaller than the threshold T_c be denoted by $\tilde{y}_i,\ i = 1, \ldots, N_t$. The maximum

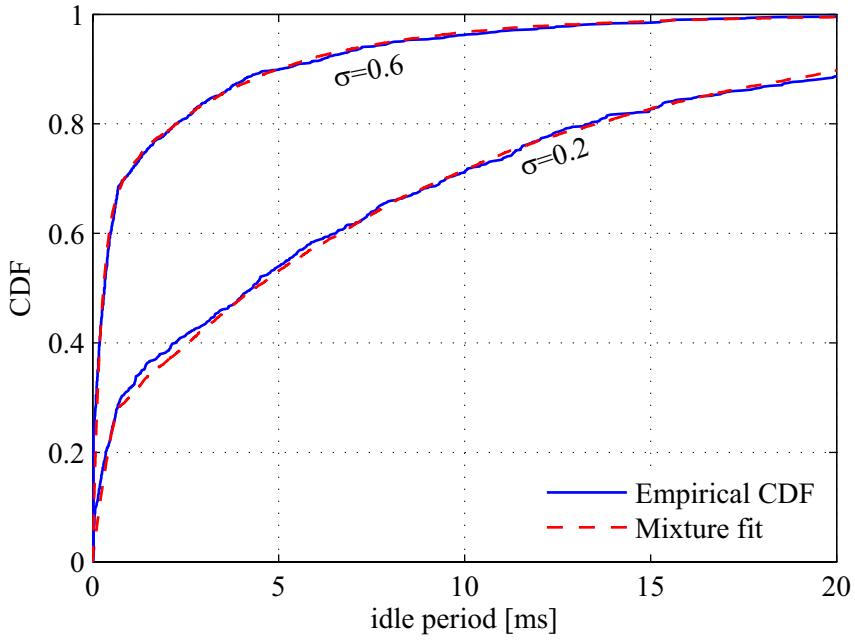


Figure 2.10: Empirical cumulative distribution functions and their mixture-based fits. The curves correspond to a WLAN traffic load of $\sigma = 0.2$ and $\sigma = 0.6$, respectively.

likelihood estimator for the parameters $\boldsymbol{\theta} = [k, \omega]$ is then given by [42]

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{N_t} \frac{f_f(y_i; \boldsymbol{\theta})}{1 - F_f(T_c; \boldsymbol{\theta})}, \tag{2.10}$$

where the denominator is due to the left-truncation of the data. The maximization in the above formula is performed numerically, using an initial value obtained by a moment estimate [43]. Having estimated one of the components of the mixture distribution (2.8), we can easily find the mixture parameter p_c because $F_c(t)$ is uniform.

The mixture distribution approximates the empirical data well as can be seen in Figure 2.10 for traffic intensities $\sigma = 0.2$ and $\sigma = 0.6$, respectively.

Fitting a Phase-Type Distribution. Our objective in fitting distributions is to obtain a concise model that approximates the empirical data for varying

traffic intensities while depending on few parameters. The mixture distribution introduced in the last section represents one such model and leads to a good fit based on the insights taken from the IEEE 802.11 standard. In this section, we consider fitting phase-type distributions to the empirical data.

Phase-type distributions are a class of distributions which model the time to absorption of a CTMC with a set of transient and a single absorbing state. These distributions are frequently used in statistical modeling [44, 45, 46], since they allow to approximate heavy-tailed behavior within the framework of CTMCs by expanding the state space. In practice the usefulness of fitting phase-type distributions depends on how many states are needed to approximate the heavy-tailed behavior well. Ultimately, we face a tradeoff between obtaining an accurate approximation and retaining a concise statistical model with few parameters.

Phase-type distributions can be classified into different groups, such as hyper-exponential, Erlang, and hyper-Erlang distributions [45]. While, hyper-exponentials (a mixture of exponential distributions) are frequently used in statistical modeling [44] they are not an appropriate choice in our case. In fact, it can be shown that using hyper-exponential distributions we can only approximate distributions that have a coefficient of variation (CoV) greater than one [45]. From the empirical data, however, we observe CoVs smaller than one.

Instead, we consider a hyper-Erlang distribution, which corresponds to the time of absorption of the CTMC shown in Figure 2.11. This class of distributions can approximate empirical distributions with arbitrary CoV and has furthermore received interest lately [46]. In particular, an efficient algorithm based on the Expectation-Maximization algorithm has been proposed in [45] and is used for fitting the distribution in this section.

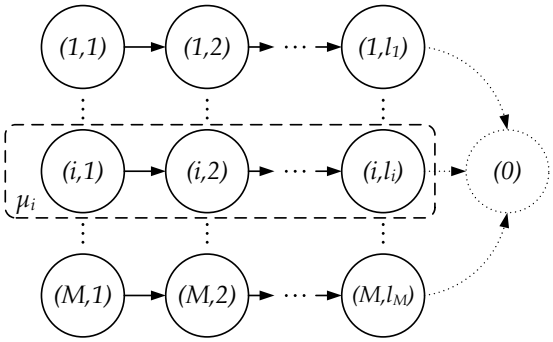


Figure 2.11: Transition diagram of a continuous-time Markov chain resulting in an overall hyper-Erlang distribution (adapted from [45]). The absorbing state is denoted by '(0)'.

To illustrate how an M-component hyper-Erlang distribution can be fitted to the empirical data, let $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_M]$ denote the rate parameters of the Erlang distributions and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_M]$ the mixture coefficients. Then the mixture distribution is given by

$$f(t; \boldsymbol{\alpha}, \boldsymbol{\mu}) = \sum_{i=1}^{M} \alpha_i f_{\mathrm{Er}}(t, \mu_i), \tag{2.11}$$

where

$$f_{\mathrm{Er}}(t; \mu_i) = \frac{(\mu_i t)^{l_i - 1}}{(l_i - 1)!} \mu_i e^{-\mu_i t} \tag{2.12}$$

is the PDF of an Erlang distribution with shape parameter l_i (assumed known). Each component of the mixture (2.11) can be viewed as generated by one of the chains in the CTMC depicted in Figure 2.11.

A maximum likelihood approach for estimating the unknown parameters $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\mu}]^T$ given observations $y_i,\ i = 1, \ldots, N$ leads to

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{k=1}^{N} \log \left[\sum_{i=1}^{M} \alpha_i f_{\mathrm{Er}}(y_k; \mu_i) \right]. \tag{2.13}$$

The maximization in the above formula is difficult since the expression involves the logarithm of a sum [45]. However, we would be able to simplify (2.13) if we

knew from which mixture each expectation were drawn from. We are not given this information but we can interpret it as missing data and use the Expectation-Maximization algorithm to estimate it jointly with the model parameters.

More precisely, let us associate every observation y_i with an index z_i indicating from which mixture it is drawn from. It can then be shown [45] that the probability mass function $p_Z(z)$ of z_i is found by Bayes' rule given some initial estimate $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}]^T$,

$$p_Z(z|y_i; \hat{\boldsymbol{\theta}}) = \frac{\hat{\alpha}_z f_{\mathrm{Er}}(y_i; \hat{\mu}_z)}{\sum_{k=1}^{M} \hat{\alpha_k} f_{\mathrm{Er}}(y_i|\hat{\mu_k})}. \tag{2.14}$$

Given the above probability mass function (PMF) we can evaluate the expected value of (2.13) and maximize this function with respect to the unknown parameters,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_z \left\{ \sum_{i=1}^{N} \log \left[\alpha_{Z_i} f_{\mathrm{Er}}(y_i; \mu_{Z_i}) \right] \right\}. \tag{2.15}$$

It is shown in [45] that the maximization can be carried out in closed-form yielding the estimators

$$\hat{\alpha}_i = \frac{1}{N} \sum_{k=1}^{N} p_Z(i|y_k; \hat{\boldsymbol{\theta}}), \tag{2.16}$$

$$\hat{\mu}_i = \frac{l_i \sum_{k=1}^{N} p_Z(i|y_k; \hat{\boldsymbol{\theta}})}{\sum_{k=1}^{N} p_Z(i|y_k; \hat{\boldsymbol{\theta}}) y_k}. \tag{2.17}$$

The iterative procedure for finding the unknown parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ can therefore be summarized at follows [45]. First, start with an initial estimate $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\mu}]^T$ and compute the PMF p_Z of the unobserved data z_i using (2.14). Then, maximize the log-likelihood function averaged over p_Z by using the closed form expressions (2.16) and (2.17). This yields a new estimate for $\boldsymbol{\theta}$ and by applying this procedure iteratively, convergence of the estimates is reached.

The number of mixture components M as well as the integer-valued shape parameter $l_i,\ i = 1, \ldots, M$ for each of the Erlang distributions remain to be

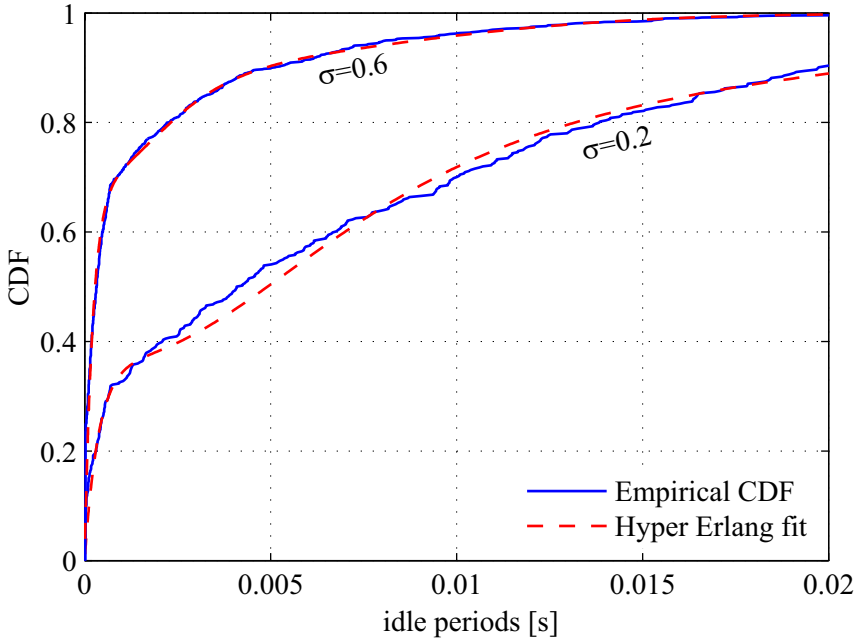


Figure 2.12: Empirical cumulative distribution functions and their hyper-Erlang based fits. The curves correspond to a WLAN traffic load of $\sigma = 0.2$ and $\sigma = 0.6$, respectively.

specified. We have explored different values and obtained the best goodness-of-fit by using $l_1 = 1$, $l_2 = 3$, and $l_3 = 6$ for $M = 3$ mixture components. The fitted distribution is shown in Figure 2.12 and approximates the empirical data well.

2.3.3 Continuous-Time Markov Chain

The semi-Markov model introduced in the previous section establishes a good fit with the empirical data. However, modeling idle and busy periods based on the distributions discussed in the previous sections complicates the derivation of cognitive medium access protocols. To improve analytical tractability, we therefore consider modeling idle and busy periods by a continuous-time Markov chain. This leads to an exponential approximation of the holding times in TRANSMIT and

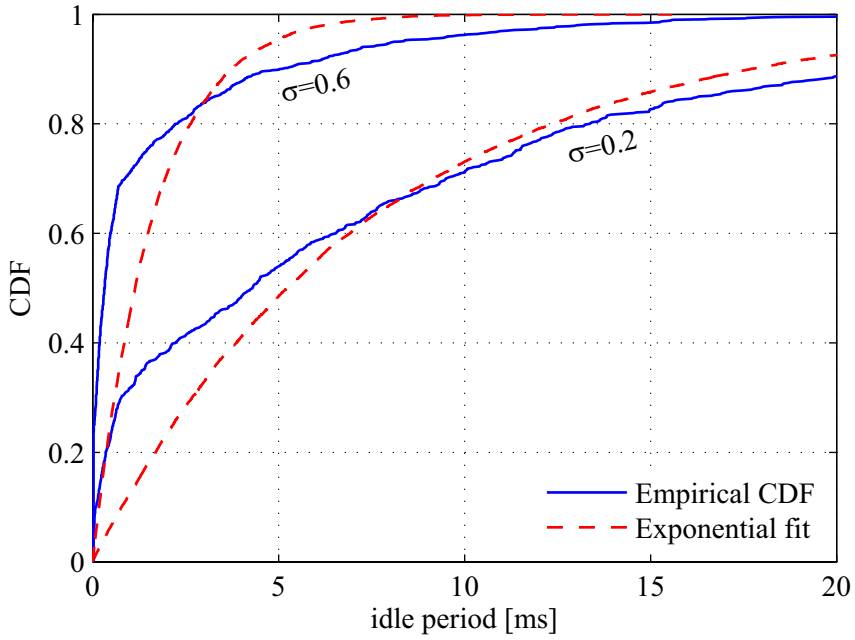


Figure 2.13: Empirical cumulative distribution functions and their exponential fits. The curves correspond to a WLAN traffic load of $\sigma = 0.2$ and $\sigma = 0.6$, respectively.

IDLE state.

The parameters of the exponential distribution are denoted λ for the IDLE state and μ for the TRANSMIT state. Note that we have previously assumed that the holding time in the TRANSMIT state is deterministic. However, a CTMC formulation requires us to fit exponential distributions and we therefore need to choose the parameter μ such that the mean of the distribution equals the deterministic busy duration.

The exponential approximation is compared with the empirical distribution in Figure 2.13. As expected, we observe a significantly larger deviation due to the simplicity of the model.

2.4 Statistical Validation

In the last section, the performance of several fitting approaches has been compared by simply plotting them against the empirical CDF. A quantitative analysis of the goodness-of-fit can be based on the Kolmogorov-Smirnov test, which discerns whether or not N independent observations are drawn from a given distribution $F(t)$ [10]:

$$\mathcal{H}_0 \quad : \quad Y_i \sim F(t),\ i = 1, \ldots, N \tag{2.18}$$

$$\mathcal{H}_1 \quad : \quad Y_i \nsim F(t),\ i = 1, \ldots, N. \tag{2.19}$$

To analyze the above hypothesis testing problem, it is useful to introduce the so-called Kolmogorov-Smirnov statistic

$$D = \max_t |F_e(t) - F(t)|, \tag{2.20}$$

where $F_e(t)$ is the empirical distribution constructed from N independent observations, as defined in (2.7). This reflects the maximum aberration of the empirical CDF from the distribution $F(t)$ and allows for a quantitative assessment of the goodness-of-fit. Nevertheless, normalization by the number of observations is necessary because as N increases, we would expect $F_e(t)$ to approximate $F(t)$ better. Consequently, we define the p-value by

$$p = \Pr(D \geq d|\mathcal{H}_0), \tag{2.21}$$

where d denotes the realization of D constructed from the data. It turns out that the p-value is independent of the distribution $F(t)$ [10] and therefore (2.21) can easily be evaluated by Monte-Carlo simulation or by using an appropriate table [47, 42]. A value of $p \approx 0.1$ is typically deemed high enough to consider the observations coming from $F(t)$.

Table 2.1: Prediction model parameters for semi-Markov model (SMM) and continuous-time Markov chain (CTMC).

		WLAN Traffic Load										
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
		Prediction Model Parameters										
Mixture	k	0.03	-0.07	-0.02	-0.01	0.01	0.00	0.13	0.00	0.15	0.07	0.19
	ω [ms]	21.8	15.1	10.2	8.14	5.48	4.81	3.43	3.63	2.59	2.42	2.17
	p_c	0.18	0.17	0.19	0.26	0.32	0.43	0.50	0.66	0.73	0.81	0.92
	μ [ms]	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Hyper-Erlang	α_1	0.32	0.22	0.33	0.41	0.52	0.62	0.65	0.71	0.79	0.96	0.93
	α_2	0.40	0.62	0.50	0.49	0.26	0.20	0.16	0.29	0.13	0.00	0.03
	α_3	0.28	0.16	0.17	0.10	0.22	0.17	0.18	0.00	0.08	0.04	0.04
	$1/\lambda_1$[ms]	0.70	1.00	0.80	0.67	0.61	0.57	0.53	0.46	0.45	0.26	0.25
	$1/\lambda_2$[ms]	3.54	3.90	4.96	3.54	3.58	2.83	2.71	1.56	2.39	1.61	0.45
	$1/\lambda_3$[ms]	8.35	5.44	0.80	0.79	0.61	0.62	0.48	1.02	0.35	0.74	0.95
CTMC	λ [ms]	23.3	11.6	7.89	5.42	3.32	2.34	1.63	1.01	0.68	0.43	0.24
	μ [ms]	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

By applying the Kolmogorov-Smirnov test, we obtain the p- and d-values shown in Table 2.2. We see that for traffic loads below $\sigma = 0.7$ (which are of primary interest to cognitive coexistence applications) the semi-Markov model with both mixture and hyper-Erlang fit is validated through the test, whereas the CTMC model is rejected. Above $\sigma = 0.7$ the contention window begins to deviate from a uniform distribution leading to a reduced goodness-of-fit.

2.5 Tracking Non-Stationary Traffic

In previous sections we have focused primarily on stationary traffic scenarios, in which the statistics of the idle and busy periods remained unaltered. While this has enabled us to find accurate characterizations by recording long packet traces, traffic is non-stationary in practice and its parameters change over time. To address this variability, we show how non-stationary traffic can be tracked by adapting the parameters of our statistical model to maintain an accurate representation.

Table 2.2: Kolmogorov-Smirnov test parameters for semi-Markov model (SMM) and continuous-time Markov chain (CTMC).

		WLAN Traffic Load					
		0.05	0.1	0.2	0.3	0.4	0.5
Mixture	d-value	0.0521	0.0442	0.0346	0.0540	0.0400	0.0784
	p-value	0.6385	0.8221	0.9678	0.5928	0.9001	0.1637
Hyper-Erlang	d-value	0.0354	0.0452	0.0306	0.0288	0.0664	0.0535
	p-value	0.9602	0.8002	0.9911	0.3310	0.6042	
CTMC	d-value	0.2040	0.0925	0.1456	0.1869	0.2477	0.2573
	p-value	$8.6{\cdot}10^{-8}$	0.0614	$3.5{\cdot}10^{-4}$	$1.3 \cdot 10^{-6}$	$2.8 \cdot 10^{-11}$	$3.9{\cdot}10^{-12}$
		WLAN Traffic Load					
		0.6	0.7	0.8	0.9	1.0	
Mixture	d-value	0.0852	0.1028	0.1336	0.2823	0.3454	
	p-value	0.1042	0.0271	0.0014	$1.6{\cdot}10^{-14}$	$1.6{\cdot}10^{-21}$	
Hyper-Erlang	d-value	0.0662	0.0728	0.1233	0.1942	0.1764	
	p-value	0.3339	0.2312	0.0041	$4.3{\cdot}10^{-7}$	$6.3{\cdot}10^{-6}$	
CTMC	d-value	0.2526	0.2881	0.3382	0.2782	0.3016	
	p-value	$1.03 \cdot 10^{-11}$	$4.1{\cdot}10^{-15}$	$1.2{\cdot}10^{-20}$	$4.1{\cdot}10^{-14}$	$1.6{\cdot}10^{-16}$	

As illustrated by our empirical results, the time scale at which we model spectrum opportunities is in the range of tens of milliseconds. This is significantly shorter than typical traffic variations and motivates to track changes based on a "sliding window" approach in which we estimate model parameters from samples within a window of recent observations and re-estimate parameters continuously to track the non-stationary behavior. This is conceptually similar to block-fading models in wireless communications in which we assume a constant channel state during short blocks and track the fading by re-estimating the channel.

We assessed the performance of this tracking approach based on observations from the HTTP traffic generator "GenSyn" [48]. This traffic generator accesses a list of specified web files in a way that mimics the behavior of an actual user. The resulting non-stationary behavior matches the characteristics that are typically observed in HTTP traffic.

Based on an approximately six second long trace of such traffic, we followed the above sliding window approach; see Figure 2.14. The parameters of the semi-

Markov model with both mixture and hyper Erlang fit were evaluated together with the simple CTMC model based on a window of the last 70 idle periods. For each realization, the Kolmogorov-Smirnov statistic was evaluated and is plotted in Figure 2.14. We also show the acceptance threshold of the Kolmogorov-Smirnov test for a significance level of 0.1. The figure shows that the semi-Markov model with both mixture and hyper-Erlang fit is able to capture the statistics well. While the exponential distribution is rejected by the Kolmogorov-Smirnov test we observe that it is fairly close to the threshold and may serve as a computationally tractable approximation.

In tracking non-stationary traffic it is important to choose the length of the sliding window adequately. In choosing this parameter, there is a tradeoff between having too few samples to accurately fit the models and choosing it too large such that traffic variations start to impact the tracking process.

2.6 Summary

In conclusion we have proposed stochastic models that can be used to predict the idle periods between the bursty transmissions of a WLAN. Our contribution involves both a measurement-based component as well as the statistical analysis of the data.

Since the measurements are based on raw data gathered by a vector signal analyzer we were able to validate the setup and guarantee accurate results using both antenna-based and RF-isolated setups. The statistical analysis focused on a simple two-state ON/OFF model in continuous-time. We developed a semi-Markov model which approximates the idle periods based on either a mixture or

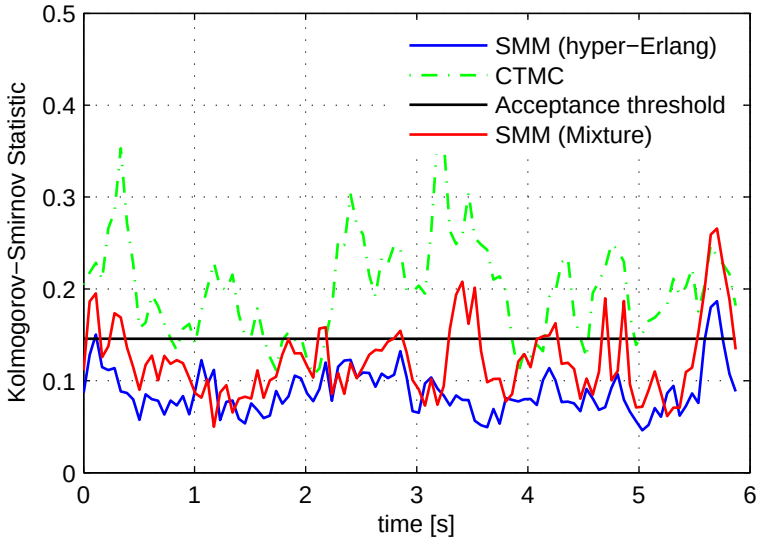


Figure 2.14: Tracking non-stationary traffic through a sliding-window approach. The Kolmogorov-Smirnov test statistics shows that by adjusting model parameters, the traffic variations can be tracked satisfactorily using a semi-Markov model with hyper-Erlang and mixture-based fits. Despite the worse fit of the exponential approximation, the continuous-time Markov chain (CTMC) facilitates the derivation of cognitive coexistence methods.

a hyper-Erlang distribution. Both models accurately characterize the empirical data as verified by the Kolmogorov-Smirnov test. A more tractable CTMC model was also introduced. While it does not approximate the data as well as the other models it has the benefit of being analytically better tractable when it comes to deriving cognitive medium access schemes.

CHAPTER 3

COGNITIVE FREQUENCY HOPPING PROTOCOL

3.1 Summary of Contributions and Related Work

The prediction framework introduced in the previous chapter is now applied to derive a cognitive frequency hopping protocol, which makes optimal use of spectrum opportunities among a set of parallel ad-hoc transmission bands. Based on periodically detecting the activity of the ad-hoc bands, we design the hopping pattern such as to maximize the cognitive radio's throughput while adhering to interference constraints with respect to the ad-hoc bands.

This problem directly relates to coexistence setups in which WLANs are being interfered with by other close-by wireless systems. In order to focus on the design of the optimal hopping pattern we simplify propagation conditions and assume a worst-case collision model in which any packet overlap between ad-hoc and cognitive radio systems leads to a packet loss. In practice this assumption approximates scenarios in which all devices are located in close proximity and there are no hidden terminals.

The design of the optimal hopping pattern is based on a slotted time structure for the cognitive radio. By detecting the ad-hoc bands' activity at the beginning of every slot, and by using the prediction framework introduced in the previous section, we can compute the collision probabilities in each of the bands and derive the optimal hopping pattern accordingly.

Analytical tractability requires us to use the CTMC prediction model as opposed to the more accurate SMM formulation. Specifically, the memoryless prop-

erty of the exponential distribution greatly simplifies the derivation because optimal decisions can be made based on the current sensing result only. Within the SMM framework, on the other hand, potentially the entire sensing history needs to be incorporated into the optimal decision making process.

3.1.1 Main Contribution

The main contribution of this chapter is the derivation of a cognitive frequency hopping (CFH) protocol, which maximizes the throughput of the cognitive radio system subject to interference constraints. We propose two different interference metrics depending on whether interference is normalized by the traffic intensity of the ad-hoc system.

The problem is solved through a decision-theoretic analysis by recasting it as a constrained Markov decision process and linear programming is used to find the optimal hopping protocol based on a standard solution approach. In addition we show that the optimal solutions possess a special structure which provides further insight into the optimal medium access.

We then extend our results to the partially observable case in which the activity of all ad-hoc bands cannot be observed simultaneously; instead we need to select which bands to sense and transmit on. The analysis of this scenario is fundamentally more challenging, because a tradeoff between exploring spectrum opportunities (by frequently switching between bands) and exploiting them (by staying in a certain band and transmitting) needs to be struck. We do not present a solution to the general case, but provide preliminary results for the case of two ad-hoc bands.

Numerical performance analysis of our algorithms is based system-level simulations for the temporal activity models discussed in Chapter 2. We examine the performance under the CTMC model, and investigate its robustness by running the algorithms on data generated from the semi-Markov model. A comparison with a blind reference scheme, which does not perform any sensing, shows a significant performance improvement by a factor of 3.5–4.5.

3.1.2 Related Work and Organization

Temporal DSA has received increasing interest in recent years. For a survey of existing architectures see, *e.g.*, [5, 6]. Among the first to address this problem, Zhao *et al.* analyze the problem of accessing slots that are left idle by primary users [49]. The optimal medium access is derived within a Markovian decision-theoretic framework and a separation principle between sensing and medium access is shown [50]. Similar setups have received increasing interest. The modeling of temporal white space in the framework of machine learning has been addressed in [51]. An experimental test bed which heuristically accesses white space in WLAN is presented in [52]. Other contributions in this area include [53, 54].

Coexistence between wireless local and personal area networks such as Bluetooth has conceptual similarities with our contribution because of the similar physical-layer setup. Different coexistence methods have been considered within the Bluetooth/WLAN coexistence framework and range from interference cancelation at the physical layer [55] to changes in MAC layer scheduling at the WLAN stations [56] and adaptive frequency hopping [57, 58].

Adaptive frequency hopping techniques are most closely related to CFH: both

schemes adapt the hopping sequence to reduce interference. The major difference lies in how interference is detected and modeled. Adaptive frequency hopping typically classifies channels as being either "good" or "bad" according to the empirical error rates of its own transmission attempts. However, WLAN medium access is not modeled explicitly and no spectrum sensing is performed. Bad channels are simply avoided by reducing the hopping set to good channels, if possible. Naturally, this approach is well suited to suppress interference that is static or slowly time varying with respect to the Bluetooth slot length. In many cases, however, WLAN packets are only slightly longer than the slot length reducing the benefit of this modeling technique. In contrast, CFH's sensing and prediction framework is well suited to account for the time variant behavior of WLAN traffic.

The remainder of this chapter is organized as follows. In Section 3.2 the system setup is introduced and measurement-based interference models are presented in Section 3.3. The optimal CFH behavior is developed in Section 3.4 for the fully observable case and results for the partially observed case are given in Section 3.5. Numerical results are presented in Section 3.6.

3.2 System Setup and Problem Formulation

The coexistence setup considered in this chapter consists of M parallel, independently evolving ad-hoc bands as shown in Figure 3.1. The cognitive radio can transmit in any of these frequency bands and dynamically hops among them based on sensing and statistical prediction. Typically, the number of ad-hoc bands M will be quite small, for example practical WLAN setups in the ISM band support $M = 3$ non-overlapping bands.

3.2.1 Physical Layer Setup

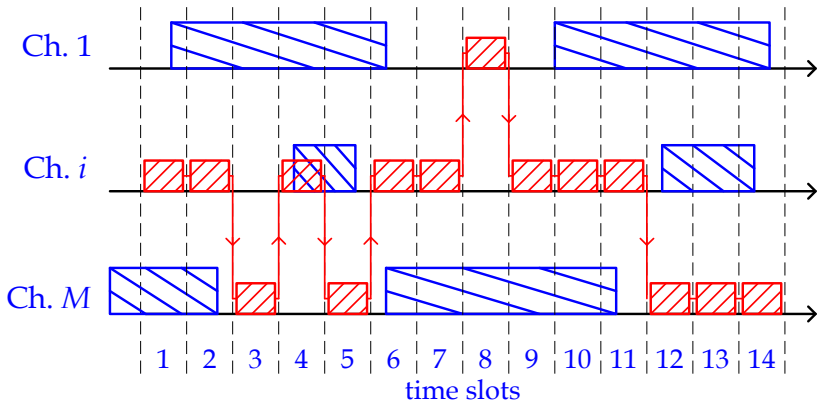
Although there are no restrictions on designing the cognitive radio, the choice of a time-slotted setup arises naturally. Since the ad-hoc bands evolve in an unslotted fashion based on CSMA/CA, a logical approach toward enforcing an interference constraint is to sense the medium periodically and transmit in a slot if it will likely remain idle according to the temporal prediction framework.

This setup is conceptually similar to the coexistence of Bluetooth and WLAN systems shown in Figure 3.1(b). Bluetooth performs frequency hopping in narrowband channels with 1 MHz of bandwidth. Therefore $N = 22$ Bluetooth channels overlap with a single WLAN band. The CFH protocol derived in this chapter could be directly applied for finding the optimal hopping behavior among the M ad-hoc bands. The hopping within the narrowband channels that overlap with the same ad-hoc band can, for example, be performed pseudo-randomly.

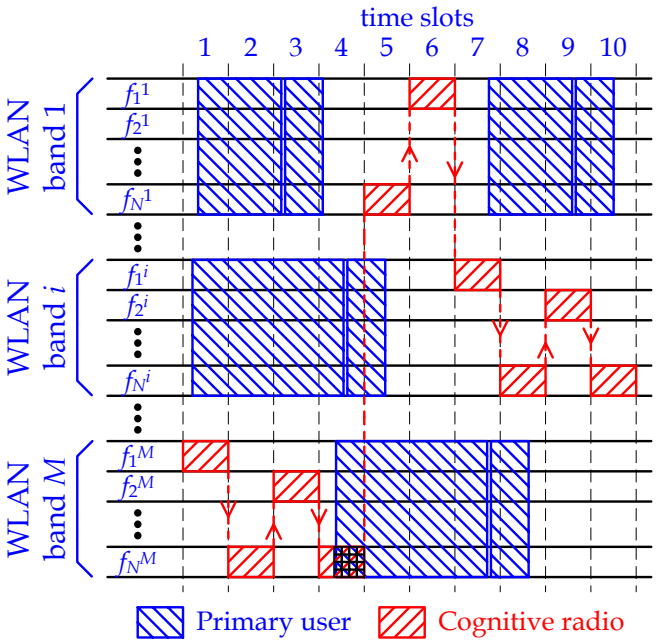
3.2.2 Cognitive Radio Operation

The operation of the cognitive radio is illustrated in Figure 3.2. At the beginning of every slot a spectrum sensor detects whether the medium is busy, either based on energy detection or by exploiting features of the WLAN standard. The binary sensing result (idle or busy) is processed by the CFH controller to determine whether it is safe to transmit, and if yes, in which band. Based on this control action, the secondary transmitter is tuned accordingly and a transmission may be initiated.

The following operations are therefore performed on a slot level: at the begin-



(a) Coexistence setup



(b) Bluetooth/WLAN coexistence

Figure 3.1: Cognitive frequency hopping setup. The time-slotted, frequency-hopping cognitive radio hops among M ad-hoc bands such as to maximize its throughput while adhering to interference constraints. As shown in Figure 3.1(b) this setup has a direct application in Bluetooth/WLAN coexistence.
Figure 3.1(b) ©2008 IEEE. Reprinted, with permission, from [19].

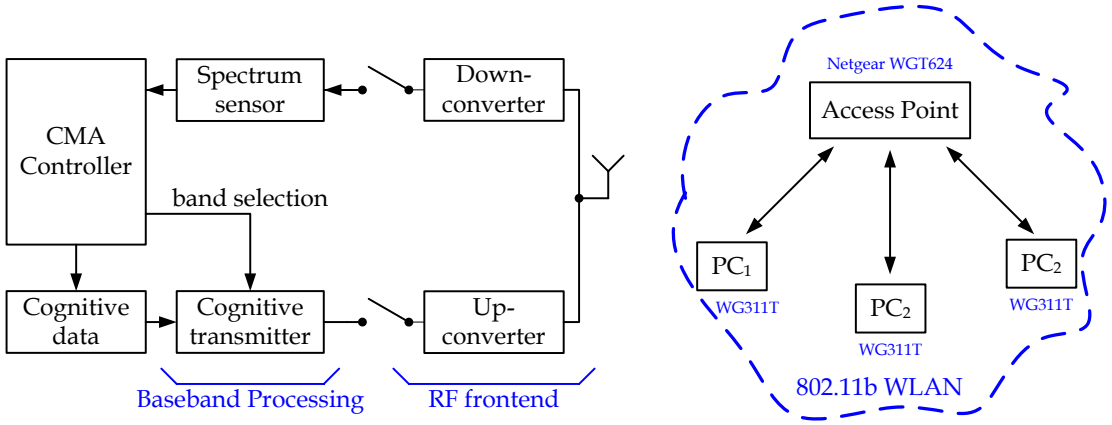


Figure 3.2: Block diagram of the cognitive frequency hopping protocol's operation. The cognitive radio is shown on the left and coexists with the WLAN on the right. ©2008 IEEE. Reprinted, with permission, from [19].

ning of every slot the medium is sensed, the CFH controller decides which band, if any, to transmit across, the transmitter is retuned accordingly, and a transmission may take place for the remainder of the slot period.

The sensing time, the run-time of the controller, and the time it takes to retune the transmitter contribute to the overhead of the system. Based on our previous results on energy detection of WLAN signals (see Chapter 2) a detection error probability of 10^{-5} can be achieved at 5 dB SNR in less than 1 µs. The sensing overhead at the beginning of every slot is therefore quite small. The CFH controller implements a randomized control policy (described in detail in this chapter), which can be implemented through biased coin flips. The delay associated with it is small as well. Finally, current technology yields a frequency retuning time on the order of 100 µs [59], which dominates the processing overhead. In our numerical evaluations we choose a slot size of $T_s = 625$ µs, which is the same as in Bluetooth [60]. We believe this is a practical choice given the similarities in the physical layer setup.

Throughout this chapter, we assume that the cognitive radio system is synchronized. In practice, maintaining identical hop sequences may be challenging,

as sensing results obtained by different nodes could potentially differ. While it goes beyond the scope of this paper to address this issue in detail, we believe that collaborative sensing techniques [61] can be used to provide hop sequence coordination. By exchanging sensing metrics across subsequent slots, terminals could perform sensing jointly and arrive at identical results. Another potential approach for maintaining synchronization is to employ acknowledgement feedback [20].

3.3 Measurement-Based Interference Model

CFH is a protocol that dynamically hops across multiple parallel ad-hoc bands in an optimal fashion. It is fundamentally based on the measurement-based prediction model introduced in Chapter 2. In addition, however, we need physical-layer coexistence models to characterize the interaction between both systems in the event of a collision. Our assumptions, corroborated by actual measurements, are presented in the following. In particular, we first evaluate whether the cognitive radio affects the ad-hoc systems' carrier-sensing. Second, we obtain empirical results for the probability that a collision between both systems leads to a packet error.

3.3.1 Impact on Carrier Sensing

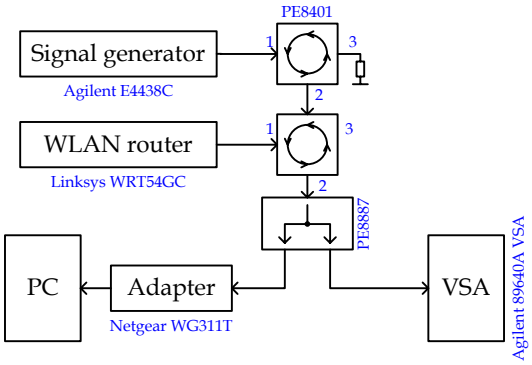
The design of the cognitive radio needs to ensure that its transmissions remain transparent to the ad-hoc system. This implies that the ad-hoc systems' carrier sensing must not be impacted by the cognitive radio's transmissions. Otherwise the hierarchical access structure would be undermined and the cognitive radio's

dynamic effect on the ad-hoc bands would render our prediction model useless, unless it incorporated the ad-hoc bands' retransmission behavior.

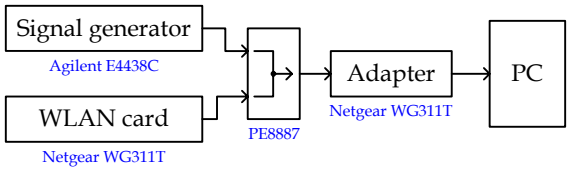
We evaluate this effect by measuring the probability that the ad-hoc system defers medium access to the cognitive radio's transmissions. The experimental setup is shown in Figure 3.3(a). It consists of an IEEE 802.11b router for generating a WLAN signal and an RF signal source for generating the cognitive radio signal (a static narrowband signal with Bluetooth's modulation parameters [60] is used). As the signal remains static in one of the channels it is possible to examine the mutual interference resulting from a specific channel. The WLAN adapter card and the signal source are connected via circulators, which couple generator and router while providing isolation in the reverse direction. Another WLAN adapter card is used to capture the received signal and a vector signal analyzer is used to verify the correct operation of the setup.

The impact of the cognitive radio transmissions is assessed by measuring the effect on the WLAN packet transmission rate. The WLAN router continuously transmits packets, and the WLAN adapter card is used to measure its average rate (by capturing packets over long periods of time). In the presence of the cognitive system, if the WLAN detected the interference, its rate would decrease as back-off periods would need to be accommodated.

Our results show that the ad-hoc bands' carrier sensing is generally not affected by the cognitive radio. The reason for this insensitivity is due to WLAN specifications which only mandate that WLAN signals are detected with a specified level of sensitivity; other signals need not be detected [39]. The standard outlines three methods to accomplish the carrier sensing, which includes the detection of the WLAN spreading sequence. While the implementation of the carrier sensing is



(a) Impact on carrier sensing



(b) Impact on packet error rate

Figure 3.3: Experimental setup for evaluating the cognitive radio's impact on the WLAN's carrier sensing and packet error rate.

vendor-specific, the above suggests that the WLAN carrier sensing will be quite insensitive to the cognitive radio transmissions. Our measurement results confirmed this conjecture and led to our assumption that the WLAN carrier sensing will not be impacted by the cognitive radio transmissions.

Furthermore, we validate our measurement approach by performing the same analysis for a different type of interfering signal, namely a WLAN-type signal using the *same* spreading code as standardized for IEEE 802.11b [39]. We observe that the adapter card is significantly more sensitive to this type of signal. The power level above which an impact occurs is determined to be approximately $-77\,\mathrm{dBm}$. This is in accordance with the 802.11b standard [39, p.58] which mandates the sensitivity to be $-76\,\mathrm{dBm}$ or better.

3.3.2 Effect on Packet Error Rate

The second component of our interference model focuses on the cognitive radio's impact on the ad-hoc bands' packet error rate. Specifically, we measure the probability that a collision between both systems leads to a packet error. The measurement setup is shown in Figure 3.3(b). It consists of a WLAN adapter card and a signal source generating a WLAN and the cognitive radio signal, respectively. The signals are combined and captured by another WLAN card running commercial packet capturing software. A vector signal analyzer is used to verify the operation of the setup.

The packet error probability is measured in the following way. A continuous stream of packets is generated and captured at the receiver to determine the packet rate in the absence of interference. Subsequently, after turning on the interferer, the rate decreases since some packets will be too distorted to be captured by the receiver. Other packets will be captured but show an invalid redundancy check. By comparing the number of successfully received packets with the interference-free case, the packet error probability can be determined.

The impact of the cognitive radio's interference depends on the channel index. Close to the center frequency a significant impact is observed for signal-to-interference ratios (SIRs) of less than 0 dB. For instance, for an offset of 3 MHz from center frequency and an SIR of -3 dB, we observe a packet error rate of 85%. If the SIR drops below -5 dB virtually every packet is lost. The impact of the interference decreases as we move away from the center frequency. This is not surprising and has previously been reported [62]. It is due to the down-conversion and filtering performed by typical WLAN receivers.

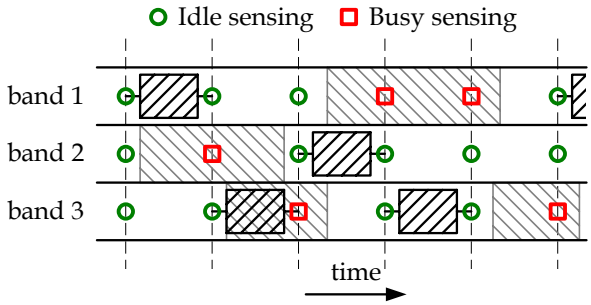
3.4 Designing Optimal Hopping Patterns: Full Observability

The cognitive access schemes presented in this paper can be categorized according to the sensing capabilities of the cognitive radio. If the spectrum sensor supports a high enough bandwidth, the state of all M channels can be observed simultaneously at the beginning of every slot. This is illustrated in Figure 3.4(a). Based on the sensing result, the cognitive controller decides in which, if any, channel to transmit.

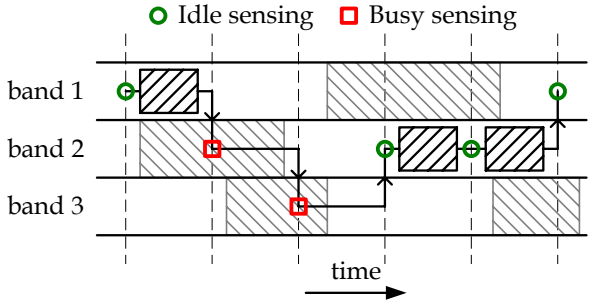
If the spectrum sensor has limited bandwidth, we assume that only one of the M bands can be sensed at a time; the state of the other bands remains hidden. Furthermore, we assume that for practical reasons, a transmission can only be initiated in the channel that has just been sensed. This assumption makes the system *partially observable* and significantly complicates the analysis. This scenario is illustrated in Figure 3.4(b).

We first analyze the fully observable case and recast the problem mathematically as a constrained Markov decision process. The standard solution technique based on linear programming [63] is briefly reviewed and we show that our problem setup admits a structured solution. The partially observable case is then addressed the subsequent section.

In order to find the optimal access strategy we need to formulate a constrained optimization problem [18]. Let each of the ad-hoc bands $i = 1, \ldots, M$ evolve as a CTMC $\{X_i(t), t \geq 0\}$ where $X_i(t) = 0$ ($X_i(t) = 1$) denotes that band i is in the idle (busy) state. The holding times for ad-hoc band i are exponentially distributed with parameters λ_i in the idle and μ_i in the busy state, respectively.



(a) Fully observable scenario



(b) Partially observable scenario

Figure 3.4: Illustration of fully and partially observable cognitive frequency hopping. Circles and squares denote idle and busy sensing results, respectively. ©2008 IEEE. Reprinted, with permission, from [19].

The generator matrix $\mathbf{G}_i$ for channel i is hence given by

$$\mathbf{G}_i = \begin{bmatrix} -\lambda_i & \lambda_i \\ \mu_i & -\mu_i \end{bmatrix}, \tag{3.1}$$

which leads to the stationary distribution

$$\eta_0^{(i)} = \frac{\mu_i}{\lambda_i + \mu_i}, \quad \eta_1^{(i)} = \frac{\lambda_i}{\lambda_i + \mu_i} \tag{3.2}$$

and transition matrix [64, p.391]

$$\mathbf{P}^{(i)} = \frac{1}{\lambda_i + \mu_i} \begin{bmatrix} \mu_i + \lambda_i e^{-(\lambda_i+\mu_i)t} & \lambda_i - \lambda_i e^{-(\lambda_i+\mu_i)t} \\ \mu_i - \mu_i e^{-(\lambda_i+\mu_i)t} & \lambda_i + \mu_i e^{-(\lambda_i+\mu_i)t} \end{bmatrix}. \tag{3.3}$$

The cognitive radio system senses the state of all ad-hoc bands at the beginning of every slot, inducing discrete-time Markov chains $\{Y_i[k], k \geq 0\}$ of sensing results

for each channel i. For notational convenience let us define the vector-valued random process $\{\mathbf{Y}[k], k \geq 0\}$ that contains the latest sensing result of each ad-hoc band,

$$\mathbf{Y}[k] = \big[Y_1[k], \ldots, Y_M[k]\big]^T. \tag{3.4}$$

It is straightforward to verify that $\mathbf{Y}[k]$ is a discrete-time Markov chain with state space $\mathbb{X} = \{0,1\}^M$. The transition matrix becomes, due to the independence of the ad-hoc bands,

$$\mathbf{P}_{\mathbf{xy}} = \prod_{i=1}^{M} \mathbf{P}^{(i)}_{x_i y_i}, \ \mathbf{x}, \mathbf{y} \in \mathbb{X} \tag{3.5}$$

and we arrive at the following expression for the stationary distribution

$$\eta_{\mathbf{x}} = \prod_{i=1}^{M} \eta^{(i)}_{x_i}. \tag{3.6}$$

Given the sensing results in each slot, the CFH controller determines in which channel (if any) to transmit. The action set is thus $\mathbb{A} = \{0, 1, \ldots, M\}$, where $a = 0$ denotes that no transmission takes place, and $a \geq 1$ represents a transmission in ad-hoc band a.

Transmitting in ad-hoc band a accrues a unit reward provided that no collision occurs. The expected immediate reward of choosing action a in state $\mathbf{y}$ thus becomes

$$r(\mathbf{y}, a) = \begin{cases} \mathbf{1}_{[y_a=0]} e^{-\lambda_a T_s}, & a \geq 1 \\ 0, & a = 0 \end{cases}, \tag{3.7}$$

where $\mathbf{1}_{[\cdot]}$ represents the indicator function and T_s denotes the slot duration.

The interference constraint can be formulated in several ways. We shall call slot k in band i busy, and denote this as $\mathcal{C}_i[k] = 1$, if

$$\exists t \in [kT_s, (k+1)T_s) \text{ s.t. } X_i(t) \neq 0. \tag{3.8}$$

The interference constraint can then be formulated as

$$D_c = \lim_{N\to\infty} \frac{\sum_{k=1}^{N} \mathbf{1}_{[A_k=i,\mathcal{C}_i[k]=1]}}{N}, \tag{3.9}$$

where A_k refers to the action taken in slot k. It is capitalized to stress that A_k is random; it depends on the current sensing result and the action (randomly) chosen by the CFH controller. The above equation corresponds to the long run fraction of slot collisions per unit time. We will refer to (3.9) as the *cumulative interference constraint*.

While the cumulative interference constraint appears to be an intuitive measure, it quantifies interference from the cognitive radio's perspective and does not take into account the traffic intensity of the ad-hoc bands. The interference metric can be better tailored to the ad-hoc bands by imposing *packet error rate constraints* for each band individually,

$$D_p^{(i)} = \lim_{N\to\infty} \frac{\sum_{k=1}^{N} \mathbf{1}_{[A_k=i,\mathcal{C}_i[k]=1]}}{N_i(NT_s)},\ 1 \le i \le M, \tag{3.10}$$

where $N_i(t)$ counts the number of transmitted ad-hoc packets in band i up to time t,

$$N_i(t) = \sum_{n=0}^{\infty} \mathbf{1}_{[S_n^{(i)}\le t]}. \tag{3.11}$$

In the above equation, $S_n^{(i)}$ denotes the arrival times of ad-hoc packets in band i and therefore (3.10) is the long-run fraction of collisions per transmitted ad-hoc packets. In short, this approximates the fraction of packets that get dropped due to the cognitive radio's interference.

Based on the above definitions, we define the expected immediate costs for the

cumulative interference metric,

$$d_c(\mathbf{y}, a) = \begin{cases} 1 - e^{-\lambda_a T_s} & \text{if } y_a = 0,\ a \geq 1 \\ 1 & \text{if } y_a = 1,\ a \geq 1 \\ 0 & \text{if } a = 0 \end{cases}, \tag{3.12}$$

and for the packet error rate metric,

$$d_p(\mathbf{y}, a) = \begin{cases} \frac{(\lambda_a + \mu_a)(1 - e^{-\lambda_a T_s})}{\mu_a \lambda_a T_s} & \text{if } y_a = 0,\ a \geq 1 \\ 1 & \text{if } y_a = 1,\ a \geq 1 \\ 0 & \text{if } a = 0 \end{cases}. \tag{3.13}$$

Having introduced rewards and costs, the CMDP can now be formally defined. We aim at finding a protocol which maximizes

$$J(\beta, \pi) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}_{\beta}^{\pi} r(\mathbf{Y}_t, A_t) \tag{3.14}$$

with respect to policy π, subject to a cumulative interference constraint

$$D_c(\beta, \pi) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}_{\beta}^{\pi} d_c(\mathbf{Y}_t, A_t) \leq \alpha, \tag{3.15}$$

or subject to packet error rate constraints,

$$D_p^{(i)}(\beta, \pi) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}_{\beta}^{\pi} \mathbf{1}_{[A_t = i]} d_p(\mathbf{Y}_t, A_t) \leq \alpha_i,\ 1 \leq i \leq M. \tag{3.16}$$

In the above formulas β denotes the initial distribution of the system, and π the policy we maximize for. The expectation operator is thus taken with respect to the probability distribution induced by π given initial distribution β [63].

3.4.1 Linear Programming Solution

It is well-known that a CMDP's optimal policy is Markovian, and therefore only depends on the sensing outcome in the current slot. Randomization needs to

be incorporated, however, due to interference constraints [65]. Consequently, the optimal policy π^* is a function that maps state-action pairs $[\mathbf{y}, a]$ to the probability of choosing action a in state $\mathbf{y}$. Since both reward and constraints can be expressed using the frequency of state-action pairs $[\mathbf{y}, a]$ the optimal policy can be found by linear programming [63]. The following theorem formulates this standard solution approach.

Theorem 1 [63, p.38] The linear program

$$\max_{\rho(\mathbf{y},a)} \sum_{\mathbf{y}\in\mathbb{X}} \sum_{a\in\mathbb{A}(\mathbf{y})} \rho(\mathbf{y}, a) r(\mathbf{y}, a) \tag{3.17}$$

subject to

$$\sum_{\mathbf{y}\in\mathbb{X}} \sum_{a\in\mathbb{A}(\mathbf{y})} \rho(\mathbf{y}, a) d_i(\mathbf{y}, a) \leq \alpha_i,\ 1 \leq i \leq M, \tag{3.18}$$

where $\rho(\mathbf{y}, a) \in \mathbb{Q}(\beta)$ and

$$\mathbb{Q}(\beta) = \left\{ \begin{array}{l} \rho(\mathbf{y}, a), \mathbf{y} \in \mathbb{X}, a \in \mathbb{A}(\mathbf{y}) : \\ \sum_{\mathbf{y}\in\mathbb{X}} \sum_{a\in\mathbb{A}(\mathbf{y})} \rho(\mathbf{y}, a)(\delta_{\mathbf{y}}(\mathbf{x}) - P_{\mathbf{x}a\mathbf{y}}) = 0 \\ \sum_{\mathbf{y}\in\mathbb{X}} \sum_{a\in\mathbb{A}(\mathbf{y})} \rho(\mathbf{y}, a) = 1,\ \rho(\mathbf{y}, a) \geq 0 \end{array} \right\} \tag{3.19}$$

is equivalent to the CMDP formulations (3.14)-(3.16).

The CMDP's optimal policy is completely determined by the state-action frequencies. After obtaining $\rho(\mathbf{y}, a)$ via the above linear program, the probability $w_{\mathbf{y}}(a)$ of choosing action a in state $\mathbf{y}$ simply becomes,

$$w_{\mathbf{y}}(a) = \frac{\rho(\mathbf{y}, a)}{\sum_{a\in\mathbb{A}(\mathbf{y})} \rho(\mathbf{y}, a)},\ \mathbf{y} \in \mathbb{X},\ a \in \mathbb{A}(\mathbf{y}), \tag{3.20}$$

provided the denominator is non-zero (arbitrary otherwise).

3.4.2 Structure of the Optimal Solutions

In this section we show that the above linear programs, under some conditions, admit structured solutions that help gain insight into the problem and may simplify the implementation of CFH.

Cumulative Interference Constraint. First, consider the case of a cumulative interference constraint and, without loss of generality, assume that $\lambda_1 \leq \cdots \leq \lambda_M$. We show that a threshold policy, which only utilizes channels with small λ is optimal. The threshold depends on the interference constraint α and the CTMC parameters of the ad-hoc bands. The solution structure is depicted in Figure 3.5.

Algorithm 1 (Threshold solution)

1. Define the maximum interference level for channel i as

$$\xi_i = \sum_{\mathbf{y}\in\mathbb{X}} 1_{[y_i=0]} 1_{[y_j=1,\forall j<i]} \eta_{\mathbf{y}} d_c(\mathbf{y}, i) \tag{3.21}$$

2. Based on the $\{\xi_i\}$, find the smallest k such that

$$\xi_1 + \cdots + \xi_k > \alpha \tag{3.22}$$

3. Adopt the following randomized policy. With probability w_i, transmit in the idle channel with the lowest λ_i, *i.e.*, in state $\mathbf{y}$ transmit in channel i if

$$y_i = 0 \text{ and } y_j = 1 \ \forall j < i, \tag{3.23}$$

and choose not to transmit otherwise. The probabilities w_i are given based on the k defined in step 2,

$$w_j = 1,\ 1 \leq j < k,\ \ w_k = \frac{\alpha - \xi_{k-1}}{\xi_k - \xi_{k-1}},\ \ w_j = 0,\ j > k \tag{3.24}$$

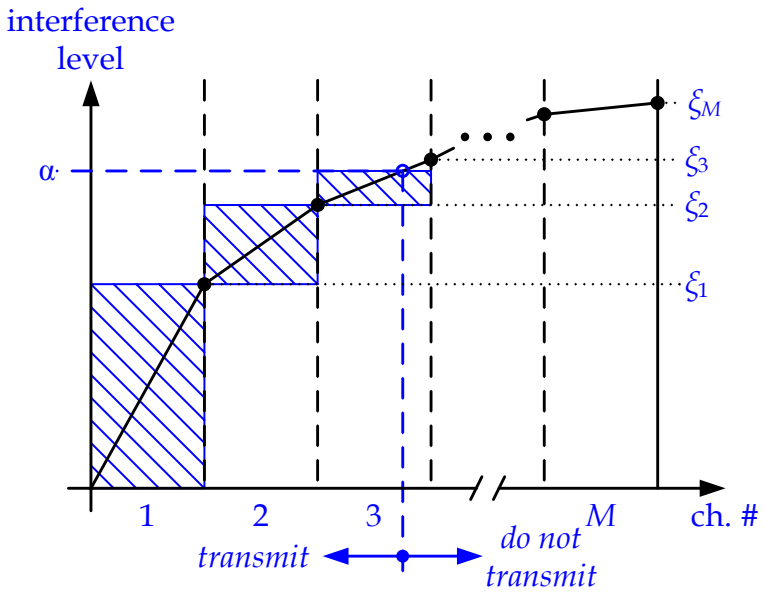


Figure 3.5: Threshold solution structure under the cumulative interference constraint. ©2008 IEEE. Reprinted, with permission, from [19].

Theorem 2 The policy induced by Algorithm 1 is a solution to the linear program (3.17)-(3.18) and therefore corresponds to an optimal policy.

Proof: To show that Algorithm 1 is a solution to (3.17)-(3.18) we first rewrite the linear program in terms of its stationary distribution. This is possible since the transition behavior does not depend on the actions a. We obtain

$$\max_{w_{\mathbf{y}}(a)} \sum_{(\mathbf{y},a)\in\mathbb{X}\times\mathbb{A}} \eta_{\mathbf{y}} w_{\mathbf{y}}(a) r(\mathbf{y},a) \tag{3.25}$$

subject to

$$\sum_{(\mathbf{y},a)\in\mathbb{X}\times\mathbb{A}} \eta_{\mathbf{y}} w_{\mathbf{y}}(a) d_c(\mathbf{y},a) \leq \alpha. \tag{3.26}$$

Due to the assumption $\lambda_1 \leq \cdots \leq \lambda_M$ and the fact that for the cumulative interference constraint, $d_c(\mathbf{y},a) = 1 - r(\mathbf{y},a)$, channel i clearly offers higher (or equal) reward at lower (or equal) cost than channel j if $i < j$. Hence it is optimal to increase the transmission probability in channel 1 until $w_{\mathbf{y}}(1) = 1$ for all $\mathbf{y}$ with $y_1 = 0$. We continue with channels $i = 2, \ldots, k$ until the interference constraint is exceeded. We then choose $w_{\mathbf{y}}(k)$ such that the constraint is satisfied with equal-

ity. We thus have the optimal policy structure outlined in (3.24) and illustrated graphically in Figure 3.5. □

Packet Error Rate Constraints. In the case of packet error rate constraints, there exists a separate interference constraint for each individual channel. Intuitively, the maximum reward would be achieved if all constraints could be made tight (otherwise transmission opportunities would be wasted). It may not be feasible, however, to tighten all constraints, given that a transmission can be initiated in at most one channel per slot. Whether the constraints can be made tight ultimately depends on how loose the packet error rate constraints α_i are chosen.

If the M constraints can be made tight, we show that the problem decouples, and the optimal policy can be found by considering each channel individually. This is the case if the following condition is met for all channels,

$$\xi_a = \sum_{\mathbf{x}\in\mathbb{X}} \frac{1_{[x_a=0]}\eta_{\mathbf{x}}}{\sum_{l=1}^{M} 1_{[x_l=0]}} \geq \frac{\alpha_a}{d_a}, \quad 1 \leq a \leq M \tag{3.27}$$

where

$$d_a = \frac{(\lambda_a + \mu_a)(1 - e^{-\lambda_a T_s})}{\mu_a \lambda_a T_s}. \tag{3.28}$$

is the expected average cost associated with a collision. We can thus adopt the following algorithm.

Algorithm 2

1. The maximum interference level ξ_i in channel i is given by (3.27). Disregarding other transmission opportunities, we transmit in band i with probability

$$w_i = \frac{\alpha_i}{d_i \xi_i}, \quad 1 \leq i \leq M \tag{3.29}$$

2. Since (3.27) is met the constraints can be made tight and we obtain the transmission probabilities

$$w_{\mathbf{y}}(a) = \frac{w_a}{\sum_{l=1}^{M} 1_{[y_l=0]}}. \tag{3.30}$$

The intuition behind (3.27) is to find a condition under which the interference constraints can be made tight by considering channels separately. In fact, a trade-off on which channel to transmit in need only be struck if $\mathbf{x}$ contains multiple zeros. The denominator accordingly normalizes by the number of transmission opportunities in state $\mathbf{x}$ and thus ensures that, although the M bands are considered separately, the probability of transmission in any given state will never exceed one.

Theorem 3 The policy induced by Algorithm 2 is a solution to the linear program (3.17)-(3.19) and therefore corresponds to an optimal policy.

Proof: To show that Algorithm 2 is a solution to (3.17) and (3.19) we again rewrite the linear program in terms of the chain's stationary distribution

$$\max_{w_{\mathbf{y}}(a)} \sum_{(\mathbf{y},a)\in\mathbb{X}\times\mathbb{A}} \eta_{\mathbf{y}} w_{\mathbf{y}}(a) r(\mathbf{y},a) \tag{3.31}$$

subject to

$$\sum_{(\mathbf{y},a)\in\mathbb{X}\times\mathbb{A}} \eta_{\mathbf{y}} w_{\mathbf{y}}(a) d_p(\mathbf{y},a) \leq \alpha_a, \quad 1 \leq a \leq M. \tag{3.32}$$

Clearly, since $r(\mathbf{y},a) \geq 0$, an optimal solution would be obtained ideally by satisfying all packet error rate constraints with equality. In general, this may not be possible, however, since we can only transmit in one of the channels at a time, even if multiple of them are idle. However, a sufficient condition for a structured solution is (3.27). In this case, $\sum_{i=1}^{M} w_{\mathbf{y}}(i) \leq 1$, and the constraint can be made tight for channel i independently of all other channels. □

3.5 Designing Optimal Hopping Patterns: Partial Observability

In the last section we assumed that the state of all M channels can be observed simultaneously. In this section we alleviate this assumption and assume that only a single channel can be observed at a time, as shown in Figure 3.4(b). Furthermore, a transmission may only be initiated in the channel that has just been sensed. The action set therefore reduces to $\mathbb{A} = \{0, 1\}$ denoting whether or not a transmission is initiated.

Partial observability severely complicates the problem because it requires to trade off the exploration of the system (by frequently sensing different bands) with the exploitation of transmission opportunities. In order to illustrate how such a tradeoff can be struck, we present some preliminary results for the special case of $M = 2$ channels.

A fundamental consideration in designing the optimal hopping pattern under partial observability is to determine which of the past observations and actions are useful for making optimal decisions. Given that all bands are modeled as CTMCs, the continuous-time Markov property leads to the conclusion that the latest sensing result of every channel is sufficient for predicting its behavior.

This is illustrated in Figure 3.6 for the special case of $M = 2$ bands. The states are labeled according to whether the current sensing result is busy or idle. A busy channel is simply denoted 'b,' whereas for an idle channel we also keep track of how many consecutive slots the channel has been sensed idle. Hence the states labeled '0,'...,'N' all correspond to an idle sensing result.

The sensing history is incorporated in brackets, where the first number reflects the currently active channel and the second index denotes the latest sensing result in the other channel, respectively. Note that the time since the other channel has last been sensed can be determined directly by the number of slots spent in the current channel, in this special case of $M = 2$.

According to the above, the setup is fully described by the triple $[y, i, x]$ where $y \in \mathbb{Y} = \{b, 0, \ldots, N\}$ denotes the current channel's sensing history, i the currently active channel index, and $x \in \{0, 1\}$ the last sensing result in the other band. The transition behavior can be understood as follows. Under action 1 the system keeps transmitting in the current band and the following transitions are possible

$$(y, i, x) \rightarrow (y + 1, i, x), \quad 0 \leq y < N \tag{3.33}$$

$$(y, i, x) \rightarrow (0, \bar{i}, 1) \tag{3.34}$$

$$(y, i, x) \rightarrow (b, \bar{i}, 1), \tag{3.35}$$

where the first line denotes the channel staying idle. The other two possible transitions correspond to the channel becoming busy and the cognitive radio thus relocating to the other band. This other band can in turn be either idle or busy and thus two different transitions can occur. The notation $\bar{i}$ represents the "other" band, that is $\bar{i} = (i \mod 2) + 1$.

Under action 0, the system switches bands and the following transitions may occur

$$(y, i, x) \rightarrow (0, \bar{i}, 0) \tag{3.36}$$

$$(y, i, x) \rightarrow (b, \bar{i}, 0), \tag{3.37}$$

denoting a relocation to the other band and finding it either idle or busy, respectively. The transition behavior is shown in Figure 3.6. In order to keep a finite

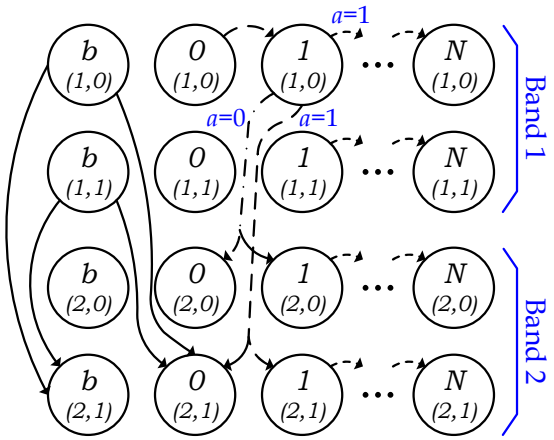


Figure 3.6: State transition model for the partially observable case. Only some example transitions are shown. The indices denote the sensing result, the active channel number, and the last sensing result, respectively. ©2008 IEEE. Reprinted, with permission, from [19].

state space, we assume that we cannot stay with any channel longer than for N slots. The optimal policy for the partially observable case can be found by linear programming, as discussed for the case of full observability.

3.6 Performance Results

In this section we present numerical results for the CFH schemes and evaluate their performance gain compared to a blind reference that does not perform spectrum sensing. The schemes are evaluated in terms of throughput and interference for varying ad-hoc traffic intensity, as it is standard in the coexistence literature [66]. The results are based on simulations using the CTMC approximation. Furthermore, we examine the robustness of this approximation by running algorithms derived from the CTMC model on data generated from the SMM. We will see that the results match closely, justifying the use of the CTMC approximation for deriving CFH.

3.6.1 Simulation Parameters

The numerical results reflect the throughput and interference behavior for varying ad-hoc traffic intensity. While performance is assessed by simulation, the model parameters (for both SMM and CTMC approximation) are taken from the experimental study presented in Chapter 2. The slot duration was chosen as $T_s = 625\,\mu\text{s}$.

3.6.2 Fully Observable Scenario

The throughput performance of CFH under full observability is shown in Figure 3.7 for both cumulative interference and packet error rate constraints, respectively. Both metrics are plotted with respect to the traffic load σ, which indexes the traffic intensity of the ad-hoc system and is normalized such that $\sigma = 0$ corresponds to an inactive primary user, whereas $\sigma = 1$ indicates that the primary system uses the channel to the maximum extent.

The plot shows the fundamental difference between cumulative interference and packet error rate constraints. In the former case, interference is modeled cumulatively and therefore the performance does not depend on the number of parallel channels as long as these have the same CTMC parameters (otherwise, the curves would reflect the threshold solution we presented earlier in this chapter). On the other hand, in the case of packet error rate constraints, there are separate constraints for each channel and therefore the throughput increases proportionally with the number of available channels. By comparing the curves for both cases, we see that for small traffic loads the packet error rate constraints are more restrictive than the cumulative interference constraint because the former conditions on the traffic load of the primary system. The opposite is true at high traffic load for

the same reason. Naturally, in both scenarios, the throughput decreases with the traffic load because primary activity increases and therefore fewer transmission opportunities are available.

In order to put the performance of CFH in perspective, we compare it to a reference scheme which does not perform any spectrum sensing. Instead, in each slot, the reference scheme simply selects one of the M bands with equal probability and initiates a transmission with probability p. Depending on p and the traffic load, the reference scheme achieves some level of throughput and interference. Clearly, as p is increased both throughput and interference increase as well. In order to draw a fair comparison with CFH, in which the transmission probabilities are designed such as to generate no more than a certain level of interference, we design p such that the interference level is the same as the constraint. We can then compare the throughput of CFH directly with the reference method.

The performance results are shown in Figure 3.8 for both cumulative interference (left) and packet error rate constraints (right). In both scenarios, CFH outperforms the reference scheme by a factor of approximately 3.5-4.5, regardless of traffic load σ.

3.6.3 Robustness to CTMC Approximation

The derivation of CFH has fundamentally relied upon modeling the temporal activity of the ad-hoc bands as a CTMC, despite the better fit that can be obtained with an SMM. Unfortunately, finding the optimal access policy based on the latter model is difficult because the memoryless property, which holds for the CTMC framework, no longer applies. Therefore, making optimal access decision may po-

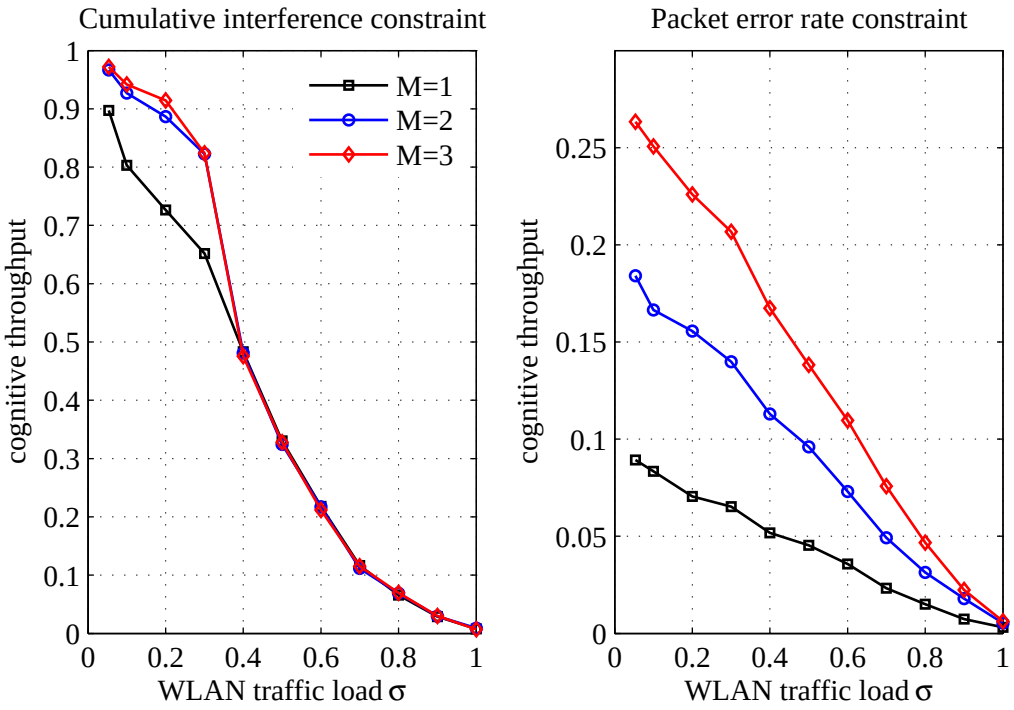


Figure 3.7: Throughput of the cognitive radio system for different numbers of ad-hoc bands. The performance is shown under a cumulative interference constraint of 0.05 (left side) or packet error rate constraints of 0.1 (right side). In both plots, the normalized throughput represents the time-averaged fraction of successful transmissions out of the total number of slots.

tentially involve the entire sensing history and not just the sensing outcome in the current slot.

Since finding the optimal medium access in the non-exponential case appears intractable, we focus on a robustness analysis and run the CFH protocol, derived based upon the CTMC model, on traces simulated according to the SMM framework. This helps to quantify the impact of deviating from exponentially distributed idle periods and enables us to check whether interference constraints remain (approximately) met.

The result in Figure 3.9 shows that the resulting abberation is fairly small. The

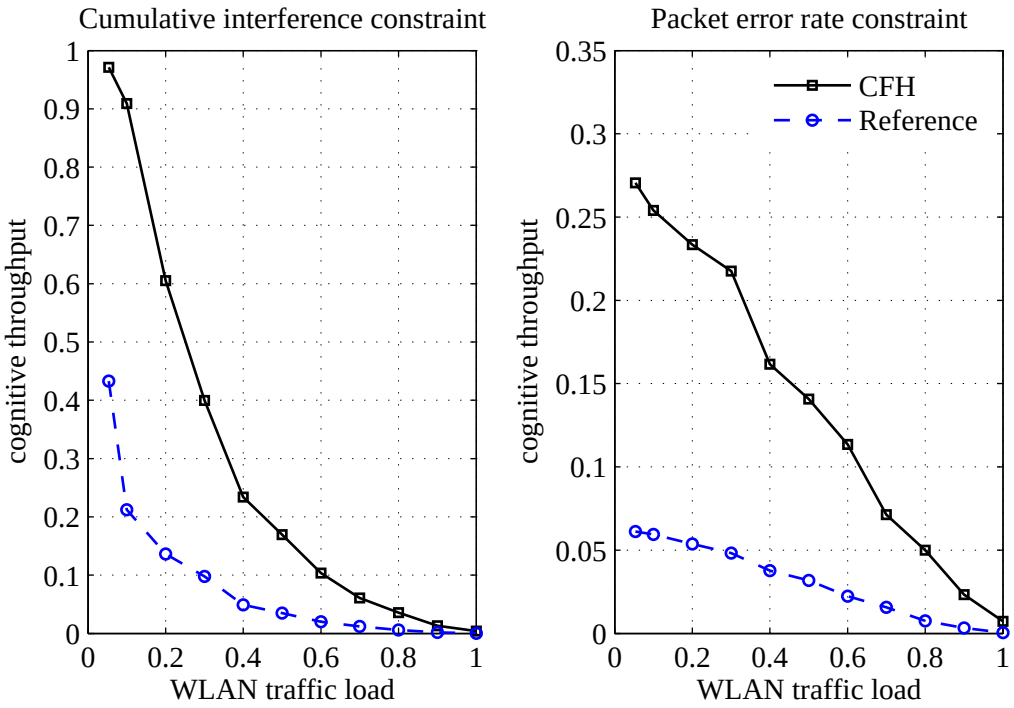


Figure 3.8: Performance of cognitive frequency hopping compared to a blind reference scheme. The blind hopper operates with constant rate and is completely oblivious of the WLAN.

throughput curves show an excellent match and the packet error rate constraint also remains approximately met.

3.7 Summary

In summary, this chapter has presented a cognitive frequency hopping method that ensures coexistence with a set of parallel ad-hoc bands by reusing the idle periods between packet transmissions. Based on approximating the temporal activity as a CTMC, the optimal hopping pattern was found through formulating a constrained Markov decision process. A conventional solution approach via linear programming was discussed and structured solutions could be obtained, which reduce computational complexity. A partially observable scenario, in which a tradeoff

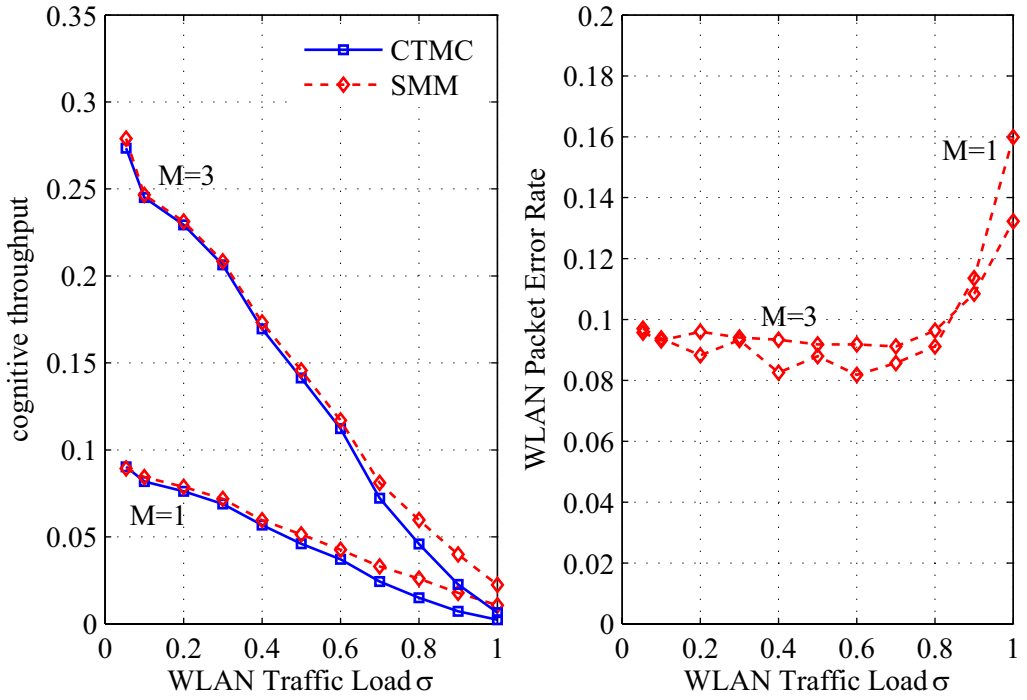


Figure 3.9: Robustness of cognitive frequency hopping with respect to deviations from the continuous-time Markov chain (CTMC) model. The performance results are based on heuristically running the methods on data generated under the semi-Markov model.

between exploring and exploiting transmission opportunities exists, was also analyzed. Numerical performance results demonstrate an interference reduction by a factor of 3.5-4.5 compared to a reference scheme without sensing and demonstrate a satisfactory robustness with respect to deviations from the exponential model.

CHAPTER 4

COGNITIVE COEXISTENCE BETWEEN INFRASTRUCTURE AND AD-HOC NETWORKS

4.1 Summary of Contributions and Related Work

The last chapter introduced a cognitive frequency hopping protocol in which a cognitive radio adapted its medium access such as to minimize interference to a set of close-by ad-hoc bands. In this chapter, we show that the sensing and prediction framework can also be useful in coexistence scenarios where the ad-hoc system has lower priority.

This chapter addresses a coexistence scenario involving two different types of networks: an infrastructure (IS) wide area network that shares spectrum with local, ad-hoc or peer-to-peer systems. Motivated by the superior communication resources of the IS system, we analyze how its flexible, centralized resource allocation can accommodate the ad-hoc links based on sensing and predicting their interference patterns. Despite adapting its resource allocation based on sensing results, the IS system minimizes interference to the ad-hoc network subject to maintaining a specified quality-of-service level for its users.

This approach is different from typical DSA formulations in which a secondary system exploits spectrum opportunities left over by a primary system, subject to the constraint that no significant interference is created. While both approaches are hierarchical, spectrum property rights in DSA mandate that secondary users adapt to spectrum licensees. In contrast, in cognitive coexistence it is possible to use the flexibility of the primary system to accommodate a secondary, lower

priority system while maintaining a desired primary system performance level. This can be viewed as a "best-effort" approach toward interference management and coexistence.

This framework is relevant to a number of practical setups. For example, the convergence of wide and local area networks has received increasing interest. The coexistence of IEEE 802.16 and IEEE 802.11 systems is such an example of practical importance [67, 68]. In cellular networks, the concept of femtocells has emerged recently in which base stations are deployed in homes to improve coverage and unload some of the traffic that otherwise would have to be supported by the cellular system [12]. Yet other applications arise in the military domain, where the coexistence of high and low priority links is of fundamental concern [69].

This chapter presents two approaches for improving the coexistence of IS and ad-hoc networks. First, we consider a scenario in which the interference channel is known to the IS terminals and derive the optimal power allocation. Second, when the interference channel is unknown, a logical approach is to separate transmissions in the time domain. We present the optimal power and transmission time allocation for this case.

4.1.1 Scenario 1: Known Interference Channel

In the case where IS terminals have knowledge of the interference channel to the ad-hoc nodes, we consider the optimal power allocation at the IS terminals, which minimizes interference to the ad-hoc nodes subject to a rate constraint for the IS users. To achieve this, we formulate a convex optimization problem and derive closed-form expressions for the optimal frame-level solution that only depend on a

set of two Lagrange multipliers. The optimal value of these Lagrange multipliers is found based on the bisection method with guaranteed convergence and low complexity. Rate constraints are then relaxed to the long-term average case and it is shown that water pouring in both frequency and time can further reduce interference.

4.1.2 Scenario 2: Temporal Interference Prediction

In the case where the interference channel from IS clients to the ad-hoc nodes is unknown, our goal is to separate transmissions in the temporal domain. We address the question of how the IS network can allocate power and transmission time judiciously such as to accommodate surrounding ad-hoc links. The problem of optimal power and transmission time allocation is again formulated as a convex program and the optimal frame-level solution is derived. Based on optimality conditions, a solution algorithm with guaranteed convergence and low complexity is introduced, and the structure of the optimal solution is studied to provide a better understanding of fundamental tradeoffs.

The frame-level problem is relaxed to the case of average-rate constraints, in which statistical knowledge of the temporal ad-hoc activity patterns and the IS channel coefficients is used to allocate resources in both frequency and time. Finally, the multi-terminal scenario is considered in which the IS base station allocates subchannels based on average interference metrics and IS terminals perform optimal power and transmission time allocation based on sensing results. A comparison with conventional subchannel allocation methods shows that heuristics may yield performance close to optimal.

4.1.3 Related Work and Organization

Optimal resource management in multiuser multicarrier wireless systems has been well-studied for both downlink and uplink cases; see [70, 71, 72] for an overview of the topic. In cognitive radio networks, optimal resource allocation is more challenging because it needs to incorporate interference constraints, which protect the primary system from harmful interference. Typical formulations aim at finding a power and subchannel allocation which maximizes the throughput of the cognitive radio system while meeting interference and power constraints. Recent work in this area includes [73, 74]. In addition to meeting interference constraints, spectrum sharing and self-coexistence within the cognitive radio network also need to be addressed. Contributions in this area include [75, 76, 77, 78]. To the best of the authors' knowledge, interference-aware resource management based on predicting temporal activity patterns has not been addressed before.

Quite different from our approach are information theoretic approaches that allow cognitive networks to communicate without interfering with the primary systems [79, 80].

4.2 Optimal Power Allocation Based on Known Interference Channel

4.2.1 System Setup

The system setup is shown in Figure 4.1. The IS network is a multicarrier system, which evolves in frames of fixed duration and whose IS clients operate on mutu-

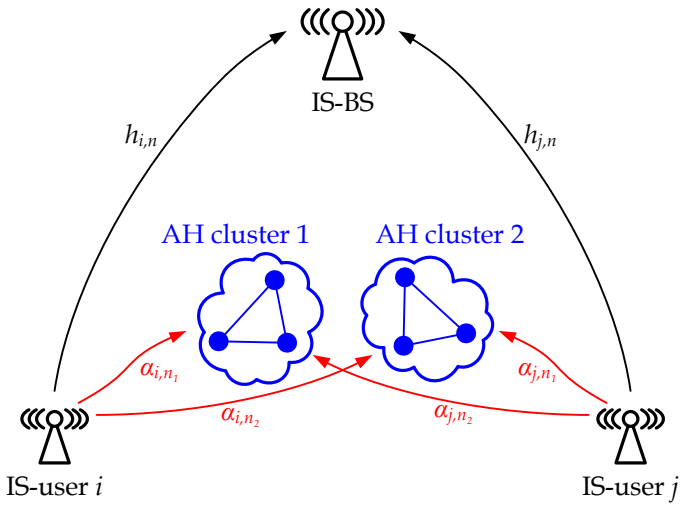


Figure 4.1: Infrastructure/ad-hoc coexistence setup for known interference channel. Infrastructure uplink transmissions across channel $h_{i,n}$ interfere with local, low-power transmissions in ad-hoc clusters 1 and 2. The interference channel is denoted by α_{i,n_1} and α_{i,n_2}.

ally exclusive subsets of subchannels, assigned by their base station. Finding the optimal, interference-aware power allocation on these subchannels is the objective of this chapter.

The ad-hoc system consists of distributed nodes, which evolve passively in the same frequency band. Ad-hoc transmissions are low power, limited to small clusters, and are assumed not to interfere with the uplink of the IS system. However, they are being interfered with through the channels $\alpha_{i,n}$. Our analysis focuses on the scenario in which uplink transmissions of the IS clients interfere with ad-hoc links; extensions to downlink scenarios are possible but not considered in this work.

Throughout this section, the interference channels $\alpha_{i,n}$ are assumed to be known. In practice there are several ways to estimate or model $\alpha_{i,n}$. For example, IS users could dedicate a portion of the frame to spectrum sensing, capture ad-hoc transmissions and (assuming channel reciprocity and that $\alpha_{i,n}$ varies reasonably

slowly) estimate the channel condition (note that only relative values are needed; any common scaling of $\alpha_{i,n}$ can be disregarded). Alternatively, if the location of ad-hoc nodes is known (as may be the case in some military scenarios), geolocation information can be combined with path-loss models to approximate $\alpha_{i,n}$.

Based on $\alpha_{i,n}$, IS users minimize interference power while maintaining an uplink rate constraint across channel $h_{i,n}$. The rate on sub-carrier n is given by

$$r_{i,n} = \log_2\left(1 + \kappa\frac{p_{i,n}|h_{i,n}|^2}{N_0}\right) = \log_2\left(1 + \beta_{i,n}p_{i,n}\right), \tag{4.1}$$

where N_0 is the noise power, κ is a normalization factor, and $\beta_{i,n}$ is introduced for notational convenience[1].

In summary, the cognitive allocation method operates as follows. In every frame, IS users request a rate R from their base station and are assigned a set of subchannels accordingly. IS users perform spectrum sensing to estimate the interference channel $\alpha_{i,n}$ and determine the optimal power allocation, which minimizes interference while maintaining rate and power constraints.

4.2.2 Frame-Level Formulation

In this section we derive the optimal power allocation for IS users, show that it admits a special structure, and use this property to derive an efficient solution algorithm.

The objective of this section is to find a power allocation $p_{i,n}$ which minimizes interference subject to the constraint that IS users meet rate and power constraints.

[1]The above formulation encompasses a channel capacity formulation (for $\kappa = 1$) as well as the case of variable-rate M-QAM in which case $\kappa = 1.5/(-\ln \mathsf{BER})$ is chosen such that a target BER is met [81].

Based on the assumption that IS users operate on pre-assigned, orthogonal sets of subchannels, the problem reduces to finding the optimal power allocation for each of the IS users individually. Without loss of generality we can therefore consider a single IS user, denote its set of subchannels $\mathbb{A}$, and for convenience drop the first index of the power allocation and channel coefficients.

The optimization problem **A-1** is then formulated mathematically as ($\mathbf{p} = [p_1, \ldots, p_N]^T$)

$$\min_{\mathbf{p}} \sum_{n \in \mathbb{A}} \alpha_n p_n \tag{4.2}$$

$$\text{subject to } \sum_{n \in \mathbb{A}} \log_2(1 + \beta_n p_n) \geq R \tag{4.3}$$

$$\sum_{n \in \mathbb{A}} p_n \leq P \tag{4.4}$$

$$p_n \geq 0, \ \forall n \in \mathbb{A}, \tag{4.5}$$

with rate constraint (4.3) and uplink power constraint (4.4). It is straightforward to show that **A-1** is a convex optimization problem since, once rewritten in standard form, both the objective function and the constraints are convex [82].

The inequalities (4.3) and (4.4) may not have a common solution because for any power constraint P there exists a sufficiently large rate constraint R that renders the problem infeasible. To avoid trivial complications we therefore assume that (4.3) and (4.4) have a solution. Fundamentally, the power constraint (4.4) makes the problem interesting; in its absence the problem reduces to classical water filling[2][82].

The optimization problem **A-1** admits a structured solution that can be developed in a similar fashion to classical water filling. The Lagrangian of the problem

[2]To see this, introduce new variables $\bar{p}_n = \alpha_n p_n$ to convert the problem into classical water filling over the equivalent channel β_n/α_n.

is given by

$$L(\mathbf{p},\gamma,\epsilon)=\sum_{n\in\mathbb{A}}\alpha_n p_n+\gamma\left(R-\sum_{n\in\mathbb{A}}\log_2(1+\beta_n p_n)\right)+\epsilon\left(\sum_{n\in\mathbb{A}}p_n-P\right), \quad (4.6)$$

where $\gamma\geq 0$ and $\epsilon\geq 0$ denote the Lagrange multipliers for the rate and power constraint, respectively. The non-negativity constraints (4.5) will be absorbed into the optimality conditions and do not require separate Lagrange multipliers.

The Karush-Kuhn-Tucker (KKT) conditions [82] for problem **A-1** consist of its constraints (4.3)-(4.5), non-negativity constraints for the Lagrange multipliers, $\gamma\geq 0$ and $\epsilon\geq 0$, the slackness conditions

$$\gamma\left(R-\sum_{n\in\mathbb{A}}\log_2(1+\beta_n p_n^*)\right) = 0 \quad (4.7)$$

$$\epsilon\left(\sum_{n\in\mathbb{A}}p_n^*-P\right) = 0, \quad (4.8)$$

and the condition

$$\left.\frac{\partial L(\mathbf{p},\gamma,\epsilon)}{\partial p_n}\right|_{p_n=p_n^*}\begin{cases}=0 & p_n^*>0\\ >0 & p_n^*=0\end{cases}. \quad (4.9)$$

Note that the non-negativity constraints (4.5) have been absorbed into (4.9).[3] The above equation can be interpreted by noting that, for p_n^* to minimize $L(\mathbf{p},\gamma,\epsilon)$, its partial derivative must vanish at p_n^* unless it lies at the boundary of the feasible set.

Substituting (4.6) into (4.9) and solving for p_n^* we obtain the solution structure

$$p_n^*=\left[\frac{\gamma}{(\alpha_n+\epsilon)\ln 2}-\frac{1}{\beta_n}\right]^+, \quad (4.10)$$

where $(\cdot)^+=\max\{0,\cdot\}$ denotes the positive part.

[3]This can be verified by introducing Lagrange multipliers for the inequality constraints. The non-negativity constraints for those Lagrange multipliers together with their slackness conditions lead to (4.9).

In order to find the optimal solution, we need to find the optimal Lagrange multipliers γ and ϵ in (4.10) such that both power and rate constraints are satisfied. With this objective, first substitute (4.10) into (4.3) and arrive at

$$\sum_{n\in\mathbb{A}} \left[\log_2 \frac{\gamma\beta_n}{(\alpha_n+\epsilon)\ln 2}\right]^+ \geq R. \tag{4.11}$$

In order to find a closed-form expression for γ, we introduce the set

$$\mathbb{P} = \left\{ n \in \mathbb{A} : \frac{\gamma\beta_n}{(\alpha_n+\epsilon)\ln 2} \geq 1 \right\}, \tag{4.12}$$

to express the γ for which (4.11) holds with equality as

$$\gamma = \ln 2 \left[\frac{2^R}{\prod_{n\in\mathbb{P}} \frac{\beta_n}{\alpha_n+\epsilon}}\right]^{1/|\mathbb{P}|}. \tag{4.13}$$

For any value of $\epsilon \geq 0$ the above expression specifies a $\gamma(\epsilon)$ such that the rate constraint (4.3) is met with equality. In the following define $\mathbf{p}(\epsilon)$ by (4.13) for a given ϵ. To fully determine the optimal solution to **A-1** we require a value of ϵ such that the power constraint (4.4) is satisfied. First note that if $\epsilon = 0$ leads to a feasible solution then $\mathbf{p}(\epsilon = 0)$ is in fact optimal. To see this, consider the slackness condition (4.8), and note that if $\epsilon = 0$ is feasible then the power constraint is not active at the optimal solution. Hence for any $\epsilon' > 0$,

$$f(\mathbf{p}(\epsilon')) := \sum_{n\in\mathbb{A}} \alpha_n p_n(\epsilon'), \tag{4.14}$$

can be no less than $f(\mathbf{p}(0))$.

Assume now that $\epsilon > 0$. Substituting (4.13) into (4.10) we obtain

$$p_n(\epsilon) = \frac{2^{R/|\mathbb{P}|}\ln 2}{\left(\prod_{i\in\mathbb{P}} \beta_i\right)^{1/|\mathbb{P}|}} \left[\prod_{i\in\mathbb{P}} \frac{\alpha_i+\epsilon}{\alpha_n+\epsilon}\right]^{1/|\mathbb{P}|} - \frac{1}{\beta_n} \tag{4.15}$$

if $n \in \mathbb{P}$ and zero otherwise. We now show that $f(\mathbf{p}(\epsilon))$ is a non-increasing function in ϵ. To see this assume without loss of generality that $\alpha_1 \leq \cdots \leq \alpha_N$. Then,

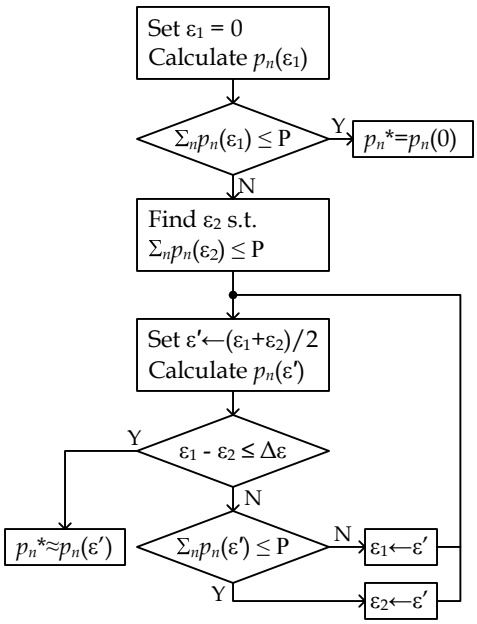


Figure 4.2: Algorithm for finding Lagrange multipliers γ and ϵ. The flow chart shows how a solution can be approximated by upper and lower bounding the optimal value of ϵ and using the bisection method.

$p_1(\epsilon)$ is clearly a non-increasing function in ϵ, and consequently less power gets allocated to the first subchannel as ϵ increases. Given the ordering of α_i and the fact that the total allocated power remains constant (the power constraint is active for $\epsilon > 0$) the interference power also increases with ϵ. We hence have that the optimal solution represents a feasible solution where ϵ is as small as possible, *i.e.*, if ϵ were decreased any further, $\mathbf{p}(\epsilon)$ would no longer be feasible. We use this result to derive an algorithm for finding the optimal ϵ.

As shown in Figure 4.2 the algorithm starts with calculating $\mathbf{p}(0)$. If this solution is feasible, *i.e.*, if it satisfies the power constraint, we know that it is optimal and the algorithm terminates. Otherwise, we find an ϵ_2 such that the power constraint is satisfied. Such an ϵ_2 exists as long as (4.3) and (4.4) have strictly feasible solutions, which we assume hereafter.

Having obtained an infeasible $\mathbf{p}(\epsilon_1)$ and a feasible $\mathbf{p}(\epsilon_2)$, the bisection method is used to find the optimal solution: first, the power allocation at the mid-point $\epsilon' = (\epsilon_1 + \epsilon_2)/2$ is computed. If it is feasible we set $\epsilon_2 \leftarrow \epsilon'$; otherwise we continue with $\epsilon_1 \leftarrow \epsilon'$. The algorithm terminates once a desired accuracy $\Delta\epsilon$ is reached.

The above algorithm converges to the optimal solution $\mathbf{p}^*$. In fact, in every step, the algorithm keeps an infeasible $\mathbf{p}(\epsilon_1)$ which represents a lower bound to the optimal solution, and a feasible $\mathbf{p}(\epsilon_2)$, which serves as an upper bound due to the fact that $f(\mathbf{p}(\epsilon))$ is non-decreasing in ϵ. Hence,

$$f(\mathbf{p}(\epsilon_1)) \leq f(\mathbf{p}^*) \leq f(\mathbf{p}(\epsilon_2)), \tag{4.16}$$

and since upper and lower bound converge as the algorithm progresses, the optimal solution is approximated arbitrarily well.

4.2.3 Average-Rate Formulation

In most applications, imposing rate constraints on a frame-level is unnecessary; instead average rate guarantees suffice. This section presents an average problem formulation which is based on knowing the long-term statistics of β or both α and β (in addition to measuring both α and β on a frame-level basis). For both cases, the additional flexibility allows for further interference reduction and leads to water filling in both frequency and time [83], [84].

Average Formulation for Random $\boldsymbol{\beta}$

First, we address the case where β is a random variable but α is a known deterministic constant. This leads to problem **A-2**

$$\min_{p_n(\boldsymbol{\beta})} \int_{\boldsymbol{\beta}} \sum_{n\in\mathbb{A}} \alpha_n p_n(\boldsymbol{\beta}) dF(\boldsymbol{\beta}) \tag{4.17}$$

$$\text{s.t.} \int_{\boldsymbol{\beta}} \sum_{n\in\mathbb{A}} \log_2(1+\beta_n p_n(\boldsymbol{\beta})) dF(\boldsymbol{\beta}) \geq R \tag{4.18}$$

$$\int_{\boldsymbol{\beta}} \sum_{n\in\mathbb{A}} p_n(\boldsymbol{\beta}) dF(\boldsymbol{\beta}) \leq P, \tag{4.19}$$

$$p_n(\boldsymbol{\beta}) \geq 0, \ \forall n, \beta, \tag{4.20}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_N]^T$ and $F(\boldsymbol{\beta})$ is the cumulative density function (CDF) of $\boldsymbol{\beta}$. Note that there are infinitely many decision variables $p_n(\boldsymbol{\beta})$ (for each realization of β). It is straightforward to see that the structure of the optimal solution is still given by (4.10). Since the optimality conditions have to be satisfied for each realization of $\boldsymbol{\beta}$ we obtain that

$$p_n(\boldsymbol{\beta}) = \left[\frac{\gamma}{(\alpha_n+\epsilon)\ln 2} - \frac{1}{\beta_n}\right]^+, \tag{4.21}$$

where γ and ϵ are Lagrange multipliers for the rate and power constraint, as before.

The optimal values of γ and ϵ need to satisfy the average constraints (4.18) and (4.19). For illustration, we derive closed-form solutions for the integral expressions assuming that all subchannels are independently flat Rayleigh fading, $\beta_i \sim \exp(1)$; in the general case the integral expressions need to be evaluated numerically.

For a specific subchannel i the rate constraint evaluates to

$$\int_{\beta_i=0}^{\infty} \left[\log_2 \tfrac{\gamma\beta_i}{(\alpha_i+\epsilon)\ln 2}\right]^+ e^{-\beta_i} d\beta_i = -\frac{1}{\ln 2}\text{Ei}\left(\tfrac{-(\alpha_i+\epsilon)\ln 2}{\gamma}\right), \tag{4.22}$$

where $\mathrm{Ei}(z) = -\int_{-z}^{\infty} e^{-t}/t\, dt$ denotes the exponential integral function [38]. The rate constraint can hence be written as

$$\frac{1}{\ln 2} \sum_{n \in \mathbb{A}} \mathrm{Ei}\left(\frac{-(\alpha_n + \epsilon) \ln 2}{\gamma} \right) \geq R. \tag{4.23}$$

In a similar way we obtain for the power constraint

$$\sum_{n \in \mathbb{A}} \frac{\gamma}{(\alpha_n + \epsilon) \ln 2} e^{-\frac{(\alpha_n + \epsilon) \ln 2}{\gamma}} + \mathrm{Ei}\left(\frac{-(\alpha_n + \epsilon) \ln 2}{\gamma} \right) \leq P. \tag{4.24}$$

Based on the expressions (4.23) and (4.24), the optimal values of γ and ϵ can again be found through the bisection method.

Average Formulation for Random $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$

If the statistics of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are known, interference can be reduced further. This leads to the optimization problem **A-3**

$$\min_{p_n(\boldsymbol{\alpha}, \boldsymbol{\beta})} \int_{\boldsymbol{\alpha}} \int_{\boldsymbol{\beta}} \sum_{n \in \mathbb{A}} \alpha_n p_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) dF(\boldsymbol{\alpha}) dF(\boldsymbol{\beta}) \tag{4.25}$$

$$\text{s.t.} \int_{\boldsymbol{\alpha}} \int_{\boldsymbol{\beta}} \sum_{n \in \mathbb{A}} \log_2(1 + \beta_n p_n(\boldsymbol{\alpha}, \boldsymbol{\beta})) dF(\boldsymbol{\alpha}) dF(\boldsymbol{\beta}) \geq R \tag{4.26}$$

$$\int_{\boldsymbol{\alpha}} \int_{\boldsymbol{\beta}} \sum_{n \in \mathbb{A}} p_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) dF(\boldsymbol{\alpha}) dF(\boldsymbol{\beta}) \leq P, \tag{4.27}$$

$$p_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) \geq 0, \ \forall n, \boldsymbol{\alpha}, \boldsymbol{\beta}, \tag{4.28}$$

in which both $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$ and $\boldsymbol{\beta}$ are random variables. This problem can again be solved by evaluating the integrals in (4.26) and (4.27) and using the bisection method.

4.2.4 Numerical Results

In this section we present numerical performance results for frame-level and average formulations. We assume that there are a total of five subchannels and that α_n, β_n are independent and identically Rayleigh distributed in both frequency and time and compare the proposed methods to a reference method in which the IS system minimizes transmit rather than interference power.

The result is shown in Figure 4.3 and plots the interference power with respect to the achieved IS rate. For better comparison the rate is shown as a fraction of the IS channel capacity and the total interference power is normalized by P and shown on a logarithmic scale. The plot shows that more interference is generated at higher rates of the IS link since there is less flexibility in allocating IS resources.

The performance ordering reflects our expectations. The reference scheme is dominated by the frame-level formulation, which is in turn outperformed by the average formulations. The average formulations show better performance because of their additional flexibility in assigning resources (the doubly average formulation outperforms the case of average $\boldsymbol{\beta}$ since additional diversity of the channel $\boldsymbol{\alpha}$ can be leveraged).

In comparing average and frame-level formulations, note that the latter does not achieve the same capacity as the average case because in some frames **A-1** will be infeasible (in such frames we maximize frame rate subject to the power constraint to ensure a fair comparison). This complication is avoided in the average formulations because water filling is performed in both frequency and time. The performance gain associated with the average schemes is large and amounts to more than 9 dB for an IS system operating at 50% of its capacity (compared to

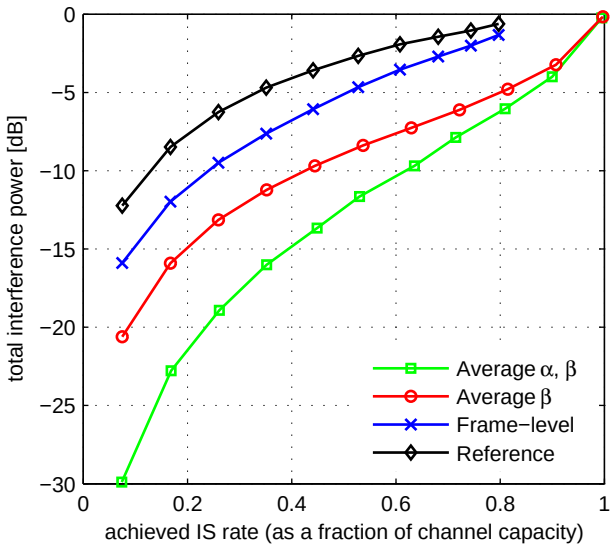


Figure 4.3: Average-rate performance result. The total interference power (normalized by the uplink power constraint P and shown in dB) is plotted with respect to the IS rate (as a fraction of the maximum IS capacity). ©2008 IEEE. Reprinted, with permission, from [21].

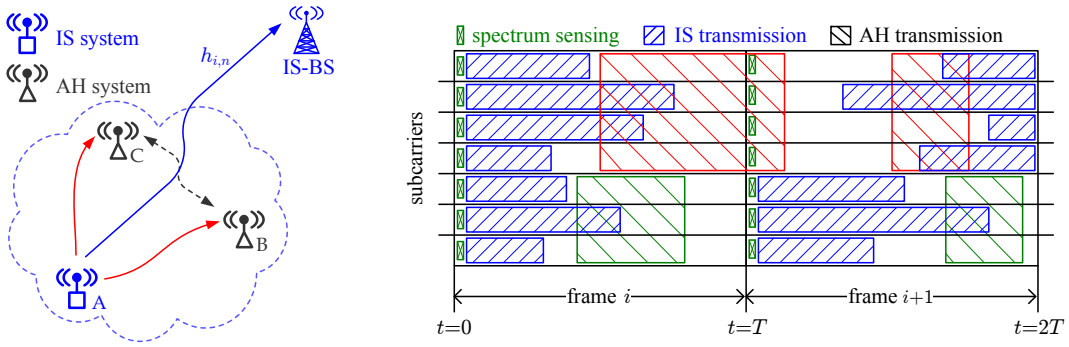
the reference scheme).

4.3 Optimum Transmission Time Allocation Based on Temporal Interference Prediction

4.3.1 Problem Formulation

This section introduces the problem formulation mathematically, describes the system setup and states important modeling assumptions. This will lay the groundwork for finding the optimal frame-level solution in the subsequent section.

The system setup is shown in Figure 4.4. We consider an IS system, which consists of a base station and a single client (the multi-terminal case will be addressed



(a) System setup (b) Time/frequency setup

Figure 4.4: Infrastructure/ad-hoc coexistence setup for temporal interference prediction. An infrastructure link allocates power and transmission time such as to minimize interference to close-by ad-hoc networks. The interference-aware resource allocation is based on detecting and predicting the ad-hoc system's temporal activity.

at the end of this chapter). The uplink transmissions of this client may strongly interfere with local transmissions of one or multiple ad-hoc networks surrounding the client. For this setup, the problem of optimally assigning power and transmission time at the IS client, such as to minimize interference to the ad-hoc links is analyzed. The time/frequency behavior of both systems is shown in Figure 4.4(b).

Ad-Hoc Network. The ad-hoc network consists of a set of ad-hoc nodes which operate in a frequency band that overlaps with the IS system. As depicted in Figure 4.4(b) there can be multiple ad-hoc networks which operate in non-overlapping bands that each overlap with a certain set of IS subchannels. It is assumed that the partitioning of the ad-hoc bands is fixed and that the temporal activity of different bands is statistically independent.

We model the time behavior of each ad-hoc band by a two-state ON/OFF continuous time Markov chain (CTMC). The holding times in both ON and OFF state are exponentially distributed with parameters μ for the ON state and λ for

the OFF state. Therefore, if an ad-hoc link is detected to be in a certain state at time t_0, then its transition matrix for time $t_0 + \tau$ is given by

$$\mathbf{P}(\tau) = \frac{1}{\lambda + \mu} \begin{bmatrix} \mu + \lambda e^{-(\lambda+\mu)\tau} & \lambda - \lambda e^{-(\lambda+\mu)\tau} \\ \mu - \mu e^{-(\lambda+\mu)\tau} & \lambda + \mu e^{-(\lambda+\mu)\tau} \end{bmatrix}, \tag{4.29}$$

which follows directly from the definition of a CTMC [64, p.391]. Therefore, the probability of an ad-hoc link being ON at time $t_0 + \tau$, conditioned on having it observed in the ON (OFF) state at time t_0 is given by the lower right (upper right) entry in the matrix above. Modeling ad-hoc links based on a two-state CTMC approximates the carrier sense random medium access typically employed in such systems and is motivated by our findings presented in Chapter 2.

Infrastructure System. The IS system operates in the same frequency band as the ad-hoc network and evolves in frames of fixed duration T. At the beginning of each frame, spectrum sensing is used to detect the ON/OFF activity of the ad-hoc bands, and based on the sensing outcome, power and transmission time are assigned as shown in Figure 4.4(b).

Throughout this chapter we assume *perfect sensing*, that is, sensing outcomes are always accurate and the overhead associated with sensing is negligible. This is a reasonable assumption because the proximity of ad-hoc terminals to the IS client results in moderate to high signal-to-noise-ratios (SNR). Conceptually, the detection task therefore becomes similar to the carrier sensing employed in systems such as IEEE 802.11 [9].

Based on the sensing result at the beginning of each frame, the IS system allocates power and transmission time on a subchannel basis. This is conceptually similar to the allocation of time/frequency resource blocks in broadband cellular

systems based on OFDMA. The case where a subset or even all subchannels need to share the same timing allocation (for example when transmissions in the entire band can only be turned on or off) in general has worse performance. Nevertheless, a similar solution approach remains applicable.

The IS system minimizes interference subject to maintaining rate requirements for their clients. The rate that is supported by a specific subchannel is modeled based on a channel capacity formulation,

$$\sum_n \rho_n \log\left(1+\kappa\frac{p_n|h_n|^2}{\rho_n N_0}\right) = \sum_n \rho_n \log\left(1+\frac{p_n\beta_n}{\rho_n}\right), \tag{4.30}$$

where $\mathbf{p} = [p_1, \ldots, p_N]^T$ denotes the power allocation, $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_N]^T$ represents the transmission time allocation, N_0 is the noise power, κ a normalization factor, and β_n is introduced for notational convenience[4].

Interference Metrics and Scheduling Assumptions. The interference between IS and ad-hoc networks is modeled by the average temporal overlap between both systems. Based on the sensing result at the beginning of a frame and knowledge of the CTMC parameters of the ad-hoc links, transmission time and power are allocated.

The allocation of transmission time consists of specifying duration and placement of the transmission within the current frame. We first show that it is optimal to transmit at the beginning (the end) of the frame if the sensing outcome is idle (busy).

Lemma 1. *Assume that a ρ fraction of transmission time needs to be allocated to a subchannel, on which the ad-hoc user's ON/OFF behavior is modeled by the*

[4]The above formulation encompasses a channel capacity formulation (for $\kappa = 1$) as well as the case of variable-rate M-QAM in which case $\kappa = 1.5/(-\ln \mathsf{BER})$ is chosen such that a target BER is met [81].

CTMC (4.29). Based on a sensing outcome at the beginning of the frame, the minimum expected overlap with the ON period of the ad-hoc user is achieved by

- *transmitting at the beginning of the frame (i.e., during* $[0, \rho T]$*) if the sensing outcome was idle and*
- *transmitting at the end of the frame (i.e., during* $[(1-\rho)T, T]$ *if the sensing outcome was busy.*

Proof. see appendix. □

Based on Lemma 1, we derive the expected time overlap between IS and ad-hoc transmissions, conditioned on the sensing result $y \in \{0, 1\}$ at the beginning of the frame. Consider a subchannel n, which overlaps with ad-hoc band $i = g(n)$. Then, the activity of ad-hoc user i is given by the CTMC $\{X_i(\xi), \xi \geq 0\}$ with parameters λ_i and μ_i. Transmitting for a ρ fraction of the frame, leads to the expected time overlap

$$\phi_{n,0}(\rho) = \frac{1}{T}\mathbb{E}\left\{ \int_0^{\rho T} \mathbf{1}_{\{X_i(\xi)=1\}} d\xi \middle| X(0) = 0 \right\} = \frac{1}{T} \int_0^{\rho T} \Pr(X(\xi) = 1 | X(0) = 0) d\xi \tag{4.31}$$

if the sensing result was idle. By substituting (4.29) it is then easy to show that

$$\phi_{n,0}(\rho) = \tfrac{\lambda_i}{(\lambda_i+\mu_i)T} \left(\rho T + \tfrac{1}{\lambda_i+\mu_i}(e^{-(\lambda_i+\mu_i)\rho T} - 1) \right). \tag{4.32}$$

In the case of a busy sensing result we obtain

$$\phi_{n,1}(\rho) = \tfrac{\lambda_i}{(\lambda_i+\mu_i)T} \left(\rho T + \tfrac{\mu_i/\lambda_i}{\lambda_i+\mu_i} e^{-(\lambda_i+\mu_i)T}(e^{(\lambda_i+\mu_i)\rho T} - 1) \right). \tag{4.33}$$

This derivation makes use of the fact that the sensing results of two IS subchannels are either perfectly correlated (if they overlap with the same ad-hoc band) or

statistically independent (if they overlap with different bands). Therefore, the prediction performance of a specific subchannel cannot be improved by using sensing results from other subchannels.

Lemma 2. *The functions $\phi_{n,0}(\rho)$ and $\phi_{n,1}(\rho)$ are strictly convex and increasing in the transmission time ρ.*

Proof. Both $\phi_{n,0}(\rho)$ and $\phi_{n,1}(\rho)$ are nonnegative linear combinations of a convex and a strictly convex function. One is linear, the other an exponential function with nonzero exponent. The monotonicity can easily be verified by differentiation. □

4.3.2 Optimal Frame-Level Allocation

Based on the system setup and interference metrics defined in the previous section we formulate the optimal frame-level allocation as a convex program, and subsequently analyze it using a Lagrangian approach.

Consider a single IS client, which minimizes the time overlap between IS and ad-hoc transmissions subject to maintaining a rate constraint across the IS channel to its base station. Mathematically, this leads to problem **B-1**

$$\min_{\mathbf{p},\boldsymbol{\rho}} \sum_n \phi_{n,y_n}(\rho_n) \tag{4.34}$$

$$\text{s.t.} \sum_n \rho_n \log\left(1 + \frac{p_n \beta_n}{\rho_n}\right) \geq R \tag{4.35}$$

$$\sum_n p_n \leq P \tag{4.36}$$

$$p_n \geq 0, \quad 1 \leq n \leq N \tag{4.37}$$

$$0 \leq \rho_n \leq 1, \quad 1 \leq n \leq N, \tag{4.38}$$

with rate constraint (4.35) and power constraint (4.36). It is straightforward to show that **B-1** is a convex optimization problem since the objective function is convex (by Lemma 2), the rate constraint (once rewritten in standard form) is convex by the perspective property [82], and all other constraints are linear.

A solution to **B-1** can be found by general solution techniques in polynomial time [82]. For this specific problem, however, it is possible to show a special structure that enables us to gain further insight into the problem.

The solution structure is obtained by introducing Lagrange multipliers γ and ϵ for the rate and power constraint, respectively. This leads to the Lagrangian

$$L(\mathbf{p}, \boldsymbol{\rho}; \gamma, \epsilon) = \sum_n \phi_{n,y_n}(\rho_n) + \gamma \left[R - \sum_n \rho_n \log\left(1 + \frac{p_n \beta_n}{\rho_n}\right)\right] + \epsilon \left[\sum_n p_n - P\right]. \tag{4.39}$$

The Karush-Kuhn-Tucker (KKT) optimality conditions are then given by the constraints (4.35)-(4.38) of **B-1**, non-negativity constraints for the Lagrange multipliers, $\gamma \geq 0$, $\epsilon \geq 0$, the slackness conditions

$$\gamma \left[R - \sum_n \rho_n^* \log\left(1 + \frac{p_n^* \beta_n}{\rho_n^*}\right)\right] = 0 \tag{4.40}$$

$$\epsilon \left[\sum_n p_n^* - P\right] = 0, \tag{4.41}$$

the condition

$$\left.\frac{\partial L(\mathbf{p}, \boldsymbol{\rho}; \gamma, \epsilon)}{\partial p_n}\right|_{p_n = p_n^*} \begin{cases} = 0, & p_n^* > 0 \\ > 0, & p_n^* = 0 \end{cases}, \tag{4.42}$$

and

$$\left.\frac{\partial L(\mathbf{p}, \boldsymbol{\rho}; \gamma, \epsilon)}{\partial \rho_n}\right|_{\rho_n = \rho_n^*} \begin{cases} > 0, & \rho_n^* = 0 \\ = 0, & \rho_n^* \in (0,1) \\ < 0, & \rho_n^* = 1 \end{cases}. \tag{4.43}$$

The conditions (4.42) and (4.43) can be interpreted by noting that, if p_n^* or ρ_n^* minimize $L(\mathbf{p}, \boldsymbol{\rho}; \gamma, \epsilon)$, their partial derivative must vanish unless they lie on some boundary of the feasible set. Conversely, if p_n^* or ρ_n^* lie on the boundary of the feasible set, $L(\mathbf{p}, \boldsymbol{\rho}; \gamma, \epsilon)$ may not decrease by moving to a point in the interior.

By substituting (4.39) into (4.42) and solving for p_n^* we arrive at

$$p_n^* = \rho_n \left(\nu - \frac{1}{\beta_n} \right)^+, \tag{4.44}$$

where $\nu := \gamma/\epsilon$ has been introduced to simplify notation in what follows. For any fixed value of ρ_n, (4.42) represents a water filling solution [85].

The optimal transmission time allocation is obtained by substituting (4.39) and (4.44) into (4.43). For an idle sensing result, $y_n = 0$, we obtain,

$$\rho_n^* = \begin{cases} \frac{1}{(\lambda_i+\mu_i)T} \log \frac{1}{1-\frac{\lambda_i+\mu_i}{\lambda_i}\gamma h_n(\nu)}, & \gamma h_n(\nu) \leq \zeta_{0,i} \\ 1, & \text{o.w.} \end{cases}, \tag{4.45}$$

where $\zeta_{0,i} = \lambda_i/(\lambda_i + \mu_i)(1 - \exp(-(\lambda_i + \mu_i)T)$ and $i = g(n)$ denotes the ad-hoc subband that overlaps with subchannel n. In the above equation we have defined

$$h_n(\nu) := [\log(\nu\beta_n)]^+ - \frac{(\nu\beta_n - 1)^+}{1 + (\nu\beta_n - 1)^+} \tag{4.46}$$

to simplify notation. Similarly, we can obtain the solution structure for the case of a busy sensing result, $y_n = 1$,

$$\rho_n^* = \begin{cases} 0, & \gamma h_n(\nu) < \zeta_{1,i} \\ 1 + \frac{\log\left(\frac{\lambda_i+\mu_i}{\mu_i}\gamma h_n(\nu) - \frac{\lambda_i}{\mu_i}\right)}{(\lambda_i+\mu_i)T}, & \zeta_{1,i} \leq \gamma h_n(\nu) \leq 1 \\ 1, & \gamma h_n(\nu) > 1, \end{cases}, \tag{4.47}$$

where $\zeta_{1,i} = \lambda_i/(\lambda_i + \mu_i)(1 + \mu_i/\lambda_i \exp(-(\lambda_i + \mu_i)T)$. Note that these closed-form solutions depend on the Lagrange multipliers only through the term $\gamma h_n(\nu)$ which does not depend on the ad-hoc activity parameters λ_i and μ_i. Further,

the transmission time allocations (4.45)-(4.47) are monotonic with respect to this term.

Iterative Solution Algorithm for γ and ν. To find the optimal power and transmission time allocation based on the above closed-form expressions, we present an algorithm for finding the pair $[\gamma^*, \nu^*]$, which corresponds to the optimal solution of **B-1**.

For any pair $[\gamma, \nu]$ the power allocation $\mathbf{p}(\gamma, \nu)$ and transmission time allocation $\boldsymbol{\rho}(\gamma, \nu)$ define the optimal solution to **B-1** with modified rate constraint

$$R(\gamma, \nu) := \sum_n \rho_n(\gamma, \nu)[\log(\nu\beta_n)]^+ \tag{4.48}$$

and modified power constraint

$$P(\gamma, \nu) := \sum_n \rho_n(\gamma, \nu) \left(\nu - \frac{1}{\beta_n}\right)^+, \tag{4.49}$$

where $\rho_n(\gamma, \nu)$ is given by (4.45) or (4.47) (depending on the sensing result). The fact that this solution is optimal for rate constraint $R(\gamma, \nu)$ and power constraint $P(\gamma, \nu)$ follows directly from the KKT optimality conditions, which are necessary and sufficient for convex optimization problems [82].

Based on the above, finding the pair $[\gamma^*, \nu^*]$ corresponding to the given rate constraint R and power constraint P could theoretically be performed by searching all pairs $[\gamma, \nu]$. In this section we show, however, that $R(\gamma, \nu)$ and $P(\gamma, \nu)$ exhibit some monotonicity which enables us to use the bisection method for finding $[\gamma^*, \nu^*]$ with guaranteed convergence and low complexity. We first study the case of keeping γ fixed and adjusting ν such that the rate constraint is met with equality. Then, we show that the allocated sum power decreases with γ.

We first consider the case of adjusting ν such that $R(\gamma, \nu) = R$ while keeping γ fixed. It is easy to verify that for any n, $h_n(\nu)$ is nondecreasing in ν. Therefore, $\rho_n(\gamma, \nu)$ increases with γ as well, as can be seen from (4.45) and (4.47). Further, since $\rho_n(\gamma, \nu)$ increases with ν for fixed γ, so does $R(\gamma, \nu)$. We can exploit this property to find the ν for which $R(\gamma, \nu) = R$ by the bisection method. First, we can find upper and lower bounds, ν_u and ν_l, for this value. These bounds are guaranteed to exist since $R(\gamma, \nu) \to \infty$ for $\nu \to \infty$ and $R(\gamma, \nu) \to 0$ for $\nu \to 0$. Once these bounds have been obtained the bisection method iteratively finds ν^* with guaranteed convergence.

Having obtained an algorithm for finding ν for arbitrary γ such that the rate constraint is satisfied, we study the behavior of the power constraint as γ is adjusted. As γ is varied, we continue to adjust ν such that the rate constraint is satisfied at all times. The pair of Lagrange multipliers is therefore given by $[\gamma, \nu^*(\gamma)]$.

The slackness conditions imply that at the optimal solution both rate and power constraints are met with equality. From (4.48) we observe that decreasing γ requires increasing ν in order to continue meeting the rate constraint. It can further be shown that decreasing γ reduces the objective function (due to the fact that $\log(\hat{\nu}\beta_n) \geq \log(\nu\beta_n)$ for $\hat{\nu} \geq \nu$, decreasing γ enables us to reduce $\rho_n(\gamma, \nu^*(\gamma))$ for at least some n).

While decreasing γ reduces the objective function, we need to find a pair $[\gamma, \nu]$ which satisfies both power and rate constraints. The total allocated sum power, as a function of γ, is given by

$$\sum_n \rho_n(\gamma, \nu^*(\gamma)) \left(\nu^*(\gamma) - \frac{1}{\beta_n}\right)^+. \tag{4.50}$$

To study the effect of reducing γ we can intuitively argue that decreasing γ requires

increasing $\nu^*(\gamma)$. However, since (4.48) remains constant, (4.50) increases since the term $[\nu^*(\gamma) - \frac{1}{\beta_n}]^+$ increases faster than $[\log(\nu^*(\gamma)\beta_n)]^+$. Therefore, reducing γ results in an allocation with higher sum power. We make this argument rigorous in the following lemma.

Lemma 3. *The sum power $\sum_n p_n$ associated with allocation $[\gamma, \nu^*(\gamma)]$ is a decreasing function in γ.*

Proof. see appendix. □

Lemma 3 enables us to find γ^* again by the bisection method. Assuming that **B-1** is feasible which we will assume hereafter, there exist bounds γ_u and γ_l such that $P(\gamma_u, \nu^*(\gamma_u)) \leq P \leq P(\gamma_l, \nu^*(\gamma_l))$. Therefore, by starting the bisection method from these points we can find the pair $[\gamma^*, \nu^*]$ with guaranteed convergence. The solution algorithm is shown in detail in Figure 4.5. The inner loop (lines 4–14) correspond to finding $\nu^*(\gamma)$, whereas the outer loop finds γ^*.

4.3.3 Properties of Optimal Allocations

Beyond simplifying solution algorithms, the structured solutions also enable us to make some qualitative statements about the optimal resource allocation.

We first investigate the ordering with respect to the IS channel coefficients β_n and assume that all other parameters are identical. This scenario is shown in Figure 4.6(a) for $N = 4$ subchannels with coefficients $\boldsymbol{\beta} = [.9, 1.1, .5, 1.5]^T$. We observe that, for any rate constraint, more transmission time is allocated to the subchannel with higher channel coefficient, *i.e.*, $\beta_i \geq \beta_j \Rightarrow \rho_i \geq \rho_j$ provided all other parameters and the sensing results are identical. On an intuitive level, the

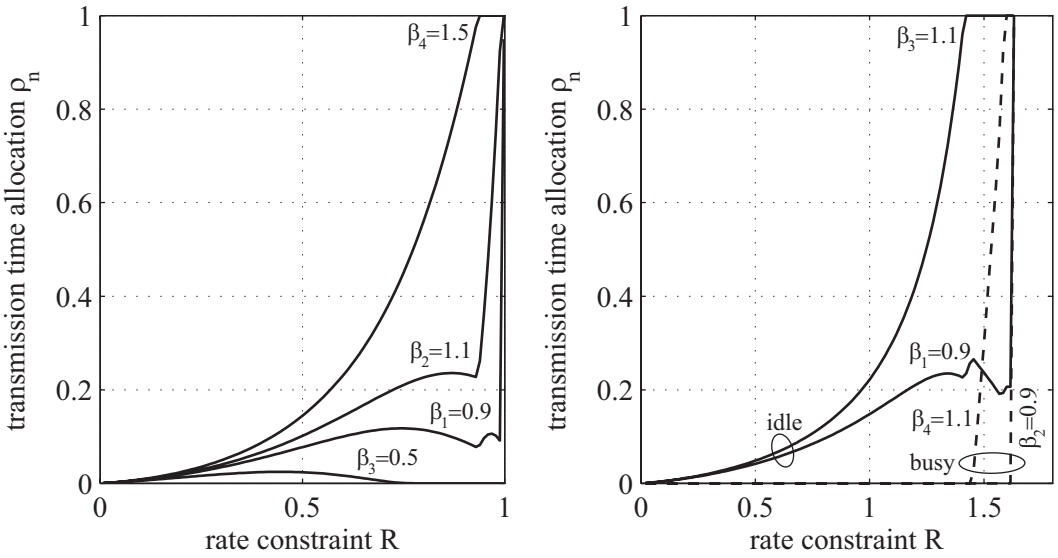
Algorithm 1: Solution Algorithm

1 **Initialization.** Obtain bounds $\nu_l, \nu_u, \gamma_l, \gamma_u$;
2 **repeat**
3 $\quad \hat{\gamma} \leftarrow (\gamma_u - \gamma_l)/2$;
4 $\quad$ **repeat**
5 $\quad\quad \hat{\nu} \leftarrow (\nu_u - \nu_l)/2$;
6 $\quad\quad$ Find time allocation $\rho_n(\hat{\nu})$ using (4.45);
7 $\quad\quad$ Find power allocation $p_n(\hat{\gamma}, \hat{\nu})$ using (4.44);
8 $\quad\quad$ Compute achievable rate $r(p_n, \rho_n)$ using (4.30);
9 $\quad\quad$ **if** $r(p_n, \rho_n) \geq R$ **then**
10 $\quad\quad\quad \nu_u \leftarrow \hat{\nu}$
11 $\quad\quad$ **else**
12 $\quad\quad\quad \nu_l \leftarrow \hat{\nu}$
13 $\quad\quad$ **end**
14 $\quad$ **until** $0 \leq R - r(p_n, \rho_n) \leq \epsilon_R$;
15 $\quad$ Find time allocation $\rho_n(\hat{\gamma}, \hat{\nu})$ using (4.45);
16 $\quad$ Find power allocation $p_n(\hat{\gamma}, \hat{\nu})$ using (4.44);
17 $\quad$ **if** $\sum_n p_n \geq P$ **then**
18 $\quad\quad \gamma_l \leftarrow \hat{\gamma}$
19 $\quad$ **else**
20 $\quad\quad \gamma_u \leftarrow \hat{\gamma}$
21 $\quad$ **end**
22 **until** $0 \leq P - \sum_n p_n \leq \epsilon_p$;

Figure 4.5: Algorithm for finding the optimal Lagrange multipliers γ and ν for problem (4.34)-(4.38). The inner loop (lines 4–14) find $\nu^*(\gamma)$ which satisfies the rate constraint (4.35). The outer loop determines γ^*, which satisfies the power constraint (4.36)

result captures the fact that in channels with high β_i we can achieve the same rate in a shorter transmission duration using the same amount of power. Mathematically, the result follows from the monotonicity of the optimal solution and $h_n(\nu)$.

Similar to the IS channel, the optimal transmission time allocation can be ordered with respect to the sensing results. If channels have the same IS channel coefficient, $\beta_i = \beta_j$, but different sensing results (in one of the channels the ad-hoc system is active, in the other it is not) then it is preferable to allocate more transmission time to the idle channel, *i.e.*, $\rho_i \geq \rho_j$. This is illustrated in Figure 4.6(b)



(a) Ordering by IS channel quality

(b) Ordering by sensing outcome

Figure 4.6: Structure of the optimal transmission time allocation for varying rate constraint. The solution can be ordered with respect to IS channel coefficients or sensing outcomes.

for $N = 6$, $\boldsymbol{\beta} = [.9, .9, 1.1, 1.1]^T$, and $\mathbf{y} = [0, 1, 0, 1]^T$. It is also interesting to note that some transmission time is allocated to frames with busy sensing results even when idle frames are not yet used to the maximum extent.

4.3.4 Optimal Average Resource Allocation

Problem formulation **B-1** requires that rate and power constraint are met in every frame, even if sensing outcome or IS channel quality are disadvantageous. In practical systems, satisfying rate constraints at the frame-level is usually unnecessary; it suffices to maintain average rate constraints across time. This less stringent requirement can be used to further reduce interference by allocating less transmission time during frames with adversarial channel/interference conditions, while compensating for the rate decrease during frames with better conditions. Ulti-

mately, this leads to an improved resource allocation across both frequency (the subchannels of the IS system) and time (consecutive frames of the IS system).

This section introduces such an average rate formulation by averaging across the temporal activity of the ad-hoc network and subsequently extends the formulation to random IS channel coefficients. Further, this section introduces two reference schemes that help to put the performance of the optimal resource allocation in perspective.

The average-rate formulation requires associating probabilities with all possible sensing outcomes. While there are a total of N subchannels available, the sensing outcomes for subchannels that overlap with the same ad-hoc band will be identical. Therefore, for M subbands, there are a total of 2^M possible sensing outcomes. Let the set of all possible sensing outcomes be represented by $\mathbb{Y} = \{0,1\}^M$ where $\mathbf{y} = [y_1, \ldots, y_M]^T \in \mathbb{Y}$ denotes the sensing outcome per subband.

Problem **B-2** of optimally allocating power and transmission time then becomes

$$\min_{\substack{p_{n,\mathbf{y}} \\ \rho_{n,\mathbf{y}}}} \sum_{\mathbf{y}\in\mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} \phi_{n,y_{g(n)}}(\rho_{n,\mathbf{y}}) \tag{4.51}$$

$$\text{s.t.} \sum_{\mathbf{y}\in\mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} \rho_{n,\mathbf{y}} \log\left(1 + \frac{p_{n,\mathbf{y}}\beta_n}{\rho_{n,\mathbf{y}}}\right) \geq R \tag{4.52}$$

$$\sum_{\mathbf{y}\in\mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} p_{n,\mathbf{y}} \leq P \tag{4.53}$$

$$p_{n,\mathbf{y}} \geq 0, \quad \forall \mathbf{y} \in \mathbb{Y}, 1 \leq n \leq N \tag{4.54}$$

$$0 \leq \rho_{n,\mathbf{y}} \leq 1, \quad \forall \mathbf{y} \in \mathbb{Y}, 1 \leq n \leq N, \tag{4.55}$$

where $\eta_{i,0} = \mu_i/(\lambda_i + \mu_i)$, $\eta_{i,1} = \lambda_i/(\lambda_i + \mu_i)$, and due to the independence of the ad-hoc subbands, $\eta_{\mathbf{y}} = \prod_{i=1}^{M} \eta_{i,y_i}$. Note that this optimization problem has 2^M as many decision variables because we power and transmission time allocation will be different for every possible sensing outcome. The fact that the decision variables

grow exponentially with M is not of major concern, because M (the number of parallel ad-hoc bands) is typically quite small (in the order of one to five).

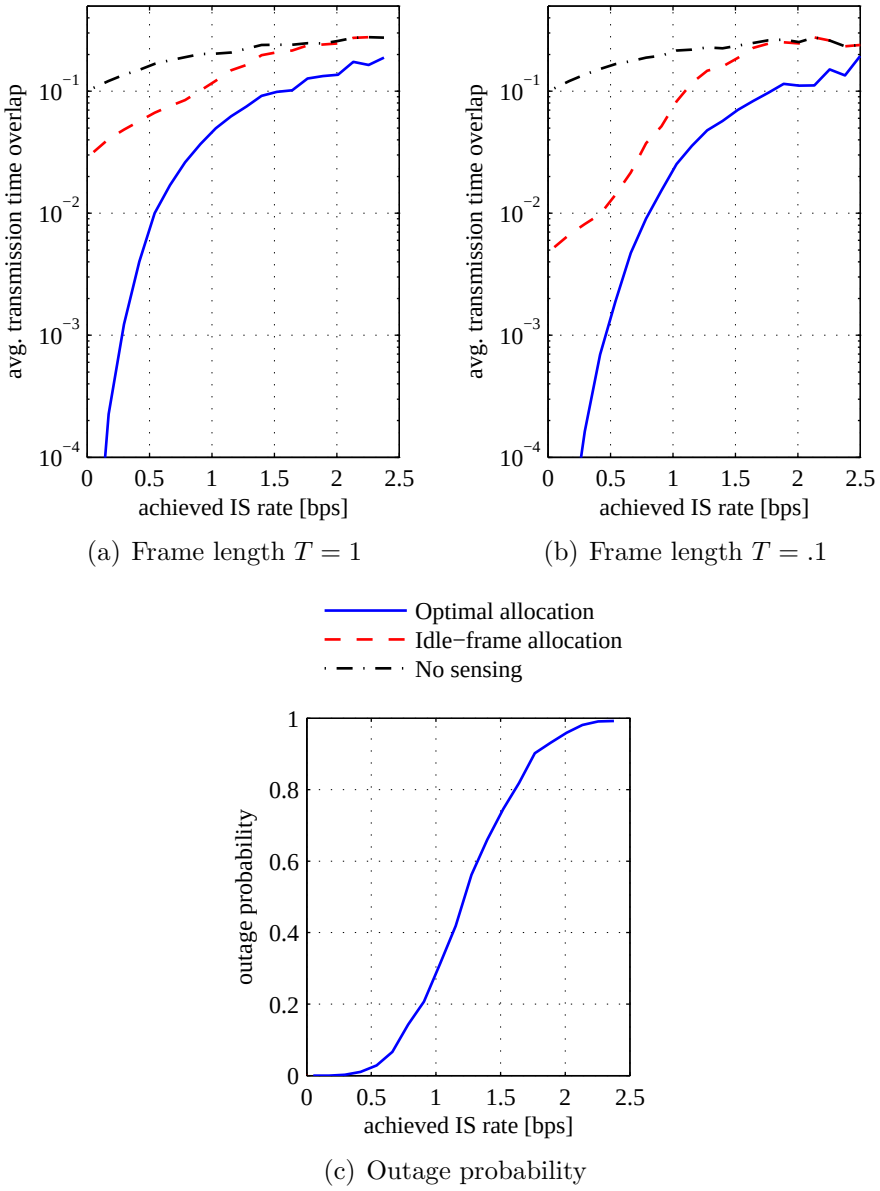
Problem **B-2** can be solved similar to problem **B-1**. In particular, by forming the Lagrangian, introducing Lagrange multipliers γ and ϵ, and taking the derivative with respect to the decision variables, we obtain a similar solution structure as in (4.44) and (4.45)-(4.47).

Reference Schemes

Having incorporated random sensing outcomes into our formulation, we introduce two reference schemes in order to put the performance of the optimal resource allocation in perspective.

No Sensing Case. As a first benchmark, consider an approach that allocates power but does not perform any transmission time optimization. This case corresponds to conventional resource management in IS systems, which simply allocates power to make best use of the uplink channel. Mathematically this is formulated as minimizing $\sum_n p_n$ subject to the constraints (4.35)-(4.38). We assume that for any subchannel with $p_n > 0$, the subchannel is used for the entire frame duration. Unused carriers for which $p_n = 0$ are not allocated any transmission time.

Idle-Frame Allocation. Another possible reference scheme performs spectrum sensing but allocates resources in a suboptimal way. Specifically, consider allocating the entire frame by setting $\rho_n = 1$ for all idle subchannels while completely avoiding busy subchannels by setting $\rho_n = 0$ for all n with $y_n = 1$. In the average rate formulation, this method can be formulated mathematically as minimizing



(a) Frame length $T = 1$

(b) Frame length $T = .1$

(c) Outage probability

Figure 4.7: Performance of optimal average-rate resource allocation and comparison with suboptimal reference schemes.

$\sum_{\mathbf{y}\in\mathbb{Y}} \eta_{\mathbf{y}} \sum_n p_{n,\mathbf{y}}$ subject to (4.52)-(4.55) and the additional constraint that subchannels with busy sensing outcome are never allocated. Note that the above optimization problem may be infeasible even when **B-2** is feasible because we are imposing the additional restriction of never transmitting during busy frames. To ensure that the reference scheme is always feasible when **B-2** is, we force allocation

to busy channels if the resulting optimization would otherwise be infeasible.

Numerical Results

This section presents numerical performance results for the optimal average-rate resource allocation and compares them to the reference schemes introduced in this section. The results are obtained for $N = 5$ subchannels and a single ad-hoc subband $M = 1$. The prediction parameters are $\lambda = \mu = 1\,\mathrm{s}^{-1}$ and the IS channel coefficients are flat Rayleigh fading and statistically independent. We further assume a block fading scenario in which the IS channel varies slowly compared to the frame duration.

The performance for fixed IS channel and random sensing results is shown in Figure 4.7(a) for $T = 1\,\mathrm{s}$ and in Figure 4.7(b) for $T = 0.1\,\mathrm{s}$. The plot shows the average transmission overlap between IS and ad-hoc network versus the achieved IS rate (note that the achieved IS rate and not the rate constraint is plotted). The performance results are averaged over 100 realizations of the IS channel. Since the IS channel is not modeled statistically, it is inevitable that for some realizations of $\boldsymbol{\beta}$ problem **B-2** is infeasible. The outage probability, which is identical for all three schemes, is therefore shown in Figure 4.7(c) to put the results in perspective. Typical outage probabilities of approximately 10% correspond to an IS rate of about 0.7 bps. At this rate, plots (a) and (b) show that a significant performance gain is achieved by performing sensing-based transmission time allocation.

The performance ordering reflects our expectations. The idle-frame allocation scheme outperforms the no-sensing case but shows a quite significant gap with respect to the optimal allocation, especially for low IS rates. Further, all curves show increasing interference as the IS rate increases. This is expected, since high

IS rates prevent the IS system from being able to accommodate the ad-hoc links. The plots also show that idle-frame allocation and no-sensing scheme converge for high IS rates, because allocating only idle frames is almost always infeasible (and therefore busy frames typically need to be used as well).

By comparing Figure 4.7(a) ($T = 1$) and Figure 4.7(b) ($T = 0.1$), we observe that while the performance of the optimal scheme does not change significantly, the idle-frame reference performs much better. This is intuitive, because by reducing the frame length, it is easier to "fill up" the idle periods of the ad-hoc network. The performance of the no-sensing scheme remains unaltered and is the same in both figures.

4.3.5 Optimal Allocation for Random IS Channels

Interference can be reduced by relaxing per-frame rate constraints to the long-term average. The previous section assumed a fixed $\boldsymbol{\beta}$ and averaged over the random behavior of the ad-hoc links (and therefore over the random sensing outcomes). This section further extends the analysis to the case of random IS channel coefficients $\boldsymbol{\beta}$. The optimization problem **B-3** is given by

$$\min_{\substack{\mathbf{p}(\mathbf{y},\boldsymbol{\beta}) \\ \boldsymbol{\rho}(\mathbf{y},\boldsymbol{\beta})}} \int_{\boldsymbol{\beta}} \sum_{\mathbf{y}\in\mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} \phi_{n,y_{g(n)}}(\rho_n(\mathbf{y},\boldsymbol{\beta}))dF(\boldsymbol{\beta}) \tag{4.56}$$

$$\text{s.t.} \int_{\boldsymbol{\beta}} \sum_{\mathbf{y}\in\mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} \rho_n(\mathbf{y},\boldsymbol{\beta}) \log(1+\tfrac{p_n(\mathbf{y},\boldsymbol{\beta})\beta_n}{\rho_n(\mathbf{y},\boldsymbol{\beta})})dF(\boldsymbol{\beta}) \geq R \tag{4.57}$$

$$\int_{\boldsymbol{\beta}} \sum_{\mathbf{y}\in\mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} p_n(\mathbf{y},\boldsymbol{\beta})dF(\boldsymbol{\beta}) \leq P \tag{4.58}$$

$$p_n(\mathbf{y},\boldsymbol{\beta}) \geq 0, \quad \forall \mathbf{y} \in \mathbb{Y}, 1 \leq n \leq N \tag{4.59}$$

$$0 \leq \rho_n(\mathbf{y},\boldsymbol{\beta}) \leq 1, \quad \forall \mathbf{y} \in \mathbb{Y}, 1 \leq n \leq N,, \tag{4.60}$$

where the decision variables $\mathbf{p}(\mathbf{y}, \boldsymbol{\beta})$ and $\boldsymbol{\rho}(\mathbf{y}, \boldsymbol{\beta})$ correspond to the power and transmission time allocation that is used for sensing outcome $\mathbf{y}$ and IS channel condition $\boldsymbol{\beta}$ and $F(\boldsymbol{\beta})$ is the cumulative distribution function of $\boldsymbol{\beta}$.

By again forming the Lagrangian and computing the derivative with respect to the decision variables, it is easy to show that the structured solutions (4.42) and (4.45)-(4.47) again hold. Therefore, it is again possible to express the allocation as a function of the Lagrange multipliers $[\gamma, \nu]$. The rate constraint can then be evaluated by

$$\int_{\boldsymbol{\beta}} \sum_{\mathbf{y} \in \mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} \rho_n(\mathbf{y}, \boldsymbol{\beta}) [\log(\nu \beta_n)]^+ dF(\boldsymbol{\beta}) \tag{4.61}$$

and the allocated sum power is given by

$$\int_{\boldsymbol{\beta}} \sum_{\mathbf{y} \in \mathbb{Y}} \eta_{\mathbf{y}} \sum_{n} \rho_n(\mathbf{y}, \boldsymbol{\beta}) \left(\nu - \tfrac{1}{\beta_n}\right)^+ dF(\boldsymbol{\beta}). \tag{4.62}$$

While the above integrals can only be evaluated numerically, it is possible to again find the optimal solutions via the bisection method.

Numerical Results

The performance for average IS channel coefficients and random ad-hoc behavior is shown in Figure 4.8 which compares the solution of **B-3** with the same reference schemes. We can observe that by exploiting the channel variability and allocating across frequency and time, we can further reduce interference. Otherwise, the performance trends are similar to those of Figure 4.7. Note that idle-frame allocation does not achieve the same channel capacity as the optimal scheme because it only transmits in frames with an idle sensing result.

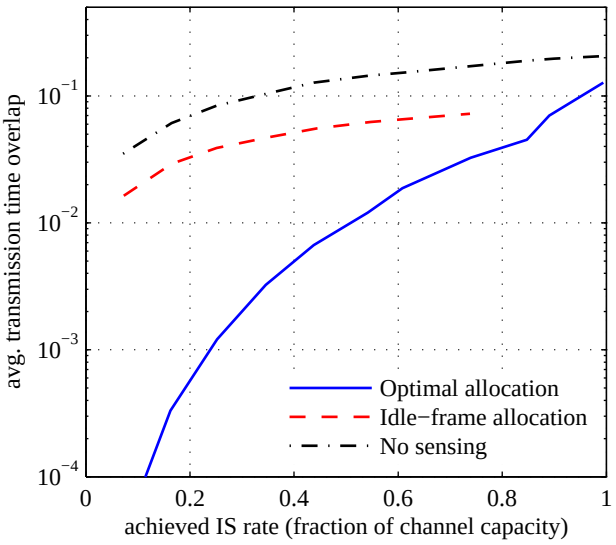


Figure 4.8: Performance result for random ad-hoc behavior and random IS channel coefficients.

4.3.6 Allocation for Multiple IS Users

The previous section derived the optimal power and transmission time allocation assuming that an orthogonal set of subchannels has already been assigned to each IS user. This enabled us to consider each of the terminals individually and perform resource allocation based on local sensing results.

In practice, the IS base station needs to assign subchannels to each of the IS users without knowing what their sensing outcomes will be. We therefore consider the problem of optimal subchannel allocation based on minimizing average interference metrics. This leads to a similar formulation as compared to the average interference case in Section 4.3.4. Once a subchannel allocation has been computed and fed back to the IS users, they can use the locally available sensing results to optimize their medium access.

The optimal subchannel allocation is a combinatorial problem, which is com-

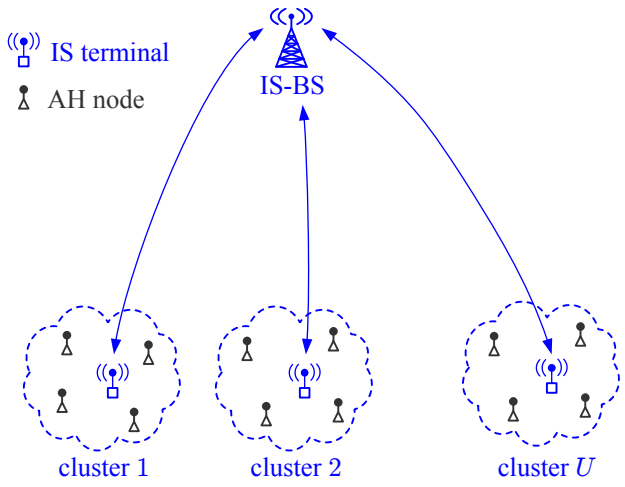


Figure 4.9: System setup for the multi-terminal case.

putationally more challenging than problems **B-1** through **B-3**, which could be analyzed based on convex optimization. While a general analysis of this problem goes beyond the scope of this work, we find the optimal solution for fairly small problem instances by exhaustive search. A comparison with heuristic allocations suggests that efficient greedy subchannel allocation algorithms developed for related problem setups can be adopted to this problem and yield a performance close to optimal.

Optimal Subchannel Allocation

The problem of optimal subchannel allocation involves assigning orthogonal sets of subchannels to each terminal, such that the overall interference is minimized; see Figure 4.9. A mathematical formulation can be based on problem **B-2**. Specifically, define $f(\mathbb{A})$ as the optimal value of **B-2** where the summations over subchannels are restricted to $n \in \mathbb{A}$,*i.e.*, the subchannels on which a specific IS user operates. Define $f(\mathbb{A}) := \infty$ if **B-2** is infeasible.

Assume that the IS base station is serving a total of U users. The problem of optimal subchannel allocation is then formulated as

$$\min_{\mathbb{A}_u} \sum_{u=1}^{U} f(\mathbb{A}_u) \tag{4.63}$$

$$\text{s.t.} \mathbb{A}_1 \cup \cdots \cup \mathbb{A}_U = \{1, \ldots, N\} \tag{4.64}$$

$$\mathbb{A}_i \cap \mathbb{A}_j = \emptyset \quad i \neq j, \tag{4.65}$$

where $\mathbb{A}_i$ denotes the set of subchannels assigned to terminal i. Due to (4.64) and (4.65) the subchannel allocations are mutually exclusive and collectively exhaustive. Note that the above problem does not require knowledge of the sensing outcome at the individual terminals. The base station only requires knowledge of the CTMC parameters λ and μ, as well as knowledge of the IS channel coefficients $\boldsymbol{\beta}$.

The above problem is difficult to solve due to its combinatorial nature and conventional subchannel allocation methods are not easily extended to incorporate the additional dimension of allocating transmission time. For small problem instances, however, the optimal allocation can be found by exhaustive search.

Suboptimal Algorithm

The problem of optimal subchannel allocation in multicarrier systems has been well-studied in the absence of transmission time allocation. Standard methods typically minimize the total transmit power subject to rate constraints. In our setup, this can be formulated mathematically as minimizing $\sum_{u=1}^{U} \sum_{n \in \mathbb{A}_u} p_n$ subject to rate and power constraints for the individual terminals. While the resulting optimization problem is still combinatorial, efficient approximation techniques have been developed with close-to-optimal performance [70].

Allocating subchannels in this manner can be used as an effective heuristic. Since good channel quality results in lower average transmission time, we conjecture that conventional subchannel allocation may be a good approximation to the optimal interference-aware subchannel allocation. Numerical results show that this is indeed the case in the scenarios we have examined.

Numerical Results

Numerical results for the multi-terminal case are shown in Figure 4.10. The total average transmission time overlap (summed over all IS users) is plotted with respect to the rate constraint for each individual IS users (constraints are assumed to be identical). The performance trends are the same as in the case of a single IS user. For low rate requirements we can effectively mitigate interference by assigning resources judiciously. On the other hand, as rate requirements become more stringent, there is less flexibility in accommodating the ad-hoc links. The scenario plotted in Figure 4.10 corresponds to $U = 3$ terminals, $N = 5$ subchannels, and flat Rayleigh fading IS channel coefficients.

The performance of the optimal and suboptimal subchannel allocation schemes is very similar regardless of the rate constraint. This suggests that minimizing the total transmission power is a reasonable approximation to the optimal subchannel allocation.

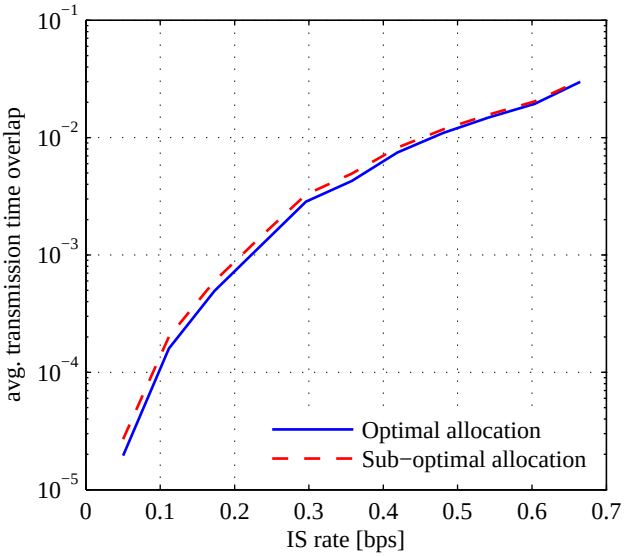


Figure 4.10: Performance result for the multi-terminal case. The sub-optimal allocation of subchannels achieves almost achieves the optimal performance found through exhaustive search.

4.4 Summary

In summary, we addressed cognitive coexistence among infrastructure and ad-hoc networks. Based on the superior resources and flexibility of the infrastructure system we developed resource allocation methods in which the infrastructure system adapts to the ad-hoc network despite retaining priority. The hierarchial structure is incorporated through minimizing interference subject to rate constraints for the infrastructure clients.

We have analyzed two coexistence scenarios within this framework. First, we addressed the case where infrastructure terminals have knowledge of the interference channel and derived the optimal power allocation. Second, we considered the case of unknown interference channels and improved coexistence by allocating power and transmission time such as to minimize the temporal overlap between both systems.

The analysis of both scenarios was based on convex optimization. By initially formulating the problems at the frame-level, insights into the solution structure could be obtained. The methods were then extended to the case where rate constraints are only met on the long-term average. This additional flexibility enables allocations across both frequency and time and leads to further interference reduction.

APPENDIX 4.A

PROOF OF LEMMA 1

We denote the IS transmissions within the current frame by a finite set of closed and disjoint time intervals $\mathbb{I}_k = [a_k, b_k]$ where each $\mathbb{I}_k \subseteq [0, T]$ corresponds to a contiguous transmission of the IS user. Clearly, this formulation incorporates possible pauses between IS transmissions. We also require $\sum_k b_k - a_k = \rho T$, because a total of ρT transmission time needs to be allocated.

First, consider the case of an idle sensing result at the beginning of the frame, say at time $t = 0$. Then according to (4.31) and (4.29), the expected time overlap is given by

$$\frac{1}{T}\sum_k \int_{a_k}^{b_k} \Pr(X(\xi) = 1 | X(0) = 0) d\xi = \frac{1}{T}\sum_k \int_{a_k}^{b_k} \frac{\lambda}{\lambda + \mu}\left(1 - e^{-(\lambda+\mu)\xi}\right) d\xi. \tag{4.66}$$

Since the integrand is strictly increasing in ρ, the above expression is minimized by transmitting contiguously during the time interval $[0, \rho T]$.

In the case of a busy sensing result, an equivalent approach leads to a strictly decreasing integrand and therefore it is optimal to transmit during the time interval $[(1-\rho)T, T]$ in that case.

APPENDIX 4.B

PROOF OF LEMMA 3

In Section 4.3.2 we have defined the sum power corresponding to the pair of Lagrange multipliers $[\gamma, \nu]$ as $P(\gamma, \nu)$. Further, we showed that by keeping γ fixed and varying ν it is possible to find a $\nu^*(\gamma)$ for which the rate constraint is satisfied with equality. To simplify notation let us now define $P(\gamma)$ as the sum power associated with $[\gamma, \nu^*(\gamma)]$.

The proof that $P(\gamma)$ decreases with γ proceeds by contradiction. First, we note that $\gamma \to 0$ implies $P(\gamma) \to \infty$ due to the structure of the optimal solutions (4.45)-(4.47). Assume now that $P(\gamma)$ is not monotonically decreasing. Then, because $P(\gamma)$ is continuous, there exist two different values of γ, say γ_1 and γ_2, such that $P(\gamma_1) = P(\gamma_2)$.

Based on the KKT conditions stated in Section 4.3.2, it is easy to verify that both γ_1 and γ_2 correspond to optimal solutions of Problem **B-1** with rate constraint $R(\gamma_1, \nu^*(\gamma_1))$ and power constraint $P(\gamma_1, \nu^*(\gamma_1))$. Further, from the structure of the optimal solutions it is clear that the transmission time allocations associated with γ_1 and γ_2 must be different, that is, $\boldsymbol{\rho}(\gamma_1) \neq \boldsymbol{\rho}(\gamma_2)$. This is a contradiction, however, because Problem **B-1** has a strictly convex objective function and therefore at most one optimal solution.

CHAPTER 5

COGNITIVE FREQUENCY HOPPING TEST BED

5.1 Summary of Contributions and Related Work

In Chapter 3 we demonstrated that coexistence between wireless local and personal area networks can be improved through cognitive frequency hopping. The derivation of the protocol is based on the CTMC prediction framework and a set of assumptions on how these two systems interact. For example, we assume that the ad-hoc network does not defer to the cognitive radio's transmissions and that its temporal activity remains unaltered as long as collisions occur infrequent enough.

In order to demonstrate proof-of-concept and to validate some of these model assumptions, we develop an experimental test bed which implements the CFH protocol in real-time. The implementation details of this test bed are presented in this chapter and the measurement results are compared to our analytical findings. Based on these results, we show that there is a close match between theory and experiment.

5.1.1 Main Contribution

This chapter describes the implementation of the test bed in detail and outlines the measurement methodology that is used to obtain experimental coexistence results. The experimental approach corroborates our work and helps to identify areas that may becomes pitfalls in a complete implementation of the protocol.

Besides demonstrating proof-of-concept and validating model assumptions, the

test bed helps to understand better the dynamic interaction between the cognitive radio and the ad-hoc network. An assumption in our analytical work is that collisions between both systems do not alter the statistical behavior of the ad-hoc network. This assumption is important to ensure that our CTMC prediction model remains applicable and is justifiable as long as collisions occur fairly infrequent.

While studying the dynamic interaction between both systems is challenging from an analytical perspective, the test bed enables us to gain valuable insights into this problem. Based on the measurement methodology developed in this chapter, we can configure our setup such as to both draw a direct comparison with our analytical work (since there is no interaction between both systems we will refer to this case as the open-loop setup) as well as examine the dynamic interaction between both systems (by using the closed-loop setup).

The experimental approach requires us to focus on key aspects of the CFH protocol. As such it is necessary to clarify some limitations due to hardware and complexity constraints. First, the test bed is limited to a single cognitive transmitter; no receiver has been implemented. This limitation does not affect the ability to measure CFH's impact on the WLAN but removes the requirement of maintaining synchronization between CFH nodes. While methods such as collaborative sensing [7] or acknowledgement feedback [19] are applicable, their implementation in real-time is difficult and goes beyond the scope and budget of this work. Similarly, multi-user aspects of the cognitive radio are not addressed. However, well-studied concepts could be applied, such as Bluetooth's piconet structure [60], where a master node acts as a central controller.

Mutual interference among collocated radio systems crucially depends on the underlying propagation conditions. In this work, we focus on the worst-case col-

lision model in which any overlap in frequency and time results in a packet drop. This approximates a setup in which both devices are in close proximity and furthermore lends itself to a practical implementation.

Finally, we stress that while Bluetooth will be used frequently for conceptual comparisons, CFH is fundamentally based on periodic spectrum sensing, a task that may be hard to accommodate in current Bluetooth receivers (especially due to their low-cost design). The findings of this paper should not be seen as a straightforward extension of the standard but rather as indicators of the importance which cognitive radio concepts may play in future wireless devices.

5.1.2 Related Work and Organization

Since theoretical contributions related to CFH have already been discussed in Chapter 3, we focus on related experimental work. In [52] Jones *et al.* present a test bed for a cognitive radio that shares spectrum with WLAN channels. The developed protocol, however, is primarily based on heuristics.

The remainder of this chapter is organized as follows. Details on the test bed implementation are presented in Section 5.2 and are followed by a discussion of the measurement methodology in Section 5.3. The experimental performance results are presented in Section 5.4 and a comparison with our analytical findings is drawn.

5.2 Test Bed and Experiment Design

In this section, we describe the test bed implementation, compare the experimental design with our analytical setup, and discuss fundamental design objectives.

The test bed has been developed with the objective of validating some of the implicit assumptions made in our analytical work. This specifically includes dynamic effects between both systems that are difficult to characterize analytically. Our analytical work focuses on designing transmission probabilities, given a stochastic model for the WLAN, such that interference constraints remain met. The idea behind this modeling approach is that as long as the interference constraints are sufficiently tight, the residual interference caused to the WLAN will not impact its temporal behavior. While analytical approaches for justifying this assumption are difficult, our experimental results enable us to confirm its validity.

A block diagram of the test bed is shown in Figure 5.1. The implementation is based on a Sundance software defined radio (SDR) development kit, consisting of processing and data acquisition modules. Radio-frequency (RF) signals are down-converted using a commercial WLAN transceiver and up-converted using an Agilent vector signal generator. The baseband processing is done entirely on the SDR board. The cognitive radio's slot structure is implemented using an accurate timer, which triggers periodic interrupts with a period of $T = 625\mu s$. The analog-digital converter (ADC) is triggered at the beginning of each slot. A 1 µs block at a rate of 72 MHz is captured and passed on to the energy detector, which computes the signal's energy and compares it with a threshold. This results in a binary sensing result (idle/busy), which is used by the CFH controller to determine the cognitive radio's medium access. Depending on the outcome of the stochastic control a transmission may be initiated by using a programmable signal generator.

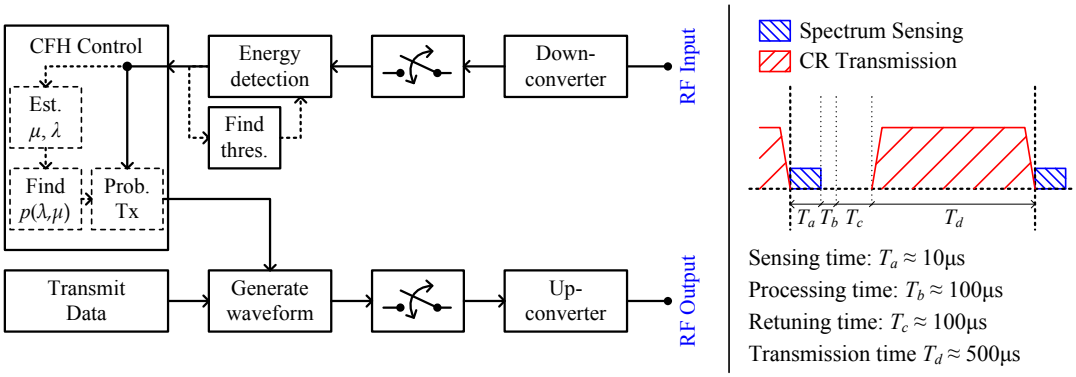


Figure 5.1: Block diagram of the test bed (left) and slot operation and timing (right).

Upon digital-to-analog (DAC) conversion, the signal is up-converted by the RF front-end.

The above operations introduce processing delays that reduce the actual transmission time per slot. Typical values for these delays are shown in Figure 5.1. Spectrum sensing relies on blocks of less than 10 µs, and is almost negligible compared to T. The processing time for the sensing result and the CFH controller together amount to roughly 100 µs in our implementation. The time it takes to re-tune the transmitter to a different channel amounts to approximately 100 µs (this delay does not occur in our setup, however, as we only deal with a single WLAN channel). The remainder of the slot can be used for the cognitive radio's transmission.

The baseband processing can be categorized into three parts, namely (i) the spectrum sensor, (ii) the CFH controller, and (iii) the CFH transmitter. In the following each component is discussed in more detail.

5.2.1 Spectrum Sensor

Spectrum sensing plays a key role in cognitive radio systems and the challenges associated with reliably detecting signals at very low SNR have been the subject of much investigation. Compared to some DSA setups, however, the burden of reliable spectrum sensing is reduced in this cognitive coexistence setup. In contrast to DSA schemes that orthogonalize systems by sufficient spatial separation (and therefore require the ability to detect weak primary signals), typical SNR values faced in this cognitive coexistence setup will be substantially larger as both systems operate in close proximity of each other.

Thanks to the fairly high SNR conditions, energy detection can be used efficiently with very little complexity. Energy detection is mathematically formulated as a binary hypothesis testing problem on a set of N samples that either represent just noise, or a signal in noise, respectively. As discussed in Chapter 2, this leads to

$$\begin{aligned} \mathcal{H}_0 \quad &: \quad Y_i = V_i, \ i = 1, \ldots, N \\ \mathcal{H}_1 \quad &: \quad Y_i = S_i + V_i, \ i = 1, \ldots, N, \end{aligned} \tag{5.1}$$

where Y_i denotes the complex baseband samples, V_i are noise samples, $V_i \sim \mathcal{CN}(0, \sigma_0^2)$, and S_i denotes the signal samples drawn from a complex Gaussian, $S_i \sim \mathcal{CN}(0, \sigma_1^2)$. This hypothesis testing problem is standard [37]. The optimal Neyman-Pearson detector is given by

$$T(\mathbf{y}) = \sum_{i=1}^{N} |y_i|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{5.2}$$

where the threshold γ needs to be chosen such that the probability of false alarm (*i.e.*, erroneously declaring a busy channel) is no larger than a specific value. The Neyman-Pearson detector then yields the optimal probability of detection.

In theory, when (5.1) holds exactly, the threshold γ can be determined analytically by finding closed-form expressions for the probability of false-alarm and detection. In the experimental domain, however, numerous other factors need to be taken into account. A fundamental problem in this work is the fact that hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ cannot be observed isolated from each other. The WLAN adapter cards cannot be configured to transmit a signal continuously and therefore, without a means of synchronizing test bed and adapter cards, some samples will be drawn during idle periods. This leads to a mixture distribution of $\mathcal{H}_0$ and $\mathcal{H}_1$ where the mixture coefficients depend on the WLAN's traffic intensity, *i.e.*, the long-run amount of idle and busy periods.

Empirical observations of the sufficient statistic (5.2) are plotted in Figure 5.2. We expect to observe a mixture of chi-square distributions because both $T(\mathbf{y}|\mathcal{H}_0)$ and $T(\mathbf{y}|\mathcal{H}_1)$ are chi-square distributed. More than two mixture components may be necessary, however, due to slightly different power levels of the WLAN terminals. Indeed, the empirical CDF can be well approximated by a mixture of three chi-square distributions,

$$f(x) = \sum_{i=1}^{3} p_i f_i(x; \alpha_i, \beta_i), \tag{5.3}$$

where $\sum_{i=1}^{3} p_i = 1$ and

$$f_i(x; \alpha_i, \beta_i) = x^{\alpha_i - 1} \frac{\beta_i^{\alpha_i} e^{-\beta_i x}}{\Gamma(\alpha_i)} \tag{5.4}$$

represents the PDF of a Gamma distribution with shape parameter α_i and rate parameter β_i. An Expectation-Maximization algorithm [86] is used to find the model parameters of (5.3) and the fitting result is shown in Figure 5.2. A good match with the empirical data is observed. We also found that busy and idle mixture components have very different rate parameters. This illustrates the very different energy levels present in idle and busy sensing slots.

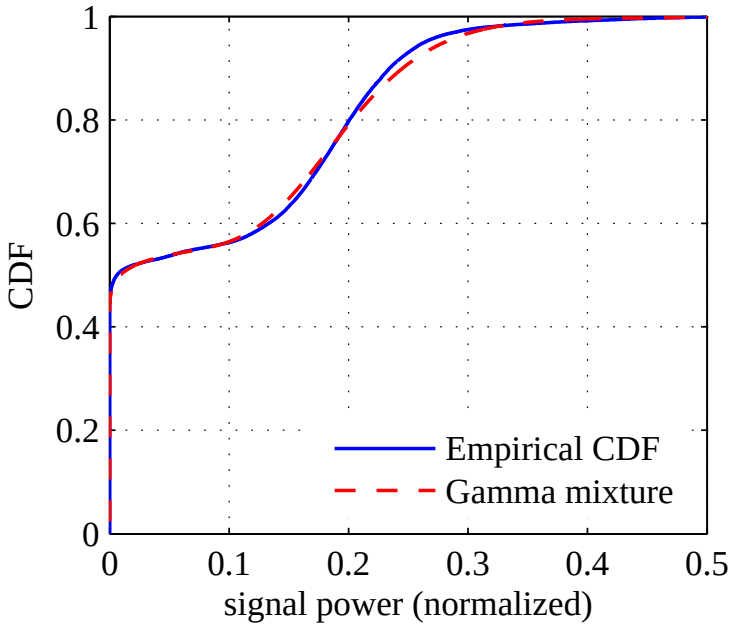


Figure 5.2: Experimental decision statistic for energy detection and its Gamma mixture fit. A good match between the model and the empirical data is observed.

Numerical performance analysis demonstrates that the test bed's spectrum sensor works reliably. By choosing the decision threshold appropriately a detection probability of 98.5% can be realized with less than 1% false alarms. The good performance of energy detection is, of course, due to the moderate to high SNR conditions, which make spectrum sensing similar to the carrier-sensing employed in systems such as IEEE 802.11.

5.2.2 Cognitive Controller

The role of the cognitive controller is to initiate a transmission probabilistically, whenever an idle sensing result is observed. Transmissions are never initiated after a busy sensing result because this would lead to a collision with high probability. A full implementation of CFH encompasses the tracking of traffic variations and

estimation of the CTMC model parameters λ and μ, the computation of the optimal transmission probabilities $p(\lambda, \mu)$, and stochastic control action (by using a random number generator). Due to the complexity associated with estimating λ and μ, the experimental test bed uses a fixed transmission probability p. Despite this limitation, the performance of CFH can be evaluated without much loss of generality. By measuring the number of successful CFH transmissions and simultaneously determining the number of WLAN packet errors, both throughput and interference can be quantified. Thus, even though the transmission probability remains fixed, it is possible to project the system's behavior for varying transmission probability.

5.2.3 Cognitive Transmitter

If a cognitive radio transmission is initiated, it lasts for the remainder of the slot duration. For the CFH operation it is not relevant what specific transmission scheme is used. For optimal usage of the white space between consecutive packets the cognitive radio could use the same frequency bands as the WLAN. Motivated by the conceptual similarity with Bluetooth/WLAN coexistence, however, we designed the transmitter to resemble that of Bluetooth. It therefore transmits in narrowband channels of 1 MHz and similar modulation parameters [60]. Gaussian Frequency Shift Keying (GFSK) with a time-bandwidth product of 0.3 was used at a symbol rate of 1 MSps.

The test bed's transmitter is implemented based on a programmable signal generator which is integrated into the acquisition module, and can be triggered in software. Data contained in an internal buffer is then transferred to the DAC and played back in an infinite loop.

5.3 Measurement Methodology

The previous section described the test bed's implementation. As CFH relies on sensing and predicting packet collisions, its performance naturally depends on the specifics of the coexistence setup, such as propagation conditions, traffic intensity, system parameters, etc. This section is devoted to describing the measurement methodology that underlies the performance assessment.

5.3.1 Hardware Setup

The experimental coexistence setup is depicted in Figure 5.3 and consists of the WLAN system (composed of an access point and three workstations), the cognitive radio, as well as a vector signal analyzer which was used to monitor the operation of the system.

Two fundamentally different configurations are considered. In the *open-loop* setup the cognitive radio's output is not fed back to the WLAN system but only used to determine the packet error rate. While this does not reflect what would occur in practice, this setup enables us to draw a direct comparison with our analytical results. In the *closed-loop* setup, on the other hand, interference impacts the WLAN and leads to frequent retransmissions. We analyze the cognitive radio's impact and relate the results to the open-loop setup.

WLAN Configuration

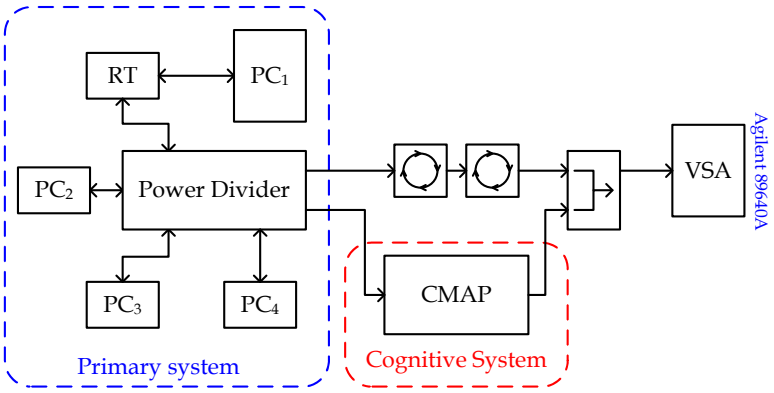
The WLAN consists of commercial off-the-shelf IEEE 802.11 devices and includes an access point and three workstations with adapter cards. The access point is

connected to a fourth workstation using a wired ethernet connection. All wireless devices are configured identically to operate in channel 6 (corresponding to a center frequency of 2.437 GHz) and use a transmission rate of 5.5 Mbps.

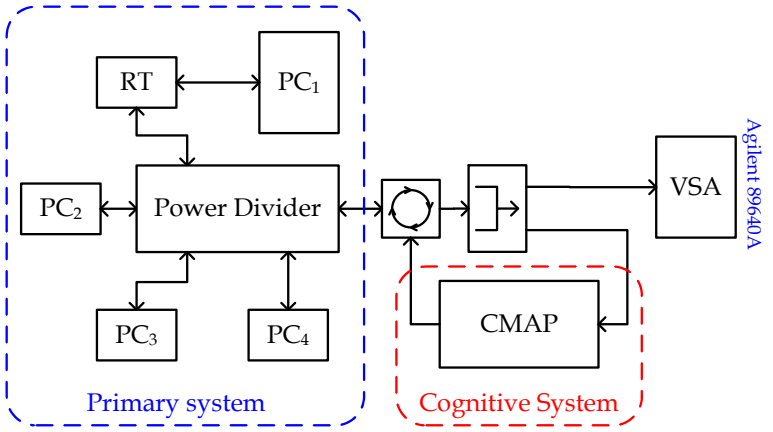
The wireless devices' RF outputs are all connected to a resistive power divider using coaxial cables. This isolates the transmissions from the environment and reduces interference that could otherwise result from unrelated transmissions in the unlicensed bands (measurements were taken in an office building with a number of unrelated WLAN access points). In addition, this configuration removes any propagation effects and allows for repeatable results. While the propagation conditions encountered in practice will deviate from this idealized setup, our results correspond to a worst-case scenario in which packet overlaps inevitably result in collisions.

Open-Loop Setup. The open-loop configuration is shown in Figure 5.3(a). While the combined WLAN signals serve as the input to the cognitive radio, its output is not fed back to the WLAN system. Instead it is combined with the WLAN signal and detected by a separate workstation computer with a WLAN adapter card. The two circulators isolate the output of the test bed and prevent it from impacting the WLAN.

The workstation capturing the combined signal of the WLAN and the test bed uses an adapter card in promiscuous mode together with commercial WLAN analysis software to detect WLAN packets. The output power level of the cognitive radio is set large enough such that a collision between both systems will prevent the adapter card from successfully receiving the packet (either an error will be displayed or the packet will be missed altogether, depending on whether the synchronization



(a) Open-loop setup



(b) Closed-loop setup

Figure 5.3: Open-loop and closed-loop measurement setups.

preamble or only the payload is affected). The rate of successful packet reception is thus measured and, by comparison with settings of the traffic generators, the rate of packet losses can be inferred.

Closed-Loop Setup. The closed-loop setup is depicted in Figure 5.3(b). The test bed again receives the combined WLAN signal, but its output is now connected to the same power-divider as the WLAN devices. Therefore, packet collisions lead to packet drops at the WLAN devices themselves (and not at a reference station as in the open-loop setup). Packet loss will therefore initiate WLAN retransmis-

sion, which may in turn affect the test bed's performance. The closed-loop setup therefore allows for dynamic interaction between both devices.

In the closed-loop setup, PC_1, which is connected to the WLAN access point through a wired connection, runs traffic analysis software. The traffic generators are configured such that PC_1 is the intended receiver, and hence the successful portion of the WLAN traffic (including any retransmissions that may occur) is measured. By comparing with the case of no interference, the packet loss rate is inferred.

5.3.2 Traffic Characteristics

WLAN transmissions are generated by using a traffic generator on each workstation. The Distributed Internet Traffic Generator (D-ITG) [36] is used and allows to specify packet lengths, the distribution of inter-arrival times, and transmission rates. The traffic properties form an important component of the measurement methodology.

This work focuses on UDP traffic with constant payload of 1024 bytes and plots performance with respect to traffic intensity. Specifically, we define the WLAN traffic load σ, which is normalized such that $\sigma = 0$ corresponds to an inactive WLAN and $\sigma = 1$ represents a WLAN operating at full capacity.

Experimentally, the WLAN traffic load σ is measured as follows. First, the rates of all traffic generators are set to a value that is too large to be supported by the WLAN (even in the case of no interference) and by measuring the actual rate, the WLAN capacity is found. Then, the settings of the traffic generator are normalized accordingly, leading to values $0 \leq \sigma \leq 1$. For example, given the traffic

and propagation settings in our setup, a maximum of approximately 450 packets per second could be supported by the WLAN. Configuring each of the traffic generators to transmit at a rate of 50 packets per second therefore corresponds to $\sigma = 1/3$.

The measurements focus on the average throughput and interference for stationary traffic scenarios with different intensities σ. Measurement results are compared with simulations using the model parameters in Table 5.1. These parameters were obtained based on statistical analysis, as discussed in Chapter 2. The results can be extended to non-stationary traffic setups, provided that parameters of the traffic model are tracked over time (see Section 2.5).

5.3.3 Measurement Process

The measurements are performed in the following manner. First, with the cognitive radio portion of the test bed turned off, the WLAN traffic generators are adjusted such that a specific traffic load σ is realized. The cognitive radio is then enabled and the successfully received WLAN packets are counted. By comparing with the nominal packet rate in the interference-free case, the packet error rate can be obtained. At the same time, the cognitive radio keeps statistics of the number of initiated and successful slot transmissions. Lacking a CFH receiver, a successful CFH slot transmission is defined as initiating a slot transmission *and* observing an idle channel at the beginning of the next slot. Due to the reliable spectrum sensing performance and the fact that, in our setup, WLAN packets are always longer than the slot duration T, this is a valid metric.

5.4 Performance Results

This section presents experimental performance results and compares them to simulations results obtained using the SMM and CTMC models discussed in Chapter 2. For the open-loop setup, where theoretical and experimental results should coincide, we observe an excellent match. In the closed-loop setup, where mandatory re-transmissions affect the experimental performance results, we introduce a simple heuristic formulation that allows us to approximate the WLAN's behavior based on open-loop results.

The performance assessment in this section focuses on the cognitive radio's throughput and the interference it causes (in terms of WLAN packet errors induced by the interference). Clearly, both performance metrics are interrelated: by increasing the cognitive radio's transmission probability p we can increase throughput at the expense of a larger number of collisions and vice versa. By measuring both metrics, we can get a better insight on how aggressive the cognitive radio transmission policy can be without significantly affecting WLAN performance.

We compare the performance of CFH with a blind reference scheme that neither detects nor predicts WLAN activity but obliviously initiates transmissions with probability p_r regardless of the state of the medium. Our results demonstrate that CFH introduces a significant performance gain and increases cognitive radio throughput while reducing WLAN interference.

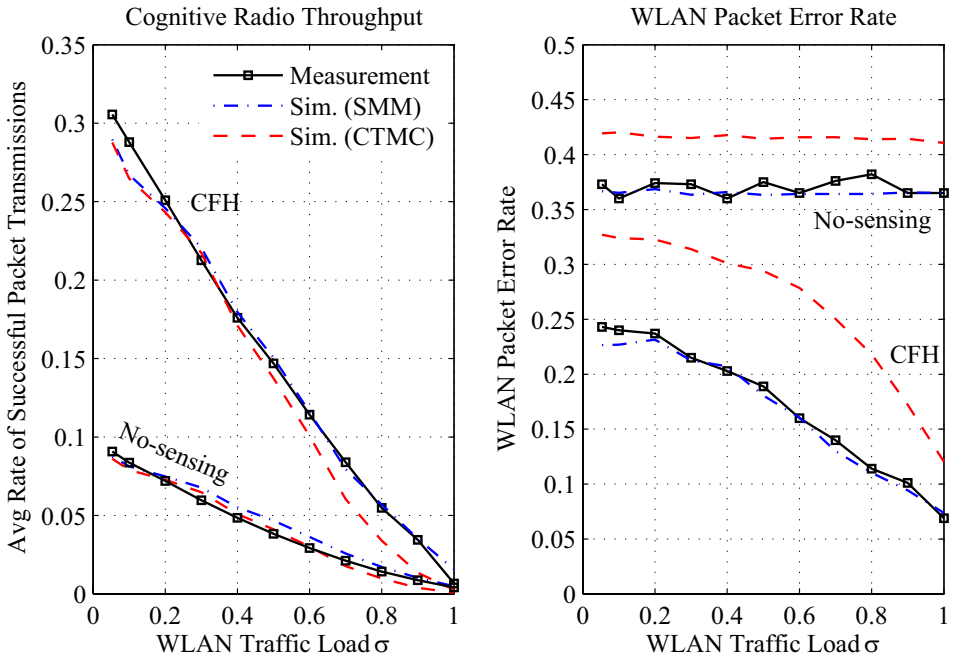


Figure 5.4: Open-loop performance result. The cognitive radio throughput (left) and WLAN packet error rate (right) are shown. The measurement results match closely with semi-Markov model (SMM) based simulations. The results obtained under the continuous-time Markov chain (CTMC) model show a larger deviation.

5.4.1 Open-Loop Measurement Result

The open-loop measurement result is shown in Figure 5.4. The left panel shows the cognitive radio throughput in terms of the expected number of successful slot transmissions per unit time and the right panel shows the packet error rate of the WLAN. We stress that the experimental curves have been obtained by counting the rate of successfully received WLAN packets and the packet error rate is computed by comparing with the average number of packets that should have been received during that time period. The results are compared with simulations based on the SMM and CTMC model. The SMM-based simulation also incorporates the fact that the cognitive radio does not use the entire slot for transmission and should therefore approximate the measurement results with good accuracy.

The SMM-based simulation results indeed show an excellent match with the experimental results. The throughput curves shown in the left panel of Figure 5.4 almost coincide (the largest aberration amounts to less than two percentage points) and for the packet error rate (shown on the right) we observe a maximum abberation of two percentage points. By comparison, the CTMC model is less accurate. While the throughput results match fairly well, the predicted packet error rate deviates significantly. From a modeling standpoint, this is not surprising because the exponential approximation does not capture the WLAN's contention behavior.

Similar performance trends are observed for the reference scheme without spectrum sensing. As shown in Figure 5.4, measurement and SMM-based simulation again match very well while the CTMC approximation shows noticeable deviation in terms of predicted packet error rate.

5.4.2 Closed-Loop Measurement Result

The results for the closed-loop setup are shown in Figure 5.5. Here, the WLAN terminals are strongly interfered with by the cognitive radio and therefore initiate retransmissions whenever a packet is dropped. If a retransmission attempt is ultimately successful, the packet is counted as successfully received (the retransmission packets themselves are not counted toward the WLAN traffic load).

Compared to the open-loop setup, the throughput of the cognitive system stays approximately the same, but the interference to the WLAN changes drastically due to the retransmission behavior. Up to approximately $\sigma = 0.9$, no packet loss is observed. At $\sigma = 1.0$, it increases to roughly 5%.

The non-cognitive reference scheme exhibits a similar behavior. No packet

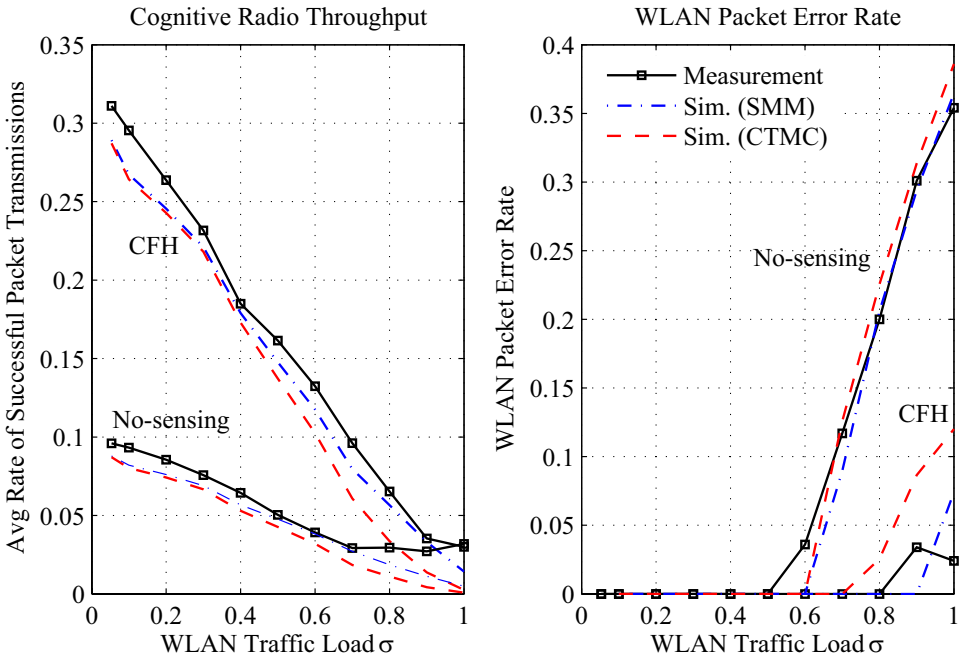


Figure 5.5: Closed-loop performance result. Cognitive radio throughput (left) and WLAN packet error rate (right) are shown. The packet error rate includes retransmissions and is therefore smaller than in the open-loop case.

errors occur for $\sigma = 0.5$ or below, and an approximately linear increase is observed for higher values of σ. The packet error rate reaches a maximum of roughly 35% at $\sigma = 1.0$.

The reason for the reduced packet errors lies in the WLAN's retransmission behavior. If a packet transmission fails, the standard mandates that a retransmission be initiated. At low WLAN rates, it is likely that these retransmissions will be successful because the medium is predominantly idle. At high rates, however, it may no longer be possible to accommodate such retransmissions, leading to an increasing packet error rate.

In order to compare experimental and theoretical results, we consider a simple approximation that incorporates the retransmission behavior. For simplicity, we

assume that packets are retransmitted indefinitely until they are received successfully. A traffic load σ encountering a collision probability q leads to the cumulative traffic load

$$\sigma' = \sigma(1 + q + q^2 + \cdots) = \frac{\sigma}{1 - q}. \tag{5.5}$$

We assume that as long as the traffic load σ' is below the capacity of the WLAN channel $\bar{\sigma}$, no packet loss occurs because retransmissions can be accommodated. However, for $\sigma' > \bar{\sigma}$ this is no longer possible and there will be a packet error rate of approximately

$$\frac{\sigma' - \bar{\sigma}}{\sigma'}. \tag{5.6}$$

This simple heuristic approximation is used to compare closed and open-loop results. Clearly, in order to compute (5.6) only open-loop performance result (obtained by measurement or simulation) are needed. Based on these it is possible to approximate the packet error rate in the closed-loop setup. The performance curves gathered in this way are shown in Figure 5.5 and again show cognitive radio throughput (left) and packet error rate (right). The measurement curves are obtained directly by measurement from the closed-loop setup. The simulation curves are obtained based on open-loop simulations with the SMM and CTMC models and applying (5.6). We observe a good match with the measurement results for the closed-loop case suggesting that the simple heuristic predicts the closed-loop behavior quite accurately.

Lastly, the results show that even at high traffic load, there exists a residual throughput of the cognitive radio system. This appears to be the result of WLAN's retransmission behavior, which due to frequent collisions enlarges some contention windows to accommodate other stations. As can be seen in the left panel of Figure 5.5, this results in a residual throughput of approximately 0.03 even when

the WLAN is fully loaded.

5.5 Performance Trends and Tradeoffs

Based on the performance results presented in the previous section, we discuss some tradeoffs and challenges which may arise in a complete implementation of such a system.

The choice of system parameters in this chapter is motivated by facilitating a comparison between theory and practice. Consequently, we focus on a worst-case propagation scenario in which any time overlap between transmissions results in packet errors. This approximates the case in which the devices are located in close proximity.

An important design parameter in trading off throughput versus interference is the transmission probability p. Clearly, cognitive radio throughput increases with p, but so does the interference that is inflicted upon the WLAN. A natural design approach is to choose p based on the traffic parameters such that a specific constraint on the WLAN packet error rate is met with equality. In the multi-channel case the optimal vector of transmission probabilities can be found through decision-theoretic analysis, which leads to a linear programming solution with fairly low implementation complexity [19]. Further, transmission probabilities only need to be recomputed whenever the prediction parameters change significantly; on a slot level, the random access can be implemented by storing the vector of transmission probabilities in a lookup table.

A tradeoff arises in selecting the slot length. If there was no overhead associated

with sensing and re-tuning the cognitive radio, reducing the slot duration would enable us to more efficiently "fill up" the idle periods between packets. Due to this overhead, however, choosing a small slot duration leads to small cognitive radio payload and consequently reduced performance. Because of conceptual similarity we have chosen a slot length of 625 µs, which equals the value used in Bluetooth. Further, we have found that around this value, the throughput of the cognitive radio changes very little.

Throughout this chapter we have focused on stationary traffic scenarios with varying WLAN traffic intensity. This enables us to measure throughput and interference by recording long packet traces and computing time averages. Nevertheless, we believe that our results will extend to the non-stationary traffic scenarios observed in practice. In fact, the time scale of our prediction model is in the order of tens of milliseconds, which is much smaller than the typical time scale associated with changes in usage patterns. Tracking non-stationarities by adapting model parameters is therefore a viable approach (see Chapter 2 for further details).

While our comparison between theory and experiment shed some light on implementation aspects, our work represents the first steps toward developing a fully functional prototype. Synchronization is one aspect that goes beyond the scope of this work. Due to the dependence on local sensing results, synchronization is more difficult to achieve than in related AFH setups. Collaborative sensing concepts are a possible solution approach. By exchanging sensing metrics, the detection process can be coordinated, and the stochastic control actions can be synchronized by using identical random seeds. Methods such as acknowledgement feedback are also applicable [20].

Table 5.1: Prediction model parameters for modified semi-Markov fit. The idle periods are approximated by a mixture of two generalized Pareto distributions, $F_m(t) = p_1 F_f(t; k_1, \omega_1) + (1 - p_1) F_f(t; k_2, \omega_2)$.

		WLAN Traffic Load										
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
		Prediction Model Parameters										
CTMC	λ [ms]	23.3	11.6	7.89	5.42	3.32	2.34	1.63	1.01	0.68	0.43	0.24
	μ [ms]	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Mixture	k_1	0.75	-0.39	-0.33	-0.40	-0.52	-0.3	-0.09	0.02	0.12	0.14	0.07
	ω_1 [ms]	29.6	18.6	10.7	8.14	5.48	4.81	3.43	3.63	2.59	2.42	2.17
	k_2	0.75	-0.39	-0.33	-0.40	-0.52	-0.3	-0.09	0.02	0.12	0.14	0.07
	ω_2 [ms]	0.07	0.34	0.32	0.34	0.41	0.29	0.22	0.18	0.15	0.12	0.11
	p_1	0.11	0.33	0.27	0.34	0.41	0.54	0.62	0.76	0.82	0.88	0.95
	μ [ms]	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

5.6 Summary

In conclusion, we have studied the benefit of sensing and predicting temporal WLAN activity within a cognitive coexistence framework. Based on predicting activity patterns, we were able to efficiently use idle periods for cognitive radio transmissions, while constraining interference to the WLAN. The implementation of a real-time test bed was presented and enabled a comparison of theory and experiment. A good fit between analytical and measurement results was observed and model assumptions were validated. The results demonstrate that a promising performance gain can be achieved within this framework.

CHAPTER 6

SUMMARY AND CONCLUSIONS

In conclusion, this thesis introduces the concept of *cognitive coexistence*, which aims at reducing interference among heterogeneous wireless networks through sensing-based radio resource management. Based on detecting temporal activity and interference patterns, it enables interfering systems to implicitly coordinate their spectrum access such that coexistence is improved. Cognitive coexistence therefore contributes to orthogonalizing interfering radio systems dynamically, and promotes an efficient way of sharing spectrum among heterogeneous systems. In light of the rapid growth of wireless communication devices, this addresses an important problem in the design of emerging wireless technologies.

We study the benefits and challenges of this approach for two important coexistence scenarios. First, we analyze how a frequency hopping cognitive radio can reduce interference to a set of parallel ad-hoc bands by optimally taking advantage of idle periods between ad-hoc transmissions. Motivated by conceptual similarities with Bluetooth/WLAN coexistence, we compare the protocol to adaptive frequency hopping methods which have received significant attention recently. Our analytical and numerical performance results demonstrate that cognitive coexistence methods can reduce interference effectively and significantly outperform reference methods without sensing capabilities. Second, we analyze the coexistence of infrastructure and ad-hoc networks. Based on the superior communication resources of the infrastructure system we propose an interference-aware resource management protocol which reduces interference to the ad-hoc network subject to maintaining a specified quality-of-service level for infrastructure users. In practice, this approach may contribute to incorporating peer-to-peer functionality into

future cellular networks and ultimately provide a framework for prioritizing users based on their quality-of-service requirements.

In both scenarios, the derivation of cognitive coexistence protocols is fundamentally based on a simple ON/OFF continuous-time Markov chain model for the ad-hoc system's behavior. Through extensive experimental studies we show that this model approximates the behavior of practical systems while being simple enough to facilitate analytical derivations of resource management methods. In addition, the understanding of practical limitations and the validation of model assumptions is facilitated by a real-time test bed for the cognitive frequency hopping protocol.

This dissertation presents first steps toward leveraging the potential of cognitive coexistence techniques. Nevertheless, by comparing the proposed methods to reference schemes without sensing capabilities, our findings clearly illustrate the merits of this framework and accentuate the importance that such concepts may play in future standards. Going forward, our results motivate further studies of cognitive coexistence. In particular, future work may incorporate more elaborate prediction of interfering systems such as models in which the idle periods are not exponentially distributed, or models that go beyond a simple ON/OFF characterization and incorporate knowledge of interference channel conditions or other aspects of the primary system. Also, the case in which multiple interfering systems perform simultaneous adaptation based on sensing results has not been addressed.

BIBLIOGRAPHY

[1] J. Mitola III and G. Q. Maguire, Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications Magazine*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[2] J. Mitola, "Cognitive radio – model-based competence for software radios," Licentiate Thesis, Royal Institute of Technology (KTH), Stockholm, Sep. 1999.

[3] R. V. Prasad, P. Pawelczak, J. A. Hoffmeyer, and H. S. Berger, "Cognitive functionality in next generation wireless networks: standardization efforts," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 72–78, Apr. 2008.

[4] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[5] Q. Zhao and B. M. Sadler, "Dynamic Spectrum Access: Signal Processing, Networking, and Regulatory Policy," *IEEE Signal Processing Magazine*, vol. 55, no. 5, pp. 2294–2309, May 2007.

[6] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[7] A. Sahai, N. Hoven, S. Mishra, and R. Tandra, "Fundamental tradeoffs in robust spectrum sensing for opportunistic frequency reuse," *submitted to IEEE J. Select. Areas Commun.*, 2006.

[8] F. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II–The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1417–1433, Dec. 1975.

[9] ANSI/IEEE Standard 802.11, 1999 Edition (R2003), "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE/SA Standards Board, Tech. Rep., 1999.

[10] S. M. Ross, *Simulation*, 3rd ed. Academic Press, 2002.

[11] F. H. P. Fitzek and M. Katz, *Cellular Controlled Peer to Peer Communications: Overview and Potentials*. Springer, 2007, ch. 2, pp. 31–59.

[12] H. Claussen, L. T. Ho, and L. G. Samuel, "An Overview of the Femtocell Concept," *Bell Labs Technical Journal*, vol. 13, no. 1, pp. 221–246, 2008.

[13] S. Geirhofer, L. Tong, and B. M. Sadler, "A Measurement-Based Model for Dynamic Spectrum Access," in *Proc. IEEE Conference on Military Communications (MILCOM)*, 2006.

[14] ——, "Dynamic Spectrum Access in WLAN Channels: Empirical Model and Its Stochastic Analysis," in *Proc. First International Workshop on Technology and Policy for Accessing Spectrum*, 2006.

[15] ——, "Dynamic Spectrum Access in the Time Domain: Modeling and Exploiting Whitespace," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, May 2007.

[16] ——, "Measurement-Based Models for Cognitive Medium Access in WLAN Bands," Cornell University, Adaptive Communications and Signal Processing Group (ACSP), Technical Report ACSP-TR-02-07-02, Feb. 2007. [Online]. Available: http://acsp.ece.cornell.edu/papers/ACSP-TR-02-07-02.pdf

[17] ——, "Cognitive Medium Access: A Protocol for Enhancing Coexistence in WLAN Bands," in *Proc. IEEE Global Communications Conference (GlobeCom)*, 2007.

[18] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Optimal Dynamic Spectrum Access via Periodic Channel Sensing," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2007, pp. 33–37.

[19] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive Medium Access: Constraining Interference Based on Experimental Models," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 95–105, Jan. 2008.

[20] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Opportunistic Spectrum Access via Periodic Channel Sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 785–796, Feb. 2008.

[21] S. Geirhofer, L. Tong, and B. M. Sadler, "A Cognitive Framework for Improving Coexistence Among Heterogeneous Wireless Networks," in *Proc. IEEE Global Communications Conference (Globecom)*, Nov. 2008.

[22] ——, "Interference-aware OFDMA resource allocation: A predictive ap-

proach," in *Proc. IEEE Military Communications Conference (MILCOM)*, Nov. 2008.

[23] ——, "Cognitive coexistence between infrastructure and ad-hoc systems," *submitted to IEEE Transactions on Wireless Communications*, 2008.

[24] S. Geirhofer, J. Z. Sun, L. Tong, and B. M. Sadler, "CMAP: A Real-Time Prototype for Cognitive Medium Access," in *Proc. IEEE Military Communications Conference (MILCOM)*, Oct. 2007.

[25] ——, "Cognitive Medium Access in WLAN Bands: A Real-Time Testbed," in *Proc. Asilomar Conference on Signals, Systems and Computers*, 2007.

[26] ——, "Cognitive frequency hopping based on interference prediction: Theory and experimental results," *submitted to ACM Mobile Computing and Communications Review (MC2R)*, 2008.

[27] K. Park and W. Willinger, Eds., *Self-Similar Network Traffic and Performance Evaluation*. Wiley Interscience, 2000.

[28] Z. L. Zhang, V. J. Ribeiro, S. Moon, and C. Diot, "Small-time scaling behaviors of internet backbone traffic: an empirical study," in *Proc. IEEE Conference on Computer Communications (INFOCOM)*, vol. 3, Mar. 2003, pp. 1826–1836.

[29] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless LAN," in *Proc. ACM SIGMETRICS*, 2002.

[30] J. Yeo, S. Banerjee, and A. Agrawala, "Measuring traffic on the wireless medium: experience and pitfalls," University of Maryland, College Park, Tech. Rep., 2002.

[31] C. Phillips and S. Singh, "Analysis of WLAN traffic in the wild," in *Proc. IFIP-Networking*, 2007.

[32] M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, and J. Zahorjan, "Measurement-based Characterization of 802.11 in a Hotspot Setting," in *Proc. ACM SIGCOMM*, 2005.

[33] C. Na, J. K. Chen, and T. S. Rappaport, "Measured Traffic Statistics and Throughput of IEEE 802.11b Public WLAN Hotspots with Three Differ-

ent Applications," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3296–3305, Nov. 2006.

[34] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[35] A. Motamedi and A. Bahai, "MAC Protocol Design for Spectrum-agile Wireless Networks: Stochastic Control Approach," in *Proc. IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Apr. 2007, pp. 448–451.

[36] S. Avallone, A. Botta, D. Emma, S. Guadagno, and A. Pescape, "D-ITG V.2.4 Manual," University of Napoli "Federio II", Tech. Rep., Dec. 2004.

[37] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer-Verlag, 1994.

[38] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions*, M. Abramowitz and I. A. Stegun, Eds. National Bureau of Standards, 1972.

[39] ANSI/IEEE Standard 802.11b-1999 (R2003), "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4GHz band," IEEE SA Standards Board, Tech. Rep., 1999.

[40] R. B. D'Agostino and M. A. Stephens, *Goodness-of-fit techniques*. Marcel Dekker, Inc., 1986.

[41] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[42] A. C. Cohen, *Truncated and Censored Samples. Theory and Applications*. Marcel Dekker, Inc., 1991.

[43] S. Kotz and S. Nadarajah, *Extreme Value Distributions. Theory and Applications*. Imperial College Press, 2000.

[44] A. Feldmann and W. Whitt, "Fitting Mixtures of Exponentials to Long-Tail Distributions to Analyze Network Performance Models," in *Proc. IEEE Con-*

ference on Computer Communications (INFOCOM), vol. 3, Apr. 1997, pp. 1096–1104.

[45] A. Thümmler, P. Buchholz, and M. Telek, "A Novel Approach for Fitting Probability Distributions to Real Trace Data with the EM Algorithm," in *Proc. International Conference on Dependable Systems and Networks (DSN)*, Jun. 2005, pp. 712–721.

[46] R. El Abdouni Khayari, R. Sadre, and B. Haverkort, "Fitting World-Wide Web Request Traces with the EM-algorithm," *Performance Evaluation*, vol. 52, no. 2, pp. 175–191, Apr. 2003.

[47] F. J. Massey Jr., "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

[48] P. E. Heegaard, "GenSyn - a Java based generator of synthetic Internet traffic linking user behaviour models to real network protocols," in *Proc. ITC Specialist Seminar on IP Traffic Measurement, Modeling, and Management*, 2000.

[49] Q. Zhao, L. Tong, and A. Swami, "Decentralized Cognitive MAC for Dynamic Spectrum Access," in *Proc. First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 224–232.

[50] Y. Chen, Q. Zhao, and A. Swami, "Joint Design and Separation Principle for Opportunistic Spectrum Access in the Presence of Spectrum Errors," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[51] A. N. Mody, S. R. Blatt, D. G. Mills, T. P. McElwain, McElwain, N. B. Thammakhoune, J. D. Niedzwiecki, M. J. Sherman, C. S. Myers, and P. D. Fiore, "Recent Advances in Cognitive Communications," *IEEE Communications Magazine*, vol. 45, no. 10, pp. 54–61, Oct. 2007.

[52] S. D. Jones, N. Merheb, and I.-J. Wang, "An experiment for sensing-based opportunistic spectrum access in CSMA/CA networks," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 593–596.

[53] S. Huang, X. Liu, and Z. Ding, "Opportunistic spectrum access in cognitive radio networks," in *Proc. IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2008, pp. 1427–1435.

[54] A. Motamedi and A. Bahai, "MAC Protocol Design for Spectrum-agile Wireless Networks: Stochastic Control Approach," in *Proc. IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Apr. 2007, pp. 448–451.

[55] M. Ghosh and V. Gaddam, "Bluetooth interference cancellation for 802.11g WLAN receivers," in *Proc. IEEE International Communications Conference (ICC)*, May 2003, pp. 1169–1173.

[56] N. Golmie, "Bluetooth Dynamic Scheduling and Interference Mitigation," *Mobile Networks and Applications*, vol. 9, pp. 21–31, 2004.

[57] N. Golmie, N. Chevrollier, and O. Rebala, "Bluetooth and WLAN Coexistence: Challenges and Solutions," *IEEE Transactions on Wireless Communications*, vol. 10, no. 6, pp. 22–29, Dec. 2003.

[58] N. Golmie, R. E. Van Dyck, A. Soltanian, A. Tonnerre, and O. Rébala, "Interference Evaluation of Bluetooth and IEEE 802.11b Systems," *Wireless Networks*, vol. 9, no. 3, pp. 201–211, May 2003.

[59] Maxim MAX2810 Datasheet, http://datasheets.maxim-ic.com/en/ds/MAX2820-MAX2821A.pdf.

[60] Bluetooth Special Interest Group, "Specification of the Bluetooth System," Nov. 2004.

[61] R. Tandra and A. Sahai, "Fundamental limits on detection in low SNR under noise uncertainty," in *Proc. International Conference on Wireless Networks, Communications and Mobile Computing*, 2005.

[62] J. Zyren, "Extension of Bluetooth and 802.11 Direct Sequence Interference Model," Harris Semiconductor, Study Group Contribution IEEE 802.11-98/378, Nov. 1998.

[63] E. Altman, *Constrained Markov Decision Processes*. Chapman & Hall/CRC, 1999.

[64] S. I. Resnick, *Adventures in Stochastic Processes*. Birkhäuser, 1992.

[65] M. L. Puterman, *Markov Decision Processes. Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.

[66] C.-F. Chiasserini and R. R. Rao, "Coexistence Mechanisms for Interference Mitigation in the 2.4-GHz ISM Band," *IEEE Transactions on Wireless Communications*, vol. 2, no. 5, pp. 964–975, Sep. 2003.

[67] L. Berlemann, C. Hoymann, G. Hiertz, and B. Walke, "Unlicensed Operation of IEEE 802.16: Coexistence with 802.11(A) in Shared Frequency Bands," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2006.

[68] X. Fu, W. Ma, and Q. Zhang, "The IEEE 802.16 and 802.11a Coexistence in the License-Exempt Band," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2007, pp. 1944–1949.

[69] J. L. Burbank and W. T. Kasch, "IEEE 802.16 Broadband Wireless Technology and Its Application to the Military Problem Space," in *Proc. IEEE Military Communications Conference (MILCOM)*, vol. 3, Oct. 2000, pp. 1905–1911.

[70] I. Wong and B. Evans, *Resource Allocation in Multiuser Multicarrier Wireless Systems*. Springer Science+Business, 2008.

[71] K. Kim, Y. Han, and S.-L. Kim, "Joint Subcarrier and Power Allocation in Uplink OFDMA Systems," *IEEE Communications Letters*, vol. 9, no. 6, pp. 526–528, Jun. 2005.

[72] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[73] T. Peng, W. Wang, Q. Lu, and W. Wang, "Subcarrier Allocation Based on Water-filling Level in OFDMA-based Cognitive Radio Networks," in *Proc. ACM WiCom*, Sep. 2007, pp. 196–199.

[74] P. Wang, M. Zhao, L. Xiao, S. Zhou, and J. Wang, "Power Allocation in OFDM-based Cognitive Radio Systems," in *Proc. IEEE Global Communications Conference (Globecom)*, Nov. 2007, pp. 4061–4065.

[75] J. Acharya and R. D. Yates, "A Framework for Dynamic Spectrum Sharing between Cognitive Radios," in *Proc. IEEE International Communications Conference (ICC)*, Jun. 2007, pp. 5166–5171.

[76] A. T. Hoang and Y.-C. Liang, "A Two-Phase Channel and Power Allocation

Scheme for Cognitive Radio Networks," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2006.

[77] W. Wang and X. Liu, "List-coloring based channel allocation for open-spectrum wireless networks," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Sep. 2005, pp. 690–694.

[78] H. Zheng and C. Peng, "Collaboration and fairness in opportunistic spectrum access," in *Proc. IEEE International Conference on Communications (ICC)*, May 2005, pp. 3132–3136.

[79] N. Devroye, P. Mitran, and V. Tarokh, "Limits on Communications in a Cognitive Radio Channel," *IEEE Communications Magazine*, vol. 44, no. 6, pp. 44–49, Jun. 2006.

[80] A. Jovicic and P. Viswanath, "Cognitive Radio: An Information-Theoretic Perspective," in *Proc. IEEE ISIT*, Jul. 2006, pp. 2413–2417.

[81] X. Qiu and K. Chawla, "On the Performance of Adaptive Modulation in Cellular Systems," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 884–895, Jun. 1999.

[82] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[83] A. J. Goldsmith and P. P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[84] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. International Communications Conference (ICC)*, Jun. 1995, pp. 331–335.

[85] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[86] Z. Liu, J. Almhana, V. Choulakian, and R. McGorman, "Traffic Modeling with Gamma Mixtures and Dynamical Bandwidth Provisioning," in *Proc. IEEE Communication Networks and Services Research Conference (CNSR)*, May 2006.