
A Dissertation Presented

by

Rahman Doost-Mohammady

to

The Department of Electrical and Computer Engineering

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in the field of

Computer Engineering

Northeastern University
Boston, Massachusetts

December 2014
Abstract

The limited availability of usable wireless spectrum and the ever increasing demands of high bandwidth data transfer raise concerns on whether current spectrum access regimes can match future communication requirements. Moreover, most desirable frequency ranges with good channel characteristics are already licensed, and purchasing new licenses for small operators is often prohibitively expensive. This thesis proposes methods for achieving efficient spectrum access through devising protocols for identifying and sharing unused spectrum, analyzing the theoretical bounds of these protocols, and implementing these solutions in practical medical and vehicular environments.

A significant portion of the thesis is focused on opportunistic spectrum access within licensed frequency bands, where cognitive radios transmit on frequencies without interfering with the primary users in them. First, a cooperative sensing method based on reinforcement learning technique is designed to efficiently detect spectrum opportunities. After identifying portions of the available spectrum, a channel allocation technique is devised for the cognitive radios with quality of service provisioning. The supporting analytical framework is constructed using Markov process and ensures that radios opportunistically using the licensed spectrum meet their latency and throughput requirements. The analytical framework is tested through traces collected in the wireless medical telemetry service (WMTS) band, and reliability enhancements in possible hospital application areas are quantified. A mobile architecture composed of vehicular networks is also investigated, where spectrum databases that provide the spectrum availability information are included in the network design.

Apart from re-use of the licensed spectrum, this thesis investigates full duplex channel access scheme for improving throughput. Full duplex enables simultaneous transmission and reception on the same channel for a single radio, which promises doubling of the throughput. An analytical framework for the performance of CSMA/CA based channel access for full duplex enabled network of nodes is formulated. The closed form expressions for the average throughput and packet collision probability in such a network is analytically derived and verified through comprehensive simulations.
I would like to thank my adviser, Prof. Kaushik Chowdhury, for his great support, encouragement, and guidance throughout my study period. I feel indebted to him for all his help during these years to move toward being an independent researcher.

I also would like to thank Prof. Stefano Basagni and Prof. Miriam Leeser for being on my committee and their invaluable suggestions and critiques which led to improvements in my research.

I would like to give my special thanks to my colleagues and collaborators: Dr. Marco Di Felice and Yousof Naderi. I learned a lot from them and enjoyed collaborating with them on various research projects during my studies some of which are part of this thesis.

Finally, I would like to thank my family and especially my parents for their unconditional love and support. Even thousands of miles apart, they have been present through every step of my life and this thesis is dedicated to them.
## Contents

1 Introduction .......................................................... 9
   1.1 Motivation ......................................................... 9
       1.1.1 Cognitive Radio: A solution to statically-allocated underutilized spectrum ......................................................... 12
       1.1.2 “More than one at a time” communication paradigms for efficient spectrum access ......................................................... 15
   1.2 Contributions ....................................................... 17
   1.3 Dissertation Outline ............................................. 19

2 Opportunistic Spectrum Access Using Cognitive Radios: Applications and Challenges ............ 21
   2.1 Cognitive Radios in Wireless Medical Telemetry Networks ................................................. 22
       2.1.1 Motivation ......................................................... 22
       2.1.2 Overview of WMTS Limitations ........................................ 24
       2.1.3 Spectrum Study of the WMTS Band ........................................ 25
       2.1.4 Power and Spectrum Allocation Framework ........................................ 30
       2.1.5 Network Management Framework ........................................ 35
       2.1.6 Performance Evaluation ............................................ 38
       2.1.7 Open Research Challenges ............................................ 38
   2.2 Cognitive Radios in Vehicular Networks ................................................. 41
       2.2.1 Motivation ......................................................... 41
       2.2.2 Applications of CRVs .............................................. 42
       2.2.3 Characteristics and Features of CRVs ........................................ 43
       2.2.4 Classification of Existing Schemes for CRVs ........................................ 44
       2.2.5 Open Research Issues .............................................. 48
       2.2.6 Spectrum Database Assisted Cognitive Radio Vehicular Networks ......................... 49

2

3.1 Related Work ................................................. 61
3.2 System Model and Problem Formulation .................... 62
3.2.1 Network Model ........................................... 63
3.2.2 Delay analysis for streaming type allocation ............ 64
3.2.3 Delay Analysis for Non-streaming type allocations ....... 69
3.2.4 Frequency Allocation Algorithm ......................... 72
3.3 Performance Evaluation ...................................... 78
3.3.1 Wireless Medical Telemetry: A Case Study using Real-World QoS
Constraints ................................................................ 78
3.3.2 Mapping of our QoS Framework for WMTS Bands ........ 79
3.3.3 Characterizing the WMTS Bands ......................... 80
3.3.4 Simulation results .......................................... 81
3.3.5 Streaming nodes .......................................... 82
3.3.6 Non-streaming nodes .................................... 84

4 Efficient Spectrum Sensing: Key to Opportunistic Spectrum Access

4.1 Related Work and Motivation ................................ 89
4.1.1 Standard deviation or mean as a channel quality metric? . 93
4.1.2 Range of collaboration ................................... 94
4.1.3 Intervals between collaboration events .................. 95
4.2 CLICK: A Cooperative Reinforcement Learning Scheme .... 95
4.2.1 Overview ................................................. 95
4.2.2 Stage 2 - Intra-node Learning ......................... 96
4.2.3 Stage 3 - Inter-node Cooperation ................... 97
4.3 Leveraging Cooperation at the Link Layer .................. 99
4.3.1 MAC Protocol Design .................................. 99
4.3.2 Cooperation Benefits in the MAC Protocol .......... 101
4.3.3 PU protection and Switching Minimization ............ 102
4.4 Performance Evaluation ..................................... 102
4.4.1 PU Protection Analysis ................................ 104
4.4.2 Analysis of MAC Protocol using CLICK ............. 104

5 Full Duplex: Doubling the Spectrum Opportunity

5.1 Related Work and Motivation ................................ 109
5.2 Network Architecture ....................................... 110
5.3 Analytical Model for Full Duplex MAC ................... 112
5.3.1 Analysis from a client’s perspective .......................... 115
5.3.2 Analysis from the AP’s perspective .......................... 119
5.3.3 Steady-state probabilities ................................. 121
5.3.4 Throughput Analysis .................................. 124
5.4 Model Validation ............................................. 125
  5.4.1 Simulation Setup ..................................... 125
  5.4.2 Simulation Results ..................................... 126
  5.4.3 Full duplex Gain Evaluation .............................. 129

6 Conclusion and Future Work .................................. 133
  6.1 Conclusion ............................................. 133
  6.2 Future Work ............................................ 135

Publications .................................................. 137

Bibliography .................................................. 138
# List of Figures

1.1 The cognitive radio cycle. ................................................. 13

2.1 Spectrum ranges for WMTS bands with varying levels of access priority . 23
2.2 Spectrum survey of the channel 37 (a), lower $L$ (b) and upper $L$ (c) bands. 26
2.3 Frequency domain signal in the UHF Channel 37 vs. the noise floor. .... 28
2.4 Frequency domain signal in the Lower L-band vs. the noise floor. ....... 28
2.5 Spectrum Sensing result via energy detection on UHF Channel 37. ....... 30
2.6 PDF fitting for the “off” time duration of the bins centered at $608.8516$ MHz and $613.0569$ MHz. .......................................................... 31
2.7 Flowchart of the proposed transmission parameter optimization framework 36
2.8 Performance comparison of dynamic spectrum allocation vs. static allocation based on, average unused bandwidth (a), and average interfered bandwidth (b) ......................................................... 39
2.9 Three different deployment architectures for CRVs: (a) vehicle to vehicle only, (b) multiple local BSs, and (c) centralized BS serving vehicles. .... 42
2.10 Classifications of existing works on CRVs. ............................ 45
2.11 (a) The map and (b) the spectrum sensing accuracy at three different locations A (moderately spaced structures), B (open area), C (congested downtown) ......................................................... 49
2.12 Different modes of operation in a vehicular CR network ............... 50
2.13 Cost of each channel information access vs. BS density ............... 57
2.14 Silent mode probability change vs. BS density ....................... 58
2.15 Access cost function derivation vs. BS density ...................... 59
2.16 Average sensing error probability vs. BS density .................... 59
2.17 Average number of channel information accesses per year vs. BS density 60
2.18 Average annual revenue received at each BS vs. BS density ........ 60
3.1 An example of channel assignment for 2 streaming nodes with 3 backup channels. Random PU arrivals and departures trigger SU channel movements and the corresponding Markov process state transition. 65

3.2 Finite quasi-birth-death process representing $M$ main channels with PU arrival rate $\lambda$ and PU "on" rate $\lambda_{on}$, and $N$ backup channels with PU arrival rate $\mu$ and PU "on" rate with $\mu_{on}$. 66

3.3 (a) Traversing order of the 2D-Histogram during allocation. (b) A histogram $H_3$ shown with PU arrival and departure shown in one dimension. PU channels are aggregated with $s = 3$ to form the histogram. 74

3.4 (a) Exponential CDF fitting for PU inter-arrival time of three sample bins centered at 608.028 MHz, 1395.691 MHz and, 1428.897 MHz. (b) Exponential CDF fitting for PU ON time of three sample bins centered at 608.028 MHz, 1395.691 MHz, and 1428.897 MHz. 80

3.5 (a) Incurred queueing delay both in theory and simulation for 5, 10 and 20 streaming channels while varying number of backup channels for the medium group, i.e., $\lambda = 0.024$, $\lambda_{on} = 0.1$, $\mu = 0.1$, $\mu_{on} = 0.1$. (b) Number of backup channels $N$ vs. number of main channels $M$ to keep the average queueing delay below 500 s shown for three PU activity groups of table 3.4 both in theory and simulation. 83

3.6 Comparison of amount of allotted spectrum versus streaming request load resulted by algorithm 3, simplified of algorithm 3 based on 1-D histogramming method, simple method with no histogramming, and theoretical estimate. 85

3.7 Average delay vs. PU arrival rate $\lambda$ with varying number of nodes with mean packet inter-arrival time of 120 sec contending over a single channel. 87

3.8 Average delay vs. PU arrival rate $\lambda$ with varying packet arrival rate $\gamma$ for 30 nodes contending over a single channel. 87

4.1 The variation in the mean and the standard deviation in the received power and pilot signal are shown for indoor location for TV channels 21 − 51. 90

4.2 The variation in the mean and the standard deviation in the received power are shown for roof-top location for TV channels 21 − 51. 91

4.3 The time variation in the received signals at indoor (broken lines) and outdoor locations (bold lines). 92

4.4 The pairwise difference of the mean power for near (X-Y) and far (X-Z) separation distances. 93

4.5 The maximum drop in the mean of the received signals, and its breaching the noise floor, for sensing time $t_{sense}^k$ for the channels $k = 21 − 30$. 94
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6</td>
<td>The MAC protocol design that implements CLICK</td>
<td>100</td>
</tr>
<tr>
<td>4.7</td>
<td>Overhead of CLICK and $K$ by $N$ scheme.</td>
<td>102</td>
</tr>
<tr>
<td>4.8</td>
<td>The probability of correct PU detection for 15 nodes (a) and 30 nodes (b),</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>while the convergence behavior for CLICK is in (c).</td>
<td></td>
</tr>
<tr>
<td>4.9</td>
<td>The throughput of CLICK MAC for different values of sensing time interval (a),</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>and sensing schemes (b). The link delay is given in (c).</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Representation of a Star topology with full duplex nodes.</td>
<td>108</td>
</tr>
<tr>
<td>5.2</td>
<td>Cases of full duplex transmissions initiated by (a) the client node and a</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>packet reply by the AP, (b) the client node and BT broadcast by the AP, and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) the AP and packet reply by the client.</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>Embedded Markov chain representing the states of head-of-line packet in</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>full duplex enabled CSMA/CA.</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>The effect of number of clients and hidden terminals on system throughput</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>for the ring topology.</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>The effect of contention window size on system saturation throughput for</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>the ring topology.</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>System throughput versus network size in random topology</td>
<td>129</td>
</tr>
<tr>
<td>5.7</td>
<td>System throughput versus contention window in random topology</td>
<td>130</td>
</tr>
<tr>
<td>5.8</td>
<td>FD gain versus network size</td>
<td>131</td>
</tr>
<tr>
<td>5.9</td>
<td>FD gain versus backoff window length</td>
<td>131</td>
</tr>
<tr>
<td>5.10</td>
<td>FD gain versus number of hidden terminals in ring topology</td>
<td>132</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Commonly Used Wireless Medical Telemetry Products . . . . . . . . . . 25
2.2 Spectrum Sensing Techniques for CRVs . . . . . . . . . . . . . . . . . 42
2.3 Empirically measured traffic parameters from [1] . . . . . . . . . 51
2.4 Type of wireless equipment in vehicle and their probability distribution. . 51
2.5 Simulation Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . 57

3.1 List of Channel Allocation with QoS provisioning works in the literature. 63
3.2 List of notations used in frequency allocation algorithm. . . . . . . 73
3.3 Application specifications in a hospital case-study [2] . . . . . . . . 81
3.4 Groups of main and backup channels with various PU activity. . . . . 82
3.5 Simulation Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . 86

5.1 List of notations used in the analysis . . . . . . . . . . . . . . . . . . . 116
5.2 Simulation Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . 125
Chapter 1

Introduction

1.1 Motivation

Wireless spectrum is a costly resource, worth billions of dollars of transactions each year [3], and its availability limited in part by the constraints of commercially feasible telecommunications infrastructure and the governing laws of electromagnetic propagation as defined by Physics. The inherent relationships posed by the length of antennae to the usable center frequency of transmission, the ability of electromagnetic waves to penetrate walls and other opaque objects, as well as the ability for the electronics in the receiver to respond to rapid changes at high frequencies has led to the creation of a set of desirable frequencies starting from the low MHz range and bounded at the higher end at several GHz. Since the first successful demonstration of long-distance wireless propagation by Marconi in 1902, the methods of frequency spectrum allocation have not seen a drastic change, despite spectacular advances in wireless technology and a burst of new applications. Spectrum regulators around the world, through the power vested in them by the national governments, issue licenses to wireless users including the military, major cellular companies, media channel networks etc., while leaving some portions unlicensed for the use of home and enterprise wireless networks. This thesis revisits the concept of static spectrum allocation, instead arguing that dynamically allocating spectrum and opportunistic transmission in licensed but unused frequency bands is a feasible method for ensuring efficient spectrum utilization. It also addresses spectrum efficiency concerns in traditional networks, where the operational spectrum cannot be directly changed. In such cases, higher utilization efficiency can be achieved by transmitting more information in terms of bits per unit of time. Towards this, our work also explores the feasibility of full duplex communication, where the radio may send and receive at the same time, breaking free from a receive-only or transmit-only operation and effectively doubling the informa-
tion rate in certain cases. In this following, we begin by describing why the problem of spectrum efficiency is timely and important, which serves as the motivation for the work undertaken in this thesis.

Despite the interest in obtaining and using wireless spectrum, several measurement campaigns in recent years have shown that the legacy method of spectrum assignment has left a significant portion of the spectrum unused over large extents of time. According to a report published by the spectrum efficiency working group [4] of the Spectrum Policy Task Force, in 2002 only 700 MHz in the spectrum below 1 GHz was being used in the major US cities of Atlanta, New Orleans and San Diego. Also only four out of 18 TV channels in the UHF band in Washington, D.C. were being continuously used, although all of them were allocated. Most of these unused channels were left as protective guard bands [5]. Yet another study of the spectrum in the $0 - 6$ GHz range by the Berkeley Wireless Research Center in downtown Berkeley [6] revealed less than 30% utilization of the spectrum below 3 GHz, and around 0.5% in the range 3 – 6 GHz.

The unlicensed (ISM) bands are being used by millions of wireless devices for streaming video and essential data communication in only 85 MHz of the RF spectrum in the 2.4 GHz band. An increasing number of new product developments today have led to expanding their use as unlicensed operators to the 5 GHz ISM band. While this new spectrum band has partially alleviated the spectrum availability concerns, there are persistent questions raised on whether the spectacular growth in wireless applications and the projected requirement in spectrum availability for unlicensed use will scale proportionately. The pressing societal need of affordable and ubiquitous internet connection has been highlighted in the US President’s ”Broadband Plan” [7] in 2009 with a vision of enabling 100 million American homes with broadband Internet access, with at least 100 Mbps download and 50 Mbps upload rates, by 2020. This plan largely relies on the availability of fast wireless connections, requiring up to an additional 500 MHz of spectrum. As an enabling step, this plan makes recommendations for updating rules for wireless backhaul spectrum to increase capacity in urban areas and range in rural areas. Part of the recommendations in the plan is quoted below, which serve as motivations of our work:

- **Make 500 MHz of spectrum newly available for broadband within 10 years, of which 300 MHz should be made available for mobile use within five years.**

- **Enable incentives and mechanisms to repurpose spectrum to more flexible uses.** Mechanisms include incentive auctions, which allow auction proceeds to be shared in an equitable manner with current licensees as market demands change. These would benefit both spectrum holders and the American public. The public could benefit from additional spectrum for high-demand uses and from new auction rev-
enues. Incumbents, meanwhile, could recognize a portion of the value of enabling new uses of spectrum. For example, this would allow the FCC to share auction proceeds with broadcasters who voluntarily agree to use technology to continue traditional broadcast services with less spectrum.

- Ensure greater transparency of spectrum allocation, assignment and use through an FCC-created spectrum dashboard to foster an efficient secondary market.

- Expand opportunities for innovative spectrum access models by creating new avenues for opportunistic and unlicensed use of spectrum and increasing research into new spectrum technologies.

Since the release of this visionary national level plan, the FCC has made several rulings on the opening of white spaces for secondary use, including the most recent directives issued in September 2010 and April 2011 that provide the authorization for dynamic spectrum access. In the ruling of September 2010 [8], it authorizes the use of TV band devices (TVBD) as radio devices that can operate on unused ultra high frequency (UHF) TV bands, also called as the TV whitespace. It establishes the rules and regulations under which these devices can operate. Furthermore, spectrum databases are introduced as the enabler of TVBDs to find unused spectrum based on their geographic location and time. Also when spectrum databases are not accessible, the process of identifying free spectrum or spectrum sensing, is allowed under strict operational limitations for detecting the unused TV bands.

The FCC defines several important modes of operation, described below, under specific assumptions of location awareness, mobility, and spectrum database access capability. Thus, TVBDs are categorized as follows:

- **Fixed** TV Band devices (TVBDs) are installed at fixed locations, and their coordinates are known to themselves and to the public.

- **Mode II** device is portable and has internal geolocation capabilities, along with the access capability to the spectrum database.

- **Mode I** is a portable device with no geolocation capability and no direct access to the database. This type of device must contact its neighboring Mode II or Fixed device to obtain the list of available TV channels.

- **sensing-only** is a FCC certified portable device unable to access the spectrum database through either a Fixed or Mode II device, and must rely on local spectrum measurements.
This thesis is focused on devising protocols for radios, analyzing their performance, and studying deployment scenarios that can leverage the TV whitespace and other licensed frequency bands that are not always utilized. At the device level, it relies on software defined radios that can sense the environment and adaptively react to changes in spectrum use, akin to the feature of automated and operator-less cognition. We call this emerging technology as **cognitive radio** (CR), which resides at the intersection of inter-disciplinary research involving hardware design, wireless communications theory, network protocol stack, computer science, among others.

### 1.1.1 Cognitive Radio: A solution to statically-allocated underutilized spectrum

The research community took the first steps toward modeling and investigating opportunistic spectrum access through CR technology, before the policy of TV band re-use was formally articulated by the FCC. First coined by Mitola [9], CR was initially defined as a radio that performs model-based reasoning and inference to enhance the effectiveness of the data communications and delivery. Based on this definition, a so-called “cognition cycle” was introduced as the core inference mechanism. CR uses the cognition cycle to interact with the environment by observing it and detecting radio frequency (RF) stimuli, such as channel activity. With these RF inputs, the CR is able to modify its transmission plans, makes future operational decisions, and acts on those decisions while learning from the outcomes of past choices. Formally defined by Haykin [10], CR is an intelligent wireless communication system that is aware of its environment, and learns from it. It also adapts its internal states to statistical variations in the observed RF signal by making changes in certain operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time, to achieve reliable communications and efficient utilization of the spectrum.

The **cognitive cycle** from [11] describes the inter-relationship between the various roles the CR must play to effectively detect and use the spectrum. We see four different states in the cycle shown in Figure 1.1. In the following we interchangeably refer to the licensed users as primary users (PUs) of a spectrum, as they have priority of access and own their respective channels. The CR is a secondary user or SU in these channels.

- **Spectrum sensing**: Here, the CR analyzes the received signals on different channels to check whether licensed users are present on those channels.

  This function has been extensively studied in the literature and various techniques has been proposed for accurate detection of the PUs. Depending on the extent of *a priori* knowledge of the PU signal, three broad classes, i.e. energy detection, feature
Collaborative spectrum sensing has also been considered in many works [14–16], where spectrum sensing results obtained by individual CRs are fused together to infer the presence or absence of the PU signal with a higher level of accuracy. In Chapter 4 of this thesis we propose a collaborative fusion spectrum sensing framework specifically focused on the detection of digital TV (DTV) signals measured indoors and outdoors, under varying environmental factors.

- **Spectrum decision**: This function is initiated when the sensing step concludes and a set of unused channels are identified. Among the detected empty channels, the CR decides on which one to use for communication thus ensures higher service quality and fewer interruptions.

Generally, spectrum decision algorithms are sought for two types of applications, namely real-time and best effort [17]. In the real-time case, minimum variance in the quality of service (QoS) vector is desired. In the best effort case, the maximum achievable capacity in terms of bit/sec/Hz for either an individual or a group of CRs
is sought. In chapter 3, we propose algorithms to achieve maximum capacity in terms of the number of CRs with diverse QoS requirements that can be allocated to a set of channels with heterogeneous PU statistical behavior. This research stems from the very fundamental question of how QoS can be achieved when bandwidth resources are not exclusively allocated in the case of CRs as well as answering the question of what the sufficient bounds of required PU spectrum are to satisfy a certain QoS measure for a given set of CRs.

- **Spectrum mobility**: The CR performs an occasional spectrum handoff in this state by moving to a new channel determined by *spectrum decision*. It also tunes itself with new operating parameters to adapt to the new channel characteristics for best service quality. The spectrum mobility state is entered when the PU occupies the current channel or when CR’s quality of service drops due to mobility or excessive interference from other CRs operating on the same channel. A special case for such a scenario is discussed in Chapter 2, where a CR network architecture is proposed for addressing the needs of wireless medical telemetry.

- **Spectrum sharing**: As the licensed channel may be chosen opportunistically by multiple CRs in a distributed setting, each CR coordinates with its neighbours and performs various functions such as packet scheduling, channel rendezvous, collaboration for spectrum sensing, etc. to ensure coexistences among the CRs.

Various MAC protocols for packet scheduling are proposed in CR networks. As representative examples, C-MAC [18], OS-MAC [19] and HC-MAC [20] are works that fall under the three broad classical categories of random-access, time-slotted and hybrid protocols, respectively [11].

Although most of protocols consider a common control channel (CCC) for enabling channel rendezvous among CRs during spectrum handoff, it is not considered the best solution due to inflexibility, limited scalability and security issues. In this regard, a few techniques for blind rendezvous are studied in the literature [21, 22].

In Chapter 4, we propose a MAC protocol to enable collaboration among CR nodes, which facilitates effective data communication, timely exchange of the sensing results, and also results in lower sensing time.

While the research community has made major strides towards the development of the constituent functions of the cognitive cycle, there still exists a gap between the theory defining CR operation and its actual practical deployment. This thesis describes two application areas of CR, namely in wireless medical telemetry service (WMTS) and in vehicular networks, in Chapter 2 with an effort to bridge this gap.
In context of medical telemetry, the CR technology helps in avoiding interference from regularly-occurring but higher priority utility telemetry signals and neighboring DTV leakage signals by switching to channels that ensure reliable delivery of telemetry data. The additional constraint of preventing any harmful electromagnetic interference (EMI) to sensitive medical devices that are widespread in hospital environment is also incorporated. The recent developments in vehicular networks have enabled a new class of in-car entertainment systems and enhanced the ability of emergency responders, by integrating high bandwidth communication technology. However, assuring spectrum availability in highly congested scenarios for multimedia applications, and obtaining new spectrum under sudden disaster-induced outages, remain a concern. Cognitive radio enabled vehicles (CRVs) are able to use additional spectrum opportunities outside the designated band, such as fixed channels in the 5.9 GHz range specified in the IEEE 802.11p standard for vehicular networks. This opens up several new research areas for integrating and predicting mobility patterns with spectrum usage, devising cooperative spectrum sensing strategies, accounting for channel variations in narrow, urban streets flanked by tall buildings, and finally, exploring new architectures where vehicles operate under varying degrees of freedom, from complete autonomous spectrum-aware operation to limited support from roadside base stations.

As the second major contribution, we will explore the topic of efficient spectrum access within the same frequency band. This occurs when the frequency-agile CR radios are unavailable within the network, but significant improvements in throughput are still possible by increasing the number of bits sent and received over the channel at a given time. This leads us to the “full duplex” paradigm, as explained next.

1.1.2 “More than one at a time” communication paradigms for efficient spectrum access

Under the classical wireless channel access regime, devices monitor the channel for activity, and unable to detect any, they capture the channel for the duration of the data transfer. While one radio is transmitting, others back-off from using the same channel themselves, owing to the well known problems of interference and collision. The physical structure of the wireless transceiver has traditionally supported either transmission or reception at a time owing to this reason, as both functions cannot be simultaneously undertaken. Ultimately, this results in the at most one at a time paradigm in the commonly seen wireless infrastructure deployed today. Moreover, in widely used transmission standards, such as the IEEE 802.11 family, the mechanisms for packet scheduling at the link layer employ back off counters so as to stagger the channel access for competing nodes, and thereby reduce the possibility of collision among them. However this further impinges on spectrum
efficiency, leaving the channel unused during the back-off time of the nodes in a saturated network\(^1\).

Full duplex wireless (FD) [23–25, 100] allows a radio to send and receive data on the same channel simultaneously, which promises massive improvements in channel capacity and shifts the communication paradigm in design of networking protocols to “more than one at a time” approach. Before the inception of FD, the default assumption was that radios can only undertake half duplex communication, i.e., they may either transmit or receive [26]. Consequently, the design of the protocol stack, especially the channel access mechanism at the link layer, was influenced by this key assumption. As a result, any simultaneous use of the channel by more than one node within interference range in the same network could cause collision of the transmitted packets. A very common problem arising from this characteristic in half duplex networks is the *hidden terminal* problem when a node is a within the range (neighbor) of an intermediate node but not of its other neighbors. In this case, any concurrent transmission of these nodes to the intermediate node will cause collision. This, in turn, results in a waste of bandwidth resources, and the need for retransmission by all of the contending nodes who suffered packet loss. With the advent of the FD technology and the ability to transmit and receive at the same time on the wireless channel, this problem of simultaneous channel access can be mitigated to some extent. Having two FD nodes at two sides of a link to transmit and receive to each other at the same time could further improves the spectrum efficiency by preventing the *hidden terminal* problem. Recent works on building such FD systems [27, 28] indicate progress towards practical realizations of this technology.

In this thesis, we focus on protocol design and analysis of FD networks. We formulate the first analytical model of a CSMA/CA based full duplex MAC protocol for a wireless LAN network composed of an access point serving mobile clients. There are two major contributions in this part of the thesis: First, our Markov chain-based approach results in closed form expressions of throughput for both the access point and the clients for this new class of networks. Second, our study provides quantitative insights on how much of the classical hidden terminal problem can be mitigated through full duplex. We specifically demonstrate that the improvement in the network throughput of up to 15% over the half duplex case. Our analytical models are verified through packet level simulations in ns-2. Our results also reveal the benefit of full duplex under varying network configuration parameters, such as number of hidden terminals, client density, and contention window size.

\(^1\)In saturated networks, nodes have packets to transmit at all times.
1.2 Contributions

The main contributions of the thesis are as follows:

- We propose the first learning-based spectrum selection scheme and MAC protocol (called CLICK) derived from realistic measurements on actual TV bands. CLICK is a cooperative reinforcement learning framework that allows CRs to learn about the channel availability over time. Based on real TV channel measurements, we show that because of the different channel perspective of CRs, even at very close distances to each other, they need to collaborate on detecting empty channels for better accuracy. Therefore, we define a novel reliability metric based on the standard deviation of channel received power of the PUs, instead of the classically used mean power level, and demonstrate its superiority for urban spectrum selection scenarios. CLICK enables nodes to reliably come up with spectrum decisions also by considering the level of trust that can be assigned to the readings of the collaborating nodes. These are unique features of CLICK which is not present in any existing work. We implemented CLICK in a distributed network using a CR MAC protocol (CR-MAC) that manages the spectrum and power selection to minimize both the effect on the PUs and the other neighboring CR users. We demonstrate that CLICK can achieve better throughput and better PU protection by better choice of channels and shorter sensing times compared to other existing schemes.

- WMTS bands in a medical environment are composed of heterogeneous devices with different bandwidth and QoS requirements and hence there is a need for a general communication framework for optimum utilization of these bands. We demonstrate the first application of cognitive radio in the area of wireless medical telemetry service (WMTS). We define new methods for allocating communicating nodes to WMTS channels based on trace-driven studies in hospitals, which also satisfy QoS requirements for streaming and non-streaming traffic. We cast the problem of frequency and power allocation in these bands as an optimization problem, subject to the constraints of the medical telemetry and typical hospital applications. We further this research by solving a more general problem of channel allocation to CRs from a scheduling standpoint. We propose the novel idea of backup channel allocation as a technique for guaranteeing the quality of service for CRs such that they have seamless access to a spectrum channel at all times without sacrificing their delay requirements. Inspired by wireless medical telemetry where various type of wireless devices with different QoS requirements are present in the network and also various PU channel types with varying statistical activity pattern, we develop the first comprehensive analytical framework based on queueing theory and Markovian analysis.
to calculate the theoretical delay in accessing the PU prone spectrum depending on
the required CR QoS, with guidelines on how to select channel(s) for streaming and
intermittent data transmission types of CR. Based on these analyzes, we devise a
spectrum allocation algorithm for CR with different QoS demands. Our method fo-
cuses on allocating the least amount of the overall available spectrum and serves as
many CRs as possible such that the QoS requirements are met. At last, we use our
measurements of WMTS bands, undertaken at the Massachusetts General and Beth
Israel hospitals in Boston, as our PU model and a realistic scenario for heterogenous
QoS classes of CR telemetry devices as the input to the proposed algorithm and
demonstrate the effectiveness of our framework by showing how the spectral effi-
ciency of our methods closely matches our theoretical predictions. The techniques
presented in this research can be readily reused in other areas of CR application,
such as cellular and vehicular networks.

• We formulate an analytical expression for allocation (location, number) of roadside
access points based on practical scenarios where vehicles are equipped with one
or more of the following: cellular connection, GPS, local sensing capability. We
propose the first cross-disciplinary approach for this problem, combining FCC rules
on spectrum databases, economic considerations and data available on vehicular
traffic. By adhering to the most recent FCC ruling for TVBDs, we provide a cost-
revenue analysis, given the environmental parameters such as vehicular density, BS
road coverage and the tradeoffs between local vs. database assisted sensing, thereby
making this a potential tool for future network design.

• We formulate the first Markov-chain based analysis for access point driven CSMA/CA
networks using full duplex communication. While extensive research has been con-
ducted for half duplex networks, this investigation is focused on the completely dif-
ferent and emerging scenario where transmission and reception occur concurrently,
bringing about major changes to the classical and well known Bianchi framework.
Closed form expressions for the calculation of network throughput in various topolo-
gies and CSMA/CA MAC parameter setting is provided. We consider ring topology
and random topology in which the number of nodes in the network and the length of
CSMA/CA backoff window size is given. In the case of ring topology, the number
of hidden terminals in the network can be chosen as well. Apart from considering
full duplex, our work contributes to the existing works on MAC layer analysis in
the presence of hidden terminals by separately analyzing both uplink traffic flows
from clients to the AP, and downlink traffic from the AP to client. This is in contrast
to [29–31] where only uplink traffic is considered, and the AP is assumed to be
only receiving. The consideration of both uplink and downlink traffic is imperative
because full duplex is based on active bi-directional links. Our theoretical findings are verified through packet level simulations in ns-2, which has been considerably modified from its stock installation to incorporate the full duplex operation.

1.3 Dissertation Outline

These are unique features of CLICK which is not present in any existing work. The first part of this thesis focuses on cognitive radio driven efficient spectrum access, protocol design, analysis of its use in the licensed bands, and several practical applications. We also investigate full duplex wireless in the second part by proposing an analytical framework for the performance gain of full duplex versus classical half duplex networks.

In Chapter 2, the application of CR in two specific types of wireless networks is investigated.

- **Wireless medical telemetry**: We overview the research challenges in the use of CR in wireless medical telemetry. We specifically address one of them, namely frequency and power allocation for wireless medical telemetry CR devices that need to find the best set of channels and power to guarantee their SNR and rate requirements.

- **Vehicular networks**: we provide a taxonomy of the existing literature on this fast emerging application area of CRV networks, as well as highlight the research issues that need to be addressed for realization of such networks. We also propose a cost-benefit analysis that finds the roadside infrastructure density to provide spectrum information to vehicles.

Chapter 3 addresses the problem of spectrum resource management for co-located CRs with both streaming and intermittent data. With the objective of using the least amount of spectrum per user, we provide analysis for achievable QoS for groups of users given the PU statistical behavior on certain channels. Using these we propose guidelines on how to allocate a large spectrum of PU channels with different activity patterns to users with different QoS requirements.

In Chapter 4 we look into the problem of gathering the spectrum information from various nodes in a distributed architecture. This is due to the fact that individual measurements at CRs give a partial view of the spectrum environment based on the local sensing range, which are in turn affected by channel uncertainties and location-specific fluctuations in signal strength.

Chapter 5 provides a rigorous theoretical framework for analyzing FD networks using CSMA/CA as the channel access mechanism at the link layer. We consider a general scenario of an AP controlled wireless LAN, with the AP at the center, surrounded by the...
serviced clients in a *star* topology. We derive closed form expressions for the probability of successful transmission and throughput separately for the AP and clients, while considering the effect of hidden terminals on these performance metrics.
Chapter 2

Opportunistic Spectrum Access Using Cognitive Radios: Applications and Challenges

In this chapter, we describe two different application areas beyond TV whitespace that showcase possible improvements through CR technology. These areas are: i) wireless medical telemetry, ii) vehicular networks. We argue that in both cases, the problems of spectrum scarcity, presence of higher priority users in the operational band, and interference from undesirable sources can be solved using CR technology.

In section 2.1, we propose the use of CR in order to dynamically utilize the WMTS frequencies based on the activity patterns of the high priority users, and the quality of service constraints of the patients’ data, while ensuring protection to existing higher priority transmissions and the safe operation of sensitive medical equipment. This section also discusses the current state of the art and the major challenges in the implementation of this new CR-assisted medical telemetry paradigm.

Section 2.2 provides a taxonomy of the existing literature on the fast emerging application area of cognitive radio enabled vehicular (CRV) networks, as well as highlights the research issues that need to be addressed for realization of such networks. In addition, it develops an analytical framework for determining the design parameters of vehicular networks with unlicensed access to white spaces when FCC mandated spectrum databases are to be the source of spectrum availability information at each location for the vehicles.
2.1 Cognitive Radios in Wireless Medical Telemetry Networks

2.1.1 Motivation

Healthcare is facing increasing costs in the U.S., with one of the contributing factors being infrastructure investment. Hospital spending by itself is projected to accelerate nearly 7.3% till 2019. However, projected benefits from information technology and networking alone may touch 81 billion annually [32], giving further impetus for modernizing the data networks that carry the digital patient information. Wiring all devices to transport this patient data has been a rather simplistic solution, and the situation is often described as “patients becoming trapped in a sheer impenetrable net of wires and tubes, often resembling a spaghetti” [33].

Since June 2000, the FCC has reserved wireless spectrum for medical telemetry in the (i) current digital television (DTV) channel 37 between 608 – 614 MHz, (ii) lower-L band (1395 – 1400 MHz), and the (iii) upper-L band (1427 – 1432 MHz), together referred to as the WMTS band [34], as shown in Figure 2.1. Yet, unfortunately, the same ruling states raises several concerns of interference, access priority of medical devices, and limitations on multimedia transmission.

CR is prescribed to solve fundamental problems of spectrum scarcity, low utilization efficiency, interference, and high-bandwidth communication [35], all present in wireless medical telemetry systems. In the healthcare environment, medical transceivers equipped with a frequency-agile front-end (i.e., the CR users) may opportunistically use portions of the WMTS band, without affecting adversely the operation of the priority users as well as the legacy medical equipment in operation. We envision a centralized network, where the CR devices (simply referred to as devices or nodes subsequently) interface with important medical equipment and forward traffic over a single hop to the central base station (BS). The BS then forward the traffic to the end destination, such as an overhead monitor, medical data repository, or to a distant doctor’s office through out of band transmissions or through the backend wireline network. The BS also periodically performs spectrum sensing over the entire WMTS band to detect any potential interference caused by transmissions in neighboring TV channels to the DTV channel 37, and also monitors for other transmission activities from utility telemetry or government operated installations on the L band. After determining the effective residual spectrum, it allocates this spectrum to requesting medical devices based on their needs. Strict adherence to the limits on the allowed electromagnetic interference (EMI) for sensitive equipment using power control mechanisms that limits transmission range, ease of deployment of multiple types of sensing devices with heterogeneous performance constraints, and resilience to external
Figure 2.1: Spectrum ranges for WMTS bands with varying levels of access priority
interference and protection from self-congestion are some of the unique features of our proposed architecture.

2.1.2 Overview of WMTS Limitations

In this section, we summarize the major concerns impacting medical telemetry and describe the need for further research through the following three case studies.

- **Case I. Effect of DTV interference on medical telemetry:** Hospitals in (i) Kansas, (ii) New Jersey, and (iii) Ohio have documented recent outages in the WMTS bands owing to increased transmission powers in the adjacent DTV channels 36 and 38 [36]. Here, the effective bandwidth in the channel 37 was reduced to a third, and valuable time was lost for manually re-surveying the spectrum and requesting the offending DTV operators to regulate their power. These incidents demonstrate that a static licensed approach is unsuitable for guaranteeing connectivity, and dynamic spectrum access is required.

- **Case II. Need for high bandwidth intra-hospital communication:** According to Dr. Julian Goldman, a Massachusetts General Hospital (MGH) anesthesiologist, many fatalities caused by human error could be prevented if medical devices are seamlessly connected. The operating room of the future moves towards this strongly connected paradigm that stresses on high bandwidth availability both within the operating room and outside [37]. However, as FCC rules prohibit multimedia transmission in the WMTS bands, one possible option is using vacant DTV channels for audio and video. Interestingly, prior to constituting the WMTS bands, medical devices were only allowed to operate on an unlicensed basis on vacant DTV channels 7–13 and 14–46 (i.e., 174–216 MHz and 470–668 MHz).

- **Case III. Navigating the maze of complex spectrum access rules:** In Figure 2.1, we depict how portions of the WMTS spectrum have varying priority access rights. As an example, military and governmental agencies have a priority access in the lower-L band spectrum with a number of operational radars, such as the FAA Air-Route Surveillance Radars (ARSR-1, -2, -3, and -4) and the Air Force ANFPS-117 and -24 radars [38]. In addition, the upper-L band and the lower-L bands are used by non-medical telemetry companies on a priority access and equal-access basis, respectively [34]. Medical devices must defer access to high-priority users in the L bands and also deal with (unwanted) interference when access rights are equal. This can be done by adding dynamic spectrum access capability to these devices to use WMTS bands as efficiently as possible while protecting the rights of priority users by not interfering their transmission.

We would also like to mention that similar to the FCC mandated spectrum databases for the DTV channels, the American Society for Healthcare Engineering of the American Hospital Association (ASHE/AHA) is tasked to serve as the exclusive WMTS fre-
Table 2.1: Commonly Used Wireless Medical Telemetry Products

<table>
<thead>
<tr>
<th>Company</th>
<th>Product</th>
<th>Spectrum Access</th>
<th>Standard/Protocol</th>
<th>Capacity (Number of nodes)</th>
<th>Network Type</th>
<th>Access Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE</td>
<td>Apex Pro CH</td>
<td>FM (Channelized)</td>
<td>-</td>
<td>438</td>
<td>WMTS (all)</td>
<td>Centralized</td>
</tr>
<tr>
<td>GE</td>
<td>Apex Pro FH</td>
<td>Frequency hopping</td>
<td>-</td>
<td>640</td>
<td>WMTS (all)</td>
<td>Centralized</td>
</tr>
<tr>
<td>Philips</td>
<td>1.4 GHz Intellivu</td>
<td>Smart Hopping (Proprietary)</td>
<td>DECT</td>
<td>1028</td>
<td>1.4 GHz (L-band)</td>
<td>Centralized</td>
</tr>
<tr>
<td>Welch Allyn</td>
<td>FlexNet</td>
<td>OFDM</td>
<td>802.11a</td>
<td>1500</td>
<td>2.4 GHz</td>
<td>Centralized</td>
</tr>
<tr>
<td>Draeger</td>
<td>Infinity TruST</td>
<td>DSSS</td>
<td>802.11b/g</td>
<td>600</td>
<td>2.4 GHz</td>
<td>Centralized</td>
</tr>
<tr>
<td>Mindray</td>
<td>Panorama</td>
<td>FM (channelized)</td>
<td>-</td>
<td>500</td>
<td>WMTS (all)</td>
<td>Centralized</td>
</tr>
</tbody>
</table>

frequency coordinator. One of the reasons is that home usage of WMTS is forbidden as every provider must first provide the channel usage data to the ASHE/AHA database before transmitting. However, if a large hospital reserves an entire width of the available channels, say, the entire lower-L band, but uses a limited portion of this capacity, this results in an acute shortage of available channels. In a densely packed hospital area, such as Boston’s Longwood area, this is an existing problem, which can only be addressed through dynamic spectrum access.

2.1.3 Spectrum Study of the WMTS Band

Our spectrum studies were conducted at two of the largest hospitals in the metropolitan Boston area, which are among the largest medical care complexes in the US. We used a USRP2 device with a WBX antenna manufactured by Ettus Research LLC, and our setup was calibrated in the laboratory before the onsite experiments with the Agilent N9000 signal analyzer. Our measurements were undertaken over a one week period, each session lasting for four hours, in which we obtained $6.25M$ samples per second. Figures 2.2(a), 2.2(b) and 2.2(c) show a sample two hundred second period of activity in channels 37 and the L-bands that is extracted from our measurements. The overall aim of this study is to characterize the channels forming the WMTS band, and establish site-specific activity patterns of the priority users.

Observations

For each given portion of the WMTS band, we measure the variation in power (z-axis) with time and frequency (x-y plane).

- Channel 37: A common observation throughout Figure 2.2(a) is that the existing transmissions exhibit narrow band behavior, and they are mostly contained within a few Kilohertz. This is in line with the specification of the telemetry systems shown in [39]. Temporally, two types of transmissions are seen in the figures, the ones that are intermittent, such as the signal at 609 MHz up to the first 150s, and the rest are
Figure 2.2: Spectrum survey of the channel 37 (a), lower $L$ (b) and upper $L$ (c) bands.
streamed, such as the continuous peaks present at 610 MHz. The streamed signals may correspond to the telemetry indicating the fault-free operation of a medical device, e.g., an infusion pump. The intermittent signals could be alarms, device status report messages, among others.

- **L-Band**: We observe from Figure 2.2(c) that the upper-L-band is somewhat different from the other two bands. The upper-L band shows a wide-band usage with occasional intermittent frequent changes. This activity pattern matches that of non-medical utility telemetry companies that are operating around this area. Moreover, the presence of a continuous peak around 1429 MHz in the available wedge in the spectral graph, indicates the medical telemetry, as it is closely aligned with the shape and duration of the peaks observed for channel 37 (where no utility telemetry is allowed) in Figure 2.2(a). Finally, Figure 2.2(b) shows continuous high power received pulses at the far end of the frequency for the lower-L band, which can only be attributed to active radar pings in the neighborhood.

Table 2.1 lists some of the popular medical telemetry systems used in hospitals. Our interaction with the wireless personnel in one of the two hospitals in which we did our measurements indicated that the GE Appex Pro CH was the most frequently used system. Our plots confirm a static allocation used in these devices since all the transmission activities are confined on their own assigned channel.

**Noise Floor Determination**

Obtaining a precise measurement of the noise floor is critical in correct decision making, a key concern in the WMTS bands that relies on life-saving communication. The noise floor that we measure includes the thermal white noise incurring at the radio frequency (RF) input plus the noise added by the radio receiving chain due to noise figure of the intermediate frequency (IF) components. We use the USRP2 as a measurement tool, and terminate its RF input with a 50 Ω SMA terminator. We then tune the device to the three portions of the WMTS bands (37, upper and lower-L) successively, and collect $5 \times 10^5$ samples each time. We calculate the FFT from these samples using a bin size of 1024. Each bin records the noise power at a resolution of $\frac{6}{1024}$ MHz for the channel 37, and $\frac{5}{1024}$ MHz for the L bands. At the end, the individual power values in each bin are averaged to obtain the reliable noise floor. Note that this method produces bin-specific noise level, unlike a channel-wide noise level commonly assumed in the literature. We show a plot of the measured noise floor for the channel 37 in Figure 2.3 and for the lower-L band in Figure 2.4.
Figure 2.3: Frequency domain signal in the UHF Channel 37 vs. the noise floor.

Figure 2.4: Frequency domain signal in the Lower L-band vs. the noise floor.
In our measurements, we observe a continuously occurring peak at the center frequency of the channel, which is due to signal leakage from the oscillator at the radio front-end. This artifact is observed for any USRP2 device\(^1\), and is added to the incoming signal. This will cause a problem in correct detection of the present signals in the band when a fixed threshold is used, i.e., the center frequency will always and falsely indicate a present signal. By using the above bin-specific noise floor method, the effect of such RF hardware imperfections can also be eliminated from the spectrum sensing. Figure 2.3 shows a true frequency domain signal in the channel 37 with several peaks versus noise floor. Our sensing method for the WMTS band, which overcomes the hardware imperfections, involves the following steps:

- Subtract the antenna gain from the bin power (6.15 dB in our case).
- Subtract a fixed threshold from the bin power to compensate for signal and noise floor variance (3 dB in our case).
- Compare the observed bin power with the previous power of the noise floor and infer the sensing results.

Figures 2.3 and 2.5 how these steps are undertaken to extract the true signal from the measured signal. All eight peaks in the signal have been detected in this process, and more importantly, the leakage peak issue described earlier, is not included in the clean, final signal. This is because during noise measurement, the center peak (i.e., error in the noise floor) is greater than the center peak in the signal after subtracting antenna gain and threshold value from the value contained in the FFT bin. Since this center peak is the internal hardware artifact, its power is constant, with or without the antenna on the radio, and it can hence be easily suppressed through the above steps.

**Inference of Channel Activity**

Our spectrum and power allocation framework relies on obtaining an a priori estimate of the WMTS channel activity at a given location. The longer a specific portion of the channel is expected to be free, the better candidate it becomes for use, as it can support a longer duration of medical telemetry transmissions. From the spectrum measurements, shown in figures 2.2(a), 2.2(b) and 2.2(c), at each time instant, an output of a 1024-point FFT is calculated with a resolution of 6 KHz at each frequency bin. With this measurement setting, the activity at each bin can be analyzed independent of other bins based on the

\(^1\)We tested with 5 different radios, each with the same result and this is widely reported in the USRP2 Community’s mailing list.
“on” and “off” times of the user of that bin. We use a simple probabilistic model of the channel, derived with the help of the exponential distribution, to approximate the “on” and “off” times at each bin. Figure 2.6 shows the histogram of “off” times at two separate bins (out of 1024 bins, one is shown with a bold line and the other with a broken line) within the channel 37 and their fitted exponential PDF. Similar results were obtained in all the other bins in the channel 37 and the L-band. The resulting family of activity models is then utilized to estimate the channel availability in the optimization framework described in Section 2.1.5.

2.1.4 Power and Spectrum Allocation Framework

In this section, we shall provide a sketch of the optimization framework that carefully assigns the transmission power and the available portions of the WMTS channel. We formulate the problem as a two-step process. The first involves identifying the power and modulation for the requested transmission, and then, we identify the specific frequency range to which it may be allotted.

Determining Transmit Power and Modulation

Let the occupancy function $P_{on}(f)$ represent the probability that frequency $f$ is occupied by legacy medical users without dynamic spectrum access ability or utility meter applications (both denoted as PUs) at frequency $f$. From Section 2.1.3, we obtain two functions $\lambda_{on}(f)$ and $\lambda_{off}(f)$ as the mean “on” and “off” exponential durations, respectively.
Based on the PU activity at each frequency \( f \), the sensing durations may be different. One possible method for determining the sensing time \( t_s \) is given in [40].

We begin with the BS computing the power \( p \) for the given vector \((D_{ij}, d_i, R_{avg}, \ell, BER, EMI)\) submitted by a node. We consider a single-carrier communication method with MPSK modulation. The combined effect of the modulation in the signal and the signal power must satisfy the BER constraints given in the QoS vector. At the same time, the transmit power of the node is constrained by the EMI threshold level of the sensitive equipment [41]. For a node \( i \),

\[
p_i \leq \left( \frac{d_i EMI_i}{c} \right)^2 \tag{2.1}
\]

where \( c \) is constant and \( EMI_i \) is the lowest EMI threshold level. The probability of symbol error for MPSK modulation is given as follows [42]:

\[
P_E = 2Q \left( \sqrt{\frac{2E_s}{N_0}} \sin \frac{\pi}{M} \right), \tag{2.2}
\]

where \( \frac{E_s}{N_0} \) is symbol-energy to noise-power spectral density, \( Q(x) \) is the complemen-
atary error function and is defined as
\[ Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left( -\frac{u^2}{2} \right) du, \]  
(2.3)

Also \( \frac{E_s}{N_0} \) is related to \( \frac{E_b}{N_0} \) (bit-energy to noise-power spectral density) via the following relation:
\[ (\log_2 M) \left( \frac{E_b}{N_0} \right) = \left( \frac{E_s}{N_0} \right) \]  
(2.4)

Furthermore, \( \frac{E_b}{N_0} \) is function of SNR and rate \( R \) [43]:
\[ \left( \frac{E_b}{N_0} \right)_{dB} = \left( \frac{S}{N_0} \right)_{dB-Hz} - R_{dB-bit/sec} \]  
(2.5)

The \( \frac{E_s}{N_0} \) in 2.2 can be replaced with \( (\log_2 M) \left( \frac{E_b}{N_0} \right) \).

Considering free-space path loss for transmitted power \( p \), the received power at the receiving node is:
\[ p_r = p \times \alpha D_{ij}^{-\beta} \]  
(2.6)

Bit error rate \( P_b \) and symbol error rate \( P_E \) for the case of MPSK are related as [42]:
\[ P_b = \frac{1}{k} P_E \]  
(2.7)

Combining the above equations, we get:
\[ P_b = \frac{2}{k} Q \left( \sqrt{2m \frac{p_i \alpha D_{ij}^{-\beta}}{N_0 R_p} \sin \frac{\pi}{M}} \right), \]  
(2.8)

where \( m = \log_2 M \). To achieve the given \( BER \) \( (p_b) \),
\[ p_i \geq [Q^{-1}(\frac{1}{2m p_b})]^2 \frac{N_0 R}{2m \alpha D_{ij}^{-3\beta}} \]  
(2.9)

This is the second constraint for transmit power to be assigned to the application. The selected modulation \( (M) \) must satisfy these two constraints integrated in the following optimization problem:
Given: $D_{ij}, d_i, R_p, P_b, EMI$

To find: $m$

To Minimize: $P_{tx}$

Subject to:

$$P_{tx} \leq \left( \frac{d_iEMI}{c} \right)^2$$  \hspace{1cm} (2.11)

$$P_{tx} \geq \left[ Q^{-1} \left( \frac{1}{2m} P_b \right) \right]^2 \left( \frac{N_0 + I}{2mD_{ij}^2} \right)$$  \hspace{1cm} (2.12)

The above optimization problem find the minimum transmission power level for the transmission of device $i$ in order to avoid exceeding EMI level for the sensitive medical equipment (2.11), and also at the same time, satisfy the BER requirement for the corresponding application (2.12).

Finally, we assume the following simple relation between the allocated bandwidth and the rate of the application:

$$B = \frac{R}{m}$$  \hspace{1cm} (2.13)

**Determining Frequency Range**

To find the best channel in the WMTS band, let the availability function $A(f)$ indicate whether a frequency $f$ is currently assigned for medical telemetry (i.e., $A(f) = 0$) or currently used by the PU (i.e., $A(f) = 1$). With this function, the BS must solve the following optimization problem to get the best channel for the submitted application.
Given: $R_p, R_a, \ell, L_p$

To find: $f_u, f_i$

To Minimize: $B = f_u - f_i$

Subject to:

\[ \int_{f_i}^{f_u} A(f) = 0 \]  \hfill (2.15)

\[ \int_0^{\infty} t\lambda_{\text{off}}^{\text{max}}(f)e^{-\lambda_{\text{off}}^{\text{max}}(f)t} dt \geq \frac{L_p}{R_a} \]  \hfill (2.16)

\[ \lambda_{\text{off}}^{\text{max}}(f) = \max_{f_i} \lambda_{\text{off}}(f) \]  \hfill (2.17)

\[ \max_t s(f)|_{f_i}^{f_u} = t_s^{\text{max}}(f) \leq \ell \]  \hfill (2.18)

\[
\frac{T}{T + t_s^{\text{max}}(f)|_{f_i}^{f_u}} \times (f_u - f_i)m \geq R_a
\]  \hfill (2.19)

\[ T = \int_{f_i}^{f_u} \int_0^{\infty} t\lambda_{\text{off}}(f)e^{-\lambda_{\text{off}}(f)t} dt df \]  \hfill (2.20)

where $\ell$ is the Latency in the QoS vector and $L_p$ is the packet length. We define the constraints in (2.38)-(2.20) as follows:

- The first constraint (2.38) on choosing the band is to make sure that it is currently free and not occupied by any other user (either primary or secondary). If the band spans several bins, it means that the summation of the availability function on all bins must be zero.

- When a band containing several bins is picked, the average time left to the first estimated appearance of a PU must be longer than the time needed to send at least one packet (2.39). At each bin, $\lambda_{\text{off}}$ represents the inverse of the average of the exponential distribution or average time left to next appearance of a primary user. (2.17) finds the bin with lowest such value among the potential bins to be picked as the solution.

- Since we derive different required sensing times for each bin from [40], among the potential solution bins, the highest sensing time must be picked for the whole band. However, this value must be lower than the latency $\ell$ permitted by the application (2.18).
• The effective rate of the application considering the sensing duty cycle and the modulation level must be at least equal to the average required rate by the application (2.19).

• Finally, (2.20) calculates the average free time for the potential band.

The node can now immediately begin its transmission with these parameters returned by the BS.

2.1.5 Network Management Framework

In this section, we shall provide a sketch of a framework with the help of the flowchart shown in Figure 2.7 handles nodes’ requests for transmission using the optimization framework in section 2.1.4 and monitor their quality of service by keeping an eye on the interference in the channel and also protecting EMI-sensitive equipments by controlling the transmission power of the nodes when they move around.

Step 1: Initial Spectrum Sensing

We assume a 1-hop architecture, in which devices wishing to transmit patient information to a central base station (BS) must first obtain the transmit power and specific start-end frequencies in the WMTS band that it may use, by sending a request message on the common control channel (CCC) to the BS. Here, the entire WMTS band is continuously sensed with a very fine frequency resolution of 6 KHz by the BS. After each sensing period which according to our measurements is in the order of a few milliseconds, the empty and occupied frequency bins on the frequency scale are determined using energy detection. If any earlier assigned spectrum is found to be reclaimed, then those ongoing CR medical data transmissions are inserted back in the pool of non allotted but requested services. At this time, the BS also checks for any new incoming message on the CCC from the member devices, and based on the contents of the message, one of the steps 2 – 4 is chosen. In the current setting, we assume an out of band CCC presumed to be on any unlicensed band, e.g. ISM band.

Step 2: New Request for Communication

If the incoming message at the BS contains a new request for communication by a node, we first identify the maximum permissible power that can be assigned to that node. During this procedure, first of all, the $L$-band is checked for availability. This band has the lowest preference as the radio frequency (RF) electromagnetic interference (EMI) caused
Figure 2.7: Flowchart of the proposed transmission parameter optimization framework
by transmission below 800 MHz (i.e., in channel 37) is half that of comparatively higher frequencies (i.e., frequencies of the $L$-band). Then, a list of nearby $^2$ EMI sensitive devices are queried for their measured values based on the current transmission environment. The transmission power of the requesting device is then gradually increased to a maximum that still contains the EMI within safe limits. Additionally, Based on this amount of power, node’s distance to the base station and approximating path loss and fading, the signal to noise ratio (SNR) can be estimated. The BS decides whether the resulting SNR will guarantee the application specified outage ratio (in terms of packet error rate) for the corresponding node. If not, and after all frequency bands are explored with a similar negative outcome, the incoming request is dropped and a notification is sent to the doctor for in-person monitoring. If the power assignment has a feasible solution, then the frequency assignment is undertaken (which is also the action performed on a drop in service quality, and explained in Step 3).

Step 3: Drop in Quality of Service

A drop in service quality indicates that the current bandwidth is insufficient to meet the application demands, possibly owing to congestion or errors caused by non PU entities. In such cases, the BS attempts to find an alternate portion of the frequency. Additionally, if the power assignment from Step 2 is feasible, then a spectrum assignment procedure is carried out. In this step, according to the requested rate from the node, the bandwidth value is calculated. Then, the empty bins (identified in Step 1) are exhaustively searched for a contiguous band with maximum expected available time (statistics are drawn from the PU activity model estimated in Section 2.1.3). Upon determining a suitable frequency range, the node can be immediately asked to begin transmission by the BS.

Step 4: Movement Detected

In case a node detects persistent mobility, say through a pre-installed grid of RFID tags and readers in the hospital floor, the impact of the new location is assessed. This can be a key issue during hours where multiple patients may congregate at a common location. The immediate response of the BS in this case is to lower the transmit power, and if that results in an unacceptable estimated SNR, then it searches for a slot in channel 37. Steps 2 and 3 are now repeated in that order to find the optimal assignment, failing which a doctor alert is issued.

$^2$It is shown in [44] that hazardous EMI to medical devices (depending on the type of device) above 5 meters radiated by common mobile radios, like cellular phones, is highly unlikely.
2.1.6 Performance Evaluation

In this section, we show an indicative performance of our proposed method assuming a network of static nodes, and preliminary results are shown in figures 2.8(a), 2.8(b). Our simulations were run in MATLAB, based on a service area of $140 \times 140m^2$ with 7 types of medical telemetry applications, including Device Telemetry, Diagnostic Telemetry, Telemetry Alarm, Clinician Notifier, BCMA, Infusion Pump Status and Infusion Pump Alarm, randomly deployed in the area, whose specific latency, bandwidth and error rate requirements are listed in [45]. Also, 20 EMI-sensitive devices were randomly deployed in the area. Our simulations are conducted starting with the vector $(60, 21, 22, 20, 19, 81, 18)$ that based on [45] gives a realistic number of active transmitters of each telemetry application listed above respectively, and then scaling up the numbers by a load factor (i.e., an integral multiplier) at every run. We evaluated two metrics for performance comparison against static frequency allocation of these devices (where no spectrum switching is performed and each node transmits on a fixed frequency), average unused WMTS bandwidth and average interfered bandwidth to the higher priority users due to the dynamic spectrum assignment. Figure 2.8(a) shows that the proposed dynamic spectrum access achieves a significant increase in the efficient use of unused bins and a graceful reduction of residual capacity with increasing load. Figure 2.8(a) shows that interference averaged over the various telemetry devices are tremendously reduced. Specifically, it can be seen in both figures that interfered bandwidth and unused bandwidth stay respectively constant for the static allocation, a direct impact of the rigidity of the allocation when the performance requirements are not met and channels are not re-allocated.

2.1.7 Open Research Challenges

We identify the following research challenges that need to be addressed for deploying a feasible CR enabled medical telemetry network:

- **Identifying low-power priority users**: Priority users such as utility telemetry have priority access in portions of the $L$-band, but the detected peak power for these applications may still be in the range of $-100$ to $-70$ dBm. Thus, clearly distinguishing the presence of these users and establishing a dynamic noise floor is a pre-requisite for successful operation of the system. The noise floor includes the thermal white noise incurring at the RF input plus the noise added by the intermediate frequency (IF) components. Our own method for countering the effect of noise was to detect the spurious power in the USRP2 by terminating its RF input with a $50\,\Omega$ SMA terminator, and then tuning it to each of the WMTS bands. However, automatic noise floor diagnostic tools in software need to be devised for seamless operation of the
Figure 2.8: Performance comparison of dynamic spectrum allocation vs. static allocation based on, average unused bandwidth (a), and average interfered bandwidth (b)
Meeting application requirements: One of the key requirements of the electronic medical record (EMR)-based information access for hospitals is the specified packet delivery rate of 99.95% for patient data that includes the formats of X-ray, CT, PET, MRI, ultrasound, and any other test results [45]. Additionally, various types of telemetry data, e.g., EMR images and numerical data wirelessly requires peak data rates values of 4, 100 Kbps and 49.2 Kbps, respectively [45], with a maximum allowed latency of 200 ms. With such strict bounds, an end to end reliable delivery mechanism is needed. The field of transport protocol design for cognitive radio networks is still in a nascent stage, and rapid strides need to be made in this direction to ensure life-critical communication is provided with error recovery and congestion control capabilities. The access method described in this section looks at a single hop (i.e., device to BS), but distributed architectures will involve careful design of the end-to-end protocols. A similar cognitive radio architecture with power allocation and Link-Layer based QoS provisioning has been presented in [46], although dynamic spectrum access capability is not foreseen in the model.

Predicting mobility and impact on EMI: In our work, the mobility is self-detected by a telemetry device. Instead, with assistance from the network, the mobility of each user can be predicted and the change in the EMI at the various sensitive equipments can be calculated in advance. As an example, Kalman filtering could be used with restrictions on the number of degrees of freedom, given that hospital corridors are linear and motion is generally along a single dimension. Moreover, accurate channel-specific EMI formulations and data from non-networking related sources, such as break times etc., need to be integrated in the framework, which will allow a fine-grained prediction of the impact of simultaneous transmissions.

Video/audio transmission: Current FCC regulations do not allow the use of the WMTS bands for audio/video transmission. One possible solution is to leverage the vacant DTV channels 21 – 51 for this purpose. There is a growing thought towards enabling high bandwidth video access to doctors within operating rooms to allow them to visualize not only feeds from the operating table, but also inputs from distant experts without the clutter of wires [37]. The recent ruling by the FCC on spectrum databases provides guarantees on detecting available spectrum in these permissible DTV bands, which will help realize the video and audio needs of the future.

Note: Channel 37 is assigned for WMTS and unavailable for TV use
2.2 Cognitive Radios in Vehicular Networks

2.2.1 Motivation

After a decade of research and investments in vehicular networks, the large-scale deployment of distributed systems based on inter-vehicular communication appears as a concrete possibility in the next few years. As an example, in the US and Europe, the bands at 5.85 – 5.925 GHz of the wireless spectrum has been reserved for both inter-vehicular and vehicle to roadside base station (BS) communication, while the transmitting operations in these bands have been regulated by the IEEE 802.11p and IEEE 1609.4 standards.

At the same time, several applications for drivers’ safety (e.g., adaptive cruise control systems) and traffic monitoring (e.g. traffic information and monitoring systems) have been already implemented and evaluated in realistic urban environments. One of the concerns, however, is the need for strict bandwidth and delay thresholds to support these vehicular applications, especially in light of different quality of service (QoS) requests arriving from multiple different sources, and in urban scenarios where the devices contending for the channel are significantly high [47]. The resulting problem of spectrum scarcity for 802.11p-based vehicular applications has been demonstrated in [48]. This growing spectrum scarcity problem will likely become more acute, given the spurt in high-bandwidth multimedia applications (e.g., video-streaming) for in-car entertainment, and for driver-support services, like multimedia enabled assistance.

CR, an enabling technology for opportunistic spectrum use, directly benefits various forms of vehicular communication. In such a cognitive radio equipped vehicular (CRV) network, each vehicle implements spectrum management functionalities to (i) detect spectrum opportunities over digital TV frequency bands in the UHF range, (ii) decide the channel to use based on the QoS requests of the applications, (iii) transmit on it, but without causing any harmful interference to the licensed owners of the spectrum. However, CRVs have many unique characteristics that involve additional considerations than merely placing a CR within a vehicle. As an example, unlike static CR systems, the spectrum availability perceived by each moving vehicle changes dynamically over time as a function not only of the activities of the licensed or primary users (PUs), but also based on the relative motion between them. Thus, spectrum measurements need to be undertaken over the general movement path of the vehicles, leading to a path-specific distribution, instead of focusing on the temporal axis alone. While this calculation is non-trivial, the CRV network can also leverage the constrained nature of motion, i.e., along linear and pre-decided paths corresponding to streets and freeways. At busy hours, or in urban areas, spectrum information can be exchanged over multiple cooperating vehicles, leading to an advance knowledge of the spectrum availability. This also allows the following vehicles to adapt
their operations and undertake a proactive response, which is infeasible both in static or non-stationary scenarios with random motion.

We envisage CRV networks to fall under three broad classes shown in Figure 2.9. In the first example, such networks could be formed between vehicles only that rely on cooperation for increasing accuracy. The second class deals with periodic interactions between vehicles and roadside BSs, where the latter acts as a repository of data that is used by subsequently passing vehicles. Finally, a completely centralized network is possible, which undoubtedly simplifies the operation, but brings in the concern of whether individual nodes are able to reach this central BS through direct broadcasts.

We next describe the possible applications of CRVs, which will undoubtedly result in higher levels of adoption and widespread use of this technology.

### 2.2.2 Applications of CRVs

The particular choice of transmission frequency, the bandwidth available for transmissions and the interference caused in that range are important factors influencing applications.
In the following, we describe how CRVs will change existing and emerging vehicular applications.

- **Vehicle to Vehicle (V2V) Communication:** In high traffic areas, delays caused by accidents, road blockages, road repairs and slow traffic can be avoided by communicating average velocity, acceleration, brake status, all of which require periodic exchange of data with neighboring vehicles. These systems generally operate in the bands of 5.9 GHz. Practical systems created by Honda and Volkswagen-led consortium [49] have their transmission ranges limited to a few dozen meters, which also impacts the distance at which a corrective action is undertaken. There is, hence, a motivation to use lower frequencies in the sub-GHz range as the signal propagates much further, increasing the effectiveness of the response significantly.

- **Entertainment and Information Systems:** In-car streaming video entertainment options, as well as driver assistance through real-time feeds on traffic, weather, and visual inputs from external cameras are finding increasing commercial acceptance. These applications have strict bandwidth and QoS requirements. While recent work has approached both theoretical and practical aspects of multimedia delivery in vehicles [47], there are inherent scalability limitations of using fixed range or unlicensed bands alone, further motivating the use of CR technology.

- **Public Safety Communication:** The breakdown of the public safety communications infrastructure has occurred repeatedly in large scale natural disasters, such as the recent hurricane Katrina wherein public safety personnel has to resort to non-electronic means of communication. CRVs will allow distributed spectrum access in the un-congested licensed frequencies, which is especially useful for mobile public safety personnel that operate in the field during such outages, and also those beyond the reach of the fixed infrastructure installations.

### 2.2.3 Characteristics and Features of CRVs

In this section, we describe the characteristic features of CRVs that play a vital role in their design, and also differentiate them from general purpose CR networks.

- **Integration with spectrum databases.** Recent FCC rulings foresee the creation of spectrum databases that specify two modes of device operation [8]. Mode II devices have geo-location and database access to maintain a continuously updated spectrum occupancy list, while Mode I devices periodically query Mode II devices for spectrum updates. In context of CRVs, each vehicle must hence have a dedicated out of band radio interface to directly query the database (for Mode II), or multiple roadside
BSs must be placed, akin to mile markers, that feed this information to an on-board radio operating in Mode I. The strict requirements that Mode I devices must receive updates every 60 s influences the placement and density of roadside BSs.

- **Impact of mobility.** The FCC ruling also specifies a sensing-only mode, in which a CR device relies on local sensing results, once it is rigorously certified [8]. In a CRV, a vehicle may collect multiple sensing samples at different locations, but inside the same PU activity region. Since the collected samples might exhibit different degrees of correlation based on the characteristics of the environment (e.g., presence of buildings) and the speed of the vehicle, merging these data points correctly is important [50]. Thus, spectrum management in CRVs cannot be performed without considering traffic conditions and vehicular speed, among others.

- **Role of cooperation.** In CRVs relying on sensing-only mode, cooperation among vehicles can be leveraged through their predictable mobility pattern to perform enhanced spectrum sensing and decision. Here, a vehicle is aware of the spectrum resources available on its path in advance, i.e., before arriving in the area of interest by leveraging the spectrum information provided by other vehicles. However, the set of cooperating neighbors may dynamically change as an effect of the varying topology, thus impacting the performance of the cooperation scheme over time. A comparative description of the characteristics of sensing performed by a single vehicle locally (Per-vehicle), using spectrum databases (Geo-location), leveraging inputs from multiple neighbors (Cooperative), and relying on assistance from local BSs (Cooperative with BSs) are summarized in Table 2.2.

- **Presence of a common control channel (CCC).** In CRVs, a control channel is already provided in the 5.9 GHz band through the 802.11p protocol, which addresses some of the concerns present in general CR networks. However, there is still the possibility that the CCC in the 5.9 GHz band becomes easily saturated in congested scenarios (e.g., peak hours of traffic) as indicated in [48]. Hence, control messaging must be minimized in CRV networks, or additional control channels must be identified.

### 2.2.4 Classification of Existing Schemes for CRVs

Research on CRVs has mainly focused on spectrum sensing and spectrum access. The classification of existing schemes is shown in Figure 2.10.
Existing works on CRVs

Spectrum Sensing
(Detecting Presence of PUs)

Spectrum Access
(Link Layer Transmission)

Cooperative Geo-location Per-vehicle

PU protection PU protection
alone with QoS

with RSUs with vehicles

Figure 2.10: Classifications of existing works on CRVs.

Spectrum Sensing

This key component of CRV networks ensures that the spectrum availability is correctly detected, which is a challenge in highly mobile scenarios.

- **Per-vehicle sensing techniques** In this approach, CR vehicles sense the TV band by using any traditional sensing technique proposed in the literature of CR systems: i.e. energy-detector, matching-filter or cyclo-stationary techniques [35]. Since each CR performs spectrum sensing and decision autonomously, the implementation complexities and the network support are minimal. At the same time, there are several concerns on the accuracy provided by per-vehicle sensing techniques. While the thresholds imposed by IEEE 802.22 standard on the CR receiver sensitivity are very low (i.e. -116 dBm for TV bands), the sensing output in presence of obstructed environments and high mobility might be easily biased by propagation phenomena like fading or shadowing effects.

- **Geo-location based techniques** Recent FCC directives [8] suggested the utilization of geo-location database as alternative or complementary technique to sensing-only schemes. Geo-location database can provide information about the bands, including the exact types of PUs, their locations and their specific protection requirements. As a result, the CR-vehicles do not need to sense for the presence of PUs, and can adjust their transmitting parameters not to interfere with the licensed users. This solution is suitable for CRVs, since each vehicle might be likely equipped with self-localization systems (e.g. GPS device). Moreover, the digital maps might be easily integrated
with PU information provided by a spectrum allocation database (e.g. the TV Query service in US). However, there are important concerns on the implementation of geo-location techniques, i.e. the costs for building the database, the coverage area of the service, the requirements for the CR vehicles. For this purpose, some recent works [52] investigate the possibility to use geo-location information in conjunction with sensing (cooperative or not) techniques. One advantage of this joint technique is that CR vehicles might still be able to access the licensed TV spectrum in regions where the database information are incomplete or missing. In section 2.2.6, we lay out an cost-benefit analytical framework to obtain a density of the infrastructure installation along roads based on the desired service covered area.

- **Cooperative techniques with the help of external BS** A new type of cooperative sensing approach is described in [53]. Here, the network uses a stationary BS that provides coordination instructions to the passing vehicles, which can then undertake the final sensing results locally. The BS continuously gathers information of the PU occupancy at its location through energy detection, which is a fast but coarse result, and leaves fine-grained sensing (e.g., feature detection) to the CR-vehicles. The advantage of this method is that any change in government policy of new regulatory standards, as well as PU parameters can be easily loaded at the BS for further adapting the operation.

- **Cooperative techniques between “any” neighboring vehicles** A problem with a centralized fusion center or BS is pointed out in [54], where the authors argue that each vehicle may have a different view of the spectrum usage, based on its location. Instead, they propose a belief propagation method that requires each CRV to periodically send out its respective belief of the presence of a PU. Each vehicle combines these belief vectors with its own, to generate a new belief. This is then passed on, and after several iterations, the network is envisaged to enter into a steady state. Several issues need to be further explored in this work, such as the speed of convergence in the network, the practical and theoretical bounds of belief techniques, the extent of belief propagation with respect to vehicle velocity, among others.

- **Cooperative techniques between “selected” neighboring vehicles** The problem of choosing the extent of cooperation in a vehicular environment based on specific concerns of correlated readings, and a priori spectrum availability detection at future anticipated locations, is addressed in [55]. The road is divided into short segments, allowing a vehicle in the current segment to gather spectrum information up to \( h \) segments ahead from vehicles in front. Thus, when the vehicle determines its own spectrum to use for the next portions of the road, it broadcasts its choices, thereby
allowing subsequent vehicles to also adjust their own parameters. A correlation-aware weighted average is used at each vehicle to combine the results of the local sensing activity with the information coming from other CR-vehicles.

**Spectrum Access**

Once the sensing results are known, correctly choosing the spectrum and the appropriate method of using it has a manyfold impact, which is the focus of this section.

- **Licensed/primary user (PU) protection alone** A range of metrics, such as spectrum with the highest (i) data rate, (ii) product of rate and channel utilization, and (iii) product of rate and expected vacant channel duration are defined in [56] as factors that influence spectrum selection and access. More relevant to CRV networks is the consideration of the inter-vehicular communication duration, i.e. the maximum extent of time for information exchange between communicating CRV pairs, and how to ensure protected contours of PU-DTV transmissions are avoided.

- **Spectrum access with QoS support** A three-pronged approach is proposed in [57] for clustered vehicles, involving selection and access of shared channels, exclusive-use channels, and cluster size control under the dual constraints of meeting the QoS specifications and PU protection. The shared channels belong to the licensed bands, and used for inter-cluster communication and may not be always available for use by the CRVs. Conversely, the exclusive-use channels are typically reserved portions of the spectrum, such as short range transmission frequencies specified in the IEEE 802.11p standard, and used for intra-cluster networking. There are two constrained Markov decision process (CDMP) formulations in this work, one for the spectrum access based on the queue lengths (short term strategy), and a joint bandwidth and cluster size optimization (long term strategy).

- **Spectrum access with delivery guarantees only** When a complete QoS provision is unavailable or not required, issuing guarantees for data delivery will assist in reliable transmission of public safety information. This is indeed a problem in classical IEEE 802.11p based networks that have well defined control channel (CCH) and service channel (SCH) durations of 50 ms each. Safety message delays should be less than 200 ms for adequate response time of the drivers, and coupled with the need for message repetition, the window of 50 ms appears inadequate [48]. Hence, the authors propose a feedback-loop method where local BSs assist vehicles in securing additional spectrum. The CR-vehicles periodically provide feedback to the BSs about the spectrum usage, allowing the incremental addition of new spectrum to the network pool to the extent that channel contention is below a pre-decided threshold.
2.2.5 Open Research Issues

In the following, we discuss some important research issues relative to the impact of the vehicular mobility, security aspects, and the evaluation methodology for CRVs.

Impact of Vehicular Mobility on Spectrum Management

Mobility has both positive and negative implications for CRVs. On the one hand, the Doppler spread caused by mobility might result in inaccurate detection of occupied frequencies, which may cause harmful interference to the PUs. On the other hand, a moving CRV node can collect signal samples at different locations along its path, thus increasing the spatio-temporal diversity of the samples, and reducing the risk of incorrect decision caused by shadowing effects. In [50] the authors investigate the performance of spectrum sensing with mobile CR sensors through a theoretical study, and they show that the accuracy of sensing increases with the vehicle speed, due to the increased spatio-temporal diversity. Also, they demonstrate the existence of an interesting trade-off between sensing scheduling and cooperation among nodes. When CRV nodes are moving at high speed, it is more efficient to sense multiple times than wait for predetermined intervals between two sensing instants, since the observations are likely not correlated. In the converse case, when CRV nodes are moving slowly, then it is better to cooperate with other nodes than sensing the channel with high frequency. However, this work does not take into account the planned paths and the constrained nature of the movement caused by the road topology. For instance, the collaborative spectrum framework described in [55] allows a vehicle to gather information about spectrum opportunities over an enlarged area which includes the current location and also past and future locations. Our sample experiments undertaken with a USRP device placed on top of a moving vehicle in [55] also demonstrate that sensing accuracy is a function of both specific locations and speed. These results are summarized in Figure 2.11 for three locations: A that has moderately spaced buildings of medium height, B which a bridge location with open expanse on either side, and C that is a busy downtown street in Boston.

Despite these preliminary results, there are still many unanswered questions on the impact of vehicular mobility on sensing performance, which are summarized as follows: What is the optimal technique to balance cooperation and spectrum scheduling for a moving CRV node? How can the predictable mobility of vehicles be exploited to increase the spectrum awareness? What is the impact of mobility parameters (e.g. speed and direction) on the sensing performance?
Security aspects of CRVs

Cooperation among vehicles improves the spectrum sensing accuracy. At the same time, it also poses several concerns from the point of view of trust and security [15]. CR radios may develop genuine hardware faults, or turn malicious on being compromised, i.e., the users can falsify sensing reports like reporting the presence of a PU to guarantee the exclusive access on a channel. Existing schemes proposed for CR networks address the problem by assigning different weights to each node based on its trust, and then combining local decisions based on the weight of each node [15]. The situation is significantly more involved in mobile environments like CRVs, where the neighbor set might change over time. In other words, identifying a malicious vehicle which sends false sensing reports while moving might require fast detection and correction, as compared to the static case. Additionally, the implementation of collaborative spectrum sensing in CRVs poses privacy concerns of end-users. For instance, a potential attacker might track the identify and movements of a driver by eavesdropping the periodic spectrum information broadcast by the vehicle.

In conclusion, the research questions related to security in CRVs involve: How can the system detect the presence of malicious CRVs which broadcast fake sensing reports while moving? and How can the privacy of cooperating vehicles be protected?

2.2.6 Spectrum Database Assisted Cognitive Radio Vehicular Networks

The recent rulings by the FCC mandating the creation of spectrum databases [8] have reduced the erstwhile complete dependence on spectrum sensing alone to infer the primary
user (PU) activity in the licensed bands. Now, cognitive radio (CR) devices can query dedicated databases for correct and updated knowledge of the actively used channels in their neighborhood. While this is a significant step towards realizing practical CR implementations, the new resulting network architecture and associated connectivity rules raise several challenges in the seamless operation of the network. This section takes the first steps towards an analytical formulation for the placement of spectrum database query points, and is aimed to serve as a design tool for creating such database assisted networks for mobile or vehicular scenarios. We analyze in depth the cost-benefit tradeoff of the resulting network, while factoring in diverse considerations of vehicular density, road dimensions, querying costs, permissible error bounds, among others.

These network components are shown in Figure 2.12 for a road length $L$, and explained in further detail below:

Figure 2.12: Different modes of operation in a vehicular CR network

We assume that vehicles are equipped with (i) geolocation ability through GPS (or any other form of accurate localization tool), (ii) 3G connectivity that allows sending queries requesting spectrum information, and (iii) local spectrum sensing hardware, with varying probabilities. A combination of these features directly result in the mode of operation described above. As an example, a device with geolocation and 3G connectivity assumes the role of a Mode II device. The main contribution in this section is the formulation of an analytical optimization framework that gives the covered percentage of the road by fixed BSs spread uniformly for a given length of a roadway, while taking into consideration a probabilistic estimate of the CR devices that facilitate database access, and those who must rely on such services. This framework finds the optimal solution subject to the constraints of permissible error in estimating the available spectrum and more than zero revenue received at each BS, while minimizing the cost of operation of the network.
Table 2.3: Empirically measured traffic parameters from [1]

<table>
<thead>
<tr>
<th>Time of Day</th>
<th>Traffic type</th>
<th>λs [veh/m]</th>
<th>λt [veh/s]</th>
<th>fit goodness</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:00 am - 03:00 am</td>
<td>low</td>
<td>0.0039</td>
<td>0.1192</td>
<td>(2.65, 2.42)</td>
</tr>
<tr>
<td>10:00 am - 12:00 pm</td>
<td>normal</td>
<td>0.0250</td>
<td>0.7276</td>
<td>(3.13, 9.27)</td>
</tr>
<tr>
<td>15:00 pm - 17:00 pm</td>
<td>high</td>
<td>0.0728</td>
<td>0.7813</td>
<td>(8.09, 9.67)</td>
</tr>
</tbody>
</table>

Table 2.4: Empirically measured traffic parameters from [1]

<table>
<thead>
<tr>
<th>Vehicle Equipment</th>
<th>Spectrum Sensing</th>
<th>Internet Connection</th>
<th>Geolocation</th>
<th>Device Mode</th>
<th>Case Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>probability</td>
<td>x₁</td>
<td>x₂</td>
<td>x₃</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 - x₁)(1 - x₂)(1 - x₃)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Mode I when connected to BS or a Mode II, otherwise silent</td>
<td>I</td>
</tr>
<tr>
<td>(1 - x₁)(1 - x₂)x₃</td>
<td>-</td>
<td>-</td>
<td>√</td>
<td>Mode II when connected to BS or a Mode II, otherwise silent</td>
<td>II</td>
</tr>
<tr>
<td>(1 - x₁)x₂(1 - x₃)</td>
<td>-</td>
<td>√</td>
<td>-</td>
<td>Mode I when connected to BS or a Mode II, otherwise silent</td>
<td>III</td>
</tr>
<tr>
<td>(1 - x₁)x₂x₃</td>
<td>-</td>
<td>√</td>
<td>√</td>
<td>Always in Mode II</td>
<td>IV</td>
</tr>
<tr>
<td>x₁(1 - x₂)(1 - x₃)</td>
<td>√</td>
<td>-</td>
<td>-</td>
<td>Mode I when connected to BS or a Mode II, otherwise Sensing-Only</td>
<td>V</td>
</tr>
<tr>
<td>x₁(1 - x₂)x₃</td>
<td>√</td>
<td>-</td>
<td>√</td>
<td>Mode I when connected to BS or a Mode II, otherwise Sensing-Only</td>
<td>VI</td>
</tr>
<tr>
<td>x₁x₂(1 - x₃)</td>
<td>√</td>
<td>√</td>
<td>-</td>
<td>Mode I when connected to BS or a Mode II, otherwise Sensing-Only</td>
<td>VII</td>
</tr>
<tr>
<td>x₁x₂x₃</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>Always in Mode II</td>
<td>VIII</td>
</tr>
</tbody>
</table>

Table 2.4: Type of wireless equipment in vehicle and their probability distribution.

Proposed Optimization Framework for BS Placement

The opening of White spaces for unlicensed use is expected to create a boost across all sectors of the economy in the coming years [58]. This is due to the fact that a lot of applications could benefit from the availability of white spaces. CR vehicular networks constitute an emerging application area for opportunistic access of licensed spectrum for vehicle to vehicle communication, streaming entertainment, and various driver-assist services. Recent measurement data collected in [1] on a single day on the 5-lane I-80 highway between Emeryville, CA and Berkeley, CA, provided information on various aspects of vehicular traffic, including inter-vehicle spacing, inter-arrival time and vehicle speed. In our analysis, we rely on the availability of such information for a given length of the road, which can be gathered from off-line studies. Here, inter-vehicle spacing is the distance of any arbitrary car in the road to its closest car in any lane of the road, irrespective of their respective directions. Inter-arrival time is the arrival time of any car at the observation point relative to the arrival of the previous car. The result of the data analysis in [1] suggests that inter-vehicle spacings and inter-arrival times are exponentially distributed. However during low traffic time when the traffic is less than 1000 veh/hr, the accuracy of the fit is the highest. Table 2.3 shows the exponential mean λs and λt for inter-vehicle spacing and inter-arrival time at 3 different times of the day outlined in [1], which we use in our analysis.

The overall aim of this analysis is to provide guidelines for the uniform placement of
the BSs that will provide information about the availability of DTV channels, and thereby support the mobile CR enabled vehicles. The BSs are fixed, directly connected to spectrum databases, and have transmission radius of $R_t$. The actual number of deployed BSs along the road directly influences the density of deployment $\alpha$. Our optimization framework determines this $\alpha$ so as to minimize the cost of spectrum information access and also maximize the accuracy of spectrum information, irrespective of the method it is obtained either from the database or via local spectrum sensing (in the event where getting the database information is impossible).

We assume that each CR user operates in a saturated state, constantly looking for free white space to transmit data to peers. To help access the spectrum information, each device can be potentially equipped with three different options, i.e. (i) Spectrum Sensing Module, (ii) Internet connection through 3G, 4G or other related technologies, and (iii) GPS for accurate location determination at each point in time, as shown in Table 2.4 with independent probabilities $x_1$, and $x_2$, and $x_3$, respectively. Depending on the available options on each CR device, and also its position on the road relative to other network components, we identify the operation of the device under various classes as follows:

**Case I.** In case I where none of the three options are available for the CR, the device can only be in Mode I. So it has to get spectrum information from nearby CR devices in Mode II or, when in range, from a Fixed BS. Probability of having at least one Mode II CR device in range around an arbitrary node is given by Lemma 1:

**Lemma 1.** If the inter-vehicle spacing has an exponential distribution of mean $\lambda_s$, and probability of a vehicle being in Mode $\Delta$ is $P_\Delta$, then the probability of having at least one Mode $\Delta$ device in range $R$ is $1 - e^{P_\Delta R \lambda_s}$.

**Proof:** With an exponential distribution of the inter-vehicle spacing with mean $\lambda_s$, the number of vehicles in range $R$ of an arbitrary vehicle has a poisson distribution with mean $R \lambda_s$ [59]. Thus, the probability of having at least one vehicle in Mode II, is calculated as follows:

$$A = \sum_{i=1}^{\infty} \frac{(R \lambda_s)^i e^{-(R \lambda_s)}}{i!} [1 - (1 - P_\Delta)^i]$$  \hspace{1cm} (2.21)

In the above equation, the probability of having at least one Mode $\Delta$ device is $(1 - (1 - P_\Delta)^i)$, given $i$ nodes in the range. Using taylor series expansion, $A$ can be reduced to:
\[ A = e^{-(R\lambda_s)} \left[ \sum_{i=1}^{\infty} \frac{(R\lambda_s)^i}{i!} - \sum_{i=1}^{\infty} \frac{((1 - P\Delta)R\lambda_s)^i}{i!} \right] \]
\[ = e^{-(R\lambda_s)} [e^{R\lambda_s} - 1 - e^{(1-P\Delta)R\lambda_s} + 1] \]
\[ = 1 - e^{-P\Delta R\lambda_s} \]  
(2.22)

Given Case I, probability of being Mode I, is:

\[ P_{MII|CaseI} = \alpha + (1 - \alpha)A \]  
(2.23)

Additionally, we choose to call the condition when the CR user is not any of the FCC specified modes (i.e., I, II or sensing-only), as the silent mode. With respect to case Case I, such a probability is derived as:

\[ P_{silent|CaseI} = 1 - \alpha - (1 - \alpha)A \]  
(2.24)

Case II. In this case only the GPS is installed and no Internet connection exists. Consequently, the CR device is in Mode II only when it is in the range of a BS or yet another Mode II device, from which it can obtain spectrum database access. When no database access by any means is possible, the CR device is technically in Mode I. However, since the FCC rulings do not permit the sharing of channel information among other Mode I devices, the CR user must go into silent mode until it moves to a position where database access is possible. Probability of being in Mode II, in this case is:

\[ P_{MII|CaseII} = \alpha + (1 - \alpha)A \]  
(2.25)

Note that the probability of silent mode is similar to that in Case I.

Case III. The CR user is in Mode I and the probability of silent mode similar to Case I.

Case IV. In this case, where both internet connection and GPS is available, the CR user is always in Mode II.

\[ P_{MII|CaseIV} = 1 \]  
(2.26)

Case V. In this case, only spectrum sensing module is installed, the CR user switches between Mode I and sensing-only modes. Mode I occurs when there’s a nearby BS or Mode II device to transfer the information, and otherwise it becomes sensing-only. The probabilities of Mode I and sensing-only correspond to probabilities of Mode I and silent mode in Case I respectively.
Case VI. Here, both GPS and the sensing module are installed, and the device switches between Mode II and sensing-only mode. The probabilities of Mode II and sensing-only corresponds to probabilities of being in Mode II and silent mode in Case II, respectively.

Case VII. In this case, the CR user switches between Mode I and sensing-only mode similar to case V. The probability of being in each mode is exactly the same as Case V.

Case VIII. Finally, in case VIII, the CR user is always in Mode II, the same as Case IV.

Summing up on the Mode I probabilities on different cases, the cumulative probability of a CR user being in Mode I is:

\[ P_{MI} = P_{MI|CaseI} P_{CaseI} + P_{MI|CaseIII} P_{CaseIII} + P_{MI|CaseV} P_{CaseV} + P_{MI|CaseVII} P_{CaseVII} \]

\[ = [\alpha + (1 - \alpha)A](1 - x_3) \]  

(2.27)

Similarly, the probabilities of being in Mode II, sensing-only mode, and silent mode, respectively, are:

\[ P_{MII}^1 = (1 - x_2)x_3[\alpha + (1 - \alpha)A] \]  

(2.28)

\[ P_{MII}^2 = x_2x_3 \]  

(2.29)

\[ P_{MII} = (1 - x_2)x_3[\alpha + (1 - \alpha)A] + x_2x_3 \]  

(2.30)

\[ P_{sense} = x_1(1 - x_2x_3)[1 - \alpha - (1 - \alpha)A] \]  

(2.31)

\[ P_{silent} = (1 - x_1)(1 - x_2x_3)[1 - \alpha - (1 - \alpha)A] \]  

(2.32)

Here, \( P_{MII}^1 \) is the probability of being a Mode II with no direct Internet access, but through another Mode II or fixed BS. \( P_{MII}^2 \) is the probability of Mode II with direct access.

A Mode II CR user with no direct database access potentially acts as relay node to forward queries from another Mode II CR user to a nearby fixed BS, or a nearby Mode I device. Since the chain of Mode II relay nodes can induce delay to the spectrum information queries originating from a Mode II or Mode I CR user, we put a constraint on the number of relay nodes for both limiting the delay and also for the sake of analysis simplicities. The constraint is that a Mode II device with no direct database access can only relay data to a mode I CR user and only when it’s connected to a BS. Given this constraint, \( P_\Delta \) in equation (2.22) will be \( x_2x_3 + (1 - x_2)x_3\alpha \). Also looking at direct-access and BS-access mode cases separately, \( P_\Delta \) must be replaced with \( x_2x_3 \) and \( (1 - x_2)x_3\alpha \):

\[ A^1 = 1 - e^{-(1-x_2)x_3\alpha R \lambda_s} \]  

(2.33)

\[ A^2 = 1 - e^{-x_2x_3 R \lambda_s} \]  

(2.34)
where $A^1$ and $A^2$ are the probabilities of having at least one Mode II device in the range of a BS-assisted database access and an arbitrary node with direct database access respectively.

The cost of channel information access depends on the source of the channel information. We assume for each query the requesting CR user will be charged $s_1$ by the Fixed BSs and $s_2$ by the directly connected Mode II CR users.

Hence, the cost of channel access for users in Mode I is:

$$C_{MI} = \alpha s_1 + (1 - \alpha)[A^1 s_1 + (1 - A^1)A^2 s_2] \quad (2.35)$$

The first term in the equation above is for the time when Mode I device is in the range of a BS. The second term, however, considers the probability that there is a Mode II device, with preferably indirect (because of lower cost) or direct access, in the range of the requesting Mode I device. The cost of channel access for Mode II users with BS access is the same as Mode I users’ access cost, and the cost of channel access for users in Mode II with direct access is $s_2$.

The average cost of channel information access for users can be expressed in general terms as follows:

$$C_{user} = P_{MI} C_{MI} + P_{MII}^1 C_{MII}^1 + P_{MII}^2 C_{MII}^2 \quad (2.36)$$

The average revenue received for a channel information query made by an arbitrary user is:

$$C_{Rev} = [P_{MI} + P_{MII}^1] \alpha s_1 \quad (2.37)$$

The density of fixed base stations will directly determine the average cost of channel access for a road of given traffic type. The optimized value for minimized cost can be determined by the following optimization problem:

Given: $R, \lambda, x_1, x_2, x_3, s_1, s_2$

To find: $\alpha$

To Minimize: $C_{user}$

Subject to:

$$P_e \leq P_{e_{lim}} \quad (2.38)$$

$$C_{Rev} N_{av} - C_{annual} \geq 0 \quad (2.39)$$

- The probability of sensing error $P_e$, due to spectrum sensing can be calculated as follows:

$$P_e = P_{sense}[P_{md} + P_{fa}]$$

$$= x_1(1 - x_2 x_3)[1 - \alpha - (1 - \alpha)A][P_{md} + P_{fa}] \quad (2.40)$$
The constraint (2.38) limits this error within a threshold $P_{\text{lim}}$, where $P_{\text{md}}$ and $P_{\text{fa}}$ are the probabilities of missed detection and false alarm, respectively.

- In constraint (2.39), $N_{\text{av}}$ is the average number of vehicles querying a specific fixed BS on the road and is calculated as follows:

$$N_{\text{av}} = \lambda_{\text{av}}^t \times 86400 \times 365 \times [P_{MI} + P_{MII}^1] \quad (2.41)$$

where $\lambda_{\text{av}}^t$ is the average inter-arrival time at any BS along the road. This is with the assumption that every Mode I and Indirect Mode II vehicle will ping the BS once it passes by it. Also $C_{\text{annual}}$ in (2.39) is the annual cost of maintaining a fixed BS on the road. In here we only consider the maintenance cost of each BS, since it accounts for the major part of the BS cost (around 75% [60]).

The derivative of $C_{\text{user}}$ can help solve the above optimization problem. Simple derivation will lead to:

$$C'_{\text{user}} = (1 + (1 - \alpha)(1 - x_2)x_3 R \lambda R_s)[(s_1 - s_2)\alpha$$

$$+ (1 - \alpha)(2A_1s_1 + A_2s_2)A$$

$$+ \alpha s_2A^2 + (1 - \alpha)s_2AA^2] \quad (2.42)$$

In the next section, we numerically show that the derivation of $C_{\text{user}}$ is always positive, so the function has its minimum point at one of the two ends of $\alpha$ range which is confirmed by the simulation result of the cost function.

**Performance Evaluation**

In this section, we perform a numerical analysis of the optimization framework given in the previous section and show how different metrics vary with the increasing BS density. We simulate the given mathematical formulation within MATLAB while fixing inputs of the optimization problem as shown in table 2.2.6. Parameter $s_2$ is the cost of internet access through 3G or similar technologies and its value is calculated based on the current price of $25 for 2GB data traffic and a exchange of 20KB data at each incident of database access. Also $s_1$ is assumed to be half of $s_2$. Probability of mis-detection and false alarms for TV channel sensing are taken to be both 0.05.

Figure 2.13 shows the average cost of access to spectrum information per request for different traffic situations. Interestingly, in low and normal traffic scenarios, the cost of access goes up with increasing BS density. This can be observed from Figure 2.14, where for these two traffic types, the silent mode probability in low BS density is relatively high. Moreover, the increase of BS density only helps increasing the number of requests to BS,
Table 2.5: Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>5%</td>
</tr>
<tr>
<td>$x_2$</td>
<td>20%</td>
</tr>
<tr>
<td>$x_3$</td>
<td>40%</td>
</tr>
<tr>
<td>$P_{md}$</td>
<td>0.05</td>
</tr>
<tr>
<td>$P_{fa}$</td>
<td>0.05</td>
</tr>
<tr>
<td>$R$</td>
<td>500m</td>
</tr>
<tr>
<td>$s_1$</td>
<td>0.0125 cent</td>
</tr>
<tr>
<td>$s_2$</td>
<td>0.025 cent</td>
</tr>
</tbody>
</table>

while the likelihood of finding a direct-access Mode II device is low, although it costs more than a BS to retrieve spectrum information from these devices. In high traffic scenarios, the probability for finding a Mode II device is high, and this helps keeping the silent time low, even if the BS density is low as well. Figure 2.13 also confirms the argument made in the previous section about the local minimum at either end of the cost function domain. Figure 2.15 shows the derivation of $C_{user}(\alpha)$ function in different traffic types which is positive all over the range as argued in the previous section.

![Figure 2.13: Cost of each channel information access vs. BS density](image)

Figure 2.13: Cost of each channel information access vs. BS density

Figure 2.16 shows the variation of average sensing error probability occurred with
Figure 2.14: Silent mode probability change vs. BS density

different BS density. It is observed that this probability in low traffic types is significantly higher than other traffic situations, since more of Sensing-only mode will be experienced.

Figures 2.17 shows the average number of queries made to each BS given the density of BS in the road. Also Figure 2.18 shows the annual revenue received from the operation of each BS, given the BS density on the road.

The minimum feasible value for \( \alpha \) can be determined by setting the right thresholds for the constraints in (2.38) and (2.39) and even additional ones may be incorporated that are imperative from a network designer’s point of view.
Figure 2.15: Access cost function derivation vs. BS density

Figure 2.16: Average sensing error probability vs. BS density
Figure 2.17: Average number of channel information accesses per year vs. BS density

Figure 2.18: Average annual revenue received at each BS vs. BS density
Chapter 3


The effort to identify additional wireless spectrum with markedly reduced congestion has led to the radically different concept of CR, wherein individual radios identify portions of the spectrum, and opportunistically transmit when the licensed or primary users (PUs) are not currently active [61]. When multiple different CRs identify the same set of available channels, the task of allocating them adequate portions of the spectrum is a non-trivial task. Not only must the individual QoS demands be met, but also the new assignment of spectrum to one set of users may in-turn impact adversely the performance of a different set of users that have already been allotted the same portion of the spectrum. The key questions that this chapter aims to address are: (i) under which conditions must CRs be allowed to share spectrum, and when must they be assigned completely exclusive spectrum? (ii) what must be the size of the spectrum chunks that can be allotted per application? (iii) how can CRs meet their QoS needs through reservation of an optimal amount of backup spectrum, in case they are interrupted by the PU’s return? Finally, To highlight the practical aspect of this research, we present a case study of a practical problem that affects the medical community in the Boston area. We demonstrate how our approach can benefit the wireless medical telemetry service (WMTS), through measurements and using stored traces of spectrum usage from extensive spectrum surveys conducted at hospital sites.

The motivation of this chapter stems from a need to have a rigorous mathematical framework that considers the spectrum usage activity of the PUs, the latency and bandwidth requirement of the CRs, and returns an efficient spectrum allocation scheme. At a high level, we devise separate analytical formulations for streaming and non-streaming
categories of applications. Within each of these categories, depending on the packet arrival rate at a given node ($\gamma$), the required bit rate ($R$), the link-layer successful packet transfer time ($\ell$), and the packet length ($L_p$) chosen for the application, further grouping is possible. Before initiating transmission, the CR informs the above four parameters to the controlling BS, which undertakes a centralized resource allocation and informs each CR which PU channel(s) must be used to satisfy its required QoS. To meet the rate requirement $R$, we utilize channel aggregation and find the best set of contiguous PU channels, such that the cumulative bandwidth suffices for that CR. To satisfy the delay requirement $\ell$, a critical concern for both categories of applications, we use a priori statistical knowledge of PU activity in terms of inter-arrival and active (or “on” time). In our approach, each group of CRs (constructed on the basis of similarity in the above metrics) is assigned a precisely calculated number of backup channels. The selection of the number of backup channels is an important factor in our design - too few may lead to long-term service disruptions, while too many make inefficient reservation of the spectrum. Moreover, our model formulation introduces a subtle but important change: the backup channels can be used by the PU on a demand-basis, or simply by other CRs of the group who may have needed to give back their initially assigned spectrum to a returning PU.

### 3.1 Related Work

While spectrum sensing has received a lot of attention over the past several years, the problem of QoS provisioning for the SUs merits more research, as ultimately, applications will drive future adoption of CR technology. For continuous traffic generating SUs, and with exact knowledge of the channel gains for the entire licensed spectrum and the PU activity in them, a Markovian framework is presented in [62] that derives the queueing delay performance of the SU packets. This approach assigns SUs to channels on which they experience the best channel gain, after which a PDF for packet delay for each SU is derived. In [63,64], call drop and call blocking probability in a secondary network is studied via a Markov chain analysis and based on exponential inter-arrival time of PUs. Along similar lines, [65–67] formulate different call admission strategies for ensuring that QoS, expressed in terms of call dropping and blocking rate, is achieved. Through a Markovian analysis, they minimize the dropping rate while attempting to meet a user-defined call blocking rate constraint. For non-continuous traffic, the transmission delay and packet drop performance under unslotted CSMA/CA is analyzed in [68]. However, all these works assume that there is identical statistical behavior of the PUs on all channels, which does not reflect practical observations.

In this chapter we consider a general case with heterogeneous PU activity in the li-
Table 3.1: List of Channel Allocation with QoS provisioning works in the literature.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Allocation Objective</th>
<th>Target QoS metric</th>
<th>PU heterogeneity</th>
<th>SU heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[65–67]</td>
<td>Call Dropping Prob.</td>
<td>Call Blocking Prob.</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[69]</td>
<td>Fairness in Airtime Share</td>
<td>Call Blocking Prob.</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[70, 71]</td>
<td>Spectrum Utilization</td>
<td>Call Blocking Prob.</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Our work</td>
<td>Spectrum Utilization</td>
<td>Blocking delay</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

censed spectrum. Our model is further complicated by considering different types of traffic - streaming and non-streaming, each of which may have further fine-grained requirements of latency, bandwidth, among other QoS features. Our analytical work involves devising two different Markov-chain based frameworks for these two traffic types. In the streaming traffic, the average wait time for a streaming node to access a free licensed channel is derived and matched with the QoS-specified delay that serves as the permissible upper bound. For the non-streaming traffic, the average wait time of a single packet transmission is derived under the assumption of CSMA/CA between multiple contending SUs. In summary, the heterogenous spectrum-usage and channel definitions, and the inclusion of a spectrum allocation algorithm that ensures that the user-specified QoS needs are met, differentiates our work from the existing state-of-the-art. Table 3.1 lists other works in the literature that propose joint QoS provisioning and channel allocation with entries that marks their respective target QoS metrics, the assumptions made on the heterogeneity of PU channels and also the heterogeneity of SU demands.

The rest of this chapter is organized as follows: Section 5.2 explains in detail the analysis framework, and the channel allocation scheme. Section 3.3.1 describes a case study for medical telemetry using our approach. Section 3.3 provides a comprehensive simulation study.

3.2 System Model and Problem Formulation

3.2.1 Network Model

In this section we describe the various network entities, details of the QoS assumptions, and the overview of our approach.

- **Central BS**: The BS accepts new data transmission requests made by the SUs. Each node submits a QoS vector in the form \((\gamma, R, \ell, L)\) to the BS. In this vector, \(\gamma\) indicates the packet arrival rate of the nodes, assuming a Poisson arrival distribution. For the streaming case, \(\gamma\) is set to \(\infty\). Also \(R\) is the required rate, \(\ell\) is the packet delay constraint and \(L\) is the length of each packet the node is transmitting. The BS groups these requests based on their QoS requirements. For e.g., the non-streaming nodes are separated from their
streaming counterparts. If nodes vary in any one of these parameters, i.e., different arrival rates, latencies, or packet lengths, they form their own individual subset where all these requirements are the same for the nodes of that particular subset. We assume that PU arrivals and departures are accurately detected at the BS and communicated to the SUs. The BS has the knowledge of statistical PU arrival rates.

- **Streaming SUs**: Each SU for a given streaming group is assigned its own channel since this channel will be continuously utilized for streaming data. Hence, such a node will not have to contend with any other SU nodes for channel access. Despite this, the operation of the SU could be disrupted when the PU takes over the channel. Therefore, to guarantee the continuous operation of a group of streaming SUs on their channels, a set of backup channels is identified for the group. These backup channels will be used by SUs when their own default channel is occupied by the PU. Our approach identifies the number of backup channels that needs to be assigned per group in such a way that the average packet queueing delay for nodes of that group is below that of the threshold specified in the QoS vector. Note that these backup channels are also contained within the licensed spectrum, and consequently, may also be claimed by PUs. Hence, marking them of future use does not guarantee their availability. We begin the analytical formulation under the limiting assumption of fixed PU arrival rates $\lambda$ and $\mu$ for the default and the backup channels, respectively, in Section 3.2.2. We relax this assumption in Section 3.2.4, for the case of heterogeneous PU arrival and departure rates.

- **Non-streaming SUs**: For non-streaming groups, multiple SUs, provided they have the same QoS requirements, are assigned to a single channel and allowed to contend for the spectrum, provided long-term delay and packet transmission rate threshold are met. For this, the number of nodes assigned to a given channel, and using classical CSMA/CA at the link layer, must be carefully decided, as we show in Section 3.2.3.

### 3.2.2 Delay analysis for streaming type allocation

For a network composed of $M$ streaming nodes, we need to identify the lowest number of backup channels $N$, such that the affected streaming node is ensured continuous use of the spectrum with average delay below or equal to $\ell$. In determining $N$, an underestimation results in an increase in the blocking probability for the streaming nodes, while an overestimation results in inefficient use of the spectrum. If a node’s main channel is occupied, and all $N$ channels are busy, then it must await in a queue for either one of the backup channels or its own original channel to become available. Our approach involves modeling this system as a queueing problem, where the $M$ nodes are customers that randomly arrive at a queue serviced by $N$ backup channels as servers. Considering the delay constraint $\ell$ of the QoS vector, the mean queueing time must be kept below this threshold.
Since the number of available backup channels (here, servers in the queueing problem) is varying owing to the random arrival and departure of PUs on these channels, we model the problem with a two-dimensional continuous-time Markov process with state space $S = \{(m(t), n(t)) : 0 \leq m(t) \leq M, 0 \leq n(t) \leq N\}$, where $m(t)$ is the number of nodes out of their original channel due to PU presence and either seeking a backup channel or operating on one (i.e., nodes that had to vacate their earlier default channel), and $n(t)$ is the number of backup channels not occupied by PUs (they may, however, be used by SU nodes) at an arbitrary time $t$. Fig. 3.1, shows an example, with a network of $M = 2$ SUs assigned with 2 default channels - channel 1 and 2 shown in the top half of the figure. There are 3 backup channels- channels $x$, $y$, and $z$ shown in the bottom half. The notations in the parenthesis denote the current state of the system. For e.g., in slot $[t_0, t_1]$, the channel 1 is used by the PU (thus displacing SU 1), while SU 2 still has access to channel 2. Since one SU is out of its default channel, $m(0) = 1$. Looking at the backup channels, we find that SU 1 is using channel $x$, while channels $y$ and $z$ are occupied by other PUs. Hence, the number of backup channels not used by PUs is $n(0) = 1$. Thus, the state at time $t_0$ is defined as $(1, 1)$.

Upon arrival of a PU on its default channel the affected SU will switch to any available backup channel (see instants $t_1, t_8$ in Fig. 3.1). On the other hand, if the default channel becomes available (even if the SU is operating on a perfectly fine backup channel) it immediately resumes using it. For e.g., in instant $t_5$, when the PU vacates its default channel 2, the SU 2 leaves the backup channel $z$ and returns back to channel 2. If the SU is dis-

---

**Figure 3.1**: An example of channel assignment for 2 streaming nodes with 3 backup channels. Random PU arrivals and departures trigger SU channel movements and the corresponding Markov process state transition.
Figure 3.2: Finite quasi-birth-death process representing \( M \) main channels with PU arrival rate \( \lambda \) and PU "on" rate \( \lambda_{on} \), and \( N \) backup channels with PU arrival rate \( \mu \) and PU "on" rate with \( \mu_{on} \).

lodged from its own default channel and no backup channel is available, it will wait in the queue until one of them is available. For e.g., at time \( t_1 \), SU 2 vacates its default channel 2, and all other channels are either occupied by the PUs (default channels 1 and 2, and backup channels \( y \) and \( z \)) or by other SUs (backup channel \( x \)). Thus, it must now enter into a wait state, behind any already existing SUs in the wait queue.

Assuming a fixed Poisson PU arrival rate \( \lambda \) and departure rate \( \lambda \) for all \( M \) default channels, and \( \mu \) and \( \mu_{on} \) as the corresponding rates for all \( N \) backup channels, our two-dimensional Markov process is shown in Fig. 3.2.

Starting from the state \((0, 0)\), any one (out of \( N \)) backup channels can become available for use if the PU exits, which occurs with the rate \( N \mu_{on} \). Likewise, the arrival of a single PU (with the rate \( \mu \)) will result in the transition to state \((0, 0)\) from state \((0, 1)\). Similarly, the chain can be extended to the terminal state \((0, N)\) in the horizontal plane, and the state \((M, 0)\) in the vertical plane. In general:

- the transition rate to state \((i + 1, j)\), i.e., when a streaming node requests a backup
channel, is \((M - i)\lambda\).

- the transition rate to state \((i - 1, j)\), i.e., when a node currently served by a backup channel, or in the queue waiting for one, reclaims its own default channel due to PU leaving, is \(i\lambda_{on}\).

- the transition rate to state \((i, j - 1)\), i.e., a backup channel becoming free for use, is \(j\mu\).

- the transition rate to state \((i, j + 1)\) or a backup channel becoming occupied by the PU, is \((N - j)\mu_{on}\).

The Markov process of Figure 3.2 has all the properties of a quasi-birth-and-death (QBD) process [72] where it has \(M\) levels and \(N\) phases at each level. It is straightforward to show that the transition matrix \(Q\) of the QBD process in Figure 3.2 has the following form:

\[
Q = \begin{pmatrix}
A_0^{(0)} & A_1^{(0)} & 0 & \cdots \\
A_0^{(1)} & A_1^{(1)} & 0 & \cdots \\
\vdots & \vdots & \ddots & \ddots \\
0 & \cdots & A_2^{(k)} & A_1^{(k)} & A_0^{(k)} & \cdots \\
0 & \cdots & \vdots & \ddots & \ddots \\
\end{pmatrix}
\]

where each of the matrices \(A_0^{(k)}\), \(A_1^{(k)}\) and \(A_2^{(k)}\) are given as:

\[
A_0^{(k)} = (M - k)\lambda I_{(N+1)},
A_1^{(k)} = T + \tilde{A}_1^{(k)},
\tilde{A}_1^{(k)} = -((M - k)\lambda + k\lambda_{on})I_{(N+1)},
A_2^{(k)} = k\lambda_{on}I_{(N+1)}.
\]

\(T\) is the transition matrix specific to each level of the two dimensional Markov process:

\[
T = \begin{pmatrix}
-N\mu_{on} & N\mu_{on} \\
\vdots & \vdots \\
k\mu & -k\mu - (N - k)\mu_{on} & (N - k)\mu_{on} \\
\vdots & \ddots & \ddots \\
N\mu & \vdots & \ddots & \ddots \\
\end{pmatrix}
\]

67
and $I_{(N+1)}$ is the identity matrix of size $(N+1)$-by-$(N+1)$. Matrix $Q$ represents an inhomogeneous (level-dependent) finite QBD process since arrivals to and departures from level $k$ depend on $k$ (in this case it is a function of $k$). Since the properties of this type of Markov process are relatively unexplored because of its generality [73], no closed form expression for the steady state probability vector exists. In other words, to obtain the steady state probability vector $\Pi$, the set of equations given by $\Pi Q = 0$ and normalization condition $\sum_{i,j} \pi_{i,j} = 1$ needs to be solved. However, several numerical algorithms are proposed to accelerate the computation of stationary distribution of the process [74] that can be used for faster convergence. The irreducibility of matrix $Q$ is trivially deduced, and therefore, a steady-state probability vector $\Pi$ exists. The set of equations given by $\Pi Q = 0$ and normalization condition $\sum_{i,j} \pi_{i,j} = 1$ for our QBD process can be expanded as the following:

\[
\begin{align*}
\Pi_0 T + \Pi_0 \tilde{A}_1^{(0)} + \Pi_1 A_2^{(1)} &= 0, \\
\Pi_0 A_0^{(0)} + \Pi_1 T + \Pi_1 \tilde{A}_1^{(1)} + \Pi_2 A_2^{(2)} &= 0, \\
\Pi_{k-1} A_0^{(k-1)} + \Pi_k T + \Pi_k \tilde{A}_k^{(k)} + \Pi_{k+1} A_2^{(k+1)} &= 0, \\
& \quad k = 1 \ldots M - 1, \\
\Pi_{M-1} A_0^{(M-1)} + \Pi_M T + \Pi_M \tilde{A}_1^{(M)} &= 0, \\
\sum_{i=0}^{M} \sum_{j=0}^{N} \pi_{i,j} &= 1, \\
\Pi_1 &= (\pi_{i,0}, \pi_{i,1}, \ldots, \pi_{i,N}).
\end{align*}
\]  

(3.1)

By solving this set of equations, the average length of the queue can be simply calculated as the following:

\[
L = \sum_{i>j} (i - j)\pi_{i,j}.
\]  

(3.2)

(3.2) shows that there will be nodes waiting in the queue (hence contributing to the queue length) when the number of nodes forced out of their original channels is more than the number of available backup channels.

Subsequently, by using Little’s Law, the average waiting time in the queue is obtained:

\[
W = \frac{L}{\lambda_e} = \frac{L}{\sum_{i,j} (M - i)\lambda \pi_{i,j} + \sum_{i\geq j} j\mu_{on} \pi_{i,j}}.
\]  

(3.3)

Above $\lambda_e$ is the effective queue arrival rate, i.e., the rate at which nodes join the queue to get served by a backup channel. The rate expression has two parts as shown in the denominator in (3.3). The first part is the queue arrival rate from nodes requesting a backup
channel right after PU arrival at their main channel \((t_1\) instance in Fig. 3.1). The second part concerns a node that is already operating on a backup channel, but has to leave it and wait its turn in the queue because the PU arrives at this specific backup channel, and no other backup channel is immediately available (for e.g., the \(t_3\) instance in Fig. 3.1, where SU 2 is operating on backup channel \(x\), and this too gets claimed by a returning PU).

The above formulation is specific to the case when fixed PU arrival and departure rates are assumed for each of the set of default and backup channels. For a general case of unequal PU arrival and departure rates for all of the main and backup channels, any mathematical analysis become intractable as for arbitrary values of \(M\) and \(N\) a representation of any Markov process can not be easily conceived. This is due to exponential increase in the number of states as \(M\) and \(N\) increase. In fact, it is proved in [75] that the general problem of job-server queueing with any number of servers more than two, when servers go out of service (as in our case) with heterogenous rates is mathematically intractable. In section 3.2.4, by using the same formulation, we solve the general problem of heterogenous arrival and departure rate by means of an approximation algorithm.

### 3.2.3 Delay Analysis for Non-streaming type allocations

We let mon-streaming nodes share a channel, as they do not transmit continuously. Though this improves spectrum utilization efficiency (compared to issuing each node a dedicated but seldom accessed channel), the need to ensure that packing excessive number of these SUs within a channel does not lower the performance below their QoS threshold. Formally, given a maximum possible transmission rate \(R\) for a given PU channel, with the PU arrival rate of \(\lambda\), the maximum number of nodes \(K\) that can be serviced in that channel needs to be calculated. We assume that all these nodes have the common QoS vector \((\gamma, R, \ell, L)\).

- **IEEE 802.11 DCF preliminaries:** In the 802.11 DCF model, a CSMA/CA mechanism is employed. Each node with a packet to transmit contend with other nodes for channel access during a contention window. The contention period starts when the nodes senses the channel Idle for duration of Distributed Inter-Frame Spacing (DIFS) as defined in IEEE 802.11 standard. Then, a back off timer is initially set with a randomly chosen time in the range \([0, W]\) and starts counting down. At the end of each slot time of the backoff timer, the node senses the channel and if it detects a transmission on the channel, it freezes the timer until the transmission is over. When the timer hits zero, the node attempts a packet transmission which may be successful (with an ACK is received before \(ACK\) Timeout time) or result in a collision with other transmissions on the channel (No ACK). If the transmission is successful, the receiver of the packet tries sending an ACK message to the transmitter node after a duration of Short Inter-Frame Spacing (SIFS), which is the time needed for a node’s radio to switch from receiving mode to transmission mode. In
case of a collision, the node doubles the range of backoff window, i.e. $[0, W_1]$ where $W_1 = 2W$ and repeats the contention process. This process will go on in case of further collisions each time doubling the range up to $m^{th}$ stage so that at stage $i$, $W_i = 2^iW$. After the $m^{th}$ stage, the contention window range will not increase further. The contention process continue until the packet is successfully transmitted.

- **Our revisions to the 802.11 DCF model:** We revise the Markov processes described in [76, 77] to capture the behavior of such a network by incorporating the presence of the PU in the network. Therefore, collisions in the network are caused either with PU or other SUs attempting transmission at the same time. The fundamental assumption in [76] is that the probability of a packet encountering a collision $p$, after a transmission attempt is fixed over time and independent of the transmitting node. Even with the presence of PU in our revised model, this assumption holds true, for the SUs, since PU transmission equally affects the activity of all nodes, e.g. freezing their back-off counter. This does not violate the assumption that the probability of transmission attempt $\tau$ and probability of collision $p$ for each node is independent of that of others. However note that the dependence of any SU transmission attempt on the PU transmission, of course, remains. Based on derivations given in [77], $\tau$, the probability that an arbitrary node starts transmission at a randomly chosen slot is:

$$\tau = \frac{2(1 - 2p)q}{q(W + 1)(1 - 2p) + Wp(1 - (2p)^m) + 2(1 - q)(1 - p)(1 - 2p)}.$$  \hfill (3.4)

where $q$ is the probability of having packets to transmit. In the following, we provide our extensions to the model in [77] with revised formulation of the variables used in (3.4) taking the PU activity into account in network of $K$ SUs. The probability of collision for any node either with $(K - 1)$ other nodes or the PU is given as follows:

$$p = 1 - (1 - \tau)^{K-1}(1 - P_{on}).$$  \hfill (3.5)

$P_{on}$ is the probability of the arrival of PU with poisson arrival rate $\lambda$ during a time necessary for a node transmission to be successful ($T_s$) and is simply given as:

$$P_{on} = 1 - e^{-\lambda T_s}.$$  \hfill (3.6)

Probability $q$, which is the probability that there is at least one packet to be transmitted at each slot, can be approximated with the following relation as a function of $\gamma$, specified by the QoS vector [77]:

$$q = 1 - e^{-\gamma E_s}.$$  \hfill (3.7)

$E_s$ is the average slot time spent by the channel in any state including an idle-channel slot, successful transmission ($T_s$), SU collision ($T_i$), or PU-SU collision ($T_I$), which results
in interference to PU. The idle state includes both fixed slot time with length $\sigma$ during which the nodes decrement their backoff counter and also the times of PU appearance on the channel which freezes the back off timer for all SUs. Now, $E_s$ is given as:

$$E_s = (1 - e^{-\lambda\sigma})(\frac{1}{\lambda_{on}} + \sigma)$$

$$+ e^{-\lambda\sigma}\{(1 - P_t)\sigma + P_tP_sT_s + P_tP_tT_I + P_t(1 - P_s - P_I)T_l\}. \quad (3.8)$$

In the above, the first term indicates a fixed backoff slot during which a PU appears on the channels. This state occurs with probability $1 - e^{-\lambda\sigma}$ and lasts for $(\frac{1}{\lambda_{on}} + \sigma)$ in average. In case of no PU activity. In case of PU appearence, either an idle (no activity on the channel) or SU transmission attempt can occurs where the latter takes only $\sigma$ and the latter will take $T_s$, $T_I$ and $T_l$ in case of successfull transmission, PU-SU collision and SU collision respectively. Each of these occurs with the following probabilities: $P_t$ is the probability that at least one node attempts a transmission at an arbitrary slot, $P_s$ is the conditional probability that a given packet transmission on the channel is successful, and $P_I$ is the probability that any transmission attempt of SUs collide with the PU. Clearly, the complement of sum of $P_s$ and $P_I$ will indicate the probability of SU collision. These probabilities are given as

$$P_t = 1 - (1 - \tau)^K, \quad (3.9)$$

$$P_s = \frac{K\tau(1 - \tau)^{K-1}(1 - P_{on})}{P_t}, \quad (3.10)$$

$$P_I = P_{on}, \quad (3.11)$$

$$T_s = \frac{L}{R} + SIFS + \frac{L_{ack}}{R} + DIFS, \quad (3.12)$$

$$T_I = \frac{L}{2R} + \frac{1}{\lambda_{on}}, \quad (3.13)$$

$$T_l = L + AckTimeout. \quad (3.14)$$

Here, $L$ is the packet length, $R$ is channel rate and $AckTimeout$ is the permissible timeout duration for the acknowledgement to arrive. For fixed $K$ (number of contending SUs on the channel) and known $W$ (initial backoff length) and $m$ (number of backoff stages), values of $\tau$, $p$ and $q$ can be obtained numerically by solving the nonlinear system of equations comprising (3.4)-(3.14). Finally, the mean packet transmission delay is derived as below [78]:

$$\Delta = \sum_{i=0}^{\infty} \frac{2^{\min(i,m)}W - 1}{2}^i E_s + \sum_{i=1}^{\infty} ip^i(1 - p)T_i + T_s.$$
This can be simplified to:
\[
\Delta = E_s \left[ \frac{W (2p)^m}{2 - 2p} - 1 + \frac{W (p)^m}{2 - p} - \frac{2}{2 - p} - 1 \right] + \frac{p}{1 - p} T_l + T_s. \tag{3.15}
\]

Using the above derivations, we find the solution to the problem of how many nodes with a given QoS vector may be assigned to a single channel. We formulate a simple optimization problem that maximizes \( K \) for a given channel (and repeated over multiple channels) such that \( \Delta \leq \ell \), using the expression from (3.15).

To constrain the amount of interference to PU, we assume the probability of SU-PU collision cannot exceed a threshold \( P_{th} \), or we need \( P_I \leq P_{th} \). Using (3.11) and (3.6), this inequality can be simplified as:
\[
T_s \leq \frac{1}{\lambda} \log \frac{1}{1 - P_{th}} \tag{3.16}
\]

Using (3.12) and assuming fixed \( L_{ack} \) for all nodes from any QoS class, we obtain a constraint on the packet length \( L \) of SUs operating on the channel:
\[
L \leq R \left( \frac{1}{\lambda} \log \frac{1}{1 - P_{th}} - DIFS - \frac{L_{ack}}{R} - SIFS \right) \tag{3.17}
\]

The above inequality will set a constraint on the type of QoS classes that be allocated to certain PU channels. We will use as a metric in our channel allocation algorithm in the next section.

### 3.2.4 Frequency Allocation Algorithm

In the previous sections, we obtained the formulations for deriving the analytical QoS, given a set of licensed channels that the SUs use. In this section, we describe a greedy heuristic algorithm that is used to allocate channels, assuming heterogeneous PU behavior in them, to the SUs such that their QoS thresholds are met. The general class of such resource allocation problems are NP-hard [79], and hence, we seek a low-complexity heuristic approach.

At first, we present Algorithm 1 that allocates channels to streaming nodes using the analysis of section 3.2.2. Then, we propose Algorithm 2 for non-streaming nodes based on the analysis given in section 3.2.3. Finally, Algorithm 3 uses Algorithm 1 and 2 to allocate available channels to input nodes of various types and QoS requirements. Table 5.1 lists the notations used in the presented algorithms.
Table 3.2: List of notations used in frequency allocation algorithm.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>Number of PU channels required for a request in the range $s_{\text{min}}, \ldots , s_{\text{max}}$</td>
</tr>
<tr>
<td>$H_s$</td>
<td>PU arrival rate histogram of all channels of length $s$</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>Set of current channel allocation requests by nodes</td>
</tr>
<tr>
<td>$\Psi_{\text{str}}$</td>
<td>Set of all streaming requests for channels of length $c$</td>
</tr>
<tr>
<td>$\Psi_{\text{nst}}$</td>
<td>Set of all non-streaming requests for channels of length $c$</td>
</tr>
<tr>
<td>$\Psi^c$</td>
<td>Set of all requests of for channels of length $c$</td>
</tr>
<tr>
<td>$\ell(\Psi)$</td>
<td>Set of delay components of the QoS vector for all requests in set $\Psi$</td>
</tr>
</tbody>
</table>

**Algorithm 1 for streaming nodes**

To allocate channels to $m$ streaming nodes in the same QoS class, we utilize the analytical model given in section 3.2.2 to find $m$ main channels of length $s$ and a minimal number of backup channels so that their delay requirement $\ell$ is met. However, that model demands channels with identical PU arrival rate $\lambda$ and PU departure rate $\lambda_{\text{on}}$ for the main and also for any potential set of backup channels. Therefore to deal with spectrum of channels with heterogeneous arrival and departure rates $\lambda$ and $\lambda_{\text{on}}$, we construct a 2-dimensional histogram out of all $\lambda$ and $\lambda_{\text{on}}$ values of the existing channels. The PUs that are placed within a histogram bin of width $w_{b1}$ and $w_{b2}$ (for the given range of $\lambda$ and $\lambda_{\text{on}}$ values) are treated to have alike PU arrival and departure rates. In this regard, the arrival and departure rates of the channels in the same histogram bins are approximated to the upper bound arrival rate and lower bound departure rate of than bin (worst case for all the channels of the same bin).

Formally, let the minimum and maximum PU arrival rates in the set of all existing channels be $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ respectively. Also let $\lambda_{\text{min}}^{\text{on}}$ and $\lambda_{\text{max}}^{\text{on}}$ be the respective minimum and maximum PU departure rates. Fixing the bin widths at $w_{b1}$ and $w_{b2}$, we get $B_1 = \lceil \frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{w_{b1}} \rceil$ and $B_1 = \lceil \frac{\lambda_{\text{on}}^{\text{max}} - \lambda_{\text{on}}^{\text{min}}}{w_{b2}} \rceil$ each being the number of x-axis and y-axis bins of the histogram respectively.

Our proposed algorithm for streaming nodes works on the 2D-Histogram $H_c$ as its input. It determines the systemic order in which groups of channels are chosen from $H_c$ as main channels and backup channels and then assigned to groups of streaming nodes. Given $m$ streaming nodes, our algorithm performs with a greedy allocation strategy, starting with the bins indicating least PU activity (best channels for SU operation) that have the smallest arrival rate $\lambda$ and the largest departure rate $\lambda_{\text{on}}$. Thus, we start from the upper left corner of the histogram and sweep all the bins in the zig-zag order, shown in Fig. 3.3(a).

As some bins are empty and do not contain any channels with matching $\lambda$ and $\lambda_{\text{on}}$ ranges, the algorithm finds the first non-empty bin in the order of bins determined in Fig. 3.3(a). At the beginning of each iteration, the algorithm starts with the first non-empty bin, and without loss of generality, we refer to such a bin as $b_1$ and the number of
channels in it as $|b_1|$. At start, Algorithm 1 (line 1) considers the allocation of $M$ channels as main channels, namely the minimum of $m$, which is also the number of given nodes and number of channels in bin $b_1$, or $M = \min(n, |b_1|)$. It then searches for the appropriate set of backup channels on the histogram that satisfies the delay requirement $\ell$ for $M$ default channels, based on the analytical formulation of section 3.2.2.

To find the best choice of a set of backup channels for the selected $M$ main channels, we iterate over the remaining bins of the histogram starting from $b_1$ onwards (lines 5-17). At the $i$th iteration, we set $N_i = |b_i|$. We also refer to Eq. (3.3) as $D = W(M, N_i, \lambda, \lambda_{on}, \mu, \mu_{on})$, which returns the mean delay for $M$ default channels with arrival and departure rate $(\lambda, \lambda_{on})$, and $N_i$ backup channels with representative arrival and departure rates $(\mu, \mu_{on})$ of bin $b_i$. If $D > \ell$, we continue the iteration over next bin $b_{i+1}$. If $D \leq \ell$, then, bin $b_i$ contains a potential set of backup channels for the $M$ main channels in $b_1$. In this case, the smallest subset of channels in $b_i$ that satisfies the delay requirements for $M$ nodes needs to be found. For this purpose, a binary search is performed to find an $N_{opt} \leq N_i$. When a set of feasible and optimized main and backup channels with mean delay $D \leq \ell$ is found, the histogram is updated to exclude these channels as they are allocated. The same process is repeated for any remaining set of $m - M$ given nodes. If at the end of iterations, no bin with potential set of backup channels is found, we decrement $M$ and repeat the iterations until a set of backup channel is found for $M$ main channels in $b_1$. As the complexity of algorithm 1 is dependent of distribution and range of the hist-
Algorithm 1 Allocating spectrum to streaming nodes with required delay $\ell$ and $s$ PU channels

1: function ALLOCATE_STREAMING($m, s, \ell, \mathcal{H}_s$)
2:     while $m > 0$ do
3:         set $M = \min(|b_1|, m), \lambda = \lambda_{b_1}, \lambda_{on} = \lambda_{b_1}^{on}.$
4:     while $M > 0$ do
5:         for each $b_i$ in $\mathcal{H}_s$ in ascending order do
6:             $N_i = |b_i| (|b_i| - M$ if $i = 1)$.
7:             $\mu = \mu_{b_i}^{b_i}, \mu_{on} = \mu_{b_i}^{on}.$
8:             if $W(M, N_i, \lambda, \lambda_{on}, \mu, \mu_{on}) < \ell$ then
9:                 Search for smallest subset
10:                of $b_i$ with size $N_{opt}$ that
11:                $W(M, N_{opt}, \lambda, \lambda_{on}, \mu, \mu_{on}) < \ell.$
12:                Allocate $M$ ch. in $b_1$ and
13:                $N_{opt}$ ch. in $b_i$ as backup
14:                to $M$ out of $m$ nodes.
15:                $m = m - M.$
16:                update $\mathcal{H}_s$.
17:                Break
18:            else
19:                Increment $i$ and Continue search.
20:        end if
21:    end for
22:    if no set of ch. were found so that $D < \ell$ then
23:        decrement $M$.
24: end if
25: end while
26: end function
ogram, the QoS requirement of SUs and specific results of the analytical model, it is challenging to compute a general average complexity for it. However, the worst case complexity of algorithm 1 can be obtained as follows: In the worst case, all the histogram bins are searched for a feasible solution and this feasible solution is always found in the last bin. Let the total number of channels be \( P \) and the total number of bins be \( H = B_1 B_2 \), as \( B_1 \) and \( B_2 \) defined earlier. Also the complexity of solving set of \( n \) linear equations using gaussian elimination is \( O(n^3) \) [80], therefore to solve (3.1) for \( i \) main and \( j \) backup channels, we have a complexity of \((i + 1)^3(j + 1)^3\). Considering a uniform histogram (again the worst case in search for channels), the number of channels in each bin would be:

\[
P_1 = P_2 = \ldots = P_H = \frac{P}{H}.
\]  

(3.18)

Then, the complexity in the worst case can be easily shown to be \( O(P^7) \):

\[
\sum_{i=1}^{P_1} \sum_{j=P_i,\ldots,P_H} (i + 1)^3(j + 1)^3 = \frac{1}{4}(H - 1)(\frac{P}{H})^4(\frac{P}{H})^3 + 6(\frac{P}{H})^2 + 13(\frac{P}{H})^2 + 12.
\]

**Algorithm 2 for non-streaming nodes**

Given \( n \) non-streaming nodes, similar to algorithm 1, a 2D-histogram of nodes is used and channels are picked by traversing the bins of the histogram as shown in Fig. 3.3(a). Based on the results of section 3.2.3, one channel can be allocated to several nodes. Therefore, at each bin of the histogram \( b_i \), each channel \( c \) is examined with (3.17) to check whether it can operate on than channel given the packet length \( L \) in its QoS vector and then the number of nodes it can accommodate so that their delay \( \ell \) is satisfied is determined. Here, we refer to such quantity as \( n = N(\lambda, \lambda_{on}, \ell, L) \), which is obtained by doing a binary search over range \([0, n]\) and using Eq. (3.4)-(3.15) to find the maximum number of nodes \( n_{opt} \), a channel \( c \) can be allocated to. We iterate over bins and the channels within bins as long as any nodes are left and for each channel, we allocate as many nodes out of \( n \) as possible (see lines 4-14 of Algorithm 2). Then, the allocated channel is removed from the histogram and the remaining number of SUs is also updated. This procedure continues until all \( n \) SUs are accommodated.

The worst case complexity of the algorithm, since it iterates over all channels, is \( O(P E \log n) \), \( P \) being the number of channels. Also \( \log n \) is due to the binary search that is performed to find the maximum number of nodes that can allocated to each channel. This is multiplied by \( E \), which is the complexity of solving the non-linear equation obtained from Eqs. (3.4)-(3.15) for number of nodes during the binary search.
Algorithm 2 Allocating spectrum to non-streaming nodes with required delay $\ell$ and $s$ PU channels

1: function ALLOCATE_NONSTREAMING($n, s, \ell, L, \mathcal{H}_s$)
2: $i = 1.$
3: while $n > 0$ do
4:   $\lambda = \lambda^{c_1}, \lambda_{on} = \lambda^{c_1}_{on}.$
5:   if $\mathcal{N}(\lambda, \lambda_{on}, \ell, L) > 0$ then
6:     $l = \min(\mathcal{N}(\lambda, \lambda_{on}, \ell, L), n)$
7:     assign $l$ nodes of $n$ to $c_i$.
8:     $n = n - l$.
9:   update $\mathcal{H}_s$.
10: end if
11: Increment $i$.
12: end while
13: end function

Algorithm 3 for general nodes

Algorithm 3 uses the previous algorithms to allocate nodes from various QoS classes, namely various rate and delay requirements. For nodes that need higher bandwidth than a single PU channel, multiple contiguous PU channels can be aggregated to ensure sufficient available bandwidth is sufficient for the required rate. Let $\Psi_s$ denote a set of nodes that need $s$ aggregate channels to satisfy its data rate. Also $\Psi_s^{nstr}$ is a subset of $\Psi_s$ with only non-streaming type nodes. Equivalently, $\Psi_s^{str}$ is for streaming type nodes. We also refer to $\ell(\Psi_s)$ as the set of delay values in the QoS vector of nodes in set $\Psi_s$.

At first, requests are sorted in descending order based on the required number of PU channels $s$. This sorting will be independent of the streaming or non-streaming nature of the requests and their delay requirements. The channel allocation procedure begins with the maximum $s = s_{max}$ and is repeated for descending values, down to minimum $s = s_{min}$. Starting the allocation from $s_{max}$ is aimed towards minimizing the number of unused fragments at the end of allocation procedure for efficient use of the spectrum [81, 82].

At each iteration over values of $s$, the set of delay values for each of the $\Psi_s^{str}$ and $\Psi_s^{nstr}$ are formed and sorted in ascending order. In the next steps, non-streaming and streaming requests of $\Psi^s$ are processed respectively by algorithms 1 and 2 with ascending order of their delay requirement. In other words, the nodes with lower delay requirement are allocated first due to their stricter QoS. Also at each iteration, we choose to allocate non-streaming nodes prior to streaming ones due the additional constraint for the packet length $L$ of non-streaming QoS given in (3.17).
To construct $H_s$, the entire available spectrum must first be divided in channels of length $s$, and the availability of each channel, with respect to its allocation status, must be evaluated. An intuitive example of this concept is shown at the lower part of the Figure 3.3(b) where $s = 3$. For each aggregate channel of $s$ PU channels, the overall $\lambda$ would be the maximum of all $\lambda$ values of individual PU channels. Moreover, the overall $\lambda_{on}$ for the channel will be the minimum of all $\lambda_{on}$ values, as it indicates the rate at which the whole aggregate channel is vacated by the PU. The overall $\lambda$ and $\lambda_{on}$ for all aggregate channels will be used in forming the histogram $H_s$.

### Algorithm 3 Frequency allocation algorithm

1: for $s = s_{max}$ to $s_{min}$ do
2: Let $\Psi^{s}_{str}$ and $\Psi^{s}_{nstr}$ be sets of streaming and non-streaming nodes with $s$ required bins.
3: $D_s = \ell(\Psi^{s}_{str}), D_s = \{\ell_1, ... , \ell_p\}, \ell_1 < ... < \ell_p$.
4: $D_{n} = \ell(\Psi^{s}_{nstr}), D_{n} = \{\ell'_1, ... , \ell'_q\}, \ell'_1 < ... < \ell'_q$.
5: Form $H_s$ of the available channels of length $s$.
6: for $j = 1$ to $q$ do
7: allocate\_nonstreaming($\Psi^{s}_{nstr}(\ell'_j), s, \ell'_j, L, H_s$).
8: end for
9: for $i = 1$ to $p$ do
10: allocate\_streaming($\Psi^{s}_{str}(\ell_i), s, \ell_i, H_s$).
11: end for
12: end for

### 3.3 Performance Evaluation

#### 3.3.1 Wireless Medical Telemetry: A Case Study using Real-World QoS Constraints

In this section, we study a real world scenario where the channel allocation framework of section 5.2 is used to efficiently solve the problem of dynamic spectrum allocation in the WMTS bands. Although the FCC has allocated the WMTS bands for medical use, there are several issues that impair free access. First, there are no effective regulations protecting medical telemetry in channel 37 from the harmful interference caused by the power leakage from DTV transmissions in the adjacent channels 36 and 38. In fact, there are many documented cases of interruptions in hospital communication due to this DTV interference [36]. This adjacent channel interference effectively narrows the use of this
channel (that represents almost 40% of all WMTS bandwidth). Given the critical nature of hospital communication, this breach must be immediately detected and corrective actions taken [83]. A second cause for concern is the non-uniform access rights in the $L$ bands. Portions of these bands are shared by utility metering telemetry and government radar installations, which have priority or primary access right. Thus, the medical telemetry devices must be aware if these primary users (PUs) are present, and choose different portions of the spectrum, if indeed this is so.

In a medical environment composed of heterogeneous devices with different bandwidth, QoS, and access priority requirements, the problem of frequency allocation in these bands where interference from different sources is common is a challenging task. In the following, we provide the mapping of our analytical framework to this practical problem, and provide comprehensive simulation results in the subsequent section for this specific scenario.

### 3.3.2 Mapping of our QoS Framework for WMTS Bands

The algorithm of section 3.2.4 can allocate small portions of the spectrum dynamically within the WMTS band to devices based on the type and duration of transmission, thereby increasing the potential for frequency re-use and the resulting channel capacity. The algorithm is particularly useful because in WMTS, the bandwidth for each device is relatively small (in the order of several KHz) and therefore the number of devices using the WMTS band could be relatively high (thousands). This necessitates a very efficient algorithm that can quickly and efficiently allocate channels to all these devices. Also, medical telemetry involves transmitting scalar data at set duty cycles, one-shot alarms, streaming information, among others, each with different bandwidth, latency requirements that must be jointly considered [45] which also fits the general streaming and non-streaming categories discussed in earlier sections. In deploying new nodes, the existing legacy medical telemetry transmissions that are not equipped with dynamic spectrum access, as well as PUs in the designated portions of the WMTS spectrum (i.e., the utility transmissions) must be protected, thereby necessitating a dynamic spectrum access-based solution.

Our analytical framework requires a statistical knowledge of the PU occupancy within the WMTS bands. In the next section we explain the methods we used to characterize the WMTS band through real experiments and extract the PU arrival and departure statistics in that band. Some preliminary measurements are described in our earlier work in [2].
To obtain a probabilistic model of channel occupancy on the WMTS channel 37 and the $L$ bands, we performed a measurement study at several hospitals in Boston’s Longwood area. We measured the spectrum usage on this channel using the USRP2 platform. The received power on every band was measured with a fine grained resolution, taking 1024-point FFT, i.e., obtaining a 6100 Hz resolution for each FFT bin (The resolution is appropriately chosen so that it fits the 6.25 KHz which is the commonly used medical telemetry bandwidth). Using the noise floor determination technique in [2], we extracted the active medical telemetry signals for each bin. Since these signals are temporally intermittent, we performed a statistical analysis at each bin on the inter-arrival and ON times of these signals. We then fit an exponential distribution function on these time samples. Therefore, the PU activity in each bin is captured with two $\lambda$ and $\lambda_{on}$ values, representing the arrival and departure rate of their respective exponential distributions. Figs. 3.4(a) and 3.4(b) show the PU inter-arrival and ON time statistics of three sample bins at all three bands within WMTS, namely DTV channel 37, lower-L band and upper-L band respectively. These statistics in Fig. 3.4(a) are fit with exponential distribution of mean 10.11, 18.75 and 10.82 with 95% confidence interval of [9.95, 10.31], [18.45, 19.35] and [10.57, 11.11] respec-
tively. Also in Fig. 3.4(b) exponential fit on measured PU ON times is undertaken with mean \(2.29, 2.39\) and \(2.08\) each with 95% confidence interval of \([2.19, 2.40], [2.39, 2.81]\) and \([1.93, 2.24]\) respectively.

The exponentially-distributed PU activity assumption made in this section will be used for efficient channel allocation for the SUs in the network. In channel 37, these measurements represent all legacy medical telemetry activity, where devices are not equipped with dynamic spectrum access methods. In the \(L\) band, the observed channel activity jointly captures both the existing legacy medical telemetry and utility metering applications.

### 3.3.4 Simulation results

In this section, we undertake a thorough simulation through ns-2 (packet level simulation for CSMA/CA based non-streaming nodes) as well as in MATLAB (for streaming nodes with continuous channel usage) to demonstrate the performance benefit in the WMTS band in terms of spectrum efficiency, as well as verify the theoretical findings on spectrum allocation from Section 5.2 for both streaming and non-streaming nodes. We also show the near-optimal spectral utilization efficiency of our greedy approximation approach in section 3.2.4. In these studies, we vary a metric called as the load factor, i.e., the number of nodes that have the same streaming requirements of latency and bandwidth. The PU activity statistics are acquired from real measurements described in section 3.3.1 from hospitals in the Boston area. Based on the activity model of the channels, we try to accommodate additional SU nodes in the empty portions of the band. We use the actual measured activity pattern on the WMTS channels, as a reference for the activity of PUs in our simulations. Also for the SU nodes, we consider 7 types of telemetry applications with specifications given in Table 3.3 as from the previous paper [2]:

<table>
<thead>
<tr>
<th>Application Type</th>
<th>Size Kb/Packet</th>
<th>Avg Rate Kb/s</th>
<th>Events per hr</th>
<th>Latency ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telemetry</td>
<td>2.6</td>
<td>12.8</td>
<td>Stream</td>
<td>500</td>
</tr>
<tr>
<td>Telemetry Diagnostic</td>
<td>5.1</td>
<td>25.6</td>
<td>Stream</td>
<td>500</td>
</tr>
<tr>
<td>Telemetry Alarm</td>
<td>1.0</td>
<td>0.1</td>
<td>10/h</td>
<td>500</td>
</tr>
<tr>
<td>Clinician Notifier</td>
<td>2.6</td>
<td>0.1</td>
<td>20/h</td>
<td>500</td>
</tr>
<tr>
<td>BCMA</td>
<td>0.4</td>
<td>0.1</td>
<td>30/h</td>
<td>500</td>
</tr>
<tr>
<td>Infusion Pump Status</td>
<td>1.0</td>
<td>1</td>
<td>Stream</td>
<td>500</td>
</tr>
<tr>
<td>Infusion Pump Alarm</td>
<td>1.0</td>
<td>0.1</td>
<td>1/h</td>
<td>500</td>
</tr>
</tbody>
</table>

Similar to [2], a realistic wireless planning of a typical hospital with total area of 18580 m² is considered where the number of operating application nodes in each row in
Table 3.3 is estimated by the values in vector \((60, 21, 22, 20, 19, 81, 18)\). We use these values as a reference for our simulation study and choose random locations for each application node in a square area of \(140 \times 140\) m. Also packet arrival events are created with poisson distribution for each non-streaming application with the given rates of the above table, and the channel allocation algorithm is run for them in MATLAB. We then perform a packet level simulation in ns-2 to verify the validity of our allocation, based on the comparison of the delay from analytical and simulation findings. For the streaming case, since there is no channel contention (each node being allotted a dedicated channel), we verify the performance and analytical derivations through MATLAB.

### 3.3.5 Streaming nodes

Three different sets of statistics \((\lambda, \lambda_{on}, \mu, \mu_{on})\) for the streaming and backup channels obtained from measurements in Section 3.3.1 and used in the following discussion are shown in Table 3.4.

<table>
<thead>
<tr>
<th>Group</th>
<th>PU activity</th>
<th>(\lambda)</th>
<th>(\lambda_{on})</th>
<th>(\mu)</th>
<th>(\mu_{on})</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>High</td>
<td>0.05</td>
<td>0.1</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>II</td>
<td>Medium</td>
<td>0.024</td>
<td>0.1</td>
<td>0.046</td>
<td>0.1</td>
</tr>
<tr>
<td>III</td>
<td>Low</td>
<td>0.005</td>
<td>1.0</td>
<td>0.011</td>
<td>1.0</td>
</tr>
</tbody>
</table>

These three groups are examples of high, medium and low usage channels by the PUs respectively. Intuitively, lower values of \(\lambda\) and \(\mu\) indicate sparse arrivals of the PU. For \(\lambda_{on}\) and \(\mu_{on}\), lower values specify longer active durations for a given arrival event. Fig. 3.5(a) compares the theoretical and simulated queueing delay incurred for three sets \((5, 10, 20)\) of streaming nodes, for the medium group. We observe that the simulation results very closely matches the analysis in section 3.2.2 on all ranges of \(N\). Fig. 3.5(b) shows the trend for the required number of backup channels to keep the queueing delay below 500 ms as the number of streaming channels are varied.

For the high group, where the number of backup channel increases almost linearly with the number of streaming channels, the simulation results indicate a slightly smaller number of backup channels than what the theoretical model predicts. This is largely due to the limited simulation time \((3600\) s\) in the case of larger set of channels. However in the case of medium and specifically low usage groups, where the number of required backup channels is lower, the difference between theoretical prediction and what simulation indicates is small. Overall, we find that the theoretical prediction used for channel allocation always keeps the average delay within the required bound.
Figure 3.5: (a) Incurred queueing delay both in theory and simulation for 5, 10 and 20 streaming channels while varying number of backup channels for the medium group, i.e., $\lambda = 0.024, \lambda_{on} = 0.1, \mu = 0.1, \mu_{on} = 0.1$. (b) Number of backup channels $N$ vs. number of main channels $M$ to keep the average queueing delay below 500 s shown for three PU activity groups of table 3.4 both in theory and simulation.
To verify the spectral efficiency of the frequency allocation algorithm in section 3.2.4, we present a theoretical estimate of the channel allocation in a heterogeneous PU usage regime for the WMTS band. Assume that for a frequency range $\mathcal{F}$, the probability of any frequency $f$ being available for secondary use is known and represented by function $P_{off}(f)$ on the domain $\mathcal{F}$. Then at any arbitrarily small frequency range $df$, the effective bandwidth is $P_{off}(f)df$. To achieve a minimum effective bandwidth $b$, we should have:

$$\int_{f_1}^{f_2} P_{off}(f)df \geq b \quad (3.19)$$

To allocate bandwidth $b$ in the most efficient manner, $f_1$ and $f_2$ in the above equation must be found in $\mathcal{F}$ in such a way that $f_2 - f_1$ is minimized. We use the same measurement statistics for WMTS presented before to get a discrete $P_{off}(f)$ over WMTS band. At each bin we have $P_{off} = 1 - \frac{\lambda}{\lambda_{on}}$ as in [40]. We started with the vector $(60, 21, 81)$ as the number of nodes for the streaming applications in Table 3.3. We measured the number of PU bins allocated to all requests using Algorithm 3, and also the theoretical estimate for the number of PU channels used by the same set of requests using 3.19. We solved 3.19 numerically by choosing $df$ as low as 300 Hz. We repeated the simulations by then scaling the number of nodes for each application given in vector $(60, 21, 81)$, by a load factor (i.e., an integral multiplier). Fig. 3.6 shows the amount of allocated spectrum, in term of number of PU channels, of our algorithm compared to the theoretical estimate. Apart from the theoretical estimate and for the sake of comparison, we also compare the spectral efficiency of our allocation algorithm in section 3.2.4 with its simplified version which uses only a 1D-histogram of PU arrival rate instead 2-D histogram of PU arrival and departure rates and also another simple algorithm which does not use any histogramming at all. The former is similar to the algorithm 1, but only uses a one-dimensional histogram of the PU arrival rate. The latter method merely uses the analysis of section 3.2.2 to find a proper set of backup channels for each individual streaming node without any grouping or histograming. Fig. 3.6 shows how closely the spectral efficiency of matches of the theoretical estimate. The simplified histogramming algorithm just mentioned performs slightly below algorithm 3 in terms of efficiency. Also the simple algorithm with individual allocation of streaming nodes is least efficient of all.

### 3.3.6 Non-streaming nodes

In order to validate the model given in section 3.2.3, we performed simulations in the ns-2 network simulator with the environmental parameters set up to closely match the assumptions used in our model. We set the parameters of Table 3.5 as inputs to both to the
Figure 3.6: Comparison of amount of allotted spectrum versus streaming request load resulted by algorithm 3, simplified of algorithm 3 based on 1-D histogramming method, simple method with no histogramming, and theoretical estimate.

Table 3.5: Simulation Parameters

<table>
<thead>
<tr>
<th>Simulation Parameter</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC header</td>
<td>28 bytes</td>
</tr>
<tr>
<td>PHY header</td>
<td>16 bytes</td>
</tr>
<tr>
<td>ACK</td>
<td>14 bytes</td>
</tr>
<tr>
<td>Payload size</td>
<td>325 bytes</td>
</tr>
<tr>
<td>Slot time</td>
<td>20 µs</td>
</tr>
<tr>
<td>DIFS</td>
<td>50 µs</td>
</tr>
<tr>
<td>SIFS</td>
<td>10 µs</td>
</tr>
<tr>
<td>ACK timeout</td>
<td>500 µs</td>
</tr>
<tr>
<td>CW_{min}</td>
<td>32</td>
</tr>
<tr>
<td>CW_{max}</td>
<td>1024</td>
</tr>
<tr>
<td>Retry limit</td>
<td>7</td>
</tr>
</tbody>
</table>

ns-2 simulator for the packet-level simulation, and also for our MATLAB implementation that was used for the mathematical analysis of the model.

In the ns-2 simulation, the nodes contend over a single channel that is occasionally occupied by a PU with a known arrival rate. We observe that the average encountered delay of the nodes in our simulations closely follows that of the mathematical analysis. Fig. 3.7 shows the average delay versus PU arrival rate $\lambda$ for the channel, while we vary the number of contending nodes from 10 to 40, each having packet arrival rate $\frac{1}{120}$. In another trial, we varied the Poisson packet arrival rate $\gamma$ for 30 contending nodes, and plotted the average delay versus the PU arrival rate as shown in Fig. 3.8. Both plots verify that simulation results closely follow the results of MATLAB analysis.
Figure 3.7: Average delay vs. PU arrival rate $\lambda$ with varying number of nodes with mean packet inter-arrival time of 120 sec contending over a single channel.

Figure 3.8: Average delay vs. PU arrival rate $\lambda$ with varying packet arrival rate $\gamma$ for 30 nodes contending over a single channel.
In CR ad hoc networks, owing to the distributed operation of the nodes, ensuring that the licensed users or primary users (PUs) of the spectrum are not affected is a challenging task. CR nodes are neither aware of the global network topology, including the locations of the PUs, and nor do they have a priori access to PU transmission schedules [11]. Thus, the CR nodes must rely on learning the local channel utilization through individual measurements, and share these results with neighboring nodes. The cooperation between nodes facilitates quick dissemination of the knowledge of the spectrum environment, and also reduces missed detection errors by merging together different data sets from close-by locations. In this chapter we explore how to select the best channels for cooperation, how to merge together spectrum knowledge gained by the nodes, and use these learnt channel utilization models during link-layer transmission.

Prior studies on spectrum utilization have revealed significant spectrum availability in the UHF band [11]. Consequently, the FCC has recently opened up the vacant spectrum in the TV channels 21 (512 MHz) to 51 (698 MHz), with the exception of channel 37, for use by unlicensed devices [84]. However, FCC specifically points out the need to ensure that the reception of both high and low power TV signals are not adversely impacted. Thus, any practical CR deployment must be preceded by a comprehensive study of the characteristics of the expected received power in these channels, the utilization in these frequencies, in both indoor and outdoor environments. In the first part of our work, we undertake the spectrum study over the complete range of the TV channels 21 – 51 and note the variation of the signal strength with distance, location, and frequency. From this study, we identify which channels exhibit reliable and time-invariant behavior at a given location. Moreover, we point out how a CR node can independently identify these reliable channels without
external administrator help, an important consideration in an ad hoc network, where nodes rely on each other’s spectrum measurements.

CR nodes periodically undertake spectrum sensing to ensure that the channel knowledge stays current, and the choice of the channel does not affect the licensed users. While most works assume on-off Markovian models based on the birth-death process, recent experiments have pointed out that the actual PU activity is captured better by long-tailed exponential distributions [85]. Thus, during actual usage, it is beneficial if the CR user samples the channels and builds its own estimation of the PU activity. We propose a Cooperative reinforcement Learning scheme for Cognitive radio networks (CLICK) to achieve this, that differently from previous works, also takes into account (i) the reliability of the channel information, and (ii) the level of trust that can be assigned to the readings of the collaborating node. From our experimental findings we observe that even CR nodes a few meters apart from each other may exhibit a widely varying channel measurement based on location-specific wave reflections. Thus, both distance between nodes, and the peculiar characteristics of the wireless channel at the measurement location are key factors during the collaboration process.

While CLICK allows each CR user to learn about the channel availability over time, choosing a specific channel for communication requires the participation of both the sender-receiver pair. In order to demonstrate the benefit and the overhead of integrating CLICK in a higher layer protocol, we propose an extension of an existing multi-channel medium access control (MAC) protocol for CR ad hoc networks. We demonstrate how the collaboration among nodes results in better choice of channels, and the need for shorter individual sensing times. Both of these factors result in higher link layer throughput and better PU detection.

### 4.1 Related Work and Motivation

Classical cooperation techniques mainly propose combining binary decisions (hard decision) made by the deployed nodes, or their spectrum sensing measurements (soft decision) to a centralized location or fusion center. As an example of the hard decision, the Bayesian hypothesis rule is applied in [86]. The authors point out that the main research problem is the need to appropriately select the decision thresholds that signal the detection event, both at the individual node and during combination at the fusion center. A similar approach using a voting rule is described in [87], where at least half of the nodes must agree on a decision. A more general formulation of collaboration based on \( K \) nodes concurring out of \( N \), also called as the \( K \) by \( N \) rule, is given in [88], and later extended in [89] to also incorporate the MAC layer throughput optimization (note that \( \frac{K}{N} = \frac{1}{2} \) in [87]).
Figure 4.1: The variation in the mean and the standard deviation in the received power and pilot signal are shown for indoor location for TV channels 21 – 51

more realistic case is presented in [90] where a distinct channel behavior between a given PU transmitter and a CR receiver is considered. Our proposed approach, CLICK, does not combine binary decisions. Instead, it merges together the values of the states and actions that define the extent of the learning within the node. These values determine the probabilistic availability of the channel. Thus, our collaborative approach predicts long-term spectrum availability information, rather than its immediate binary availability.

A distributed soft-decision based cooperation scheme for single channel ad hoc networks is proposed in [91] that is limited to forming pairs from a larger set of nodes. However, as we show in our experiments in the next section, such a collaboration between any two nodes on any channel is not guaranteed to return accurate results in a practical environment. As mentioned in [92] and independently verified in our experiments, the assumptions of equal SNR for all nodes with respect to a particular PU transmitter (assumed in [93] [87] [89]) or perfect channel behavior (assumed in [94] [14]), are incorrect. Moreover, the locations of the nodes, say indoors/outdoors, or in an elevated roof, also have bearing on the accuracy of the sensed data. CLICK appropriately weights the contributing information for each \(<\text{channel}, \text{node}>\) pair to account for these variations. Other node-only weighting schemes have been previously proposed in [95] [96], under the limited assumptions of perfect and non-fading channel behavior. Thus the above works fail to consider in a comprehensive manner both the effects of the wireless channel, and the need for location-specific weighting of the nodes in a practical setting. Moreover, as in the case of hard decision, the above schemes are useful for the current sensing instant.
only, and do not reflect on the long term suitability, in terms of CR network throughput and PU protection, when the channel is used.

We believe that for CR networks, network design must be guided by real world experiments, with parameters carefully chosen so that the PU operation is not affected. We note that signal strength, TV transmitter locations, CR user placements, constructive and destructive effects on the signal caused by reflections from structures and the terrain, frequency of the channel, among others, affect the spectrum sensing performance. Moreover, while cooperating CR users may be spread over both indoor and outdoor locations, the PU receiver antennas for the case of the TV transmissions is typically outdoors, at an elevated surface. Thus, the TV signal measured by the indoor CR users may be markedly different, and local decisions by the node based on signal strength may not capture the received power levels experienced by the PU receiver antennas. We study these effects further on the digital TV frequencies corresponding to UHF channels 21 – 51 to motivate the design of CLICK.

**Experimental Setup and Results**

The Northeastern University campus was chosen as the site for our study with experiments conducted, both indoors and outdoors, in two adjacent research buildings. In particular, the sole outdoor measurement was carried out on a high platform on the roof, which provided a line of sight (LOS) reception with a select set of TV transmission towers that were at
approximate distances from 6.8 – 7.3 miles [51]. The non-LOS transmitters ranged from 32 to well over 100 miles. To measure the signal strength, we used the Universal Software Radio Platform 2 (USRP2) equipped with WBX daughterboard.

**Setting the PU detection threshold**

In Figure 4.1, we see both the mean received power in the channels 21 – 51 (upper plot) arranged in the increasing order of the standard deviation $\sigma$ of the power values (lower plot) for the reference indoor location $X$. We verified from [51] that at the time of experimentation, all the sensed channels were occupied. We observed that for all the 30 indoor measurement sites, including $X$, each channel exhibited a channel power of at least $-87$ dBm, which we set as the detection threshold, i.e., greater than this value signaled the presence of the PU. This importance of this selection in a learning scheme was pointed out in [86], but no solution was proposed. From the mean calculations in the outdoor roof setting in Figure 4.2 (upper plot), the decision threshold is higher at $-81$ dBm, owing to the reduced reflections, and better propagation paths with some of them being along the LOS.
4.1.1 Standard deviation or mean as a channel quality metric?

Though the transmission towers are spread out in large geographical areas, the mean received power does not show significant difference over a large extent of the spectrum. We argue that while the mean power is important in setting the decision threshold, and is indeed used by energy-detection schemes at the receiver, sudden and indeterministic fluctuations of the signal about the mean may result in detection errors. From Figure 4.3 we observe that for the outdoor location (bold lines), the signal does not show significant variation. Yet, for the indoor location $X$, and especially for channels 48 and 50, the variance with time is significant compared to the channels 35 and 43. During spectrum sensing, the node relies on a small window of collected samples to measure the signal energy and compare against a threshold. Hence, large variations in the signal may not be smoothened out in the short sample window, leading to detection errors. As an example, from the indoor standard deviation plot in Figure 4.1, though the mean powers are comparable in channels 43 and 48, the standard deviation $\sigma$ in the latter is almost 200% higher. Thus, channel 43 exhibits more consistent behavior with time, and is more suited for exchanging channel information during collaboration. CLICK weights these channels with reliable readings (i.e. with lower $\sigma$) higher using a novel concept of the expert or $E$-value for a channel. We believe that simply weighting specific nodes is not sufficient in a CR environment, but the channel weights must also be considered.

Figure 4.4: The pairwise difference of the mean power for near ($X$-$Y$) and far ($X$-$Z$) separation distances
4.1.2 Range of collaboration

How large should the physical separation between two collaborative nodes be? To answer this question, we fixed the reference location, say $X$, and considered the following two sets of nodes. The first set is represented by $Y$ and composed of nodes $y \in Y$ such that their distance from $X$, is within $100$ m, i.e., $0 \leq d^{X,y} \leq 100$ m shown by the dotted circle. Similarly consider the second set of nodes given by $Z$ such that $100$ m $\leq d^{X,z} \leq 200$ m, shown by the space between the two circular regions. Let the mean received PU signal powers at the locations for the nodes in $Y$ and $Z$ be $\bar{m}_y$ and $\bar{m}_z$, respectively. The plots of the pairwise difference of these means from the mean signal power at location $X$, i.e., $\bar{m}_x$ is given by $|m_x - m_y|$ (white bars) and $|m_x - m_z|$ (shaded bars) in Figure 4.4, respectively. Results reveal that distance and orientations of the nodes in an indoor setting affects spectrum measurements considerably. Though the TV transmitters are located at distances several order of magnitude greater than, say, $d^{X,z}$, the difference in the means is caused by the varying reflective environment and paths experienced by the signals. We have repeated this three way measurements in 30 trials, each time replacing the reference node $X$. We find that for our indoor locations, the mean difference at $100$ m is contained within $3$ dBm. Thus, we limit the collaboration radius of nodes to $100$ m so that nodes assimilating the spectrum information from their neighbors are placed in a similar spectrum environment.
4.1.3 Intervals between collaboration events

The difference between the maximum standard deviation and the mean (a metric represented by $\psi$) calculated from different sample lengths is shown in Figure 4.5 for a limited set of channels from 21 – 30 owing to space considerations. The $\psi$ plot is shown by the bold broken lines, which intersects the horizontal line representing the difference between the noise floor and mean power level. Thus for shorter sample lengths, owing to shorter sensing times, the received signal is yet to converge to the true mean, and hence, $\psi$ is higher than the mean to noise floor difference. For channel 21, this intersection is depicted by the point $\{t, p\}$, where $t$ is the sensing time $t_{\text{sense}}^{21} = 0.09$. This can be interpreted as follows: for channel 21, the $t_{\text{sense}}^{21}$ should be at least 0.09 s. The minimum time between collaboration messages $t_C$ is a critical design component, as waiting too long may also result in longer convergence time. At minimum, $t_C$ must be at least $\sum_{k=1}^{\vert S \vert} t_{\text{sense}}^k$, i.e. the cumulative time for sensing all the channels.

We describe later in Section IV how the standard deviation in a channel is used to calculate a metric called as expert value for that channel, during collaboration.

4.2 CLICK: A Cooperative Reinforcement Learning Scheme

4.2.1 Overview

CLICK is a cooperative reinforcement learning scheme, and adapted from the $Q$-learning technique. It is composed of three stages:

Stage 1 -Intra-node Measurement: Each node undertakes spectrum sensing on the channels 21 – 51 of the TV spectrum, and records the mean and standard deviation of the signal power measured for all of these channels. During this stage, the node detects the presence of the PU by comparing against the receive threshold (Section 4.1), and returns a reward $r = 1$ for those channels where the PU is present.

Stage 2 -Intra-node Learning: Nodes calculate the $Q$ and expert ($E$) values defined in Sections 4.2.2 and 4.2.2 based on the channel measurements. Basically, the PU activity on the channels is captured by the $Q$-learning tables, where $Q$ is a function of the total rewards for a given state, each PU detection event earning a reward of 1. Thus, the nodes learn about the channel availability for long-term use. This learning process is gradual, and as the PU may have intermittent transmissions, or have wireless channel induced sensing errors, a considerable time is needed before a highly accurate channel usage information is gathered locally. The $E$-values qualify how reliable the $Q$-values are, based on the signal deviations of the channel in which the measurements were undertaken.

Stage 3 -Inter-node Cooperation: The nodes periodically share their learnt information
with the neighbors, while distinguishing with whom the collaboration is undertaken, and on which channels. In this stage, the different \( Q \)-table entries of the nodes are merged together after being appropriately weighted. Thus, collaboration accelerates the overall learning at a much higher rate than that possible by local measurements alone.

Stage 1 has been explained in detail earlier in Section 4.1, where the channel measurements were collected. This stage is periodically repeated to maintain the spectrum information current. The choice of sensing time \( t_{\text{sense}} \) for a given channel was also discussed (further optimization strategies to reduce the total sensing duration are presented in Section 4.3). The rest of this section describes the stages 2 and 3 in the operation of CLICK, while mainly answering the following set of questions: How are the experimental channel measurements, especially the location based uncertainties, incorporated in the collaboration scheme? Which subset of nodes from a larger neighborhood pool are suitable for collaboration, and which channels do they share their information on? What is the interval over which the collaboration is undertaken, and what is the overhead?

### 4.2.2 Stage 2 - Intra-node Learning

The learning model is defined by a Markov Decision Process (MDP) which is composed of states represented by \( S \). Each state \( s \in S \) maps to a specific TV channel numbered between \( 21 - 51 \). The set of actions \( A \) maps the transition between states, which is the formal definition of the function of channel switching. The probability with which the node switches the channel (i.e. the state), say from \( s \) to \( s' \), through the action \( a \), is calculated from the distribution function \( \pi(s, a) \in [0, 1] \). The goal of the learning process is to determine the optimal policy that returns the channel with the least \( Q \), meaning that it is free from PU activity for longer durations of time. During \( Q \)-learning, the node observes the current channel \( s \), selects a possible action \( a \), and receives a reward \( r \) from the environment for that specific action. Different from the classical \( Q \)-learning technique that only maintains the \( Q \) (or reward) value table for each state-action pair, we also propose a new metric called as the \( E \) value. Here, for a given state action pair \((s, a)\) and a node \( i \), the value \( E^i(s) \) defines how accurate the corresponding \( Q \)-value \( Q^i(s, a) \) is. The formal expression of the \( Q \) and \( E \) values, and their update equations are as follows:

**Q-Value**

\[
Q^i(s, a) = (1 - \alpha)Q^i(s, a) + \alpha r,
\]

where the reward \( r = 1 \) if PU is detected, and 0 otherwise. \( \alpha \) is a tuning parameter that decides the speed of learning. The probability \( \pi^i(s, a) \) of arriving in the state \( s \) from a
prior state \( s' \) for node \( i \) by choosing action \( a \) is given by the Boltzmann distribution,

\[
\pi^i(s, a) = \frac{e^{Q^i(s, a)/T}}{\sum_{j \in A} e^{Q^i(s, j)/T}},
\]

(4.2)

where \( T \) is the temperature parameter and adjusts the tradeoff between exploration and exploitation actions. The values of \( \alpha \) and \( T \) are taken as 0.2 and 10 \([97]\), though \( T \) itself being a progressively decreasing metric have also been explored in the learning literature.

**Expert or E-Value**

\[
E^i(s) = 1 - \min[1, \frac{\sigma}{M}],
\]

(4.3)

The \( E \)-value \( 0 \leq E^i(s) \leq 1 \) is a function of standard deviation \( \sigma \) of the signal received on a channel, represented by the state \( s \), and the mean \( M \). Each node maintains a history of the last \( H \) signal samples for a sensed channel \( s \). We use \( H = 30 \) to ensure adequate buffer space in the nodes. Thus, the higher is the deviation in the signal strengths, lesser is the reliability of the channel. Hence, the PU occupancy predicted by \( Q(s, a) \) results in lower accuracy.

**4.2.3 Stage 3 - Inter-node Cooperation**

The cooperative learning scheme allows the agents to share their \( Q \)- and \( E \)-values, in order to increase the convergence speed of the learning. The cooperation is achieved in the following three steps: (i) The sharing function (Section 4.2.3) allows each node to decide the \( Q \) values of which channels should be shared with the other nodes. (ii) The combine function (Section 4.2.3) tells the node how to combine its own stored \( Q \) and \( E \)-values, with the respective values received from other collaborating nodes.

**Sharing Function**

A given node, say \( i \), creates a list \( L \) of only those channels for which the \( E \)-value is above a pre-decided threshold \( \phi_E \), meaning that the channel measurements are reliable, and consistent over time. For these channels \( s \in L \), the respective tuples \( Q^i(s, a), E^i(s) \) are sent to the collaborating nodes. Thus, the list \( L \) that is eventually shared can be formally defined as:

\[
L = \{ < Q^i(s, a), E^i(s) > | s \in S, a \in A | E^i(s) > \phi_E \},
\]

(4.4)

where \( \phi_E \) is the decision threshold set at \( 1 - \frac{\sigma}{|M|=80} = 0.9875 \). The rationale for selecting this value is that our experimental results show in Figure 4.1, the average indoor mean
power was found to be at $-87 \text{ dBm}$ and the deviation $\sigma$ for the reliable channels was contained within 1 dB. Thus, we select $\phi_E$ of a comparative value, while maintaining a small safety factor. This sharing function is repeated every $t_C$ time steps, with the minimum value of $t_C$ derived in Section 4.1.3. While lower values of $t_C$ increase network overhead, higher values of $t_C$ lower the learning rate, and may lead to PU interference. We investigate the effect of different intervals of $t_C$ on the PU detection and network overhead in Figure 4.7(a) during the performance evaluation.

**Combine Function**

When node $i$ receives the sharing list $L^j$ from the collaborating node $j$, with $L^j = \{< Q^j(s,a), E^j(s,a) >\}$, it will update its $Q$-values by calculating the weighted average as follows:

$$Q^i(s,a) = (1 - W^{i,j}(s,a)) \cdot Q^i(s,a) + W^{i,j}(s,a) \cdot Q^j(s,a),$$

(4.5)

where $W^{i,j}$ is a weight function which decides how much the information from node $j$ will contribute to adjust the learning process of node $i$. In our formulation, $W^{i,j}$ is a combination of two weighting factors, $\epsilon$ and $\eta$:

- **Expert Weight ($\epsilon^{i,j}(s,a)$):** When a node $i$ receives $Q$-values from another node $j$, it may only integrate these new values in its own $Q$-table if the expert value or $E$-value of $j$ is greater or equal to its own. This prevents dilution of the accuracy of its own measurements on the channels during collaboration. Formally, this is expressed as the ratio of the difference in the expert weights for a given collaborating pair to the cumulative difference of the expert values for all the $K$ collaborating nodes combined, i.e.,

$$\epsilon^{i,j}(s,a) = \begin{cases} \frac{E^j(s) - E^i(s)}{\sum_{n=1}^{K} E^n(s) - E^i(s)} & \text{if } E^j(s) > E^i(s) \\ 0 & \text{otherwise} \end{cases}$$

(4.6)

Similarly, if $E^j(s) - E^i(s) > 0$, then the node $i$ will increase its expert value $E^i(s)$ as more accurate channel measurements have been incorporated. This increase is given by the weighted sum:

$$E^i(s) = E^i(s) \cdot (1 - \eta^{i,j}) + \eta^{i,j} \cdot E^j(s)$$

(4.7)

- **Trust weight ($\eta^{i,j}$):** It measures how much a given node $i$ can trust the $Q^j(s,a)$ values coming from node $j$. The difference between $\eta$ and $\epsilon$ is that the former depends on the distance between the collaborating node pair (i.e. the $< i, j >$ pair) and not on the specific channel behavior represented by the state-action values (i.e. the $< s, a >$}
pair). From Figure 4.4, neighbors closer in distance exhibit channel measurements that have a higher correlation with a node’s own observations, and can therefore be trusted. Thus,
\[
\eta^{i,j} = 1 - \frac{d^{i,j}}{R},
\]  
(4.8)

where \( R \) is the transmission range, and \( d^{i,j} \) is the distance between the nodes \( i \) and \( j \).

The cumulative weight function \( W^{i,j} \) in eq(4.5) can now be written as:
\[
W^{i,j}(s,a) = \epsilon^{i,j}(s,a) \cdot \eta^{i,j}
\]  
(4.9)

### 4.3 Leveraging Cooperation at the Link Layer

In this section we develop a simple MAC protocol to demonstrate how the cooperation between the nodes can be implemented and analyzed in a practical setting. We mainly study the following effects at the link layer:

- Most existing MAC protocols rely on the transceiver pair to select a currently vacant channel. Instead, CLICK selects the channel with the least \( Q \)-value with the aim of maximizing the long term availability, and enhanced PU protection. The benefit and tradeoff of such a selection strategy, as opposed to the classical approach of any vacant channel selection, must be evaluated.

- The collaboration imposes a load on the network, resulting in transmission delays and packet losses. The effect of transmitting this additional state-action information on link layer performance must be quantitatively measured.

- The spectrum sensing time can be reduced though collaboration, wherein a node rapidly gains a high expert or \( E \)-value by assimilating other’s state-action tables. Thus, expending a lot of time for sensing to collect its own measurements may no longer be needed, and the radio may instead be used for data transmissions to improve throughput.

#### 4.3.1 MAC Protocol Design

We assume that the CR users are equipped with two radio transceivers, which we call as the receive radio \( (R_r) \) and transmit radio \( (R_t) \) as shown in Figure 4.6. Each node also has a default receive channel which is chosen as the channel with the least \( Q \)-value, i.e. least PU
activity at its location. The receive radio $R_t$ is always tuned to this default channel, unless requested otherwise, and the choice of this channel is broadcast through hello messages periodically by the receiver. This packet also contains the $Q$ and $E$-values used by the neighboring nodes for collaboration. The transmit radio $R_t$ tunes to the receive channel of the intended recipient to begin the link layer packet transfer. Unlike the work in [98], we do not assume any network administrator assigned channel, or specialized low power radios.

The two cases for MAC layer coordination when a node is attempting a packet transfer to node $B$ are as follows:

**Channel Common to Sender and Receiver**

Consider the interaction between $A$ and $B$, with say, channel $x$ as the default receive channel for $B$. If the $Q$-value of this default channel $x$ at node $A$ is lower than the pre-decided threshold (assumed at 0.4 to have at least greater than 50% safety), then it immediately
tunes its transmit radio $R_t$ to $x$. It listens to the channel for DIFS time, sensing both the PU and the transmission of its neighboring CR nodes, before sending out the request to send (RTS) packet, as defined in the 802.11 standard. The RTS is received by the node $B$ on channel $x$, and the remainder of the interaction follows the classical clear to send (CTS)-DATA-acknowledgement (ACK) cycle. If the RTS-CTS handshake is successfully concluded, just as the data transmission begins, the free radios ($R_r$ for $A$ and $R_t$ for $B$) begin the spectrum sensing of the channels, till all the channels from $21 − 51$ are sensed.

**Different Channels between Sender and Receiver**

After successful completion of the transfer between $A$ and $B$, consider another data transmission, this time initiated by node $C$ for recipient $B$. Unlike the previous case, assume that channel $x$ has a $Q$-value greater than the permissible threshold for $C$, and it is likely that transmission in this channel will get interrupted due to PU activity in $C$’s neighborhood. Hence, $C$ now proposes a channel list $L$ to $B$ with permissible $Q$-values, and allows $B$ to pick the best channel. $B$ chooses the channel from the entries of the list $L$, which has the least $Q$-value at its own location, say channel $C_B = y$, and informs the sender $C$. The receiver radio $R_r$ for $B$ is now tuned to $y$ for the duration of the transfer, and the RTS-CTS-DATA-ACK cycle is repeated. If no permissible channel is found, the channel reply is denied to node $C$.

**4.3.2 Cooperation Benefits in the MAC Protocol**

**Sensing Time Optimization**

The time used for collecting the spectrum information results in lowering the link layer throughput as the transceiver is busy. As seen in Figure 4.6, node $B$ is engaged in spectrum sensing, and is unable to accept new requests from node $C$, leading to re-transmissions of the channel list $L$. Through collaboration, the individual sensing time can be shortened as the state table is enhanced by integrating the information contained by the neighboring nodes. This actual sensing time, $T^i_t$ for node $i$ considering all the channels in $S$ is,

$$T^i_t = \sum_{k=1}^{[S]} t^k_{\text{sense}} \left[ 1 - \frac{E^i(k) - E'^i(k)}{E^i(k)} \right]$$  \hspace{1cm} (4.10)

where $E^i(k)$ and $E'^i(k)$ are the $E$-values after collaboration, and from the individual measurements, respectively. This expression implies that the minimum sensing time $t^k_{\text{sense}}$ for a given channel $k \in S$ (obtained from measurements described in Section 4.1.3) by each
node) is scaled by the fractional increase in the expert or $E$-value at a node. This results in a total sensing time savings of $\Delta_i = \sum_{k=1}^{S} t_{sense}^{k} - T_i^{i}$.

### 4.3.3 PU protection and Switching Minimization

The $Q$-value, with time, captures not only the spectrum availability at the node location, but is also a measure of its availability in the immediate neighborhood as well. This is because, the $Q$-value table is periodically exchanged between 1-hop neighbors, and reliable channel readings integrated. Thus, choosing the channel based on its long term availability saves on repeated spectrum switching, and ensures continued communication without possible service degradation to the PUs. Moreover, as a control channel is absent to make the link layer more spectrally efficient, it is vital that the default receiver channel $R_r$ be available for a longer extent of time.

### 4.4 Performance Evaluation

In this section, we evaluate the performance of the proposed cooperation scheme CLICK and its performance at the link layer using the ns-2 simulator extended for CR networks [99]. We use real measurement traces to derive the PDF of successful PU detection for different durations of sensing time. This follows from the discussion in Section 4.1.3, where we measure the number of samples for which the received signal strength of an active TV transmitter falls below the noise floor leading to missed-detection. We consider a grid
Figure 4.8: The probability of correct PU detection for 15 nodes (a) and 30 nodes (b), while the convergence behavior for CLICK is in (c).
topology of 9 cells at the Northeastern University campus, in a \(300 \times 300\) m\(^2\) area, and the missed detection PDFs for TV channels 21 – 51 are obtained by carrying out measurements in each cell by USRP2 devices. The latter are deployed randomly, and several sets of time series measurements are saved to later simulate the real world environment. Thus, practically observed error probability and signal strength traces are used for the packet level simulation in ns-2. We restrict the analysis to a subset of TV channels 25 – 29 and 38 – 40 which we exhibit higher variance of the PU signal for different grid locations.

### 4.4.1 PU Protection Analysis

In this analysis, we evaluate the benefits and costs of cooperation and learning on the PU detection process. The sensing duration and inter-sensing intervals are fixed at 0.026 s and 0.2 s, respectively to ensure correct PU detection. For fairness, the sensing time optimization (Section 4.3.2) is disabled in this experiment. We compare the performance of CLICK with (i) Individual sensing, i.e., classical energy detection using threshold from Section 4.1 with no cooperation, (ii) \(Q\)-learning at the node level, again without cooperation, and (iii) the \(K\) by \(N\) collaboration scheme proposed in [87].

In Figure 4.7(a), we show the overhead of the \(K\) by \(N\) scheme and of two configurations of CLICK with different collaboration intervals (i.e. every 1.0 s and 2.0 s, respectively). The \(K\) by \(N\) scheme incurs the highest overhead due to the fact that each CR node must collect measurements from other nodes after each sensing interval (i.e. 0.2 s in our experiments). Figure 4.7(a) shows that the frequency of sharing actions has also a critical impact on CLICK, i.e. higher accuracy can be guaranteed with more frequent updates. However, the sharing action can be performed asynchronously and at lower rate than the sensing, and this explains the overhead reduction of our scheme. In Figures 4.8(a) and 4.8(b) we depict the probability of PU detection over time in the configurations with 15 and 30 nodes, respectively. In both figures, we notice that the cooperation accelerates the learning at a higher rate, when compared to the non-cooperative \(Q\)-learning scheme, giving about 12 – 15% improvement. Figure 4.8(c) shows the improvement on the increasing number of nodes on the convergence time of CLICK.

### 4.4.2 Analysis of MAC Protocol using CLICK

In this section, we study the benefits of CLICK integrated into our proposed MAC protocol, which we call as CLICK MAC. We consider a topology with 12 CR nodes with 6 active connections, and we vary the system load produced by each connection. The traffic type is Constant Bit Rate (CBR) with the UDP protocol at transport layer. Figure 4.9(a) shows the throughput of the proposed MAC scheme for different sensing time
intervals. Figure 4.9(a) shows that the communication performance at MAC layer is improved by an accurate setting of the sensing interval. When the sensing time is set to the minimum value, PU interference may lead to packet losses during the transmission period of CR users. However, long sensing time reduces the transmission opportunities for CR nodes. The CLICK MAC scheme with Adaptive Sensing is able to balance the tradeoff between PU protection and CR transmission opportunities, as shown in Figure 4.9(a). Figure 4.9(b) shows the throughput of the proposed MAC scheme with different underlying sensing schemes for deciding the default channel for the receiver interface $R_r$ at each node. In the case of Individual, $Q$-learning and $K$ by $N$ scheme, channel selection is performed by receiver node only. By increasing the long-term knowledge on each channel, CLICK reduces the overhead of channel switching, and the impact of PU interference. As a result, the integrated CLICK MAC scheme guarantees the highest throughput also under high traffic loads. The same benefit can be seen in terms of end-to-end delay, which is shown in Figure 4.9(c).
Figure 4.9: The throughput of CLICK MAC for different values of sensing time interval (a), and sensing schemes (b). The link delay is given in (c).
Chapter 5

Full Duplex: Doubling the Spectrum Opportunity

Wireless full duplex (FD) technology allows a radio to send and receive data on the same channel simultaneously. It promises massive improvements in channel capacity by ushering in a paradigm shift in the design of existing networking protocols [23–25, 28, 100–102]. Before its inception, half duplex communication was the de-facto standard, i.e., nodes may either transmit or receive at any given time [26]. This key assumption influenced the design of the protocol stack, especially the channel access mechanism at the link layer. As a result, any simultaneous use of the channel by more than one node within their interference range in the same network could cause the transmitted packets to collide. This, in turn, results in wastage of bandwidth resources, and brings in the requirement for retransmission by all of the contending nodes that suffered packet losses. With the advent of the FD technology and the ability to transmit and receive at the same time on the wireless channel, this problem of simultaneous channel access can be mitigated to some extent. While recent work on building such systems [27, 103] are important steps towards practical realizations of this technology, there has been very limited work on analyzing FD performance for protocol stack implementations. This chapter attempts to bridge this gap at the link layer by assuming a simple CSMA/CA channel access, and then defining a Markov chain-based theoretical model to give closed form expressions for the system performance.

For the analysis presented in this chapter, we formally define the transmission scenario shown in Fig. 5.1, wherein an access point (AP) has several associated clients, with each device equipped with a FD radio. Using terminology similar to [23], let client $k$ be the primary transmitter that sends its own packet to the AP, which now assumes the role of the secondary transmitter. The secondary transmitter, upon receiving the primary transmitter’s packet, can potentially start transmitting its own packet, thereby creating a dual link.
Figure 5.1: Representation of a Star topology with full duplex nodes.

Client \( i \) that lies within the coverage radius of client \( k \) immediately detects the transmission and postpones its own transmissions if any. Additionally, by letting the AP to transmit while it is receiving, other clients hidden from \( k \), namely \( l \) and \( m \) detect the channel as busy, and refrain from transmitting any packets themselves. Thus, the FD channel access avoids collisions and mitigates the hidden terminal problem to a considerable extent.

Despite its benefits, FD brings in several unique challenges in MAC protocol design. To realize its full potential, the intended receiver, (AP in Fig. 5.1, must have packets for the primary transmitter \( k \) at the same time. This can be determined by examining the header of the incoming packet (before the packet has completely been delivered to save on processing time) to determine the transmitter, and checking if the head-of-line (HOL) packet at the receiver is in fact addressed to the primary transmitter. Moreover, there are fairness concerns regarding channel access, as the AP piggybacks on the original contention resolution won by client \( k \). While this improves the overall channel utilization, other nodes may potentially sacrifice their own access time to allow the AP to continue its transmissions. In an attempt to mitigate this problem of fair channel access, existing works on FD [23, 27, 28] modify classical CSMA/CA and 802.11 DCF, where an adaptive back off counter is used for randomizing channel access more fairly among nodes. Although these works have evaluated the performance of their MAC through simulation and implementation, there is a lack of a rigorous analytical model quantifying their operational benefits. While there are numerous seminal works on the analysis of MAC protocols for classical half duplex networks [29–31, 76, 78, 104–107], they do not reveal insights on the performance gain of FD over classical half duplex (HD) scenarios. Moreover, the impact of various network settings, including the number of nodes, their traffic rates, number of hidden terminals, and the selection of the backoff duration cannot be obtained through trivial extensions of these half duplex models. For these reasons, we formulate a completely new theoretical model specially suited for FD in this chapter.

The rest of this chapter is organized as follows: In Section 5.1 we review related
work, and in Section 5.2, we discuss the network architecture. In Section 5.3, we give a mathematical model of the FD MAC protocol and derive closed-form expressions. In Section 5.4, we validate our model with extensive simulation results.

5.1 Related Work and Motivation

The design of efficient protocol stack for FD networks is in an early stage. In this section, we review the existing works for MAC protocol design and evaluation of FD networks. A fullduplex MAC protocol, called ContraFlow, is proposed in [108] along with the development of a prototype that includes a back-off algorithm for improving fairness. The performance evaluation is limited to networks of limited size and selected topologies. In [28], the authors proposed a distributed full-duplex MAC, called FD-MAC, which introduces features such as shared random back-off and virtual contention resolution with their corresponding implementations on the WARP platform [109]. In FD-MAC, AP switches between full duplex and half duplex to ensure that all nodes get a chance to transmit. A MAC with dynamic contention window control based on the current transmission queue length is proposed in [110] to increase the transmission opportunity of FD operation, and to balance uplink and downlink traffic. However, simulation results are limited to sparse topologies, without including scenarios involving hidden terminals. In [27], a simple CSMA/CA based MAC protocol is implemented on the WARP platform that broadcasts a busy tone (BT) by the AP to eliminate the hidden terminal problem during an empty slot. In [111], a distributed FD MAC protocol for ad-hoc networks and multi-AP networks is proposed in order to maximize FD and concurrent transmissions in the network, using additional signaling based on pseudo-random noise sequences. Janus [112] is another full-duplex MAC protocol that is centralized at the AP, and eliminates random back-off. AP collects interference information from nodes, divides transmission schedules based on the global traffic information, and send control packets at the beginning and end of each round to avoid collisions.

The above works laid the initial foundations of how protocols and hardware that support FD operation may be designed. However, they do not include an analytical framework that can be used to predict the performance of FD networks in general network settings, such as varying number of nodes and hidden terminals, contention window length, among others.

There has been some recent effort in characterizing FD’s performance from a theoretical standpoint. In [113], achievable throughput in full-duplex is characterized as opposed to other channel access schemes such as MIMO and MU-MIMO. In [114], theoretical bounds for full-duplex gain over half-duplex has been derived for various topologies as a
function of difference between transmission and interference range. It has been shown that when these two ranges are equal for a randomly deployed ad hoc network the asymptotic bound for full-duplex gain is only 28%. However, none of these works consider a mathematical modeling of a real-world FD protocol. Our work serves in bridging this gap, and we use the CSMA/CA based MAC protocol in [27] implemented on physical hardware, as the base protocol with few additional modifications to its busy tone broadcasting scenarios. In this work, we extend the model in [105], which is an accurate analytical model of a saturated IEEE 802.11 DCF network with no hidden terminals, to the full-duplex medium access network with the presence of hidden terminals.

5.2 Network Architecture

Consider the network shown in Fig. 5.1. Let \( n \) clients be connected to an AP. Each node \( i \) has a set of covered nodes \( N_i^c \) that can hear its transmission, and a set of hidden nodes \( N_i^H \) which are out of its sensing range (assume equal sensing and receiving range for all nodes). Each node, including the AP, adopts a CSMA/CA channel access mechanism with a contention window (CCW). More specifically, to access the medium, each node with a packet to transmit randomly chooses a value in the range \([0, W]\) and counts down to zero from that value during the time it senses the channel as idle. This means that if the channel gets busy during countdown, the back off timer is frozen. The countdown is resumed once the channel becomes idle again, and once this number reaches zero, the node attempts transmission.

We assume all nodes have FD capability, and hence, once a given node initiates transmission to a destination node, the latter checks whether it has a packet for the former at the HOL. If the HOL packet is destined for the source node, the destination starts sending it in FD mode to the source. Fig. 5.2(a) shows such a scenario, where the client initiates packet transmission to the AP. The latter decodes the packet header, and compares the sending node’s address to the destination of its own HOL packet. If they are the same, it enters FD mode and sends the HOL packet. If not, the AP sends a BT in order to keep the channel busy and prevent any hidden terminal of the client from transmitting and causing a collision. This scenario is shown in Fig. 5.2(b). In this case, the length of transmission and channel busy time is \( \tau_H \) time slots.

Fig. 5.2(c) shows the AP initiating a packet transmission to a client. Since all clients are sending only to the AP, the HOL packet at the client is addressed to the AP by default. Therefore, the client enters FD mode and immediately sends its packet as it begins to receive a packet from the AP. Assuming fixed packet length for both nodes engaging in the FD data exchange, the AP notifies the hidden terminals by broadcasting a BT at the end.
Figure 5.2: Cases of full duplex transmissions initiated by (a) the client node and a packet reply by the AP, (b) the client node and BT broadcast by the AP, and (c) the AP and packet reply by the client.

of its packet transmission, while the client is still transmitting. At the end of a successful FD transmission, the two nodes send their ACKs simultaneously to each other after a fixed short gap called SIFS, as per the standard CSMA/CA algorithm (the processing time needed to check the correctness of the received packet). The total number of slot times for a successful FD transmission is assumed $\tau_F$ time slots in both Figs. 5.2(a) and 5.2(c).

The combination of FD and BT does not fully eliminate the hidden terminal problem as the header transmission of the primary transmitter is susceptible to collision with some probability. This period is shown in Figs. 5.2(a) and 5.2(b) and called vulnerable period. We assume fixed header time for all the nodes in our system, consuming $\tau_V$ time slots.

As we discussed earlier, FD does not fully remove the hidden terminal problem and collisions might still occur. Furthermore, two or more nodes, hidden or covered, might start transmission at the same slot hence causing a collision at the receiver. In all these cases the length of collision will play a pivotal role in the analysis of this chapter. Because of FD, a node should be able to quickly detect a collision with another, if the two are covered by each other. We argue that this type of collision (we call it “covered node collision”) can occur only when the nodes start transmitting at the same slot. Due to a processing overhead that exists for detecting a simultaneous transmission on the medium, we assume nodes cease their transmission after the header time which takes $\tau_V$ time slots.

However, collision between hidden terminals is more tricky. Due to being hidden from each other, nodes cannot know of the collision by themselves. So the easy argument is that nodes know of such collision when they don’t receive any BT from the AP at the end of their header. But we argue that this could potentially result in an unpredictable...
succession of collision by several hidden nodes without realizing that there has just been a collision in the network. For this reason we introduce a collision notification signal sent out by the AP once it detects a collision by two or more hidden nodes. Several works [115, 116] have shown that even a half duplex AP (receiver in general) can detect collision once it occurs. These works have then used out of band communication to notify the colliding nodes once they detect it. However in FD there is no need for out-of-band communication since nodes are capable of sending and receiving on the same channel simultaneously. The same scenario is envisaged for FD nodes in [117]. Therefore, we consider the same capability in our network setting where the AP can notify the colliding nodes immediately after detecting the collision. Once transmitting nodes are notified, they cease their transmission.

From a client’s perspective, collision will also take \( \tau_V \) time slots. However from the AP’s perspective the length of collision on the channel is varying depending on the relative starting time of the second colliding client. If the two nodes start on the same slot, collision will take \( \tau_V \) time slots. If the second node start exactly after the first node has sent it’s header, then collision take \( 2\tau_V \) time slots. We can say that in average, hidden terminal collision will take \( 3\tau_V/2 \) slots.

### 5.3 Analytical Model for Full Duplex MAC

We model the network described in Section 5.2 using a discrete-time Markov renewal process \( M \) shown in Fig. 5.3. This Markov process shows the state space of each HOL packet at the node’s transmit queue. By analyzing the steady state of the Markov process, we obtain the throughput performance of nodes in the network. Our analysis is applicable for saturated conditions, i.e., every node has a packet to transmit.

For every HOL packet, every node, either AP or client, starts from state \( S \). It randomly chooses a back off counter in the range \([0, W)\) and moves to the corresponding state \( 0, ..., W - 1 \) to start the countdown. From this state, the node counts down and transits to the lower state with probability \( \alpha_t \), if it finds the channel idle in time slot \( t \). If the channel is found busy and the initial header bits are decoded to reveal that the packet is addressed to this specific sensing node, let \( \beta_t \) be the probability that this node’s HOL packet is also for the transmitter of the header (i.e., the primary transmitter). When, this condition occurs, there is a possibility of beginning FD operation. In this case, the node immediately sends the HOL packet to the primary transmitter and then directly moves to state \( S \). Otherwise with probability \( 1 - \alpha_t - \beta_t \), the node stays in the same state, and continues to sense the channel in the next time slot.

When the node reaches state \( 0 \), if it finds the channel idle (with probability \( \alpha_t \)), it at-
tempts a transmission by sending the packet header and moving to state $T$. After header transmission is completed and the transmission is successful, i.e., no collision has occurred, the node continues sending the whole packet and makes a transition to state $S$. Otherwise a transition to state $C$ takes place. In the latter case, the back off process is repeated to attempt a re-transmission of the collided packet. The probability of a successful transmission is given as $p_t$.

The transition probabilities in the above Markov chain can be expressed as:

$$P[i|i+1] = \alpha_t, P[S|i] = \beta_t, P[S|T] = p_t,$$
$$P[C|T] = 1 - p_t, P[i|C] = P[i|S] = \frac{1}{W}$$  

(5.1)

The Markov chain in Fig. 5.3 can be easily shown to be uniformly strongly ergodic if and only if the limits below exist [105, 118]:

$$\lim_{t \to \infty} p_t = p, \lim_{t \to \infty} \alpha_t = \alpha, \lim_{t \to \infty} \beta_t = \beta$$

(5.2)

and therefore, has a stationary probability distribution. It is straightforward to derive the steady-state probability distribution from the following set of equations:
\[ \pi_{W-1} = \frac{1}{W(\alpha + \beta)}(\pi_S + \pi_C) \]
\[ \pi_i = u\pi_{i+1} + \frac{1}{W(\alpha + \beta)}(\pi_S + \pi_C) \]
\[ 0 \leq i < W - 1 \]
\[ \pi_T = \alpha\pi_0 \]
\[ \pi_S = p\pi_T + \beta(\sum_{j=0}^{W-1} \pi_j) \]
\[ \pi_C = (1 - p)\pi_T \]
\[ \pi_S + \pi_C + \pi_T + \sum_{j=0}^{W-1} \pi_j = 1 \]  \hspace{1cm} (5.3)

where,

\[ u = \frac{\alpha}{\beta + \alpha} \]

By solving (5.3), the probability \( \pi_S \) is obtained as:

\[ \pi_S = \frac{W(1-u)\beta - Xu\beta(1-p)}{W(1-u)(\beta + 1) - Xu(1-\beta)} \]  \hspace{1cm} (5.4)

where \( X \) is short for:

\[ X = (1 - u^W) \]  \hspace{1cm} (5.5)

(5.6)

Also, \( \pi_C \) and \( \pi_i, 0 = 1, ..., W - 1 \) can be given as:

\[ \pi_T = \frac{Xu\beta}{W(1-u)(\beta + 1) - Xu(1-\beta)} \]  \hspace{1cm} (5.7)

\[ \pi_C = \frac{Xu\beta(1-p)}{W(1-u)(\beta + 1) - Xu(1-\beta)} \]  \hspace{1cm} (5.8)

\[ \pi_i = \frac{(1-u)(1-u^{W-i})}{W(1-u)(\beta + 1) - Xu(1-\beta)} \]  \hspace{1cm} (5.9)

The mean holding time for count down states \( i = 0, 1, ..., W - 1 \) can be obtained using geometric distribution as \( \delta_i = \frac{1}{\alpha + \beta} \). Also assuming \( \delta_T, \delta_S \) and \( \delta_C \) as mean holding times of the states T, S and C respectively, the limiting state probabilities for Markov process \( M \) are given by:

113
\[ \pi_j = \frac{\pi_j \delta_j}{\sum_{i \in M} \pi_i \delta_i} \] (5.10)

Based on this formulation, the limiting probability of state \( S, \pi_S \) is the service rate or throughput of the corresponding node’s queue [105]. Now if the packet arrival rate \( \lambda \) to each node’s queue is more than the service rate, the node is in saturated mode.

The derived steady-state probability distribution is valid for all nodes, including the AP. However, the probability values are on a per-node basis, since the Markov transition probabilities for different nodes are not the same. For example, the probability of finding the channel idle \( \alpha \) is different for each node since the packet arrival rate \( \lambda \), and the number of covered and hidden terminals are different for each node. More specifically, client nodes are always sending packets to the AP and they only receive packets from the AP. This makes the AP’s perspective of channel activity different from that of the client nodes. For the subsequent discussion, we apply the superscript per-node parameters as \( \alpha^i, \beta^i \) and \( p^j \) for node \( i, i = 1, ..., n \) and \( \alpha^{ap}, \beta^{ap} \) and \( p^{ap} \) for the AP. We derive analytical expressions for these parameters separately for clients and the AP in the rest of this section. In Table 5.1, we list the commonly used notations.

### 5.3.1 Analysis from a client’s perspective

Each node in our system has a different view of the channel when other nodes are transmitting. From an arbitrary client’s perspective, say \( i \) in Fig. 5.1, the channel can be in five different states when it is not transmitting:

- 1) Successful FD: Client node \( i \) may overhear a client \( k \) within its coverage radius initiate communication to the AP, while the AP replies in FD mode (FD1 case). Client \( i \) may also hear the AP initiating a successful FD transmission to any other client node, either covered by it \( (k) \) or hidden from it \( (m) \) (FD2 case). In all of these cases, the time taken will be \( \tau_F \).

- 2) Successful HD: When AP’s HOL packet is not for \( k \), and \( k \) initiates a transmission to the AP, client \( i \) observes the channel usage in the situation depicted in Fig. 5.2(b) (HD1 case). Also, if \( m \) initiates a packet transmission to the AP and a successful full duplex exchange begins between them, client \( i \) (hidden from \( m \)) only hears one part of the communication, which is a HD transmission (HD2 case).

- 3) BT-ACK transmission: Client \( i \) may receive only the long BT and the ACK sent by the AP, as shown in Fig. 5.2(b), to a hidden client (say, \( m \)) during and following a successful HD transmission by the client. The length of the BT-ACK is considered as \( \tau_A \).
Table 5.1: List of notations used in the analysis

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mathcal{N})</td>
<td>set of client nodes</td>
</tr>
<tr>
<td>(n)</td>
<td>total number of client nodes or (</td>
</tr>
<tr>
<td>(\mathcal{N}^i_c)</td>
<td>set of client nodes that are covered by node (i \in {\mathcal{N}, AP})</td>
</tr>
<tr>
<td>(\mathcal{N}^i_h)</td>
<td>set of client nodes hidden from node (i)</td>
</tr>
<tr>
<td>(\omega^i(t))</td>
<td>prob. of transmission attempt by node (i) at slot (t), given that channel sensed idle at slot (t-1)</td>
</tr>
<tr>
<td>(\omega^{ap}(t))</td>
<td>prob. of transmission attempt by the AP at slot (t), given that channel sensed idle at slot (t-1)</td>
</tr>
<tr>
<td>(\nu^i(t))</td>
<td>prob. of transmission attempt by node (i) in the next (\tau_V) slot until (t + \tau_V), given that channel sensed idle at slot (t-1)</td>
</tr>
<tr>
<td>(\nu^{ap}(t))</td>
<td>prob. of hearing an AP’s FD reply at (t), given that channel sensed idle at (t-1)</td>
</tr>
<tr>
<td>(\delta_j)</td>
<td>the average time spent in state (j) known as mean holding time of state (j)</td>
</tr>
<tr>
<td>(\pi_j)</td>
<td>stationary prob. of state (j)</td>
</tr>
<tr>
<td>(\tilde{\pi}_S)</td>
<td>throughput of node (i) equal to the limiting prob. of state (S)</td>
</tr>
</tbody>
</table>

4) Failed transmission: A client may sense a collision. The collision could happen in two cases: (i) when two nodes or more start transmission in the same slot, or, (ii) one or more hidden clients transmit in the vulnerable periods of a transmitting node. The length of a collision is considered to be \(\tau_V\) which is the length of a packet header, as collisions are detected during time using full duplex.

5) Idle.

The probability of sensing the channel idle by node \(i\) at each time slot \(t + 1\), \(\alpha^i_{t+1}\) is dependent on the state of the channel at the previous time slot. Hence, we have:

\[
\alpha^i_{t+1} = P_i[\text{idle at } t + 1|\text{FD at } t]P_i[\text{FD at } t] \\
+ P_i[\text{idle at } t + 1|\text{HD at } t]P_i[\text{HD at } t] \\
+ P_i[\text{idle at } t + 1|\text{BTACK at } t]P_i[\text{BTACK at } t] \\
+ P_i[\text{idle at } t + 1|\text{Collision at } t]P_i[\text{Collision at } t] \\
+ P_i[\text{idle at } t + 1|\text{idle at } t]P_i[\text{idle at } t]
\]

(5.11)

Considering the assumption of fixed header and packet length for all the nodes in the network, the successful FD, HD, BT-ACK and collision take respectively \(\tau_F\), \(\tau_H\), \(\tau_A\) and...
\( \tau_V \) slot times to finish, therefore the conditional probabilities in (5.11) would be as the following:

\[
\begin{align*}
P_i[\text{idle at } t + 1|\text{FD at } t] &= \frac{1}{\tau_F} \\
P_i[\text{idle at } t + 1|\text{HD at } t] &= \frac{1}{\tau_H} \\
P_i[\text{idle at } t + 1|\text{BTACK at } t] &= \frac{1}{\tau_A} \\
P_i[\text{idle at } t + 1|\text{Collision at } t] &= \frac{1}{\tau_V}
\end{align*}
\] (5.12)

For the last conditional probability, namely \( P_i[\text{idle at } t + 1|\text{idle at } t] \), we use \( \omega_k, \omega_{ap} \) and \( \nu_k \) and \( \nu_{ap} \) as noted in Table 5.1:

\[
P_i[\text{idle at } t + 1|\text{idle at } t] = (1 - \omega_{ap}(t+1))(1 - \nu_{ap}(t+1)) \prod_{k \in \mathcal{N}_c} (1 - \omega_k(t+1))
\] (5.13)

Apart from the conditional probabilities in (5.11), the probability of each channel state must be calculated. We start with the state of successful FD that is given by:

\[
P_i[\text{FD at } t] = \sum_{j=1}^{\tau_F} P_i[\text{FD at } t - j + 1|\text{idle at } t - j] P_i[\text{idle at } t - j]
\] (5.14)

The inner conditional probability is the summation of conditional probabilities for the two cases of FD1 and FD2 explained earlier:

\[
P_i[\text{FD at } t - j + 1|\text{idle at } t - j] = P_i[\text{FD at } t - j + 1|\text{idle at } t - j] + P_i[\text{FD at } t - j + 1|\text{idle at } t - j]
\] (5.15)

For the FD1 case, the probability that client \( i \) initiates a transmission to the AP accompanied by the AP’s simultaneous reply in FD, is given by:

\[
P_i[\text{FD1 at } t - j + 1|\text{idle at } t - j] = \sum_{k \in \mathcal{N}_c} \frac{1}{n} \omega_k(t - j + 1)(1 - \omega_{ap}(t - j + 1)) \times \prod_{l \in \mathcal{N}_c - i} (1 - \omega_l(t - j + 1)) \prod_{m \in \mathcal{N}_H} \nu_m(t - j + 1)
\] (5.16)
The terms above show the total probability for any node \( k \) in the set \( \mathcal{N}_c^i \) having a successful FD transmission. This requires node \( k \) transmitting at time slot \( t - j + 1 \) when the HOL packet of AP is for \( k \), and the AP and all nodes in \( \mathcal{N}_c^k \) are silent at that time slot and all of its hidden terminals in \( \mathcal{N}_H^k \) are silent for the duration of vulnerable period \( \tau_V \) after \( t - j + 1 \). Note that we assume the AP’s outgoing traffic to its clients is uniform, hence \( \frac{1}{n} \) is the probability that AP’s HOL packet is destined to the client node that is transmitting to FD, while probability \( \frac{n-1}{n} \) is for otherwise, when the AP sends busy tone as in the Fig. 5.2(b).

The probability for FD2 case can also be written as:

\[
P_i[FD2 \text{ at } t - j + 1|idle \text{ at } t - j] =
\omega^{ap}(t - j + 1) \prod_{l \in \mathcal{N}^i} (1 - \omega^l(t - j + 1))
\]  

(5.17)

which is the total probability of all clients keeping silent, while the AP is attempting a transmission to its intended client at slot time \( t - j + 1 \). In this case, the intended client will respond with a FD packet with probability 1 since the client nodes are always in saturated mode, and they only send packets to the AP.

We can formulate expressions similar to (5.14) for HD and BT-ACK as well. For HD, we have the following:

\[
P_i[HD \text{ at } t] = \sum_{j=1}^{\tau_H} P_i[HD \text{ at } t - j + 1|idle \text{ at } t - j]P_i[idle \text{ at } t - j]
\]  

(5.18)

\[
P_i[BTACK \text{ at } t] = \sum_{j=1}^{\tau_A} P_i[BTACK \text{ at } t - j + 1|idle \text{ at } t - j]P_i[idle \text{ at } t - j]
\]  

(5.19)

Similar to (5.15), \( P_i[HD \text{ at } t - j + 1|idle \text{ at } t - j] \) can be written as the summation of HD1 and HD2 cases.

\[
P_i[HD \text{ at } t - j + 1|idle \text{ at } t - j] =
= P_i[HD1 \text{ at } t - j + 1|idle \text{ at } t - j]
+ P_i[HD2 \text{ at } t - j + 1|idle \text{ at } t - j]
\]  

(5.20)
For HD1 we have:

\[ P_i[H D1 \text{ at } t - j + 1|\text{idle at } t - j] = \]

\[ \sum_{k \in N_c^i} \frac{n - 1}{n} \omega^k(t - j + 1)(1 - \omega^\text{ap}(t - j + 1)) \]

\[ \prod_{l \in N_{c,k}^i} (1 - \omega^l(t - j + 1)) \prod_{m \in N_{H}^i} \nu^m(t - j + 1) \]  

(5.21)

which is the total probability over any node \( k \) in \( N_c^i \) to attempt transmission at time slot \( t - j + 1 \), while (i) nodes in \( \hat{N}_{c,k}^i \) and the AP keep silent at that slot, and (ii) its hidden nodes do not transmit during the next \( \tau_V \) slots. This is multiplied by \( \frac{n-1}{n} \), which is the probability leading up to HD1 case as explained earlier. For HD2 case:

\[ P_i[H D2 \text{ at } t - j + 1|\text{idle at } t - j] = \frac{1}{n} \nu^\text{op}(t - j + 1) \]  

(5.22)

This is also the probability of hearing an AP’s FD reply to a hidden node of \( i \).

For the state of hearing the BT-ACK, similarly, we have:

\[ P_i[B T A C K \text{ at } t - j + 1|\text{idle at } t - j] = \frac{n - 1}{n} \nu^\text{op}(t - j + 1) \]  

(5.23)

which is equal to the probability of hearing a BT by the AP in response to the HD transmission of a hidden node of \( i \).

For the idle state, we get \( P[\text{idle at } t] = \alpha_t \). Consequently, for the state of collision sensed by node \( i \), we have:

\[ P_i[\text{Collision at } t] = 1 - P_i[\text{FD at } t] - P_i[H D \text{ at } t] \]

\[ - P_i[B T A C K \text{ at } t] - \alpha_t \]  

(5.24)

Now, \( \alpha_t^i \) can be obtained by combining (5.12-5.13) and (5.14-5.24).

### 5.3.2 Analysis from the AP’s perspective

From the AP’s point of view, the channel has three states when it is not transmitting:

- 1) Successful HD: The AP successfully receives a packet from a client \( k \), when the AP’s own HOL packet is destined for a node other than \( k \).

- 2) Collision: Clients that are hidden from each other could collide or two or more nodes attempt transmission exactly at the same time.
- 3) Idle: The channel remains unused.

A question that might arise here is that why no state for FD is considered here. The reason is, the above states are from the perspective of the AP when it is in channel contention (sensing) mode. Once AP enters an FD mode, it leaves contention mode and that is why it is not considered here. With the above three states, we are able to derive parameters of the Markov chain $M$ for the AP.

Now, similar to (5.11), we have:

$$
a_{t+1}^{ap} = P_{ap}[\text{idle at } t+1|\text{HD at } t]P_{ap}[\text{HD at } t]
+ P_{ap}[\text{idle at } t+1|\text{Collision at } t]P_{ap}[\text{Collision at } t]
+ P_{ap}[\text{idle at } t+1|\text{idle at } t]P_{ap}[\text{idle at } t] \quad (5.25)
$$

where $P_{ap}[\text{idle at } t+1|\text{HD at } t]$ and $P_{ap}[\text{idle at } t+1|\text{Collision at } t]$ are the same as given in (5.12).

Also, $P_{ap}[\text{idle at } t+1|\text{idle at } t]$ is exactly $p_t^{ap}$, as defined in Section 5.3 and is obtained as:

$$
P_{ap}[\text{idle at } t+1|\text{idle at } t] = p_t^{ap}
= \prod_{k=1}^{n} (1 - \omega^k(t + 1)) \quad (5.26)
$$

This is the probability of a successful AP transmission in time slot $t$, if a transmission attempt is made, and given the channel is idle at slot time $t - 1$.

The probability that the AP senses the channel in HD state is given by:

$$
P_{ap}[\text{HD at } t] = \sum_{j=1}^{\tau_H} P^{ap}[\text{HD at } t-j+1|\text{idle at } t-j]P^{ap}[\text{idle at } t-j] \quad (5.27)
$$

where,

$$
P_{ap}[\text{HD at } t-j+1|\text{idle at } t-j] = 
\sum_{k=1}^{n} \frac{n-1}{n} \omega^k(t - j + 1) \times
\prod_{l \in N_c^k} (1 - \omega^l(t - j + 1)) \prod_{m \in N_H^t} \nu^m(t - j + 1) \quad (5.28)
$$

The latter is the total probability of any client $k$ transmitting to AP at slot $t - j + 1$, when AP’s HOL packet’s destination is not for $k$, and the following conditions hold: (i) AP and
nodes in $N^c_i$ keep silent in that slot, and (ii) the nodes in $N^H_i$ keep silent for at least the next $\tau_V$ slots.

For the state of collision, similar to (5.24) we get:

$$P_{ap}[\text{Collision at } t] = 1 - P_{ap}[\text{HD at } t] - \alpha_{ap}^t$$

(5.29)

Now, $\alpha_{ap}^t$ can be obtained by substituting (5.26-5.29) in (5.25).

### 5.3.3 Steady-state probabilities

Due to symmetry, if the following conditions hold, then as $t \to \infty$, in the steady state we have $\omega^i = \omega, \nu^i = \nu, i = 1...n$.

$$|N^c_i| = n_c, |N^H_i| = n_H, \quad i = 1, ..., n$$

(5.30)

These values, as well as $\omega_{ap}$, can be written in terms of the steady state probabilities of the states of the Markov chain in Section 5.3. Based on the definition in Table 5.1, $\nu$ is the given probability that a node does not attempt a transmission in the next $\tau_V$ slots. Also $\omega$ and $\omega_{ap}$ are the respective probabilities for the node and AP for attempting a transmission in an arbitrary slot, which is equivalent to entering state $T$.

$$\begin{cases} 
\nu = \sum_{j=\tau_V}^{W-1} \pi_j \\
\omega_{ap} = \pi_{ap}^T \\
\omega = \pi_T 
\end{cases}$$

(5.31)

Also according to the definitions in Table 5.1, $\nu_{ap}$ is the probability that a node only hears an FD reply by the AP. This event is triggered when a hidden client sent a successful packet to the AP. In the steady state, this probability can be expressed as the event that any of the hidden terminals of an arbitrary client attempts a transmission that subsequently succeeds:

$$\nu_{ap} = n_H \omega (1 - \omega_{ap})(1 - \omega)^{n_c} n^H$$

(5.32)

Once the conditions in (5.30) are met, the terms of (5.11) in the steady state can be
simplified as the following (5.12-5.24):

\[
P[idle at t + 1|FD at t]P[FD at t]_{t \to \infty} = \frac{1}{\tau_F} \tau_F \left[ \frac{n_c}{n} \omega (1 - \omega^{ap})(1 - \omega)^{n_c-1} \nu^{n_H} + \omega^{ap}(1 - \omega)^{n-1} \right] \alpha = Y_1 \alpha \tag{5.33}
\]

\[
P[idle at t + 1|HD at t]P[HD at t]_{t \to \infty} = \frac{1}{\tau_H} \tau_H \left[ \frac{n_c}{n} \omega (1 - \omega^{ap})(1 - \omega)^{n_c-1} \nu^{n_H} \right] + \frac{n_H}{n} \omega (1 - \omega^{ap})(1 - \omega)^{n_c-1} \nu^{n_H-1} \alpha = Y_2 \alpha \tag{5.34}
\]

\[
P[idle at t + 1|BTACK at t]P[BTACK at t]_{t \to \infty} = \frac{1}{\tau_A} \tau_A \left[ \frac{n_c}{n} \omega (1 - \omega^{ap})(1 - \omega)^{n_c-1} \nu^{n_H} \right] = Y_3 \alpha \tag{5.35}
\]

\[
P[idle at t + 1|Collision at t]P[Collision at t]_{t \to \infty} = \frac{1}{\tau_C} \left[ 1 - \tau_F Y_1 \alpha - \tau_H Y_2 \alpha - \tau_A Y_3 \alpha - \alpha \right] \tag{5.36}
\]

\[
P[idle at t + 1|idle at t]P[idle at t]_{t \to \infty} = (1 - \omega)^{n_c}(1 - \omega^{ap})\left[ 1 - n_H \omega (1 - \omega^{ap})(1 - \omega)^{n_c-1} \nu^{n_H-1} \right] \alpha = Y_4 \alpha \tag{5.37}
\]

By replacing (5.33-5.37) in (5.11), \( \alpha \) can be obtained as:

\[
\alpha = \frac{1}{1 + (\tau_F - \tau_C)Y_1 + (\tau_H - \tau_C)Y_2 + (\tau_A - \tau_C)Y_3 + \tau_C(1 - Y_4)} \tag{5.38}
\]
For the AP in the steady state, we get from (5.25-5.29):

\[ P_{\text{ap}}[\text{idle at } t + 1|\text{HD at } t |P_{\text{ap}}[\text{HD at } t | \to \infty \]

\[ = (n - 1)\omega (1 - \omega)^{n_c} \nu^{n_H} \alpha_{\text{ap}} \]

\[ = Z_1 \alpha_{\text{ap}} \] (5.39)

\[ P[\text{idle at } t + 1|\text{Collision at } t |P[\text{Collision at } t | \to \infty \]

\[ = \frac{1}{\tau_C} [1 - \tau_H Z_1 \alpha_{\text{ap}} - \alpha_{\text{ap}}] \] (5.40)

\[ P[\text{idle at } t + 1|\text{idle at } t |P[\text{idle at } t | \to \infty \]

\[ = p_{\text{ap}} \alpha_{\text{ap}} = (1 - \omega)^n \alpha_{\text{ap}} \] (5.41)

By replacing (5.39-5.41) in (5.25), \( \alpha_{\text{ap}} \) can be obtained as:

\[ \alpha_{\text{ap}} = \frac{1}{1 + (\tau_H - \tau_C) Z_1 + \tau_C (1 - p_{\text{ap}})} \] (5.42)

Furthermore, the probability of successful transmission for a client (p) can be written as the product of the probabilities that (i) the AP and the other covered client nodes do not transmit, and (ii) the AP does not have to respond in FD mode to any transmission initiated by any of the remaining hidden nodes:

\[ p = (1 - \omega_{\text{ap}})(1 - \omega)^{n_c} \nu^{n_H} \] (5.43)

Also, \( \beta \) and \( \beta_{\text{ap}} \) in the steady state are given by:

\[ \beta = \frac{1}{n} \omega_{\text{ap}} (1 - \omega)^{n-1} \] (5.44)

\[ \beta_{\text{ap}} = \frac{1}{n} \omega (1 - \omega)^{n_c} \nu^{n_H} \] (5.45)

For \( \beta \), i.e., the probability for a client to go to FD mode during back off, we need the AP to transmit to that node. This probability is given by the product of \( \frac{1}{n} \omega_{\text{ap}} \) with the probability that no other node transmits (excluding collision cases).

To calculate \( \beta_{\text{ap}} \), i.e., the probability for the AP to respond in FD to a packet during back-off, (i) a client must transmit, and (ii) the AP’s HOL packet must also be for the same client (prob. \( \frac{1}{n} \)). To exclude collision cases, this is multiplied with the probability that no hidden node has already initiated a header transmission, and no covered node is attempting transmission in the same time slot. Overall, this is true for any of the \( n \) client nodes.

Now, replacing (5.38), (5.42-5.45) in (5.7-5.9), a system of non-linear equations based on (5.31) is obtained. These equations must be solved for \( \omega \) and \( \omega_{\text{ap}} \) and \( \nu \), which subsequently leads to the calculation of \( \alpha, \beta, p, \alpha_{\text{ap}}, \beta_{\text{ap}} \) and \( p_{\text{ap}} \).
5.3.4 Throughput Analysis

As mentioned in section 5.3, the throughput is given by the limiting state probability of state $S$, $\pi_S$.

To calculate the average throughput of clients and the AP, $\pi_S$ for each must be calculated using (5.10). The key parameters for this are the holding times of successful transmission and collision which are given by

$$\delta_T = H, \delta_S^{ap} = \tau_F - H$$

$$\delta_S = \frac{1}{n}(\tau_F - H) + (1 - \frac{1}{n})(\tau_H - H),$$

$$\delta_C = \frac{n_H \tau_V}{n + 1} + \frac{n_c + 1}{n + 1} \sigma,$$

$$\delta_C^{ap} = \sigma$$

where $\tau_F$, $\tau_H$ and $\tau_V$ are given by

$$\begin{align*}
\tau_F &= 2H + L_p + SIFS + ACK \\
\tau_H &= H + L_p + SIFS + ACK \\
\tau_V &= H
\end{align*}$$

Here, $H$ and $L_p$ are the times to send the header and payload, respectively. Every node spends time to first send the header in state $T$ every time it attempts a transmission. A subsequent successful transmission for the AP $\delta_S^{ap}$ will always involve the remaining portion of an FD packet. However for a client this time could be that required to transmit either (i) an FD packet without the header (header is considered in state $T$) in $\frac{1}{n}$ of the cases, or (ii) an HD packet without the header in the rest of the cases. Also, for collision state $C$, the AP only spends a slot time as it knows about collision after its header is transmitted. This is also true for clients when they collide with their covered nodes. However, when a collision with hidden nodes occurs, the colliding nodes only know when they receive the notification from the AP. On average, this step takes $\tau_V$ time. An average of the two cases is considered for $\delta_C$ above.

For the calculation of $\alpha$ and $\alpha^{ap}$ in (5.38) and (5.42), $\tau_C$ is used in both cases, which is the time that channel is sensed in the state of collision from a node’s perspective. For both AP and clients, this state can be a combination of sub-states when collision occurs between two covered nodes ($H$) or two hidden nodes ($\frac{3H}{2}$ in average). The following expressions for $\tau_C$ are derived from simple probabilistic manipulations.
Table 5.2: Simulation Parameters

<table>
<thead>
<tr>
<th>Simulation Parameter</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC header</td>
<td>28 bytes</td>
</tr>
<tr>
<td>PHY header</td>
<td>24 bytes</td>
</tr>
<tr>
<td>ACK</td>
<td>38 bytes</td>
</tr>
<tr>
<td>Payload size</td>
<td>1000 bytes</td>
</tr>
<tr>
<td>Slot time</td>
<td>20 µs</td>
</tr>
<tr>
<td>SIFS</td>
<td>10 µs</td>
</tr>
<tr>
<td>Channel bit rate</td>
<td>1 Mbps</td>
</tr>
</tbody>
</table>

\[
\tau_{ap} = \frac{n_cn}{n(n-1)}H + \frac{n_Hn}{n(n-1)}\frac{3H}{2},
\]

\[
\tau_C = \frac{n_c^2 + 2n_cn_H + 2 \times n_H}{n(n-1)}H + \frac{n_c^2 - 2n_c}{n(n-1)}\frac{3H}{2}
\]

5.4 Model Validation

In this section, we first present extensive simulation results to validate our model for the FD MAC protocol in Section 5.3. For various network settings, we compare the model predictions with the simulation results and show that they are in close agreement. We then analyze and discuss quantitatively how much benefit FD ushers in over the more common HD transmission scenarios.

5.4.1 Simulation Setup

We have implemented the FD MAC protocol in the ns-2 simulator [119]. Each client is driven into saturation by generating CBR packets at 1Mbps. The MAC layer rate is set at 10 Mbps and the packets generated are each 1000 bytes, which after inclusion of the MAC PLCP header and the PHY preamble, have a total size of 1052 bytes. We consider the fact that the PHY preamble is always transmitted at 1Mbps when calculating the total duration of a packet transmission. We also assume symmetric FD link in terms of data rate and ignore channel errors and the capture effect.

The clients are deployed in the transmission range of an access point, each having the sensing and transmission radius of 150 m. Additional simulation parameters used in the simulation are shown in Table 5.2.
All the simulation results presented here have 95% confidence interval within 1% relative error. The simulations in ns-2 are conducted for both ring and random network topologies as the most relevant scenarios for full duplex networks.

We evaluate the ring topology in order to merely study the hidden terminal effect on the performance of full duplex. In this topology, the average throughput of all clients are the same due to symmetry of network. In addition, the number of hidden stations can be easily controlled by varying the ring radius. On the other hand, for more realistic cases of random topologies, we study the performance of full duplex in networks with non equal client-to-AP distances and hence, different number of hidden stations for each client across the network.

In both cases, we vary the (i) number of hidden terminals, (ii) the number of clients, (iii) and the size of back-off window. In particular, we analyze the effect of such variations on (i) the saturation throughput the network. We normalize the saturation throughput for each experiments as the nominal performance metric. Normalization is done over the MAC data rate which is considered to be 10Mbps in our experiments.

5.4.2 Simulation Results

Ring Topology

At the first step, we have simulated a network of full-duplex nodes with the ring topology to meet the network conditions in (5.30). In this topology, clients are placed uniformly at equal distance from the AP in a circle. Therefore, the number of hidden terminals for all clients can be fixed by properly adjusting the radius of the circle \( R \) while keeping the transmission range constant.

Fig. 5.4 shows the normalized system throughput for such FD network as the number of clients increases. Throughput is shown for \( n_H = 0, 4, 8, 12, 16 \) and \( W = 1024 \) is considered. When nodes are close to the AP, namely hidden terminals are low, throughput is slowly going up versus number of clients. When clients are located at the edge of AP’s range, namely more hidden terminals for each client, the number of clients start to have an adverse effect on the throughput. This is visible in the case of \( n_H = 12 \) where within the shown range, throughput is almost steady.

The effect of backoff window size on system throughput is shown in Fig. 5.5 where \( n = 20 \) is assumed. The interesting observation from this figure is a relatively large effect of hidden terminals on throughput in lower backoff sizes. But as the backoff window is enlarged this effect is diminishing to very small amount in large backoff window as in \( W = 2048 \). This effect is attributed to the decrease in the probability of collision due to hidden terminals given a fixed network size.
Another observation is that, in all cases of hidden terminal presence, i.e. $n_H = 4, 8, 12$, the network experiences a peak in throughput at $W = 512$. This is due to an inherent trade off between probability of transmission attempt, channel idle time and probability of collision. In small backoff windows, probability of transmission attempt is low as channel is sensed busy most of the time and when it is not, a collision due to hidden terminals is very likely. As the backoff window is enlarged, idle channel is more common and hidden terminal collision is less likely. But after a certain point nodes start to lose idle channel opportunities by spending too much time for backoff. For $n_h = 0$ the probability of collision is much smaller than when $n_h > 0$, hence no peak is seen in this figure. However in smaller window sizes such peak appears for $n_h = 0$ as well especially when network size is large. We do not investigate this here as it is widely studied in the other works [76, 105].

**Random Topology**

We have extended our ns-2 simulation to include more general topologies with nodes randomly deployed around the AP. The same transmission parameters are used as in Table 5.2. In this deployment we use uniform distribution to place clients within the transmission range of the AP. As the density of the nodes increases, the average number of hidden terminals for each client also increases. For each random topology we have recorded the exact number of hidden terminals for each client. To be able to predict the throughput
Figure 5.5: The effect of contention window size on system saturation throughput for the ring topology.

results for such general topology from the analysis given in section 5.3 we have used a similar method as in [107]. Jang et al. has used the “back-of-the-envelope” approximation technique from [120] to apply their analysis of a specific topology in a network with hidden terminals to more general topologies, a situation that is applicable to our analysis as well. Such approximation has been shown to yield reasonable accuracy and works as the following:

- For each network node $i$ in a random topology, the number of hidden terminals $(n_H^i)$ and covered nodes $(n_c^i)$ are used to derive a system throughput value from the analysis available for the simplified topology.

- The system throughput of the random topology is approximated as the average of system throughput derived from each pair of $(n_H^i, n_c^i)$ above.

Using this method, we obtained a good approximation of the system throughput which is shown in Figs. 5.6 and 5.7. In Fig. 5.6 throughput network size is plotted for different contention window sizes. The trend in each of the curves is different depending on the contention window size. For lower value of $W = 256$, system throughput is decreasing versus the network size (together with hidden terminals since it’s average is also going up). For $W = 512$, the throughput remains almost steady whereas for higher values of $W$, the trend reverses to increasing. This effect is better shown in Fig. 5.7 where contention window size is varied to obtain throughput. We see the rise of system throughput for
various network sizes to a peak at successive windows sizes and a subsequent decline. Furthermore, we see a crossing point at $W = 512$ after which larger networks outperform smaller networks. The peak phenomenon is very similar to Fig. 5.5 where we show it is caused by hidden terminals. In that figure, we have the peak $n_h = 4, 8, 12$ at $W = 512$. Here, by going through the trace of our random topologies, we have obtained average number hidden terminals in the network for each case of $n = 8, 12, 16, 20$ to be $n_h = 0.3, 1.5, 2.4, 3.8$ respectively.

### 5.4.3 Full duplex Gain Evaluation

To understand to what extent FD capability will improve network performance in contrast to HD, we performed a separate analysis and also conducted simulations for HD networks as well. We specifically considered HD with RTS/CTS mechanism as it is comparable to FD for its capability to mitigate hidden terminal problem. We did not consider HD basic access mechanism since its performance versus RTS/CTS is widely studied in the literature [76, 105]. To have similar analysis of RTS/CTS HD to our FD analysis, we have simplified the Markov chain in Fig. 5.3 by eliminating the transitions for FD transmission. We subsequently calculated the probability $\alpha$ for clients and the AP based on which we eventually computed HD throughput. We do not present this analysis here due to the lack of space, and since it can be similarly derived as in section 5.3. We have also changed the baseline ns-2 implementation for 802.11 with RTS/CTS and adjusted it to our Markov model. Through simulations we have confirmed a match with our analysis similar to FD.
in sections 5.4.2 and 5.4.2.

Fig. 5.8 shows FD gain as network size grows for various contention window sizes. This figure gives us an intuition into how FD behavior, in the way we have considered in our protocol, achieves throughput gain versus HD. As we see for all cases, gain is dropping as network becomes larger. This is intuitive as larger network means smaller chance of a match of packet destination for client at the AP when clients transmit. Remember that in FD MAC, the AP replies in FD to a client’s transmission if its HOL packet is in fact destined for that specific client. Also for smaller contention window size, gain is dropping at a faster rate than larger window sizes. This is due to the dependency of a FD transmission to collisions. In other words, referring to 5.44, as network grows larger, collision in smaller window sizes is more likely hence decreasing the chance of FD. Eventually as network grows above a certain number of clients, probability of FD due a client’s transmission attempt becomes very small, but FD still has a positive gain due to transmissions by AP which are always replied in FD by the receiving client. Transmission by the AP are also much more likely to be successful because it has no hidden terminals. The effect of contention window size on the FD gain is better shown in Fig. 5.9 where we observe gain peaks successively occur at bigger window sizes as network size gets bigger. So basically for each network size there is a specific window in which an FD capable network benefits the most. In other words, the bigger the network is, the higher the optimal window size would be.
Both of the last two figure are from random topologies in which the pure effect of hidden terminal cannot be seen. This is because as we increase the network size, the average number of hidden terminals increases as well. For this purpose, we use the ring topology in which the number of hidden terminals can be controlled. Fig. 5.10 shows FD gain ver-
sus number of hidden terminals with various network size cases using ring topology with $W = 1024$. This figure shows a decrease in gain when more hidden terminals are added to the network.
Chapter 6

Conclusion and Future Work

6.1 Conclusion

This thesis explores various approaches for efficient spectrum utilization. We have presented motivations on why this is an important topic, of social and national importance, and surveyed some of the recent rulings by regulatory bodies in the US, such as the FCC, that allow the use of dynamic spectrum access to achieve this spectrum efficiency. Our work has been undertaken on two research fronts: First, we have explored the use of CRs that allow opportunistic frequency-agile operation, which can make better use of the licensed but not continuously used bands. Second, we analyze the case for using full-duplex radio communication paradigm, that leads to high information rates by enabling simultaneous transmission and reception. Together, we believe, these two technologies will usher in dramatic improvements in the way usable spectrum is detected and used, for both frequency-agile CRs and legacy radios that support full-duplex operation.

In chapter 2, two unique applications of cognitive radio technology were presented and evaluated. First, in the medical domain, we have proposed a dynamic spectrum access framework to efficiently utilize the WMTS bands for wireless medical telemetry applications in a hospital environment. The different priority access rights of the medical devices in portions of the WMTS frequency spectrum in the L-band, the interference from the DTV operators in channel 37, and the variable tolerance to EMI with the change in frequency result in a complex environment, where dynamic spectrum and power assignment is a candidate solution. The proposed framework assigns optimal bandwidth and power to each medical application device seeking to send data to any other location in the hospital, e.g. telemetry data from a patient to the physician’s monitoring station, while avoids interfering with primary signals of that band also protecting the EMI threshold of sensitive medical equipment. We also performed a measurement study on the WMTS band to
derive the activity model of medical telemetry signals in that band. We also envisage CR enabled devices enabling telemetry applications in ambulances and home environments, as well as utilizing rich multimedia data, both of which are not allowed under current regulations. Thus, critical life saving applications can be realized in this new communication paradigm, which will be a driving force for societal well-being in the future.

Second, we envisage that CR will help to realize high bandwidth multimedia applications in vehicular communications and survey the challenges on this path. With the emergence of spectrum databases that could provide spectrum availability data to CRs, some of the challenges of CRVs could be mitigated while opening a new window of research on how to design network of CRVs and spectrum information base stations (BS) along roads. We presented a framework for the placement of BSs, given the diverse capabilities of geolocation, querying, and local sensing capabilities that vehicles may be equipped with. The framework effectively minimizes the cost of operation, and ensures the resulting error in spectrum detection stays within bounds. To the best of our knowledge, this is the first work that directly combines the FCC regulations with the concerns of vehicular CR networks.

In chapter 3, we have formulated a channel allocation scheme for network of CRs with heterogenous QoS classes, where the spectrum PU occupancy statistics is also not the same in different channels, and identified performance bounds for this approach. We proposed two Markovian models that calculates the average delay of the nodes for general streaming and non-streaming QoS classes that are in close agreement with simulation studies. We used the aforementioned models to devise a greedy algorithm with polynomial time complexity that assigns CRs with channels, while ensuring their QoS requirements are met. We show that the spectral efficiency of the given allocation scheme converges to the theoretical bounds for a practical case study, using real-world measurement traces for wireless medical telemetry bands.

In chapter 4, we have undertaken actual channel measurements in the TV spectrum band covering channels $21 \rightarrow 51$. Using these measurements, we have drawn inferences on how cooperation in detecting spectrum opportunities could be achieved in a distributed environment. Our reinforcement learning approach CLICK takes into account channel characteristics and node location to decide which channels and nodes are suitable for collaboration. Finally, we have demonstrated the benefits of cooperation by extending a MAC protocol for CR operation.

In chapter 5, we presented an analytical model of the performance of CSMA/CA based MAC protocol for FD wireless networks. Packet-level simulations were used to validate the formulations of the model, and we demonstrated that they accurately estimate the saturation throughput and packet collision rate. Our model reveals that the FD operation is useful in mitigating classical hidden terminal problems. Moreover, we have shown
that FD achieves higher performance than HD network using RTS/CTS mechanism in almost all network configurations. Though we have addressed specifically the concerns of CSMA/CA based FD communications, this technology can be used in other channel access mechanisms as well. Our results have shown a much smaller gains for CSMA/CA based channel access that what has been inherently assumed for FD, namely a doubling of throughput. Based on this, one could conclude that centrally scheduled networks where each node is synchronized with a central authority might benefit more from the FD than a random access network such as the one considered here.

6.2 Future Work

With advances in opportunistic spectrum access, wireless technology will go through a major overhaul in the very near future. With spectrum databases going into operation, we will soon see a major increase in white space devices in the market with actual usage of these bands in networks. Especially with the 2014 release of 802.11af, an IEEE WLAN standard for devices using TV white space database, and also 802.22, an IEEE WRAN standard for cognitive radio that uses a combination of spectrum database and local sensing, we have never been closer to a truly opportunistic spectrum access technology since this concept was incepted more than a decade ago. This major change will opens up many research fronts that can be regarded as future work to this dissertation and are presented below:

- Quality of service measurement and provision methods in white space networks must be investigated especially with the increase in use by migrating users from unlicensed bands. This research must consider the dynamic availability of white space channels due to irregular appearances of higher priority white space users such as wireless microphones which might be used in nearby events. Furthermore, it is imperative to analyze through measurements and theoretical models that how efficiently these bands can be used in urban environments with the imbalance of white space availability that exists in urban and suburban environments.

- With full duplex wireless now being a reality as discussed in chapter 5, there seems to be a benefit for spectrum sensing-based opportunistic spectrum access. The performance of CR networks are largely affected by large overheads imposed on them by the need for frequent spectrum sensing and to avoid PU collisions. However using full duplex, we can perceive CRs that perform spectrum sensing while they are transmitting. There are two benefits here: i) The need for spectrum sensing slots during which all CRs in the network are forced to be silent are removed and,
ii) CRs could immediately stop their transmission once they detect a PU leading to performance improvements by less collision with PUs. As future work, we propose revisiting existing CR MAC protocols and to analyze their performance in the presence of FD. This research must consider the additional error in spectrum sensing due to non-ideal self-interference cancellation in FD.
Publications


Bibliography


