Isameldin Mohammed Suliman

PERFORMANCE ANALYSIS OF COGNITIVE RADIO NETWORKS AND RADIO RESOURCE ALLOCATION
ISAMELDIN MOHAMMED SULIMAN

PERFORMANCE ANALYSIS OF COGNITIVE RADIO NETWORKS AND RADIO RESOURCE ALLOCATION

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 12 July 2016, at 12 noon

UNIVERSITY OF OULU, OULU 2016
Abstract

Cognitive radio (CR) is becoming a promising tool for solving the problem of the scarce radio resource and spectrum inefficiency. Spectrum sensing (signal detection) enables real-time detection of spectrum holes by unlicensed secondary users (SUs) in cognitive radio networks (CRNs). In this thesis, performance analysis of CRNs and radio resource allocation are considered. A continuous time Markov chain (CTMC) based analytical model taking into account all relevant elements as well as addressing the issue of the false alarm rate (FAR) associated with the continuous sensing is developed.

In some cases, the PU can be modeled as time-slotted with constant state (transmitting or not) in each slot. In this case, assuming SU can synchronize to the slots, its intuitive to use beginning of a slot for sensing and rest (possibly) for communication. For this model, M/D/1 priority queueing scheme has been applied in this thesis to find waiting time and queue length for PU and SU.

Multiple access among SUs in a time-slotted channel is considered next. A conventional method is e.g. using a channel access probability $\psi$ in each slot similar to the slotted ALOHA. A radically new idea is introduced in this thesis: why not increase the false alarm probability $PFA$ of each SU and use it as a multiple access method?

A game theoretic approach to radio resource allocation for the downlink capacity providing fair resource sharing among mobile nodes located along a multihop link is presented. Furthermore, the problem of resource allocations in heterogeneous wireless networks is also studied.

Finally, device-to-device (D2D) communication - with localized distribution, where users tend to gather around some areas (clusters/hot-spots) within the cell such as buildings is studied. Theoretical analysis with two dimensional clustering is presented including cases with correlated clusters. Correlation in cluster selection is shown to significantly improve performance.

Keywords: cognitive radios network, cooperative games, direct communication, false alarm rate, performance analysis, resource allocation, spectrum sensing
Tiivistelmä


Seuraavaksi käsitellään monikäyttöä SU:den joukossa aikajaoteltussa kanavassa. Tavanomainen menetelmä käyttää esimerkiksi kanavapyöristämön tiedostomäärää ψ kussakin aikavälissä vasta-ten aikajaotetusta ALOHA protokollasta. Tässä väitöskirjassa esitetään radikaali uusi idea: miksei lisätä väären hälytyksen tiedostomäärää ψ kussakin SU:ssa ja käytetä sitä moniliittymämenetelmänä?

Työssä esitetään peliteoreettinen lähestymistapa radioresurssien allokointiin siten että resursit jaetaan oikeudenmukaisesti monen yhteistoiminnan linkeissä. Lisäksi tutkitaan myös resursoinnin ongelmaa heterogenissä langattomissa verkoissa.

Lopuksi tutkitaan laitteiden välistä suoraa viestintää (D2D) paikallisena jaaman kanssa, jossa käyttäjillä on tapana kasaantua solun sisällä esim. rakennuksiin. Esitetään teoreettinen analyysi kaksiulotteisessa kluisteroinnilla myös korrelaatio ryhmien yhteydessä. Osoitetaan että korrelaatio yhteydessä parantaa merkittävästi suorituskykyä.

Asiasanat: kognitiiviset radioverkot, laitteiden välinen suora viestintä, resursoinni, spektrin nuuskuminen, tulosanalyysi, väärien hälytysten tiheys, yhteistoiminnalliset pelit
Dedicated to my family
Preface

The research presented in this thesis was carried out at the Centre for Wireless Communications (CWC), Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland. I would like to thank Professor Pentti Leppänen, the former head of the telecommunication laboratory, and Professor Matti Latva-Aho, the former director of CWC and the current leader of CWC - Radio Technologies (CWC-RT), and Professor Petri Mähönen, the former research director of CWC, for giving me the opportunity to join CWC. I am also grateful to Dr. Ian Oppermann, the former director of CWC for his support and encouragement.

I am extremely grateful to my supervisor Dr. Janne Lehtomäki for his endless support, friendship, guidance and encouragement during studies, it was invaluable. His pursuit for the highest standards in research is a great source of inspiration for me. Without his ideas, advising and criticism, this thesis would not have been completed. I wish to thank Professor Kenta Umbayashi for his guidance, encouragement and unreserved help. I am grateful to the reviewers of the thesis, Professor Riku Jäntti from the Aalto University, Helsinki, Finland and Professor Ferran Casadevall from the Universitat Politecnica de Catalunya, Barcelona, Spain. Their comments have significantly improved the clarity of the thesis.

Fortunately, I was accompanied by an extended team of experts and colleagues at the CWC always willing to supervise, help, and motivate me. For this, I would like to kindly thank them all. I am also very grateful to the administrative staff from the CWC, more specifically Antero Kangas, Elina Komminaho, Hanna Saarela, Jari Sillanpää, Kirsi Ojutkangas, Mari Lehmikangas, Timo Åikäs and Eija Pajunen for their patience and kindness. I was fortunate to receive personal grants for doctoral studies from the following Finnish foundations: Nokia Foundation, the Tauno Tönning Foundation, KAUTE Foundation, Riitta ja Jorma Takasen säätö and and Ahti Pekkalan säätö. Their financial support is gratefully acknowledged.

Finally, I take this opportunity to express my warmest and deepest gratitude to my family and my wife for their love, support, patience, and encouragement.

May 28, 2016

Isameldin Mohammed Suliman
Abbreviations

ϖ  intra-cluster / inter-cluster communication factor
ψ  channel access probability
Θ  correlation factor
θ  probability of transmission in an empty slot
ζ  probability that the primary user queue is empty
χ2cdf  chi-square cumulative distribution function
χ2inv  inverse of the chi-square cumulative distribution function
B₁  direct communication call blocking probability
B₂  blocking probability of calls going through the BS
Nₖ  number of clusters
Rₖ  radius of cluster
ϑ  offered traffic
φ  the probability that two random points are within direct communication range
PU_{FTP}  primary forced termination probability
PU_{BP}  primary blocking probability
PU_{SP}  primary successful probability
SU_{FTP}  secondary forced termination probability
SU_{STP}  secondary self-termination probability
SU_{SSP}  secondary successful probability
SU_{SBP}  secondary blocking probability
η  probability matrix
D₁  primary user average packet delay
D_{max}  primary user maximum allowed average packet delay
M/D/1  Poisson arrival, deterministic service, and one service channel/server queueing model
M/G/1  Poisson arrival, generic service, and one service channel/server queueing model
I_{m-1}(.)  modified Bessel function
Γ(.-)  upper incomplete gamma function
λ  arrival rate
\( \lambda_{\text{FAR}} \) false alarm rate Poisson parameter
\( \lambda_1 \) primary arrival rate
\( \lambda_2 \) secondary arrival rate
\( \mu \) service rate
\( \mu_P \) primary service rate
\( \mu_S \) secondary service rate
\( Q_{m(...)} \) generalized \( m \)th Marcum Q-function
\( Q \) state transition rate matrix
\( \pi \) steady state probability vector
\( N_Q^1 \) average number of packets in primary user queue
\( N_Q^2 \) average number of packets in secondary user queue
\( P_D \) probability of detection
\( P_{FA} \) probability of false alarm
\( P_M \) probability of misdetection
\( T_{\text{TOL}} \) interference tolerance
\( W_Q^1 \) average primary user queueing time
\( W_Q^2 \) average secondary user queueing time
\( W^1 \) total time spent by primary packet in the system
\( W^2 \) total time spent by secondary packet in the system
\( T_D \) waiting time until the beginning of a slot
\( \rho^1 \) primary queue utilization factor
\( \rho^2 \) secondary queue utilization factor
\( \text{AWGN} \) additive white Gaussian noise
\( \alpha_{\text{SU}} \) secondary user residual interference distortion factor
\( \text{BS} \) base station
\( \text{CCC} \) common control channel
\( \text{CR} \) cognitive radio
\( \text{CRN} \) cognitive radio network
\( \text{CPC} \) cognitive pilot channel
\( \text{CTMC} \) continuous-time Markov chain
\( \text{CTMP} \) continuous-time Markov process
\( \text{DECT} \) digital enhanced cordless telecommunications
\( \text{DTMC} \) discrete-time Markov chain
\( \text{D2D} \) device-to-device
\( \text{ED} \) energy detection

12
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAR</td>
<td>false alarm rate</td>
</tr>
<tr>
<td>GB</td>
<td>guard band</td>
</tr>
<tr>
<td>ITU</td>
<td>international telecommunications union</td>
</tr>
<tr>
<td>LI</td>
<td>loop interference</td>
</tr>
<tr>
<td>m2m</td>
<td>mobile-to-mobile</td>
</tr>
<tr>
<td>MF</td>
<td>matched filter</td>
</tr>
<tr>
<td>MIMO</td>
<td>multiple-input and multiple-output</td>
</tr>
<tr>
<td>N</td>
<td>number of channels</td>
</tr>
<tr>
<td>NBS</td>
<td>Nash bargaining solution</td>
</tr>
<tr>
<td>OSA</td>
<td>opportunistic spectrum access</td>
</tr>
<tr>
<td>P2P</td>
<td>peer-to-peer</td>
</tr>
<tr>
<td>PHS</td>
<td>personal handy-phone system</td>
</tr>
<tr>
<td>PSD</td>
<td>power spectrum density</td>
</tr>
<tr>
<td>PU</td>
<td>primary user</td>
</tr>
<tr>
<td>QoS</td>
<td>quality-of-service</td>
</tr>
<tr>
<td>QP</td>
<td>quiet period</td>
</tr>
<tr>
<td>RAT</td>
<td>radio access technology</td>
</tr>
<tr>
<td>RSI</td>
<td>residual self-interference</td>
</tr>
<tr>
<td>SIS</td>
<td>self-interference suppression</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operating characteristic</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SS</td>
<td>spectrum sharing</td>
</tr>
<tr>
<td>SSA</td>
<td>smart spectrum access</td>
</tr>
<tr>
<td>SU</td>
<td>secondary user</td>
</tr>
<tr>
<td>T2T</td>
<td>terminal-to-terminal</td>
</tr>
<tr>
<td>TETRA</td>
<td>terrestrial trunked radio</td>
</tr>
<tr>
<td>TV</td>
<td>television</td>
</tr>
</tbody>
</table>
## Contents

Abstract

Tiivistelmä

Preface

Abbreviations

Contents

1 Introduction

1.1 Cognitive radio networks

1.2 Radio resource allocation and device-to-device communications in wireless networks

1.3 Aims and outline of the thesis

1.4 Author’s contribution to publications

2 Performance analysis of cognitive radio networks

2.1 Markov chain-based analysis

2.2 Related literature

2.3 System model

2.4 State transition diagram

2.5 Performance analysis

2.5.1 Balance equations

2.5.2 Recursive approach for transition rates

2.5.3 Performance metrics

2.6 Numerical and simulation results

2.7 Summary and discussions

3 The effect of false alarm rate on the performance of cognitive radio networks

3.1 Related work

3.2 System model

3.2.1 Primary user model

3.2.2 Secondary user model

3.3 Spectrum sensing model

3.3.1 Continuous spectrum sensing

3.3.2 Energy detector-based spectrum sensing
1 Introduction

The current spectrum allocation scheme consisting of fixed allocations is not efficient, as typically only small parts of the licensed spectrum are actually used at a certain time instant in a given location. In 1999, Joseph Mitola III in his landmark paper [1] defined cognitive radio (CR) technology that allows cognitive-enabled wireless terminals to opportunistically access and utilize unused spectrum bands, leading to improved spectrum efficiency [2]. The ITU has defined the CR as:

"A radio or system that senses and is aware of its operational environment and can dynamically and autonomously adjust its radio operating parameters accordingly."

CR [3–5] is becoming a promising tool for solving the problem of spectrum scarcity and underutilization. Since its introduction, there has been considerable research effort focusing on CR techniques. In CR networks (CRNs), secondary users (SUs) employ spectrum sensing (signal detection) [4, 6–8] to discover spectrum holes during the absence of licensed primary users (PUs) before attempting network access. Typically, when studying spectrum sensing, a relatively static situation is assumed. This means that channel utilization of PUs is changing slowly, e.g., in television (TV) bands. For scenarios where PU occupancy is almost static, database-based approach is also attractive. However, sensing-based approach is also applicable to situations where PU has non-continuous dynamic traffic with gaps between transmissions. The problem with sensing is that the primary user may have to face interference for a certain period of time when it rejoins the channel (which was free and CR started using it).

The goal of this thesis is to develop an analytical framework for analyzing the performance of CRNs, including sensing imperfections, and to study resource allocation in wireless networks. This thesis contributes towards a better understanding of CRNs, provides a new insight into their operation and can be used to develop more accurate CRN performance analysis models. The rest of this chapter is organized as follows: Section 1.1 provides a brief overview of CRNs, including spectrum sensing techniques, and describes performance evaluation models. In Section 1.2, radio resource allocation and device-to-device communication in wireless networks are presented. The aims and outline of the thesis are presented in Section 1.3. Finally, the author’s contributions to the original publications are summarized in Section 1.4.
1.1 Cognitive radio networks

A major demand for the coexistence of PUs and SUs in CRNs is the protection of the PUs. PUs have a license for their spectrum and can (and have the right to) normally operate (without any modification) regardless of the existence of a CRN. However, SUs are required to have some additional functionalities such as spectrum sensing in order to detect the presence of PUs to opportunistically access and operate in the network. For seamless PU operation, SUs should also have handoff capabilities to vacate the spectrum band when PUs return. The handoff capability provides a basic quality-of-service (QoS) guarantee to ongoing SU calls. There are two main approaches for enabling licensed PUs and unlicensed SUs to coexist and share radio resources in CRNs: spectrum sharing (SS) and opportunistic spectrum access (OSA) [9]. In the SS model, SUs are allowed to transmit simultaneously with PUs on the same band. However, SUs are required to adaptively control/adjust their transmission powers to limit the interference level caused to PUs. On the other hand, the OSA approach only allows SUs to access the licensed channels opportunistically when PUs are not present. In both cases, SUs are required to ensure PUs’ protection by not causing harmful interference to their communication.

In OSA with spectrum sensing, CRs use signal detection techniques to decide if a particular channel is currently used or not. If the channel is free, CRs may be allowed to transmit. If the primary user joins the channel again, the CR must leave the channel within a certain period of time. The problem with this is that the PU may have to face interference for a certain period of time when it rejoins the channel. A significant issue is the accuracy of the spectrum sensing outcome used for detecting the presence and absence of PUs. The most widely used sensing method used for detecting the presence of primary users is the energy detector (ED) [10–12]. It is simple to implement and does not require very long sensing durations. The detection is based on calculating the energy of the received samples which is compared to a threshold. If the threshold is exceeded, it is decided that a signal(s) is (are) present. Other detection methods include matched filter and cyclostationary [13] based detection. They require either more information compared to ED or longer detection durations (but can have higher detection performance). General studies on the detection methods can be found in [14].

No matter what spectrum sensing method is used, the spectrum sensing outcomes cannot be perfect and there will be some sensing errors such as misdetections and false alarms, which degrade the QoS of PUs and SUs.
The probability of false alarm ($P_{FA}$) refers to the probability of falsely classifying an unoccupied channel as occupied. False alarms result from the effect of internal receiver noise and/or external interference. The probability of misdetection ($P_{M}$) refers to the probability of incorrectly classifying an occupied channel as unoccupied. The probability of (correct) detection is $P_D = 1 - P_M$. $P_{FA}$ reduces spectrum utilization by SUs employing spectrum sensing while $P_M$ leads to SUs colliding with the PUs. Imperfect sensing significantly affects the operation of CRNs and therefore should be considered. Imperfect sensing refers to the sensing process where $P_{FA}$ and/or $P_M$ is (are) greater than zero, i.e., sometimes the sensing process results in a wrong decision about the state of the channel. The work reported [15] addresses the issue of secondary users sensing robustness in detection of the presence/absence of PU signals in uncertain low SNR environments. The authors argue that every detection algorithm suffers from similar robustness limitations in low SNR environments. It has been shown that in such an environment, the impact of modeling uncertainties can be quantified by the position of an SNR wall below which a detector will fail to be robust, no matter how long it can observe the channel. The authors finally propose simple mathematical models for the uncertainty in the noise and fading processes.

To facilitate and regulate cognitive radio operation, FCC and ECC set spectrum access rules intended to protect primary users such as TV viewers from SU interference. Applying the regulation rules prevents unnecessary interference. However, as the number of SU's grows, their aggregate interference could harm the PU. FCC allows secondary access based on geolocation. So the SUs query a TV white space database with their location. The database (with information from the broadcasters) uses some pessimistic (to protect PUs) propagation model and evaluates whether or not SU access is feasible and in which channels. As reviewers suggested, some kind of feedback (or sensing results) would be useful to support the database so it could use less pessimistic models but still protect PU receivers. One of the limitations of the FCC regulation is that the effect of aggregate interference was not fully addressed (although the protection of PUs/QoS guarantee is the responsibility of the individual/single SUs). The authors of [16] applied the FCC and ECC rules for TV-band secondary use and computed how much white space is available in Finland. They found that the current secondary spectrum usage rules do not protect TV receivers sufficiently well. Even if each individual secondary transmitter obeys them, together their aggregate interference can still disrupt the TV reception. The authors of [17] discussed the issue of capacity considerations for cognitive networks in TV white space (TVWS) and showed that multiple challenges in...
realizing utilization of the WS include system parameters of existing PUs, propagation characteristics of local terrain as well as FCC rules for the protection of PUs.

The concept of licensed spectrum access was introduced to facilitate the introduction of new applications in a frequency band while maintaining incumbent services in the band. The European radio spectrum policy group (RSPG) defines licensed spectrum access (LSA) as a complementary spectrum tool to fit under individual licensing regimes in Europe and its implementation is the responsibility of the local national regulatory authority (NRA) [18, 19]. The authors of [20] present a practical online data assimilation method for estimating the spatial correlation of the time variant primary field strength from a collection of sensing samples. The reported sensing method is used for assigning geolocation databases and offers an efficient real-time solution for state estimation in the future geolocation databases.

In this thesis, we use the sensing-based approach for detecting the PU channel occupancy. The sensing-based approach has been extensively studied in the literature, see for example in [21–23]. We would like to point out that in some cases the information about channels occupied by the PUs could be provided to the SUs. For example, there has been significant interest in the database approach where the SUs are provided information about the channels occupied by the PUs. This approach is especially suitable for scenarios where the PU occupancy status is not changing very rapidly (such as in the TV white space). The sensing-based approach utilized in this thesis is more suitable for scenarios where the PUs’ status can change rapidly and/or the PUs are not willing to maintain a database of their current and future channel utilization. For more reliable or quick detection, SUs can perform spectrum sensing in combination with a database containing information about PUs (to assist sensing). This kind of approach can be called Smart Spectrum Access (SSA) [24].

One more solution is to exploit the functionalities of the cognitive pilot channel (CPC) mechanism which allows SUs to receive related/necessary network information such as available frequency bands, spectrum occupancy, and radio access technologies (RATs), etc. [25]. Moreover, the retrieved network information can be combined with locally acquired information (e.g., by spectrum sensing) to improve the spectrum discovery procedure. Continuous-time Markov chain (CTMC) based modeling has been widely used in the literature to evaluate the performance of CRNs. It has been widely used due to its accurate modelling, as well as easy computation characteristics. The CTMC model has proved particularly effective in modelling CRNs performance. However, the performance of CRNs under a real network environment is not well
investigated. Furthermore, optimizing spectrum detection parameters in CRNs is important in CR networks.

Although the opportunistic spectrum access considered in this thesis is based on spectrum detection, other key requirements also need to be taken into account to guarantee a reliable use of secondary access. One of these factors is the reliability and robustness of the spectrum detection. For instance in [26], the authors mentioned that secondary spectrum access cannot be reliably done based on spectrum sensing due to the impact of the aggregate interference. The authors argue that the mere fact that no transmissions can be detected in a certain block of spectrum in a certain area is not a sufficient condition for commercial cognitive spectrum exploitation. We model the interference due to imperfect suppression of the secondary user signals with the residual interference distortion factor. It should be noted that we not only use full-duplex cancellation techniques but also have different bands for sensing and SU transmissions (with guard band between them). Thus, in our model the residual SU’s interference is less of a problem than usual. Another factor is the impact of the aggregate interference caused by multiple simultaneously transmitting secondary users. This aggregate interference is dealt with in [27] where the authors show how to compute it in a shadow fading environment.

In practical systems, when secondary use of a spectrum is more common, a primary user could report to a database when the interference level it receives exceeds some threshold. Then some of the secondary links operating in the nearby area have to be turned off (by commands from the database). Alternatively, the allowed operating distance (for secondary use) from the PU location can be increased (reducing interference). To improve the analytical models presented in this thesis, the modelling of the aggregate interference is recommended for implementation in future works, while cooperative sensing can be used to improve the sensing performance.

1.2 Radio resource allocation and device-to-device communications in wireless networks

Radio resource allocation in wireless networks is a major research topic and efficient resource allocation of resources available to a network (obtained with sensing or not) will help in utilizing the spectrum more effectively. For example, device-to-device
D2D communication has been proposed to improve spectrum allocation in wireless network since it aims to avoid the passage of data through a central controller in the case that two nearby terminals want to communicate with each other [28]. Recent years have witnessed active research in realization of D2D. One important question for D2D is how to locate the spectrum for it. Either use a licensed spectrum and have orthogonal allocation (and lose the spectrum again) or have simultaneous transmissions for D2D and cellular communication (leading to interference). One promising approach toward solving these issues is to use CR techniques to locate free spectrum bands for D2D links. By employing CR techniques, mobile terminals can opportunistically utilize free spectrum bands and thereby save base station (BS) radio resources.

Wireless networks are expected to support a wide range of multimedia applications with different traffic characteristics and QoS requirements due to the increase in the demand for high bandwidth. These multimedia-centric applications will cause significant changes in the services, usage and traffic characteristics of the cellular networks. It can also be foreseen that D2D communications will be embedded as part of the cellular networks. D2D communications within a cellular architecture while still using a single air interface can be considered an augmentation toward more flexible network architecture than the hierarchical one. It can also be considered a requirement for some important applications like co-operative communications in both up- and downlink directions. D2D is considered to be among the alternatives for improving the performance of cellular systems due to its several advantages. To evaluate the performance of D2D communications, many parameters like the effect of resource reuse, clustering and direct communication range have to be taken into consideration.

1.3 Aims and outline of the thesis

The objective of this thesis is to develop a new performance analysis of CRNs and evaluate their operations with a performance evaluation model that include all relevant operation parameters in a single model. To gain further insight, the effect of false alarm rates on the performance of CRNs will also be investigated. Additionally, a queueing theoretical analysis of opportunistic access in CRNs was presented. Moreover, we also examine spectrum detection parameters optimization in CRNs. Another objective is to address the issue of radio resource allocation. Finally, D2D is studied using tools from the teletraffic theory.
Chapter 2, based on [29, 30], presents a CTMC-based model for analyzing the performance of CRNs. The proposed analytical model incorporates full state-dependent transition rates, multi-channel support, handoff capability and imperfect sensing by SUs looking for free channels in a single model. Formulas for primary termination probability, secondary success probability, secondary blocking probability, secondary forced termination probability and radio resource utilization have been derived. Those performance metrics are used to evaluate the performance of the CRNs. It is shown that incorporating fully state-dependent transition rates in the CTMC can significantly improve analysis accuracy and hence achieve more accurate analytical model. The results from extensive simulations confirm the validity of the proposed model.

Chapter 3 includes the results published in [31], which extends [29, 30] by addressing the issue of the sensing errors by already transmitting SUs that check for returning PUs (the model in the previous chapter considered imperfect sensing only for incoming SUs). In order to appropriately protect returning licensed PUs, SUs should continuously perform spectrum sensing during their ongoing transmissions. The FAR is defined as the average number of false alarms per unit of time during this continuous sensing and can be modeled by a Poisson process with Poisson parameter $\lambda_{FAR}$. The analytical model also includes interference tolerance among PUs and SUs as well as the impact of SUs residual self-interference (due to simultaneous sensing and transmission). The results show that high $\lambda_{FAR}$ can severely degrade PUs performance and reduce the overall system resource utilization. However, with increasing PU interference tolerance, PUs’ performance improves as well. Residual interference in SUs was found to decrease the detection probability, resulting in a reduced PU performance. Again, extensive simulations validate the analytical model.

Chapter 4 is based on [32, 33]. The research in [32] introduces a queueing-based analysis of opportunistic access. A time-slotted system is assumed, so that a SU can perform spectrum sensing at the beginning of each slot. M/D/1 priority queueing scheme is applied to evaluate the performance of the CRNs by finding waiting time and queue length. Simulations are used to validate the theoretical results. Results indicate that the performance of the SU depends on the data traffic characteristics of the PU, and under a high PU arrival rate the average waiting time and average queueing length of the SU grow especially when the combined arrival rate approaches the queue utilization factor. In [33], optimization of the detection parameters in time-slotted cognitive radios is presented. The study focuses on finding the optimum false alarm and
detection probabilities and proposes a radically new idea of handling random access with sufficiently high false alarm probabilities.

Chapter 5, based on [34, 35], considers radio resource allocation. In [34], a game-theoretic approach to radio resource allocation for the downlink is introduced. It examined the importance of fair resource sharing among mobile nodes located along a multihop link and described a novel technique for providing a resource allocation in a multihop relaying network. The resource allocation problem is formulated as a cooperative game using the Nash Bargaining Solution (NBS), which allows mobile nodes to fairly share a downlink bandwidth. Sharing the downlink capacity between multiple nodes using a noncooperative approach is inefficient when the radio resource is scarce. If upstream nodes manipulate their location at the head of the multihop link to exploit the downlink capacity, downstream nodes will suffer disproportionately. The undesirable properties can be avoided by means of a cooperative agreement in which all nodes share the radio resources equally, where downstream nodes are allowed to pay compensation to prevent upstream nodes from exploiting the downlink capacity while encouraging them to cooperate. In [35], the problem of resource allocations in wireless networks is also investigated. Wireless networks may cooperate because they cannot deal with resource allocation demands alone or because they can reduce their cost by working together. We formulate the problem as an optimization problem that allocates resources to mobile users and maximizes benefits for networks. In addition, we employ an exponential cost function to balance the load across multiple heterogeneous networks. A resource allocation model made of three wireless networks and a user is considered. Cooperative game-theory concepts are used to identify stable resource allocations, under which all three networks find it beneficial to cooperate. The results suggest that an allocation that maximizes revenues will make the cooperation attractive to all networks.

Chapter 6 focuses on device-to-device communications and is based on [36, 37]. In [36], direct communication with localized distribution, where users tend to gather around some areas (clusters/hotspots) within the cell such as buildings, is studied. Theoretical analysis of direct communication with two-dimensional clustering is presented. Additionally, analysis for direct communication with correlated clusters is presented. With correlated clusters some pairs of source and destination clusters are more probable than other pairs. Simulations confirm the validity of the analysis. In addition to the exact results, we also suggest using the point-based approximation to rapidly and easily obtain results. The numerical results show that the gains from direct communication, in terms of blocking probability and carried traffic, depend on the
offered traffic. Additionally, correlation in cluster selection is shown to significantly improve performance. Point-based approximation is shown to be very useful when the number of clusters is large. In [37], performance of D2D communications with resource reuse is evaluated by simulations. Many parameters like the effect of resource reuse clustering, and direct communication range are taken into consideration. Performance metrics such as blocking probabilities and carried traffic were used.

Finally, Chapter 7 provides the conclusions and directions for future work.

1.4 Author’s contribution to publications

The results presented in this thesis have been published in nine original publications [29–37]. Three papers [30, 31, 36] are published in peer-reviewed journals and six papers [29, 32–35, 37] are published in international conferences. Each of these papers focused on a particular topic on CRNs and resource allocations. The thesis author played an active role developing the ideas and writing all the papers with the following exception. The result published in [36] has been developed by the first author, the author of this thesis was actively involved in the discussion and provided ideas and comments. In all other papers the author of the thesis assumed the main responsibility in performing and developing the simulations software, carrying out the analysis and producing the numerical results. The co-authors contributed to the papers by providing ideas, comments and criticism.
2 Performance analysis of cognitive radio networks

In this chapter, we present a CTMC-based model for analyzing the performance of CRNs. A major limitation of the available literature is that all key performance evaluation factors, such as the effect of imperfect sensing for CRs looking for a free channel to transmit on and state-dependent transition rates, are not modeled in a single work. The proposed model differs from the existing models by accurately incorporating key elements such as full state-dependent transition rates, multi-channel support, handoff capability and imperfect sensing for incoming SUs. We derive formulas for primary termination probability, secondary success probability, secondary blocking probability, secondary forced termination probability, and radio resource utilization. The results show that incorporating fully state-dependent transition rates in the CTMC can significantly improve analysis accuracy, thus achieving a more accurate analytical model. The results from extensive Monte Carlo simulations confirm the validity of the proposed model. The results in this chapter are valid for any detector, as long as equations for detection probability $P_D$ and false alarm probability $P_{FA}$ are available for that detector type. Although we apply the widely used Poisson models for network traffic, it should be kept in mind that measurements have shown that they can lead to poor fitting to several PU types. In the next chapter, we will specifically consider the energy detector and have more detailed results for that detector type, including the effects of residual interference in simultaneously transmitting and sensing for SUs with on-going transmission.

The rest of the chapter is organized as follows: Section 2.1 describes Markov chain-based modeling approach. The related literature is presented in Section 2.2. The assumptions and system model are described in Section 2.3. Section 2.4 presents and describes the state transition diagram of the Markov chain. The detailed derivations of the performance analysis is presented in Section 2.5. The numerical and simulation results are presented in Section 2.6. Finally, Section 2.7 concludes this chapter.

2.1 Markov chain-based analysis

Markov chain-based modeling has been widely used in the performance evaluation of CRNs, e.g., [38–44]. Modeling can be based on either CTMC or discrete-time Markov
chains (DTMC). Typically, with a CTMC based-modeling it is assumed that SUs observe the system state continuously and can detect randomly arriving PUs. On the contrary, in DTMC models SUs perform the sensing periodically, relying on discrete time instants to observe the system state. Therefore they cannot instantly detect PUs arriving between sensing instants. Traditionally, these models assume perfect sensing and ignore $P_M$ and $P_{FA}$. Although it is easier to form a model with perfect sensing, this casts serious doubts about the accuracy of such a model.

Other important elements in the CTMC model include full state-dependent transition rates, spectrum handoff and multiple channels. Full state dependency means that the transition rate from one state to another depends fully on the current state by taking into account the number of PUs and the number of SUs in the channel searching process. Because of the imperfect sensing, the number of PUs and SUs will affect the channel searching results. By considering full state-dependent transition rates in CTMC, we will be able to accurately model CRNs by using the exact transition rate from each state. In this chapter, we develop a CTMC-based modeling of multi-channel multi-user CRNs with imperfect spectrum sensing. The main contributions of this chapter are summarized as follows:

1. We combine full state-dependent transition rates, imperfect sensing, multiple channels as well as handoff capability factors in a single unified CTMC-based analytical framework.

2. We derive the exact state-dependent transition rates by recursion for the CTMC model for an arbitrary number of channels. The transition rates are found by going through all the possible search sequences and taking into account the channel state that indicates whether it is occupied by a PU/SU or it is free. The state-dependent behavior of the transition rates is very important for modeling CRN with higher accuracy. Other researchers [42] have already considered similar issues but the effect of state-dependent transition rates has not been fully modeled. With perfect sensing parameters ($P_{FA} = P_M = 0$), the proposed model is reduced to one of the models presented in [45].

3. We present extensive evaluation (theory and Monte Carlo simulation) to find the effects of various parameters on the utilized performance metrics. We find that the proper selection of the secondary receivers’ operating point is critical. Additionally, the results from extensive Monte Carlo simulations confirm the validity of the proposed analytical model.
2.2 Related literature

In [45], authors analyzed the performance of opportunistic spectrum access in a multi-channel network using CTMC. Their results suggest that more secondary traffic can be supported when spectrum access is coordinated between primary and secondary systems. However, perfect spectrum sensing was assumed therein and the handoff mechanism was not supported. In [42], a multichannel system was assumed and the system state was reduced to two parameters: the number of channels occupied by PUs and the number of channels occupied by SUs. Although the authors in [42] have included sensing errors in their model, they did not consider fully state-dependent transition rates. Apart from [42], attention paid to the imperfect sensing process within a CTMC model has been relatively limited.

Imperfect sensing induces a well-known tradeoff between the $P_{FA}$ and $P_M$. Such tradeoff has been investigated in [46]. Therein, a multi-channel DTMC-based analytical framework was presented. The model was used for determining the sensing operating points by adjusting $P_{FA}$ and $P_M$ according to a given traffic load of both PUs and SUs. Primary-secondary spectrum sharing in CRNs has been addressed in [47] using a DTMC model. Their results suggest that sensing periodicity and sensing accuracy are key parameters that influence the performance of spectrum sharing. They also highlighted the importance of time-sharing between spectrum sensing and data transmission. A CTMC spectrum access analysis with channel reservation and handoff capability was reported in [48]. An accurate model based on [48] was reported in [49] where the expressions of forced termination and blocking probabilities were corrected and in [50] where correct expressions of the throughput of SUs and forced termination probability were provided.

In [51], a CTMC multi-dimensional model with handoff capability was presented where buffering has been used to store SU requests, and therefore reducing SU blocking probability and non-completion probability. In [38], performance (QoS) in a CRN involving PUs and SUs was analyzed by using the three-dimensional Markov chain. The authors proposed a reservation method to limit the forced termination probability of SUs that occurs because of the arrival of PUs with preemption priority to channels occupied by SUs. The analysis reported in [38] was refined by Kondareddy [43] to take into account the spectrum handoff capability by SUs. A similar approach was used by [52] to analyze the performance of a CRN with spectrum handoff capability. Therein, the authors derive several performance parameters such as forced termination
probability. Table 1 summarizes the literature on some key CTMC-based analytical models, providing a comparison among them. In the table, the "✓" symbol indicates that the analytical model supports the corresponding feature while the "✗" symbol indicates that the analytical model does not support the corresponding feature.

Table 1. Comparison among various CTMC-based analytical models (©2013 IEICE).

<table>
<thead>
<tr>
<th>Reference No.</th>
<th>Handoff support</th>
<th>Multi-channel support</th>
<th>Full-state dependent</th>
<th>Imperfect sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>[38]</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>[39]</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>[42]</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>[43, 48–51]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>[45]</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Proposed model</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

2.3 System model

The number of channels is denoted with $N$. These channels are shared between PUs and SUs, with PUs having priority over the SUs. These two types of users have distinct arrival and service rates that are modeled by a Poisson process. Connections of the PUs arrive with rate $\lambda_1$ and connections of SUs arrive with rate $\lambda_2$. The corresponding service rates are $\mu_1$ and $\mu_2$. PUs have free access to all channels not occupied by other PUs with no restriction (they do not consider SUs). A new PU connection will only be blocked if all channels are already occupied by other PUs. On the other hand, SUs have free access to the channels as long as they are not occupied by PUs or SUs. A new SU connection will be blocked if all channels are already occupied by SUs or classified (correctly or incorrectly) as occupied by a PU. In order to support imperfect sensing, handoff and multiple channels, we assume that:

1. A free channel is wrongly decided to be occupied with probability $P_{FA}$ and an occupied channel is determined to be free with probability $P_M = 1 - P_D$.
2. After detection of an incoming PU, the SU starts to search for a new free channel using some search order and imperfect sensing. Imperfect sensing is also assumed for newly arrived SUs searching for a free channel.
3. The search order for new free channels is random (similar to the random search in [53]). The search stops after a free channel is found or all channels are found to be occupied. In practice, it might be allowed to search all the channels a number of times before giving up.

For analytical simplicity, we also assume that:

1. SUs with ongoing calls always detect the arrival of a PU and $P_{FA}$ during an ongoing call is assumed negligible. Similar assumptions have been made previously in [42]. In the next chapter, we specifically consider ED instead of the generic detection model assumed in this chapter, and this assumption is not made again.

2. All state transitions are instantaneous, i.e., we assume a negligible time required for sensing and channel hopping. For secondary users, the transition time involves several parameters such as sensing time, frequency hopping time, spectrum handoff, etc., and depends on the number of available channels as well as the channel searching sequence. Some of these components have been commonly assumed to be small. For instance, in [21] the authors assumed that the sensing time is small relative to the busy and idle periods. The authors of [54] assumed that the sensing time is very small compared to the channel ON-OFF durations and they assumed that the channel state does not change during the sensing interval. Furthermore, in [48] the authors reported that preempted CR calls will be moved immediately to idle sub-bands elsewhere. We measure performance by the utilized metrics such as secondary forced termination probability. Even if the transition time would not be small, we expect a rather limited effect on the utilized performance metrics since most of them are defined for connections that are already accepted.

3. An SU is aware of the channels occupied by other SUs and will not use them. An SU can know about the presence of other SUs by using some common control channel (CCC) for informing other SUs about their occupied channels. Even if there is no CCC available, the SUs can (for example) utilize a pre-defined waveform with sufficient repetition intervals as part of their transmission. The purpose of this pre-defined waveform would be to enable powerful detection techniques (such as matched filters) so that we may consider an almost perfect detection of channel occupied by other SUs.

4. A PU knows the channels occupied by other PUs so that there will be no collisions between PUs [42]. The channel occupancy information is available to PUs e.g., from a PU core network.
5. In case of collisions (between PU and SU), both colliding users withdraw from the system. There are several possible collision models for modeling the effects of collisions between PU and SU. For example, in [42], due to SU misdetection, two collisions models were described. In the first scenario both SU and PU drop the channel because of the collision event. In the second scenario the PU terminates and the SU could continue its communication. These have been called collision types I and II respectively. Collision type I corresponds to collision model used in this chapter. Collision models including the concept of interference tolerance are presented in the next Chapter.

We use a two-dimensional Markov chain to model the CRN system. The system states are given by two-tuples \((i, j)\) where: \(i\) is the number of channels used for PUs’ connections and \(j\) is the number of channels used for SUs’ connections. For example state \((1, 2)\) refers to a state with one PU and two SUs. The total number of channels occupied by PUs and SUs cannot exceed \(N\). Therefore, we have the following restrictions:

\[0 \leq i \leq N, \ 0 \leq j \leq N, \ 0 \leq i + j \leq N.\]

Let \(Q\) denote the state transition rate matrix (also known as infinitesimal generator) of the CTMC. Let \(\pi\) denote the steady state probability vector with \(\pi_{i,j}\) denoting the probability that the system is in the steady state \((i, j)\), which can be interpreted as the proportion of time that the system spends in state \((i, j)\).

### 2.4 State transition diagram

By utilizing the assumptions listed previously, we find that from a given state there are a maximum of five state transition types. They are:

- **Transition Type 1**: \((i, j) \rightarrow (i + 1, j - 1)\). The number of PUs is increased by one and the number of SUs is decreased by one. This occurs if the existing SU did not find a free channel after a new PU forced it to move.
- **Transition Type 2**: \((i, j) \rightarrow (i, j - 1)\). The number of SUs is decreased by one. This transition happens in two cases. The first case is when an existing SU leaves the network because of service completion. The second case is when a new PU forces an existing SU to move, the SU collided with another PU.
- **Transition Type 3**: \((i, j) \rightarrow (i + 1, j)\). The number of PUs is increased by one. This transition happens in two cases. The first case is when a new PU arrives on a free
channel. The second case is when a new PU arrives on a channel occupied by a SU, the new PU forced the exiting SU to move, the SU found a free channel.

- Transition Type 4: \((i, j) \rightarrow (i - 1, j)\). The number of PUs is decreased by one. This transition happens in two cases. The first case is when an existing SU leaves the network because of service completion. The second case is, if after a newly arriving SU ends up after channel searching on a channel already occupied by PU.

- Transition Type 5: \((i, j) \rightarrow (i, j + 1)\). The number of SU is increased by one. This occurs if a new SU ends up after search on a free channel.

Fig. 1 illustrates the channel occupancy and state transitions over time by showing primary and secondary arrivals and departures, as well as handoff and collision events in a three-channel network. For example, let us consider the transition from state \((1, 1)\) to state \((1, 0)\), i.e., Transition Type 2. There are two possibilities for this transition. In the first possibility, the existing SU call is completed with the service rate \(\mu_2\) so that the number of SUs is reduced by one. In the second possibility, a new PU comes to the existing SU’s channel, forcing it to search for a free channel. If the existing SU ends up colliding with the existing PU then both PU and SU calls will be terminated, thus leaving only the new PU leading to the state \((1, 0)\). Because PUs are aware of other PUs, the new PU has two possible channels to use, one which is occupied by the SU and the other which is totally free. Therefore, the new PU comes to the SU channel with probability 0.5. Then the SU goes straight to the existing PU channel with probability 0.5 and the collision happens with probability \(P_M\). Another example is the transition from state \((2, 1)\) to state \((3, 0)\), i.e., it is Transition Type 1. In this case the new PU always forces the SU to leave its current channel and search for a new free channel. The SU correctly detects that the other two channels are occupied by PUs with probability \(P_D^2\) and the SU call is terminated. Thus the transition rate is \(T_{(2,1)}^{(3,0)} = \lambda_1 P_D^2\).
Fig. 2 shows the state transition diagram for a three-channel network. It shows the transition rates derived above. The other transition rates have also been derived by considering all the possible sequences of channels and detection events. It should be noted that transition from state \((i, j)\) to state \((i + 1, j + 1)\) is not possible because of the continuous nature of the CTMC. By the definition of the Poisson process, two or more arrivals cannot occur at the same time [55]. Therefore only one PU or one SU at a time can arrive with transition from \((i, j)\) to \((i + 1, j)\) or from \((i, j)\) to \((i, j + 1)\) respectively. The transition from state \((i, j)\) to state \((i - 1, j - 1)\), occurs if, due to a false alarm an existing SU leaves its current channel and collides with a PU user on another channel. However, the effect of false alarms from the continuous sensing during SU communication has been left for investigation in the next chapter.
2.5 Performance analysis

In this section, we present a detailed performance analysis of a CRN using the corresponding state transition diagram.
2.5.1 Balance equations

The balance equations can be written by considering the transition rates using the rule that input must equal output for each state [56]. Additionally,\[ N \sum_{i=0}^{N} \pi(i,j) = 0 \]
where \( \pi(i,j) \) denotes the steady state probability that the system is in state \((i,j)\). For example, for the simple case of the state \((0,0)\) the balance equation results in
\[ \pi(0,0) = \pi(1,0) \left( \mu_1 + \lambda_1 P_M \right) + \pi(0,1) \mu_2 \]
For the case of state \((1,2)\) we have the following balance equation
\[ \pi(1,2) = \pi(0,3) \left( \frac{\lambda_1 (1 + 2(1 - P_{FA})) + \pi(1,1)}{(1 + P_M + \mu_1 + \lambda_2 P_M)} \right) \]
The resulting set of simultaneous linear equations i.e., balance equations for every state and the normalization constraint, can be easily solved using MATLAB, for example.

2.5.2 Recursive approach for transition rates

We now generalize the results shown in Fig. 2 and derive transition rates for a Markov chain with an arbitrary number of channels \( N \).

Transition Type 5: Let us consider the case where the number of SUs is increased by one. It is important to note that the transition rate for this case depends only on two factors: the number of PUs and the number of free channels. In this case, channels used by other SUs are avoided. We get the recursion
\[ f(i,k) = \frac{k}{i+k} (1 - P_{FA}) + \frac{k}{i+k} P_{FA} f(i,k-1) + \frac{i}{i+k} \]
where \( i \) is the number of PUs and \( k \) is the number of free channels not occupied by PUs or SUs. Function \( f(\cdot) \) denotes the function that increases the number of SUs by 1. The transition rate is found by multiplying \( f(i,k) \) with \( \lambda_2 \), i.e., \( T^{(i,j)}_{(i,j+1)} = \lambda_2 f(i,k) \), where \( i \) and \( k \) are as defined above. For boundary values we get by induction
\[ \begin{cases} f(0,k) = 1 - P_{FA}^k \\ f(i,1) = \frac{1}{\lambda_2} f(1 - P_{FA}) \sum_{n=0}^{i} P_{D}^n \end{cases} \]
The first term in (4) refers to a situation where we randomly select a free channel (one out of \( i + k \) possibilities) and a \( P_{FA} \) does not occur so the search terminates successfully. The second term refers to the same situation except that we have a \( P_{FA} \) and thus the search continues with the number of free channels reduced by one. The third term refers to a situation where we randomly select a channel occupied by a PU and we correctly detect it as occupied so that the search continues with the number of PUs reduced by one. The detected channel is marked as occupied so that it will not be revisited during this search.

**Transition Type 4:** Let us next consider the case where the number of PUs is reduced by one. We get the recursion for the multiplying factor

\[
g(i,k) = \frac{i}{i+k} P_M + \frac{i}{i+k} P_{DG}(i-1,k) + \frac{k}{i+k} P_{FA}(i,k - 1)
\]

(6)

and the resulting transition rate is

\[
T_{(i-1,j)}^{(i,j)} = i \mu_1 + \lambda_2 g(i,k).
\]

Function \( g(.) \) denotes the function that decreases the number of PUs by 1.

**Transition Type 3:** For the case where the number of PUs is increased by one we directly get that the transition rate is

\[
T_{(i+1,j)}^{(i,j)} = \lambda_1 \frac{N-i-j}{N-i} + \lambda_1 \frac{j}{N-i} f(i,N-i-j).
\]

(7)

**Transition Type 2:** For the case where the number of SUs is decreased by one we get the transition rate as

\[
T_{(i,j-1)}^{(i,j)} = j \mu_2 + \lambda_1 \frac{j}{N-i} g(i,N-i-j).
\]

(8)

**Transition Type 1:** Finally, for the case where the number of PUs is increased by one and the number of SUs is decreased by one we get the transition rate as

\[
T_{(i+1,j-1)}^{(i,j)} = \lambda_1 \left[ \frac{j}{N-i} \left[ 1 - f(i,N-i-j) - g(i,N-i-j) \right] \right].
\]

(9)

### 2.5.3 Performance metrics

Based on the state diagram, a variety of performance metrics can be obtained to evaluate the performance of the CRN. We begin by outlining the different performance metrics of interest, notably PU termination probability, normal SU termination probability (i.e., secondary successful probability), forced SU termination probability as well as system resource utilization. All performance metrics can be directly derived from the state transition diagram.
Primary user forced termination probability: $PU_{FTP}$

We use the term primary termination probability ($PU_{FTP}$) to refer to the probability that a PU call, which has not been blocked initially, is terminated due to collisions with SUs because of misdetections. The probability that a PU call is terminated due to collisions with SUs can be found by going through the states of the state transition diagram. There are three cases in which the PU calls can be terminated. First, when a new SU arrives at a channel occupied by a PU and misdetects the presence of the PU ending up colliding with it. The second case for collision is when a PU arrives at a channel occupied by an SU, the SU has to leave the channel and search for a new free channel. However, the SU ends up colliding with another PU instead. Finally, there is the case where an incoming SU detects some existing PUs but misdetects one of them. Combining these three cases, the $PU_{FTP}$ can be derived as

$$PU_{FTP} = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-i} \pi_{i,j} \left(T_{(i,j)}^{(i-1,j)} - i\mu_1\right) + \sum_{i=0}^{N-1} \sum_{j=1}^{N-i} \pi_{i,j} \left(T_{(i,j)}^{(i-1,j-1)} - j\mu_2\right)}{A_1 \left(1 - \pi_{(N,0)}\right)},$$

(10)

where $\pi_{i,j}$ is the state probability of state $(i,j)$. It should be noted that PUs will not arrive when the system is in state $(N,0)$ since they are aware of other existing PUs.

Secondary blocking probability: $SU_{SBP}$

The secondary blocking probability ($SU_{SBP}$) is the probability that a newly arriving SU call cannot be accepted due to insufficient radio resources or inability of the SU to find a free channel with a certain $P_{FA}$. The SU’s connection requests can be blocked for several reasons depending on the state of the system; for example when all channels are occupied by PUs, secondary call blocking could happen because the SU searches all channels and detects the presence of all existing PUs. In this case the call will be completely lost. In the case where all channels are occupied by SUs, an incoming SU connection request will be immediately blocked. An SU call can also be blocked even though all channels are free if the SU could not classify any one of them as being free due to false alarms. The $SU_{SBP}$ is given in (11) and can be interpreted as the percentage of blocked SU calls over all SU arrivals.
The secondary forced termination probability ($S_{SUFT}$) is the probability of dropping an active SU call due to the arrival of a PU to a channel occupied by an SU. The analysis of $S_{SUFT}$ proceeds by noting that the SU will try to move to another free channel if possible, otherwise the call will be terminated. When a new PU call arrives at a channel occupied by an SU, the SU leaves that channel and starts searching for a new channel to handoff its call. If the search process ends without finding a new free channel, the SU call will be terminated. The handoff capability will decrease the $S_{SUFT}$ because the SU does not need to terminate its call before the search process ends. The $S_{SUFT}$ is:

$$S_{SUFT} = \left[ \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-i} \pi_{(i,j)} \left( \lambda_2 - T_{(i,j)}^{(i,j)} \right) + \sum_{i=1}^{N-1} \sum_{j=0}^{N-i} \pi_{(i,j)} \left( \lambda_2 - T_{(i,j-1)}^{(i,j)} - \mu_1 \right) + \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} \pi_{(i,j)} \left( \lambda_2 - T_{(i-1,j)}^{(i,j)} - \mu_1 \right) + \pi_{(N,0)} \left( \lambda_2 - T_{(N-1,0)}^{(N,0)} - N \mu_1 \right) + \pi_{(0,0)} \right]}{\lambda_2} \right].$$  \hspace{1cm} (11)

**Secondary successful probability: $S_{SUSP}$**

The secondary successful probability ($S_{SUSP}$) is defined as the probability that a secondary call has completed the service and the call is normally terminated. The $S_{SUSP}$ is given by

$$S_{SUSP} = \left[ \sum_{i=0}^{N-1} \sum_{j=0}^{N-i} \mu_2 j \pi_{(i,j)} \right] \frac{\lambda_2}{N \mu_2}. \hspace{1cm} (13)$$

**Primary blocking probability: $PU_{PBP}$**

Primary blocking probability ($PU_{PBP}$) is defined as the probability that an arriving PU call will be blocked because all radio channels are occupied by PUs. The primary
blocking probabilities $PU_{PBP}$ is given by

$$PU_{PBP} = \pi_{(N,0)}.$$  \hfill (14)

**System resource utilization**

The system resource utilization is given by

$$U = \frac{\sum_{i=0}^{N-i} \sum_{j=0}^{N-i} (i+j)\pi_{(i,j)}}{N}$$

and it represents the used fraction of the total network resources. It should be noted that not all resource utilization will lead to actual goodput since some calls in the system will not be successfully completed. Some of the calls will be terminated before the call completion.

### 2.6 Numerical and simulation results

To assess the accuracy of the analytical model we performed extensive simulations. The simulations were carried out with MATLAB, using an event-based approach and Poisson arrival processes. The simulation setup used the assumptions given in section 2.2, i.e., the acquisition time was negligible. However, during channel acquisition and spectrum handoff the channels were randomly searched using the specified $P_{FA}$ and $P_{D}$. In the simulations setup, the effect of the number of free channels has been taken into account during the channel searching process. For example, when a new SU arrives or an existing SU leaves its current channel due to detection of a PU arrival, it starts the channel searching process. The number of free channels is a random process which depends on several factors, such as sensing errors and PU and SU arrivals that are modeled using Poisson processes. We can get the number of free channels $k$ as $k = N - i - j$, where $N$ is the total number of channels, $i$ is the number of PUs and $j$ is the number of SUs.

Results from the analysis and simulations were compared and they match very well, thus validating the analysis. We also compared the performance of the analytical model with that of the model presented in [45] under the same network setup (perfect sensing conditions) and the results were identical. This comparison not only shows that the proposed analytical model can evaluate the performance of CRN with imperfect sensing conditions.
conditions, but also illustrates that it can be used for analyzing network performance with perfect sensing and for any number of channels. Fig. 3 shows the $P_{UPPP}$ (theoretical and simulation) versus primary arrival rates for different numbers of channels. We observe that with a small number of channels the $P_{UPPP}$ grows rapidly as the primary arrival rate increases. However, as more radio resources in term of the number of channels are made available to serve rapidly arriving primary calls, the $P_{UPPP}$ decreases. The effect of the perfect sensing condition on the $P_{UPPP}$ is shown by the dash-dotted curves. It can be observed that the $P_{UPPP}$ is greater for the perfect sensing conditions. This can be attributed to the fact that with perfect sensing, arriving PU calls are always noticed by SUs and thus there are no primary call terminations. On the other hand, with imperfect sensing there are primary terminations leading to less PUs residing in the system. Therefore a newly arrived PU will find more channels occupied by other primaries with perfect sensing than with imperfect sensing.

![Fig. 3. Primary blocking probability, simulation versus theory, $P_D = 0.95$, $P_{FA} = 0.15$, $\lambda_2 = 3.5$, $\mu_1 = \mu_2 = 4$ (C⃝2013 IEICE).](image)
Fig. 4 presents the $SU_{BP}$ versus primary arrival rate $\lambda_1$ for different numbers of channels. The figure indicates that at high primary arrival rate the channels are more often occupied by PUs reducing opportunities for SUs to access the network and therefore fewer secondary calls are accepted. With a small number of channels, the high primary arrival rate $\lambda_1$ increases $SU_{BP}$. However, with more resources available in terms of the number of channels, the effect of $\lambda_1$ is less significant because the probability of finding new channels by SUs increases and therefore reduces $SU_{BP}$. With high arrival rate of $\lambda_1$ most of the channels will be occupied by PUs. Additionally, newly arrived PUs force existing SUs to vacate their channels. Those SUs will have to search for new channels to handoff their calls. The figure also shows clearly that the $SU_{BP}$ increases quickly for a small number of channels.

![Secondary blocking probability](image)

Fig. 4. Secondary blocking probability, $P_D = 0.95$, $P_{FA} = 0.15$, $\lambda_2 = 3.5$, $\mu_1 = \mu_2 = 4$ \((\text{C}2013 \text{IEICE})\).

Fig. 5 illustrates the $SU_{SFT}$ as a function of primary arrival rate for different number of channels. We observe that as the number of channels increases, the secondary forced termination declines. This is because more radio resources are available to handle primary calls. Hence the probability that new primary calls assigned to channels...
occupied by SUs is reduced. On the other hand, those SUs which were forced to move because they detected the arriving PUs, will more likely find new empty channels, or they will have to terminate their call. With a low primary arrival rate, not all channels are occupied by SUs and with perfect sensing the number of channels occupied by PUs is even higher, reducing the number of accepted SUs leading to reduced $SU_{SFT}$. With a high primary arrival, most of the channels are occupied by PUs. SUs detect PUs’ arrival and try to hand off their calls. With perfect sensing, the SU is able to correctly classify channels are occupied, therefore increasing the secondary forced termination. The influence of the secondary arrival rate on the $PU_{FTP}$ has been investigated and the results are presented in Fig. 6. We observe that the $PU_{FTP}$ increases approximately in proportion with the secondary arrival rate. This is because an increase in the secondary arrival rate increases the proportion of the misdetection by SUs leading to collision with PUs. The increases in the $PU_{FTP}$ can also be attributed to the fact that with high $\lambda_2$ arrival rates SUs most often occupy most of the opportunistic channels. Newly arrived SUs compete for opportunistic resources, therefore increasing the probability of collision with existing PUs. For the perfect sensing condition, the secondary traffic has no effect on the $PU_{FTP}$, as expected. The perfect sensing result is given by the straight line of black diamonds along the x-axis.
Fig. 5. Secondary forced termination probability, $P_D = 0.95$, $P_{FA} = 0.15$, $\lambda_2 = 3.5$, $\mu_1 = \mu_2 = 4$ (©2013 IEICE).

Fig. 6. Primary termination probabilities versus secondary arrival rate, $P_D = 0.95$, $P_{FA} = 0.15$, $\lambda_1 = 7$, $\mu_1 = \mu_2 = 4$ (©2013 IEICE).
Let us now consider also the relationship between $P_D$ and $P_{FA}$. For illustration purposes we assume an energy detector using two samples for which performance equation is given in (51) (analysis is valid for any detector). Fig. 7 plots the probability that a SU call is normally terminated (i.e., successful call) as a function of $P_D$ for different numbers of channels. The figure shows that as the number of channels increases, the increase in the $P_D$ improves the probability of successful secondary calls. However, since the $P_{FA}$ is associated with the $P_D$, high values for $P_D$ means high values for $P_{FA}$ as well. This illustrates that increasing $P_D$ beyond a certain level (e.g., 0.95) reduces the $SU_{SSP}$ because it increases the $P_{FA}$, and therefore SUs most often falsely classify free channels as occupied by PUs. Fig. 7 also shows that $SU_{SSP}$ for $N = 3$ is lower than that for $N = 8$. This is because it is difficult for SUs to get network access with fewer radio channels, hence decreasing the $SU_{SSP}$. Given the fact that SUs have to vacate their channels to newly arrived PUs, the probability that SUs will find new channels to hand off their call is small. This increases the $SU_{SFT}$ and decreases $SU_{SSP}$.

![Graph showing the relationship between Probability of Detection $P_D$ and Probability that a Secondary Call is Successful for different numbers of channels.](image)

**Fig. 7.** Probability that an SU call is successful (i.e., normally terminated) versus probability of detection. $\lambda_1 = 7, \lambda_2 = 3.5, \mu_1 = \mu_2 = 4, \gamma = 15.56$ [dB], $P_{FA}$ from ROC (©2013 IEICE).
The effect of increasing $P_D$ on $SU_{SBP}$ is given in Fig. 8. The figure shows that $SU_{SBP}$ increases due to the improvement in the detection operation. As the $P_D$ increases, the $P_{FA}$ increases as well. Although higher $P_D$ means that incoming SUs detect the presence of PUs more accurately, the higher $P_{FA}$ means that due to false alarms more SUs will be blocked. The $SU_{SBP}$ increases more rapidly with a small number of channels. However, for a high number of channels the increase is flat. When the $P_D$ reaches a certain level, the curves grow exponentially due to the combined effect of the $P_D$ and $P_{FA}$.

![Graph showing the relationship between $P_D$ and $SU_{SBP}$]

**Fig. 8. Secondary blocking probability versus probability of detection, $\lambda_1 = 7, \lambda_2 = 3.5, \mu_1 = \mu_2 = 4, \gamma = 15.56$ [dB]; $P_{FA}$ from ROC (©2013 IEICE).**

Fig. 9 depicts the $PU_{FTP}$ versus $P_D$ for different numbers of channels. The $P_D$ is obtained as a function of $P_{FA}$. The figure highlights that the improved detection capability of SUs reduces the $PU_{FTP}$. Of course, when $P_D = 1$, the $PU_{FTP}$ is zero. Low values for $P_D$ means that during channel searching/acquisition, SUs most often misdetect the presence of PUs and collide with them, increasing the $PU_{FTP}$. Fig. 10 shows the system resource utilization as the $P_D$ is increased from 0.6 to 1. As mentioned earlier, the $P_D$ is obtained as a function of $P_{FA}$. We can observe that there is a trade-off
between resource utilization and the $P_D$. On the one hand, the resource utilization gains from improved detection probability. On the other hand, increasing the $P_D$ increases the $P_{FA}$, therefore reducing the system resources utilization. We can see from Fig. 10’s large $P_D$ increase the resources utilization up/near to the point where $P_D = 0.95$, at which point the $P_{FA}$ starts to grow and eventually reduces the system resource utilization. This reduction in the resource utilization can be clearly observed with a small number of channels. The figure also shows the improvement in the resource utilization by SU traffic. The dash-dotted lines show the resource utilization for only PU traffic.

![Fig. 9. Primary termination probabilities versus probability of detection, $\lambda_1 = 7$, $\lambda_2 = 3.5$, $\mu_1 = \mu_2 = 4$, $N = 3$, $\gamma = 15.56$ [dB], $P_{FA}$ from ROC (©2013 IEICE).]
2.7 Summary and discussions

A CTMC-based analytical model for evaluating the performance of CRNs under imperfect sensing environments was presented. The model incorporated key performance evaluation elements such as the effect of imperfect sensing, state-dependent transition rates single analytical framework to enable an accurate theoretical analysis. Formulas for performance evaluation metrics including primary termination probability, secondary success probability, secondary blocking probability, secondary forced termination probability and radio resource utilization were derived. A recursion approach that allows the CTMC state transition rates to be found for an arbitrary number of channels was used. Numerical results showed that the imperfect sensing parameters have a significant effect on the performance metrics and thus sensing errors need to be considered in theoretical modeling. Results showed that the system resource utilization increases with the increase in the $P_D$. However it was observed that there is a trade-off between resource utilization and the $P_D$. On the one hand, the resource utilization gains from improved detection
probability. On the other than, increasing the $P_D$ increases the $P_{FA}$ and therefore reduces the system resources utilization. Incorporating fully state-dependent transition rates in the CTMC was found to significantly improve analysis accuracy, thus achieving a more accurate analytical model. Results from extensive Monte Carlo simulations confirm the validity of the analytical model.
3 The effect of false alarm rate on the performance of cognitive radio networks

It was mentioned in Chapter 1 that spectrum sensing plays a significant role in enabling utilization of spectrum holes by unlicensed SUs in CRNs. Most of the related work concerning spectrum sensing has focused on sensing carried out by incoming SUs aiming at locating spectrum opportunities. However, in order to appropriately protect returning licensed PUs, SUs should continuously perform spectrum sensing during their ongoing transmissions. An important issue associated with the continuous sensing is the false alarm rate (FAR), which is defined as the average number of false alarms per unit of time and can be modeled by a Poisson process with Poisson parameter $\lambda_{FAR}$.

The frequent occurrence of false alarm events is highly undesirable since it makes it challenging for SUs to fully utilize the spectrum opportunities and severely degrades their QoS. Motivated by these issues, we develop a comprehensive CTMC analytical framework to evaluate the effect of the FAR on the performance of CRNs. A major feature of the proposed analytical framework is that it takes into account the effects of sensing errors by both incoming SUs looking for free channels to transmit on, and the already transmitting SUs expecting the presence of returning PUs. The analytical model also examines the interference tolerance among PUs and SUs as well as the impact of SUs’ residual self-interference. To our knowledge, the effect of the FAR on the operation of CRNs has not been investigated in the literature. The contribution of this chapter can be summarized as follows:

- It motivates and develops a CTMC-based analytical framework that precisely evaluates the performance of CRNs. Unlike existing approaches, the proposed model thoroughly investigates the effect of FARs on the performance of CRNs. The model also takes into consideration the SUs’ residual self-interference as well as interference tolerance among PUs and SUs.
- It models the occurrence of FAR events as a Poisson process with parameter $\lambda_{FAR}$ with a theoretical justification based on a shrinking Bernoulli process [57].
- It proposes a new performance evaluation measure, the SU self-termination probability. The proposed metric can precisely measure the percentage of SU calls that are terminated because of the FAR occurrence. The new metric also allows for measuring the SUs’ ability of utilizing spectrum opportunities.
Performance results show that high $\lambda_{FAR}$ can severely degrade PUs’ performance and reduce the overall system resource utilization. However, with increasing PU interference tolerance, the performance of PUs improves as well. SU residual interference was found to decrease the detection probability resulting in a low PU performance. Extensive simulations validate the analytical model, demonstrating excellent agreement with the theoretical results. The rest of this chapter is organized as follows. Section 3.1 presents the related work. The system model is presented in Section 3.2. Section 3.3 contains a description of the spectrum sensing model. Section 3.4 presents the CTMC-based analytical framework. In Section 3.5, we discuss performance evaluation metrics for the CRN. Section 3.6 summarizes results and provides comparison of simulation and theoretical results. Finally, Section 3.7 provides the conclusion.

3.1 Related work

Several studies have been proposed to detect returning PUs in CRNs. In [21], the authors investigate the issues of how to maximize the overall discovery of opportunities in the licensed channels and how to minimize the delay in locating an idle channel in order to minimize interference on returning PUs. Similarly, the authors of [58] presented a dynamic spectrum access mechanism in a network where SUs do not have perfect knowledge of PUs’ communication behavior. The interference issue has also been studied. However, this study only considered perfect spectrum sensing and a network with only one PU. In [59], the authors consider a preemptive priority approach for the channel access where SUs must vacate their channels whenever the corresponding PUs appear. The work presented in [60] formulates a joint spectrum sensing and access problem as an evolutionary game by considering the mutual influence between spectrum sensing and access. Although the interference problem has been addressed in these works, the problem of the FAR has not been investigated. The results reported in [61] have revealed the importance of investigating the effect of unnecessary spectrum handoff due to false alarms during spectrum sensing.

In [62], continuous-time Markovian process (CTMP) is used to model PU traffic in OSA systems. However, for analyzing the behavior of SUs, discrete time queuing was used. The underlying assumption made therein is that sensing and data transmission cannot be carried out simultaneously and therefore the SU has to periodically suspend its data transmission in order to perform spectrum sensing. The problems with this
technique are the overheads associated with the scheduling and synchronization of
the suspension periods among SUs as well as the frequent interruption in the SU’s
data transmission. Additionally, the SU can only detect a reappearing PU during the
suspension period, even if the PU reappeared before the suspension period. This work
also differs from our study because it only supports CRNs with one channel and the
assumption that the spectrum sensing is perfect.

Simultaneous spectrum sensing and data transmission approach has been studied in
[63–66]. The issue of self-interference due to transmitting and receiving in the same
band has been studied in [67, 68]. In spite of considering the problem of unnecessary
false alarms, the authors of [61, 69] did not investigate their effect on performance
metrics such as blocking and termination probabilities. Furthermore, the authors of [70]
analyzed different types of unreliable sensing for both incoming and ongoing SUs, and
their impact on the performance of CRNs without addressing the FAR effect. Most of
the existing CTMC models [39, 44, 51] do not cover all the aspects of the spectrum
sensing and CNRs operation and some important factors were not fully addressed. In the
previous chapter (corresponding to [30]), we analyzed the performance of CRNs using a
CTMC framework that supports multi-channel, spectrum handoff, full state-dependent
transition rates, and the ability to handle spectrum sensing errors. In this chapter, we
extend the analysis provided in Chapter 2 to capture the effect of the FAR and to handle
the residual self-interference within the SU’s transceivers.

3.2 System model

We consider a CRN with \( N \) number of channels in which SUs are allowed to opportunistically utilize licensed spectrum bands with the constraint that the QoS of PUs remains at an acceptable level.

3.2.1 Primary user model

We assume that the primary channels occupancy are time varying alternating between
idle and busy periods, and thus SUs must perform spectrum sensing continuously to
detect the presence of PUs. PU connections arrive at the network according to a Poisson
process at a rate of \( \lambda_1 \). The PU service rate which is assumed to be exponentially
distributed is \( \mu_1 \). We also assume that PUs can obtain primary channel occupancy
information, for example by accessing a core network that makes signaling or querying
of the PUs’ base station [71], and thus it is further assumed that PUs do not collide with each other [44]. We assume that both PUs and SUs have some interference tolerance $T_{TOL}$ of how many seconds of interference they will tolerate before withdrawing from the system. If the PU interference tolerance time $T_{TOL}$ is 0, no SU transmission is allowed [72]. We assume equal interference tolerance for both systems leading to both colliding users withdrawing from the system simultaneously. A similar assumption has been considered previously in [44].

3.2.2 Secondary user model

We assume that SU connections arrive at the network according to a Poisson process with $\lambda_2$. The SU service rate is assumed to be exponentially distributed with $\mu_2$. During the absence of PUs, SUs can opportunistically access the free channels if they are not occupied by other SUs. We also assume that SUs are capable of broadcasting control messages on a common control channel [73] to show their existence to neighboring SUs in the proximity. Therefore SUs do not attempt to access channels occupied by other SUs. Upon detection of the presence of a returning PU, an SU leaves its current channel and starts the spectrum handover process in order to find a new free channel. If the channel search process ends without finding a free channel, the SU terminates its call and leaves the network.

3.3 Spectrum sensing model

No matter which detection scheme is used for protecting returning PUs, it will lead to the occurrence of false detection of returning PUs, i.e., we will erroneously assume that a PU has returned, when in fact the PU’s channel is free. A simple way to characterize the occurrence of false alarms for already transmitting SUs is to use the average number of false alarms per time unit, similar to the rate parameter of a Poisson process. We call this parameter the false alarm rate parameter $\lambda_{FAR}$ [74, 75]. It is to be mentioned that in multi-channel systems with handoff capability (as studied in this chapter), an SU moving from its current operating channel (for example due to false detection of a returning PU) will attempt to locate another free channel to continue its ongoing data transmission.

In the spectrum sensing approach, a problem is that once the initial sensing assesses the channel to be free, the SU gains access to the channel, and a PU returning during the SU activity will suffer from interference. This issue is less significant in time-
slotted systems since, if the SU knows the PU’s time slot boundaries, it can perform spectrum sensing only during the beginning of each time slot, provided that the PU status remains constant during each slot (i.e., the PU cannot return during the time slot) [76]. In non-time-slotted systems that are considered here, various schemes have been proposed to ensure the protection of returning PUs. The simplest approach is to use a sufficiently short time for SU transmissions so that the harm to PUs is negligible. For longer periods of SU activity, the concept of a quiet period (QP) during data transmission has been introduced in IEEE 802.22 [77–79] in order to allow SUs to periodically suspend their data transmission so that the presence of PUs can be detected. Fig. 11 illustrates spectrum sensing carried out using QPs to detect the presence of primary signals. As a PU reappears, the SU must leave the channel. The problems with this technique are the overheads associated with the scheduling and synchronization of the QPs, as well as the frequent suspension of the SU’s ongoing transmissions. Additionally, an SU can only detect PUs during the QP, even if they reappeared before the QP.

Fig. 11. Quiet period sensing.

To protect reappearing PUs, SUs perform spectrum sensing on a continuous basis along with data transmission. Fig. 12 presents the concept of simultaneous spectrum sensing and data transmission. When a new SU arrives at a channel for the first time, it senses the whole PU channel with a bandwidth $W$ for $T_1$ seconds. After discovering that the channel is free, the SU starts data transmission over a bandwidth of $W - W_{GB} - W_2$. While the SU is transmitting data over the lower part of the PU channel, it continuously monitors the PU channel using the upper part with bandwidth $W_2$ to detect the presence of a returning PU. A guard band (GB) subchannel is used to prevent the leakage from the SU’s self-interference signal. The SU makes a sensing decision every $T_2$ seconds.
As a PU reappears (if correctly detected by the SU), the SU ceases its data transmission and leaves the channel.

![Fig. 12. Continuous sensing and data transmission.](image)

The spectrum sensing model used in this chapter does not consider the relative signal strengths for different SUs in the interference to the PU (and to interference tolerance). For example, recently the trend is towards sharing between radars (PU) and secondary users. There is a denied area around the radars where not even secondary operation is allowed. Then, in the next ring (further away from the PU) SU access is allowed, provided that the PU is not harmed. Thus, following recent trends in CRs, it can be assumed that SUs have some minimum distance to PUs. For a reliable PU detection, the aggregate interference produced by multiple SUs needs to be taken into account. This issue has been investigated in [27] where the authors describe how to compute the secondary system’s generated aggregate interference in a shadow fading environment. To translate the regulatory constraint on the aggregate interference to the system-and device-level design parameters, the authors of [80] developed a statistical model of interference aggregation in the spectrum sensing cognitive radio network. Alternatively, the PU could report the increased interference level to a database, and the region allowed for secondary use could be reduced (or some SU transmitters in the current allowed region could be turned off).

### 3.3.1 Continuous spectrum sensing

As illustrated in Fig. 12, and similar to the distributed (coordination function) interframe space (DIFS) operation in IEEE 802.11, the SU has to keep sensing the PU channel.
from the beginning of its transmission, since the PU can arrive at any time instant of a slot. This process forms a continuous sequence of sensing slots with length equals $T_2$. For example, if the PU appears in the middle of a time slot, then the first slot will not get full PU energy, leading to a smaller detection probability than the later full slots. Since the first $T_2$ may be wasted, we assume that the first partial slot sensing never leads to detection, i.e., the detection probability is close to zero. Hence, we should detect the PU arrival during $T_{TOL} - T_2$ seconds which corresponds to $\hat{T}_{TOL} = \left\lfloor \frac{T_{TOL} - T_2}{T_2} \right\rfloor$ slots. Although partial slot sensing can enable the SU to perform sensing immediately after the arrival of the PU, and hence have a prompt reaction to protect PUs, for the sake of simplifying the analysis, we consider only the full slot sensing by assuming that the detection process will begin from the first full time slot following the SU arrival.

One possibility for implementing continuous sensing is to leave the upper part of the PU channel empty (i.e., free from SU transmissions) \[81\]. As shown in Fig. 12, we split the PU channel into three subchannels: A) SU communication channel, B) a sufficient vacant guard band to reduce the effect of the SU’s self-interference, and C) SU sensing channel. When the PU is active, it uses the whole bandwidth (A+B+C) for its communication. The secondary user uses subchannel A) for its communication. A reappearing PU can be detected by sensing, during ongoing SU transmission from subchannel C). It is obvious that a problem here is the self-interference due to the leakage of the SU’s transmitted signal back to its sensing device. However, the emergence of a large variety of self-interference cancellation techniques \[82–84\] in the literature enabled efficient reduction in self-interference and therefore allowing radios to operate in full duplex mode. For example, the authors of \[85\] present a method for canceling a passband self-interference signal using adaptive filtering in the digital domain. Therefore, in addition to the vacant guard band and bandpass filtering, self-interference cancellation has also been assumed to remove most of the residual self-interference. In this model, $q \in \{1, 2\}$ denotes an index with the interpretation that $q = 1$ if the spectrum sensing is carried out by incoming SUs and $q = 2$ if the spectrum sensing is performed by ongoing SUs.

### 3.3.2 Energy detector-based spectrum sensing

Without loss of generality, we consider that initial and ongoing spectrum sensing are done using an energy detector \[12\] with an integrate and dump operation mode as described in \[74, 75\]. Fig. 13 illustrates the operation of the energy detector. Although
we focus on ED, the analysis techniques presented in this chapter are generic and not limited to ED, provided that the used detector can be mapped to false alarm probabilities, probability of detections and false alarm rates. Let $y_q(t)$ denote the SU received signal process. We express the incoming SU received signal process in the form

$$y_1(t) = \begin{cases} n(t) & : H_0 \\ h_{PU} s_{PU}(t) + n(t) & : H_1 \end{cases}$$

and the ongoing SU received signal process can be formulated as

$$y_2(t) = \begin{cases} h_{SU} s_{SU}(t) + n(t) & : H_0 \\ h_{PU} s_{PU}(t) + h_{SU} s_{SU}(t) + n(t) & : H_1. \end{cases}$$

In equations (16) and (17), $s_{PU}(t)$ is the PU transmitted signal, $s_{SU}(t)$ represents the leakage from the SU transmitted signal, $n(t)$ is the additive white Gaussian noise (AWGN), $h_{PU}$ is the PU channel gain while $h_{SU}$ represents the SU leakage signal gain and $t$ is the time. In the above equations, $H_0$ is the null hypothesis meaning that PU is not present in the sensed band, and $H_1$ represents the alternative hypothesis referring to the presence of the PU signal. The received signal is filtered by a bandpass filter to remove the out-of-band and self-interference noise. The filtered signal is then squared by the squaring device and applied to the integrator. The integrator output $Y_q$ (also denotes the decision variable of the energy detector) is sampled every $T_q$ seconds. Then the integrator is reset before integrating the next sample over the next $T_q$ seconds. Finally, $Y_q$ is compared with the decision threshold to decide about the presence of the PU. Let $W_q$ denote the sensed bandwidth. Let $\gamma_q$ denote the signal to noise ratio SNR and $\eta_q$ denote the energy detection threshold. According to [12]

$$Y_q \sim \begin{cases} \chi^2_{2u_q} & \text{under } H_0 \\ \chi^2_{2u_q}(2\gamma_q) & \text{under } H_1, \end{cases}$$

where $\chi^2_{2u_q}$ is a chi-square distribution with $2u_q$ degrees of freedom (i.e., the time-bandwidth product $u_q = W_q T_q$) and $\chi^2_{2u_q}(2\gamma_q)$ is a non-central chi-square distribution.
with \(2\mu_q\) degrees of freedom and a non-centrality parameter \((2\gamma_q)\). It has been shown in [86] that the probability of detection \(P_{Dq}\) and the false alarm probability \(P_{FAd}\) can be given as follows:

\[
P_{Dq} = Q_{\mu_q}(\sqrt{2\gamma_q}, \sqrt{\eta_q}),
\]

\[
P_{FAd} = \frac{\Gamma(\mu_q, \frac{\eta_q}{2})}{\Gamma(\mu_q)},
\]

where \(Q_{\mu(\cdot)}\) is the generalized \(m\)th Marcum Q-function [87] and \(I_{m-1}(\cdot)\) is the modified Bessel function of \((m-1)\)th order and \(\Gamma(\cdot)\) is the upper incomplete gamma function which is defined by

\[
\Gamma(a, x) = \int_x^{\infty} t^{a-1} e^{-t} dt,
\]

with \(\Gamma(a, 0) = \Gamma(a)\) [86].

Referring to Fig. 12, where \(W\) represents the PU channel bandwidth. To obtain the received signal energy, let \(P_{PU}\) denote the PU-transmitted signal power. Let also \(N_0\) denote the one-sided power spectrum density (PSD). Let \(T_1\) denote the incoming SU initial sensing time. Assuming that the PU signal power is uniformly distributed over the PU channel, then the incoming SU signal energy within the initial sensed area can be obtained as

\[
E_{S1} = P_{PU} T_1
\]

and the initial sensing SNR is then obtained with

\[
\gamma_1 = \frac{E_{S1}}{N_0}.
\]

Similarly, we can obtain the signal energy within the ongoing SU sensed area with a sensing time duration \(T_2\) as

\[
E_{S2} = P_{PU} T_2 \frac{W_2}{W}.
\]

Although we use guard band, bandpass filtering and self-interference cancellation to eliminate the effect of SU the leakage signal, we assume that the SU self-interference has not been fully removed. To handle the effect of any remaining self-interference, we follow the results presented in [88] to model the residual interference. Let us assume that the SU operates with a single-antenna full duplex transceiver. Let \(\alpha_{SU}\) denote the SU’s residual interference distortion factor. By using equation (17), the effective ongoing sensing SNR can be expressed as [88]

\[
\gamma_2 = \frac{E_{S2}}{N_0(1 + \alpha_{SU})}.
\]
In continuous spectrum sensing with full duplex communication, the consideration of self-interference is particularly important since the self-interference can affect the sensing outcome and degrade the SUs’ performance. Although the SU self-interference signal can have non-zero mean, it has been assumed in the majority of related works to have a zero-mean. For instance in [89], the authors mentioned that in practical full duplex systems the self-interference cannot be completely canceled, such that the signals received at each node are a combination of the signal transmitted by the other source, the residual self-interference (RSI) and the noise. They also assume that the RSI can be typically modeled as a zero-mean additive white Gaussian noise (AWGN). The work reported in [88] assumed that the Gaussian distortion and noise follow central chi-square distribution in the absence of PU signals but potentially includes RSI and noncentral chi-square distribution when the PU signal is present.

Self-interference mitigation in full-duplex MIMO relays has been investigated in [67] where the authors focused on minimizing the residual loop interference so that it can be regarded as an additional relay input noise. They assumed that all signals from the relay output to the relay input (including loop interference (LI) signal) and noise vectors have zero mean. Furthermore, the authors of [90–92] assumed that the SU self-interfering signal before carrying out self-interference suppression (SIS) to be a zero-mean random signal with self-interference channel coefficient equal one. In [93], the residual self-transmitted signal is modelled with circular symmetric complex Gaussian variables. Following the common practice in existing models, the use of the assumption that SU’s the leakage signal can be zero mean and follow central chi-square distributions is justified and can be held in order to take into account the RSI signal and perform the analysis.

It should be noted that when we use a dedicated part of the bandwidth (subchannel C) for continuous sensing, the effect of the residual interference becomes much lower than when we use the full bandwidth for simultaneous sensing and transmission. Each incoming SU correctly detects channel occupancy with probability \( P_{D1} \), and falsely classifies a free channel as occupied with \( P_{FA1} \). Similarly, each SU with ongoing calls detects the arrival of a PU with probability \( P_{D2} \), and falsely classifies a free channel as occupied with \( P_{FA2} \). The corresponding misdetection probabilities for incoming and outgoing SUs are \( P_{M1} = 1 - P_{D1} \) and \( P_{M2} = 1 - P_{D2} \), respectively. The detection probability \( P_{D2} \) refers to the probability of detecting incoming PU during the first \( T_{TOL} \) full slots of its arrival, instead of the per-slot detection probability. If the per-slot detection probability is denoted as \( z \) then \( P_{D2} = 1 - (1 - z)^{T_{TOL}} \). This does not affect the
FAR process since one per-slot false alarm event is enough to initiate the spectrum handoff and channel searching process. The modeling of partial slot sensing is left for future work.

### 3.3.3 Poisson process approximation

We model the occurrence of the false alarm at each sensing decision with the Bernoulli process. The energy detector makes only one sensing decision in each slot which results in a binary variable (0 or 1). Since the sensing decisions with only white Gaussian noise present are independent, the resulting binary output of the sensing clearly follows the Bernoulli process (i.e., independent and identically distributed process generating 1s and 0s), and the Bernoulli parameter corresponds to the probability of FAR occurrence (binary output 1) in each spectrum sensing decision.

At each spectrum sensing decision epoch $T_2$, a false alarm occurs with probability $P_{FA2}$ and does not occur with probability $1 - P_{FA2}$, independently of the decision outcome of the last sensing period. The $\lambda_{FAR}$ parameter is the product of the decision rate and the false alarm probability [74, 75, 94]. Therefore, $\lambda_{FAR}$ is given by $P_{FA2}/T_2$.

Let us assume that the sensing interval $T_2$ is short and therefore we assume that the decision rate given by $1/T_2$ is large, and that the false alarm $P_{FA2}$ is small as otherwise there would be too many false alarms for successful SU operation. Then the arrival process of false alarms can be approximated by a Poisson process as a limit of a shrinking Bernoulli process [57] with parameter $\lambda_{FAR}$.

### 3.4 Continuous-time Markov chain model (CTMC)

We consider a two-dimensional CTMC to describe the CRN system. At any time, the system state is determined by $(i, j)$ where $i$ represents the number of channels occupied by PUs and $j$ represents the number channels occupied by SUs with the restriction that $0 \leq i \leq N$, $0 \leq j \leq N$, $0 \leq i + j \leq N$. Let $i\mu_1$ and $j\mu_2$ denote the service completion time for PUs and SUs respectively. The transition rate from state $(i, j)$ to state $(h, l)$ is given by $T_{(i,j)}^{(h,l)}$. Note that the parameter $\lambda_{FAR}$ affects all state transitions from states with the number of SUs $j > 0$. As the number of channels increases, the number of states of the CTMC grows exponentially. Since the transition rates depend on the system states, the large number of states combined with the channel searching process under imperfect sensing conditions would make it non-trivial to compute the state transition
Fig. 14. Diagram for the allowable transitions to and from state (1,1) in a three-channel CRN.

rates. Because it is impractical to present a state transition diagram for a CRN with an arbitrary number of channels, we present an illustrative example in Fig. 14 which shows the allowable state transitions from and to state (1,1) in a Markov chain with three-channel.

As an example, consider a CRN with three-channel denoted by C1, C2 and C3. Let us assume that channel C1 is occupied by a PU, channel C2 is occupied by an SU, and the last channel C3 is free. The Markov chain is in state (1,1). We now explain a series of events that trigger the system to move from state (1,1) to state (0,0). State (0,0) indicates that all channels are free. On the occurrence of $A_{FAR}$, the SU leaves C2 and starts the channel searching process. There are two channel selection possibilities for the SU for continuing its data transmission. The SU can first select C1 with probability $1/2$
and then misdetect the presence of the PU on C1 with probability $P_{M1}$. The second possibility is to select C3 with probability $1/2$, and then falsely classify the free channel as occupied by the PU with a false alarm probability $P_{FA1}$, and finally misdetect the presence of the PU on C1 with probability $P_{M1}$. Both selections lead the SU to collide with the PU, and eventually both of them leave the network. Combining all these events, the transition rate from state (1,1) to state (0,0) can be obtained by $\frac{\lambda_{FAR}P_{M1}}{2}(1 + P_{FA1})$.

Another example is the transition from state (1,1) to state (2,0). This happens with the arrival of a PU with rate $\lambda_1$ and with probability $1/2$ to channel C2 which is occupied by the SU. The SU correctly detects the presence of the PU with detection probability $P_{D2}$ and vacates the channel. After leaving the channel, the SU has two possibilities with probability $1/2$ for each. The SU first falsely classifies the free channel C3 as being occupied by a PU with false alarm probability $P_{FA1}$, and then detects the presence of the PU in channel C1 with probability $P_{D1}$, ending up leaving the network. The other possibility is that the SU correctly detects the existence of the PU in channel C1 with probability $P_{D1}$, and then erroneously classifies the free channel C3 as occupied by a PU with false alarm probability $P_{FA1}$. The resulting transition rate is $\frac{\lambda_1P_{D2}P_{D1}P_{FA1}}{2}$.

Proceeding in a similar manner, the transition rates to and from the remaining states can be obtained.

### 3.4.1 Generalization of the CTMC

The goal of this section is to extend the results presented previously to describe a CRN with an arbitrary number of channels $N$. When $N$ is large, constructing a state transition diagram and finding solutions to the corresponding balance equations is complicated and time consuming. Similar to [30], we use a recursive method to calculate the state transition rates to and from all different states of the state transition diagram representing the CRN network. The state transition rates are used to get all the possible balance equations. Recall that the number of PUs is denoted by $i$ and the number of SUs is denoted by $j$. Let us also assume that the number of free channels is denoted by $k$ which is given by $k = N - i - j$.

It is important to note that the state transition rates presented in this Chapter are different from those defined in Chapter 2 for several reasons: 1) The inclusion of the effect of the FAR in the CTMC. 2) In Chapter 2 we assumed that ongoing SUs perfectly detect the arrival of PUs and that there are no false alarms during ongoing data transmission. However, in this Chapter we assume that SUs with ongoing connections
do not perfectly detect the arrival of PUs and also that the false alarm probability during ongoing calls is not negligible. With this assumption, ongoing SUs detect PU arrivals with $P_{D2}$ where $P_{D2}$ is an arbitrary value between 0 and 1 and that the false alarm probability during ongoing calls equals $P_{FA2}$ which is also an arbitrary value. Note that this is a more realistic assumption for practical CRNs and represents a significant improvement over work presented in the previous chapter that leads to obtaining accurate state transition rates and state probabilities. 3) Because of the false alarms during ongoing sensing, we require a new state transition that defines the transition from state $(i, j)$ to state $(i - 1, j - 1)$, with $i > 0$ and $j > 0$. In addition to state transitions because of the FAR events, we refer the reader to [30] for details concerning the other different events that trigger state transitions. All possible state transition types are described as follows:

− Transition Type 1: $(i, j) \rightarrow (i, j + 1)$. This transition defines the increase in the number of SUs by one and can be obtained by

$$f(i, k) = \frac{k}{i + k}(1 - P_{FA1}) + \frac{k}{i + k}P_{FA1}f(i, k - 1) + \frac{i}{i + k}P_{D1}f(i - 1, k), \quad (24)\,$$

where function $f(.)$ is used to define the increase in the number of SUs by 1 [30]. $P_{D1}$ and $P_{FA1}$ have been defined earlier to denote the initial sensing’s detection and false alarm probabilities. They have been used to obtain more accurate state transition rates and state probabilities in comparison to results obtained in [30]. The overall state transition rate for this case is given by $A_2f(i, k)$.

− Transition Type 2: $(i, j) \rightarrow (i - 1, j)$. This transition defines the decrease in the number of PUs by one. We use the recursive function $g(.)$ [30] to define this transition

$$g(i, k) = \frac{i}{i + k}P_{M1} + \frac{i}{i + k}P_{D1}g(i - 1, k) + \frac{k}{i + k}P_{FA1}g(i, k - 1), \quad (25)\,$$

where $P_{M1}, P_{D1}$ and $P_{FA1}$ denote the initial sensing’s misdetection, detection and false alarm probabilities respectively. The overall state transition rate for this case can be obtained by $i\mu_1 + A_2g(i, k)$ [30].

− Transition Type 3: $(i, j) \rightarrow (i + 1, j)$. This transition is given by

$$T_{(i+1,j)}^{(i,j)} = \lambda_1 \left(\frac{N - i - j}{N - i}\right) + \frac{f(i, N - i)}{N - i} \quad (26)\,$$

to reflect the increase in the number of PUs by one.
– Transition Type 4: \((i, j) \rightarrow (i, j - 1)\). The state transition rate for decreasing the number of SUs by one is given by
\[
T_{(i, j)}^{(i, j - 1)} = j \mu_2 + A_1 P_{M2} \frac{j}{N - i} + j A_{FAR} (1 - f(i, N - i - j) - g(i, N - i - j)) + A_1 P_{D2} \frac{j}{N - i} g(i, N - i - j).
\] (27)

– Transition Type 5: \((i, j) \rightarrow (i + 1, j - 1)\). The state transition rate for this case is given by
\[
T_{(i + 1, j - 1)}^{(i, j)} = \left(A_1 P_{D2} \frac{j}{N - i}\right) (1 - f(i, N - i - j) - g(i, N - i - j)).
\] (28)

– Transition Type 6: \((i, j) \rightarrow (i - 1, j - 1)\). The number of PUs is decreased by one and the number of SUs is decreased by one. This transition occurs if after the occurrence FAR, the SU ends up colliding with a PU. We get the transition rate as
\[
T_{(i - 1, j - 1)}^{(i, j)} = j A_{FAR} [g(i, N - i - j)].
\] (29)

3.4.2 Construction of state transition rate matrix and computation of the steady state probability vector

Let \(Q\) denote the state transition rate matrix (also known as infinitesimal generator) of the CTMC. Let \(\pi\) denote the steady state probability vector with \(\pi_{i,j}\) denoting the probability that the system is in the steady state \((i, j)\). When the system is in the steady or equilibrium state, the normalization condition is given by
\[
\sum_{i=0}^{N} \sum_{j=0}^{N} \pi_{i,j} = 1 \quad [95]
\] with the condition that \(0 \leq i \leq N, 0 \leq j \leq N,\) and \(0 \leq i + j \leq N\). Let \(D\) equal the total number of states in CTMC. We map the elements of the steady state probability vector \(\pi_{i,j}\) from state to index by assigning a unique integer index to identify each state. Therefore, the steady state probability vector can be represented by \(\pi = (\pi_1, \pi_2, ..., \pi_D)\) and the normalization condition is given by \(\sum_{d=1}^{D} \pi_d = 1\).

The steady state probabilities of the CTMC can be found by applying the following procedure:

– Step 1: Solve the recursive equations (24)–(29) to obtain the state transition rates.
– Step 2: Derive the balance equations using the rule that incoming transition rates to each state must equal outgoing transition rates from that state [56].
– Step 3: Use the balance equations to build the infinitesimal generator matrix \(Q\). All elements not on the main diagonal of \(Q\) represents state transition from one state to
another. The elements on the main diagonal of $Q$ make the sum of the elements in the respective row equals zero [96].

- Step 4: Apply the normalization condition $\sum_d \pi_d = 1$.
- Step 5: Solve the system of linear equations $\pi Q = 0$ to obtain the CTMC’s steady state probabilities.

Each element in the steady state probability vector $\pi$ represents the percentage of time that the system spends in that state.

Since the number of states of the CTMC grows exponentially with the number of the channels in the network it would be impossible to derive the CTMC transition rates by hand for large numbers of states. In this sense, the utilized recursive approach solves one part of this problem. However, the number of states is still exponential, which leads to higher memory and processing time requirements when the number of channels increases since the full state transition rates are used to obtain exact results. With a very large number of channels, approximation solutions with a reduced number of channel states would be beneficial. In the literature some approximation methods have been presented for CTMCs with a large number of states [96, 97]. The results presented in this Chapter have been obtained using the exact full CTMC. However, when the number of channels is very large, which brings some inefficiency, we can apply approximate solutions of large CTMCs to overcome this problem [96, 97].

### 3.5 Performance evaluation measures

In order to measure the performance of the CRN, we define several performance evaluation measures: secondary forced termination probability ($S\ U_{FTP}$), primary forced termination probability ($PU_{FTP}$) and secondary self-termination probability ($SU_{STP}$). Those performance metrics are calculated by using the steady state probabilities $\pi_{(i,j)}$ and state transition rates derived in the previous section. The reader is referred to the previous chapter (or [30]) for more details on the definition and derivation of other performance metrics such as secondary successful probability ($SU_{SSP}$), primary blocking probability ($PU_{PBP}$) secondary blocking probability ($SU_{SBP}$), as well as system resource utilization.
3.5.1 Secondary forced termination probability ($SU_{FTP}$)

The secondary forced termination probability, denoted by $SU_{FTP}$, is the probability of terminating SU calls because of the SU’s failure to find a new free channel after moving from its current channel. The $SU_{FTP}$ is calculated and defined by equation (30). It reflects the ratio of terminated SUs’ calls to total SU call arrivals $\lambda_2$.

$$SU_{FTP} = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-i} \pi(i,j) \left( T(i,j)_{(i+1,j-1)} + \sum_{i=0}^{N-1} \sum_{j=1}^{N-i} \pi(i,j) \left( T(i,j)_{(0,j-1)} - j\mu_2 \right) \right) + \sum_{i=0}^{N-1} \sum_{j=0}^{N-i} \pi(i,j) \left( T(i,j)_{(i-1,j-1)} \right) \right]}{\lambda_2}.$$  (30)

3.5.2 Primary forced termination probability ($PU_{FTP}$)

The primary forced termination probability, denoted by $PU_{FTP}$ and given by equation (31), is calculated as the ratio of terminated PU calls because of collisions with SUs to the total primary call arrivals $\lambda_1$.

$$PU_{FTP} = \frac{\sum_{i=1}^{N} \sum_{j=0}^{N-i} \pi(i,j) \left( T(i,j)_{(i-1,j)} - i\mu_1 \right) + \sum_{i=0}^{N-1} \sum_{j=1}^{N-i} \pi(i,j) \left( T(i,j)_{(i,j-1)} - j\mu_2 \right) + \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} \pi(i,j) \left( T(i,j)_{(i-1,j-1)} \right) \right]}{\lambda_1 \left(1 - \pi(N,0)\right)}.$$  (31)

3.5.3 Secondary self-termination probability ($SU_{STR}$)

Here we introduce a new performance metric that measures the secondary self-termination probability. The motivation of proposing this new metric starts with the fact that in CRNs the occurrence of false alarms are of critical importance since significant amounts of SUs’ calls could be terminated because of FARS. The new metric helps in accurately determining the percentage of terminated SU connection due to the SUs’ own errors. By identifying this metric, it allows for measuring the SUs’ ability of utilizing spectrum opportunities, and helps in designing CRNs by setting correct spectrum sensing parameters. The metric determines the ratio of terminated SU calls because of FAR occurrence to the total secondary call arrivals $\lambda_2$. As mentioned earlier, upon $\lambda_{FAR}$ arrival, an SU has to terminate its own active call if it finishes the channel.
searching process without finding a new idle channel. The \( S_{USTP} \) can be calculated as:

\[
S_{USTP} = \frac{\sum_{i=0}^{N-1} \sum_{j=1}^{N-i} \pi(i,j) \left( \bar{T}_{(i,j-1)} \right)}{\lambda_2}.
\]

(32)

where \( \bar{T}_{(i,j-1)} \) represents the portion of the transition rate from state \((i, j)\) to state \((i, j-1)\) that occurs because of the \( \lambda_{FAR} \).

### 3.6 Simulation and numerical results

In this section, we report results obtained through both theoretical analysis and simulations. We conduct simulations with MATLAB using an event-based approach and Poisson arrival processes. The parameters in simulations and theory are chosen as follows: We set the primary licensed bandwidth as 20 MHz the initial sensing bandwidth is 20 MHz, meaning that an SU carries out spectrum detection over the whole spectrum. However, the continuous sensing bandwidth is set at 2 MHz. The initial sensing time \( = 20 \mu s \). We set the primary and secondary service rates as \( \mu_1 = \mu_2 = 4 \). PU signal power is -91 dBm. The PU power has been set to a low level since SUs should be able to detect even weak PU signals. The noise level is -160 dBm/Hz. To include the effect of the residual interference signal, we set the interference distortion factor to 0.1. We present plots for different performance metrics. It can be observed from all plots that the analytical results are in excellent agreement with the simulation results, which demonstrates the accuracy and validity of the CTMC analytical model.

#### 3.6.1 ROC curves

The receiver operating characteristic (ROC) for both initial and continuous spectrum sensing is shown in Fig. 15. For continuous spectrum sensing, it can be clearly seen that better detection performance is achieved when large values of continuous sensing durations are used. However, for initial sensing, small initial sensing time \( T_1 \) is enough for good detection level. This improved initial detection performance can be attributed to the fact that incoming SUs perform spectrum sensing over the whole PU bandwidth and hence their time bandwidth product is improved.

The impact of the residual interference distortion factor \( \alpha_{SU} \) on the detection and false alarm probabilities is demonstrated by the ROC curves shown in Fig. 15. The SNR value for the initial sensing is 19 dB. However, the SNR values for the continuous
sensing vary depending on the spectrum sensing time duration $T_2$ that affects the time bandwidth product. We assume that the PU signal power is uniformly distributed over the PU channel. It can be seen that the residual interference affects the ROC curves, as the ROC performance drops significantly with increases in $\alpha_{SU}$ values. We can also see from the figure that the curve for the initial sensing with spectrum sensing time duration $T_1 = 10\mu s$ is identical with the curve for continuous sensing with spectrum sensing time duration $T_2 = 100\mu s$ and $\alpha_{SU} = 0$. This is due to the fact that their time bandwidth products are the same and equal to 200. Fig. 15 also shows the effect of the residual interference distortion factor $\alpha_{SU}$ on the false alarm and detection probabilities.

### 3.6.2 Effect of the interference tolerance $\hat{T}_{TOL}$

We start the analysis by investigating the impact of the PU interference tolerance on the performance of the CRN. We tested the cases of $\hat{T}_{TOL}=0, 1, 2, 3, 4, 5$ slots. By recalling the fact that PU detection starts from the first full $\hat{T}_{TOL}$ slots, if $\hat{T}_{TOL}=0$, PUs
tolerate only very little interference from an SU and strict interference constraints should be satisfied. This means that the SU should leave the channel immediately upon the arrival of a PU. However, in practice this is not possible since the SU user needs some sensing time to detect the presence of the PU. In Fig. 16, we compare the PU successful probability $PU_{SP}$ for different values of $T_{TOL}$ and against the PU arrival rate $\lambda_1$. We can observe from the figure the improvement in PU performance as we relax the interference constraint by increasing $\hat{T}_{TOL}$. We can also observe that the $PU_{SP}$ drops significantly when $\hat{T}_{TOL}=0$. However, when $\hat{T}_{TOL}$ is sufficiently large, PU performance starts to improve, indicating that SUs are getting enough sensing time to complete the process of detecting incoming PUs. This implies that the $\hat{T}_{TOL}$ has to be chosen in a way that meets the PUs interference constraint, and at the same time also maximizes the CRN performance. Fig. 16 also presents the impact of the primary arrival rate $\lambda_1$ on the CRN’s performance in terms of $PU_{SP}$. It can be easily seen that there are peak values in $PU_{SP}$ curves. As $\lambda_1$ increases, the $PU_{SP}$ improves. However, as $\lambda_1$ becomes

Fig. 16. Primary successful probability, $\lambda_2 = 3.5, \mu_1 = \mu_2 = 4, N = 3.$
Fig. 17. Primary forced termination probability, $\lambda_2 = 3.5, \mu_1 = \mu_2 = 4, N = 3$.

large, most of the channels will be occupied by PUs, thus increasing the possibility of collisions with SUs who move away from their channels because of the $\lambda_{FAR}$.

Fig. 17 confirms what we asserted above regarding the impact of $\hat{T}_{TOL}$ on the PU performance. While varying $\lambda_1$ from 1 to 10, the figure compares primary forced termination probability $PU_{FTP}$ for each $\hat{T}_{TOL}$ value. It can be observed that $PU_{FTP}$ monotonically decreases with increasing $\lambda_1$. It is also evident that the $PU_{FTP}$ decreases with the increase of the $\hat{T}_{TOL}$. The reason behind this trend can be attributed to the fact that employing small $\hat{T}_{TOL}$ reduces the SUs’ capability of correctly detecting incoming PUs. In such a case, SU collisions with incoming PUs increases, forcing PUs to leave the network and hence increasing the $PU_{FTP}$. In Fig. 18, we plot the secondary self-termination probability $SU_{STP}$ against $\lambda_1$ for different values of $\hat{T}_{TOL}$. The figure shows that increasing $\hat{T}_{TOL}$ can result in a significant SU performance degradation. For example, when $\hat{T}_{TOL} = 0$, SUs collide with incoming PUs, making life easier for other secondary users since they could find free channels more easily. On the one hand, when $\lambda_1$ is small, PUs will have a smaller network resource share, leaving more opportunities for SUs. For example, SUs which leave their channels due to $\lambda_{FAR}$ could find free channels and therefore reduce the $SU_{STP}$. On the other hand, when $\lambda_1$ is large,
most of the channels will be occupied by PUs, leaving smaller network resources for opportunistic access of SUs. In this case, the effect of $SU_{STP}$ is more noticeable since with FAR occurrence (which trigger SUs to leave their channels) SUs either correctly detect the presence of PUs or collide with them. In both cases this leads to an increase in the $SU_{STP}$. Intuitively PUs are better protected by employing large $T_{TOL}$ as it provides SUs with enough sensing time to complete the detection process.

### 3.6.3 Effect of the false alarm rate

Fig. 19 shows the effect of $\lambda_{FAR}$ on the primary forced termination probability $PU_{FTP}$ for different numbers of channels $N$. It can be seen from Fig. 19 that $PU_{FTP}$ curves for different $N$ have unique minimums at different $\lambda_{FAR}$. The minimum points represent the optimum sensing parameters for SUs that would strongly protect PUs against forced connection termination. According to Fig. 19, the $PU_{FTP}$ decreases with the increase in $\lambda_{FAR}$ until it reaches the minimum point, after which it starts to monotonically increase.

On the one hand, too low $\lambda_{FAR}$ reduces the PU’s performance. The degradation in PU performance is due to the fact that small values of $\lambda_{FAR}$ lead to small $P_{D2}$, and therefore
Fig. 19. Primary forced termination probability, \( \lambda_1 = 7, \lambda_2 = 3.5, \mu_1 = \mu_2 = 4, \hat{T}_{TOL} = 1 \).

Fig. 20. Secondary self-termination probability, \( \lambda_1 = 7, \lambda_2 = 3.5, \mu_1 = \mu_2 = 4, \hat{T}_{TOL} = 1 \).
existing SUs frequently collide with incoming PUs and thereby increase $PU_{FTP}$. On the other hand, with high values of $\lambda_{FAR}$ (and consequently high $P_{D2}$) it would be easy for SUs to detect incoming PUs and therefore avoid collision with them. However, too high $\lambda_{FAR}$ is not good since SUs initiate unnecessary spectrum handoffs, leading to sharp increases in $PU_{FTP}$. As shown in the figure, the gap between $PU_{FTP}$'s curves shrinks as $N$ increases. This reflects the fact that when $N$ is sufficiently large, most of the channels are occupied by PUs because $A_1 > A_2$, and the combined effect of $\lambda_{FAR}$, $P_{M2}$ and $P_{M1}$ forces some PUs to terminate their calls. With further increases in $N$, the effect of $\lambda_{FAR}$ on $PU_{FTP}$ is flat, as there would be enough radio resources to meet demands of SUs. After SUs avoid incoming PUs and initiate a handoff and channel switching process, SUs will most likely find new free channels and hence will not harm existing PUs as there are more channels to accommodate them.

Fig. 20 presents the effect of $\lambda_{FAR}$ on the secondary self-termination probability $SU_{STP}$ for different numbers of channels $N$. It can be seen that there are peak values in the $SU_{STP}$ curves. The following explanation is related to this behavior: on the one hand, when $\lambda_{FAR}$ is relatively low, $SU_{STP}$ increases with the increase in $\lambda_{FAR}$, until it reaches the peak value and then starts to decline. The degradation in SU performance can be explained by the fact that as $\lambda_{FAR}$ increases, a growing number of SUs leave their current channels. If they cannot find new free channels elsewhere they terminate their calls, and thereby increase the $SU_{STP}$. On the other hand, when $\lambda_{FAR}$ is relatively large, the detection probability $P_{D2}$ improves, which enables SUs to detect incoming PUs and move away from their channels. Hence, the proportion of SUs in the system is reduced, leading to a decrease in the $SU_{STP}$. It came as no surprise that increasing $N$ greatly influences the $SU_{STP}$. When $N$ is small, it would not be easier for SUs who initiate spectrum handoff to find new free channels to move to and the SUs have to leave the network. This results in a sharp increase in $SU_{STP}$.

In Fig. 21, for several values of $N$, we investigate the effect of $\lambda_{FAR}$ on the primary successful probability $PU_{SP}$. The plot shows when $\lambda_{FAR}$ is low; increasing $\lambda_{FAR}$ slightly increases $PU_{SP}$ until it reaches some (flat) peak points. The increase is a reflection of the improved detection $P_{D2}$. As can be observed from Fig. 22, increasing $\lambda_{FAR}$ has an obvious negative impact on the secondary successful probability $SU_{SSP}$. The reason for the reduction in $SU_{SSP}$ is that when $\lambda_{FAR}$ is high, SUs increasingly initiate unnecessary spectrum handoff processes. SUs also initiate spectrum handoff processes if they detect the arrival of returning PUs. If SUs cannot find new channels, even though there are some free channel(s), they are forced to terminate their own connections and hence
Fig. 21. Primary successful probability, $\lambda_1 = 7, \lambda_2 = 3.5, \mu_1 = \mu_2 = 4, \hat{T}_{TOL} = 1$.

Fig. 22. Secondary successful probability, $\lambda_1 = 7, \lambda_2 = 3.5, \mu_1 = \mu_2 = 4, \hat{T}_{TOL} = 1$. 
reduce $S U_{SSP}$. The results shown in Fig. 22 indicate that the SU optimal performance is when $\lambda_{FAR}$ is close to zero. However, in practice this value means SUs do not perform spectrum sensing and therefore cannot be used since too low $\lambda_{FAR}$ leads to poor PU performance. Results shown in Fig. 19 confirm that low $\lambda_{FAR}$ s are not the best for PU performance as they do not satisfy their QoS/interference constraint. Note that in order to improve network performance, there is a critical sensing tradeoff to be made. This result is a good motivation and illustrates the significant of optimizing detection parameters so that the effects of the $\lambda_{FAR}$ can be kept within limits that would not harm PUs.

Fig. 23 shows the influence of $\lambda_{FAR}$ on the system resource utilization for different numbers of channels. When $\lambda_{FAR}$ is small, SUs do not frequently initiate handoff processes and the effect of $\lambda_{FAR}$ remains flat. However, as $\lambda_{FAR}$ increases, a growing number of SUs terminate their calls if they cannot find other free channels. This explains the reduction in the system resource utilization. As illustrated in Fig. 23, PUs’ own resource utilization is lower than the overall system resource utilization. However, at high $\lambda_{FAR}$, the PUs’ and system resource utilizations get closer. This indicates that SUs
do not complete their service when $\lambda_{FAR}$ is too high and the network resource is mainly utilized by PUs.

### 3.6.4 Performance under perfect spectrum sensing

The effect of perfect sensing on system resource utilization is shown in Fig. 24 where we plot a set of curves showing network resource utilization versus the secondary arrival rate $\lambda_2$ for different numbers of channels $N$. The dash-dotted curves depict the CRN performance when SUs operate without sensing errors. It can be observed that, due to the absence of false alarms, network resources are better utilized. In this case, all unoccupied PUs’ channels can be opportunistically utilized by SUs. This effect is more noticeable when $\lambda_2$ is high. On the other hand, sensing errors can severely reduce the network resource utilization. The impact of the resource underutilization can be reflected in the huge gap between the resource utilization curves. As $\lambda_2$ increases, a growing number of SUs gain network access. However, $\lambda_{FAR}$ prevents them from using network resources because of the increasing spectrum handoff initiation, which may lead to premature connection termination. With a small $N$, the system resources will be
highly utilized since PUs and SUs will occupy most of the channels. However, when $N$ is large, the radio resources are not fully utilized and the overall system resource utilization is therefore decreased.

### 3.7 Summary and discussions

The effect of the false alarm rates $\lambda_{FAR}$ on the operation of CRNs was investigated. A CTMC-based analytical model to evaluate the performance of CRNs was developed. The proposed model not only includes sensing errors by incoming SUs, but also takes into account the misdetection and false alarm probabilities by ongoing SUs. The presented model is capable of examining other performance parameters such as the effect of interference tolerance $\hat{T}_{TOL}$ among PUs and SUs as well as the effect of SUs’ residual self-interference. Formulas for different performance metrics, including primary and secondary forced termination probabilities as well as secondary self-termination probability, were derived. Furthermore, extensive simulations to validate the accuracy of the analytical model were performed. The simulation results showed excellent agreement with the analytical results.

Results showed that $\lambda_{FAR}$ greatly influences the performance of CRNs by degrading the performance of SUs and reducing network resource utilization. Results have also shown that decreasing the interference tolerance $\hat{T}_{TOL}$ has a negative effect on the performance of PUs. $\hat{T}_{TOL}$ can also reduce the successful probability of PUs and increase their forced termination probability. A similar effect was observed with the increase in the SU residual interference distortion factor. A large amount of residual interference was found to deteriorate the detection probability leading to a reduced PU performance. The incorporation of $\lambda_{FAR}$ into the CTMC model allowed for obtaining exact and accurate state transition probabilities that improves calculation of the performance evaluation measures. The proposed analytical model gave new insight into the operation of CRNs. It can be used to develop practical and more accurate CRN performance evaluation models.
4 Queueing analysis of opportunistic access in CRNs and optimization of spectrum detection parameters

In this chapter, we extend the results obtained in Chapters 2 and 3 and apply queueing theory to analyze the performance of CRNs. The Chapter also investigates the optimization of spectrum sensing parameters. The primary/secondary behavior of the channel usage is modeled by a priority queue where PUs are assigned priority class 1 while SUs are assigned priority class 2. A modified $M/D/1$ queueing system is used to analyze the performance of the PU and the SU in terms of average waiting time and average queuing length. Queuing-theoretic results indicate that the performance of the secondary user depends on the data traffic characteristics of the primary user, and under PU the high arrival rate the average waiting time and average queuing length of the secondary user grow especially when the combined arrival rate approaches the queue utilization factor. The main goal of the spectrum sensing is to maximize the SUs’ utilization of spectrum opportunities while protecting PUs. Optimization of the sensing parameters is a challenging task in spectrum detection.

There is a tradeoff between false alarm and detection probabilities. On the one hand too large detection probability means that SUs lose transmission opportunities due to increased false alarms. On the other hand, too small detection probability means that PUs are not sufficiently protected. However, if detection probability is increased by lowering the detection threshold, then the false alarm probability will increase. The increase in the false alarm probability will decrease the misdetection probability and thus protect the primary user better. We propose a radically new concept where this reduction in transmit opportunities is used to our benefit as a way to reduce collisions among multiple secondary users competing to access the same channel. Numerical results indicate that the secondary users’ maximum throughput is obtained when the random access probabilities are equal to one and the random access is handled instead with false alarm probabilities. This chapter is organized as follows: The system model including the assumptions is presented in Section 4.1. Queueing analysis of opportunistic access in CRNs is detailed in Section 4.2. Detailed optimization of sensing parameters is given
in Section 4.3. Finally, Section 4.4 concludes the chapter by presenting the summary and discussions.

4.1 System model

We consider a time-slotted [98, 99] cognitive wireless network. The network is assumed to operate in ideal channel conditions (e.g., no noise and error-free). The primary user is the owner of the network. When the primary wishes to transmit, it is given priority over the secondary user. All packets are assumed to be one slot in duration, see Fig. 25. Typically, the beginning of the slot is used for signal detection by secondary users and the remainder of the slot can be used for secondary transmissions assuming the channel was decided to be free [100, 101]. In the case of perfect sensing, this leads to no interference to the primary user since the PU status is assumed to be constant for each slot. In this slotted system, it is possible to synchronize with the primary user transmission. We assume perfect time synchronization. Note that in practice perfect synchronization may not always be possible. However, this assumption leads to analytical results.

Fig. 25. Channel time slots (©2009 IEEE).

The queueing model presented in this chapter takes advantage of Kleinrock’s [102] mini-slotted alternating priorities (MSAP) MAC scheme that reduces the overhead associated with large numbers of SUs. We use the M/D/1 queueing model because of its simplicity and ability to accurately model the queueing system. One limitation of the M/D/1 queueing model is that any misdetection from the secondary user side would lead to collision with the primary transmission that would cause retransmission. Hence, the
number of free time slots would decrease as more secondary users collided with the primary one.

4.2 Queueing analysis of opportunistic access in CRNs

A great deal of research in the literature has focused on spectrum sensing as well as resource allocation in CRNs [100], [41]. However, few results on modelling the waiting time analysis in these networks have been reported, especially about time-slotted systems. We perform waiting time analysis of opportunistic access in a time-slotted system by utilizing queueing to model a time-slotted cognitive radio system with one primary and one secondary link (a similar model was used in [99]). Namely, the primary user has higher priority than the secondary user; therefore the secondary user has to leave the channel if the primary arrives. Primary users are assigned priority class 1 while secondary users are assigned priority class 2. We use modified a $M/D/1$ system to analyze the network performance. Poisson processes are assumed for packet arrivals so that the interarrival times are exponentially distributed. The primary user arrival rate is $\lambda_1$ and secondary user arrival rate is $\lambda_2$. Steady-state distribution of the queue lengths as well as mean delay time are derived. The theoretical results from priority queueing assume perfect sensing, i.e., the secondary user will never use the channel if it is occupied by the primary user. In practise, sensing is never perfect and there are two kinds of errors: false alarm and missed detections. We study imperfect sensing by using Monte Carlo simulations. Note that the transmission of packets can only start at the beginning of the slot, so that even if a packet arrives at the middle of the slot, it has to wait for half of the duration of the slot even if the channel is free. Infinite buffers are assumed.

4.2.1 Analysis based on priority queuing

Fig. 26 presents a wireless network with primary users (licensed) and secondary users (unlicensed or cognitive). The situation shown in Fig. 26 can also be interpreted to be an example of device-to-device (D2D) communication [103], [104]. D2D refers to a situation where traffic does not go through base station, and it is possible that the terminal(s) opportunistically use(s) free channels available for D2D communication [103], [104], [105]. Fig. 27 illustrates the priority servicing used for modelling packet arrivals and departures in the CRN. The nomenclature is given in Table 2.
The waiting time of a packet consists of three parts: time until the beginning of the next time slot, time spent in the queue waiting for the service to begin, and the average service time (transmission time). For both classes, the packets are served according to the first-come first-served discipline (FCFS), but a packet of class 2 (at the secondary user queues) may start its transmission at the beginning of a time slot only if there are no packets of class 1 (i.e., empty primary user queues) in the network. Given the fact that packets arrive according to the Poisson process and that the system time is slotted with a fixed unit time slot, it is straightforward to estimate the average time spent by a newly arrived packet waiting in a queue for the start of the next slot: on average, a packet has to spend $1/2$ slot waiting for the next slot to start. Recall the fact that $M/D/1$ (where services are deterministic and last exactly one slot regardless of user priority) is a special case of $M/G/1$, we start with the deriving expressions for the $M/G/1$, and then we obtain the expressions for the $M/D/1$ as special case. This is done by starting from the well-known analysis result of the $M/G/1$, known as Pollaczek-Khintchine
Table 2. Queuing system parameters (©2009 IEEE).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>Primary arrival rate</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>Secondary arrival rate</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Service time</td>
</tr>
<tr>
<td>$X = E[X] = \frac{1}{\mu}$</td>
<td>Average service time</td>
</tr>
<tr>
<td>$N_{Q}^1$</td>
<td>Average number of packets in PU queue</td>
</tr>
<tr>
<td>$N_{Q}^2$</td>
<td>Average number of packets in SU queue</td>
</tr>
<tr>
<td>$W_{Q}^1$</td>
<td>Average PU queueing time</td>
</tr>
<tr>
<td>$W_{Q}^2$</td>
<td>Average SU queueing time</td>
</tr>
<tr>
<td>$W^1$</td>
<td>Total time spent by PU packet in the system</td>
</tr>
<tr>
<td>$W^2$</td>
<td>Total time spent by SU packet in the system</td>
</tr>
<tr>
<td>$T_D$</td>
<td>Waiting time until the beginning of a slot</td>
</tr>
<tr>
<td>$\rho_P = \frac{\lambda_1}{\mu}$</td>
<td>PU queue utilization factor</td>
</tr>
<tr>
<td>$\rho_S = \frac{\lambda_2}{\mu}$</td>
<td>SU queue utilization factor</td>
</tr>
</tbody>
</table>

and following a similar approach as in [106]. The sojourn time of a secondary user packet depends not only on the packets found upon arrival ($N_{Q}^1$ and $N_{Q}^2$), but also on subsequent arrivals at the primary user queue. We first derive expression for the waiting time $W_{Q}^1$ for the primary user

$$ W_{Q}^1 = T_D + \frac{1}{\mu} N_{Q}^1. $$

(33)

We can eliminate the $N_{Q}^1$ from the equation by using Little’s theorem [106], $N_{Q}^1 = \lambda_1 W_{Q}^1$.

$$ W_{Q}^1 = T_D + \frac{1}{\mu} W_{Q}^1 = T_D + \rho_P W_{Q}^1, $$

(34)

where $T_D$ is the delay time that the arrived packet has to wait until the beginning of the next slot and we got

$$ W_{Q}^1 = \frac{T_D}{1 - \rho_P}. $$

(35)

For the secondary user, the waiting time $W_{Q}^2$ of a newly arrived packet depends not only on the packets found upon arrival in primary and secondary queues ($N_{Q}^1$ and $N_{Q}^2$), but also on subsequent arrivals at the primary user queue. Therefore we have to include this delay in the computation of the $W_{Q}^2$, which is given by

$$ W_{Q}^2 = T_D + \frac{1}{\mu} N_{Q}^1 + \frac{1}{\mu} N_{Q}^2 + \frac{1}{\mu} \lambda_1 W_{Q}^2. $$

(36)
and by using Little’s theorem, we obtain

\[ W_Q^2 = T_D + \frac{1}{\rho} \lambda_1 W_Q^1 + \frac{1}{\rho} \lambda_2 W_Q^2 + \frac{1}{\rho} \lambda_1 \lambda_1 W_Q^2 = T_D + \rho W_Q^1 + \rho W_Q^2 + \rho W_Q^2. \]  

(37)

\[ W_Q^2 = \frac{T_D + \rho W_Q^1}{1 - \rho}. \]  

(38)

\[ W_Q^2 = \frac{T_D}{(1-\rho)(1-\rho-\rho^2)}. \]  

(39)

Recalling the fact that the total packet delay (time spent in the system) is given by summation of the waiting time in the queue and the average service time of the packet, we can express the average delay per packet for the primary user

\[ W^1 = \bar{X} + \frac{T_D}{1-\rho_1}, \]  

(40)

and that the average delay per packet for the secondary user is given by

\[ W^2 = \bar{X} + \frac{T_D}{(1-\rho_1)(1-\rho_1-\rho_2)}. \]  

(41)

We are now interested in addressing the special case where time is slotted with the deterministic service time of one slot and the fact that packet service can only start at the beginning of a slot. We have already assumed that the service time is one slot and that a newly arrived packet has to wait for \( \frac{1}{2} \) slot before the beginning of a slot. We can substitute the corresponding values in equation (40) and equation (41) to obtain

\[ W^1 = 1 + \frac{1/2}{1-\rho_1}, \]  

(42)

and

\[ W^2 = 1 + \frac{1/2}{(1-\rho_1)(1-\rho_1-\rho_2)}. \]  

(43)

Applying Little’s formula to \( W^1 \) and \( W^2 \), we get the expected number of packets for the primary user \( N^1 \) and the secondary user \( N^2 \):

\[ N^1 = \lambda_1 + \frac{\lambda_1}{2(1-\rho_1)}, \]  

(44)

\[ N^2 = \lambda_2 + \frac{\lambda_2}{2(1-\rho_1)(1-\rho_1-\rho_2)}. \]  

(45)
We consider two scenarios to demonstrate the validity of the analysis: in the first the primary and secondary packets arrive with an equal rate so that $\lambda_1 = \lambda_2$. In the second scenario, the primary user rate is fixed while the secondary user rate is varied. Fig. 28 presents the delay for the primary and secondary users. Solid lines are theoretical results and markers denote simulation results. We can see that when the arrival rate is small, the average waiting time for the primary and secondary queues are comparable. However, as the arrival rates increase, the gap between the primary and secondary average waiting time starts to increase. For higher values of the arrival rates, the primary waiting time stays at almost the same value. Fig. 29 presents the delay for the primary and secondary users. Solid lines are theoretical results and markers denote simulation results. It can be seen that theoretical and simulation results match very well.

Fig. 28. Average waiting time for the PU and the SU. Packets arrive to each priority queue with equal rates. Solid lines represent the theoretical results and the markers (o and +) denote simulation results (©2009 IEEE).
4.2.2 Imperfect sensing

The false alarm probability is the probability that a free slot is decided to be occupied. The detection probability is the probability that an occupied slot is detected correctly as an occupied slot. Typically, there is a tradeoff between these two probabilities. The performance of the priority queueing with imperfect sensing has been evaluated using Monte Carlo simulations. Fig. 30 shows the primary user packet delay when the probability of detection $P_D = 0.7$ and the probability of false alarm $P_{FA} = 0.25$. Similarly, Fig. 31 shows the secondary user packet delay when the probability of detection $P_D = 0.7$ and the probability of false alarm $P_{FA} = 0.25$. It can be seen that the imperfect sensing conditions have affected the performance of both primary and secondary user delays.
Fig. 30. Primary user packet delay, $P_{FA} = 0.25$, $P_{D} = 0.7$ (©2009 IEEE).

Fig. 31. Secondary user delay, $P_{FA} = 0.25$, $P_{D} = 0.7$ (©2009 IEEE).
4.3 Optimizing spectrum detection parameters

As discussed earlier, the energy detector [12] is one of the most widely used detectors due to its simplicity. However, although the energy detector is optimal for detecting Gaussian signals in Gaussian noise, it is not optimal for practical signals. In energy detection, we measure the received signal power or energy and compare it to a threshold [10]. If the threshold is exceeded, it is decided that a signal or signals are present. If detection probability is increased by lowering the detection threshold, then the false alarm probability will increase. This means that there is a tradeoff between these two probabilities. Too large a detection probability means that secondary users lose transmission opportunities due to false alarms. Too a small detection probability means that the primary user is not sufficiently protected. However, in spectrum sensing the detection probability is more important than false alarm probability due to limits on allowed interference caused to the primary user.

We study optimization of the detector operating point in multiple secondary case so that secondary user throughput is maximized with the constraint that the primary user’s QoS is maintained. Specifically, the primary user packet delay is required to be less than a given limit. We assume one primary user $P$ and $M$ secondary users. Multiple access among secondary users is handled by random access similar to slotted ALOHA so that the secondary users transmit using some channel access probability $\psi$, assuming that channel was decided to be free [101]. Note that even if the slot is found to be free, due to random access probabilities it may not always be used [101]. Optimizing the used operating point of the detector (the used false alarm probability and detection probability) has been performed in [107] for a case with one primary and one secondary link. Therein, the best tradeoff between false alarm and missed detections is studied (also [108]). The packet arrival process for primary users is the Bernoulli processes [109] with mean $\lambda_1$ (packets/slot) and infinite buffers. Secondary users are assumed to always have packets to send (saturated situation). To protect the primary user, it is required that the primary user average packet delay $D_1$ is less than some upper limit, i.e, [101]

$$D_1 \leq D_{\text{max}}$$

(46)

where $D_{\text{max}}$ is the maximum allowed average delay. The primary user delay is [101]

$$D_1 = \frac{1-A_1}{\mu_1-A_1}$$

(47)
where

\[ \mu_1 = \prod_{i=1}^{M} P_{D,i} + \sum_{\nu \subseteq M, \nu \neq \emptyset} \prod_{j \in \nu} (1 - P_{D,j}) (1 - \psi_i) \times \prod_{j \notin \nu} P_{D,j}, \]  

(48)

where \( P_{D,i} \) is the detection probability of terminal \( i \) and \( M = \{1, 2, \cdots, M\} \). The probability that the primary user queue is empty is \( \varsigma = 1 - \lambda_1 / \mu_1 \). Secondary users’ saturated throughput are [101, 110]

\[ \mu_{i,j} = \theta_i \prod_{k \neq i} (1 - \theta_k) \varsigma, \]  

(49)

where \( \theta_i = \psi_i (1 - P_{FA,i}) \) is the probability of transmission in an empty slot by secondary users. Probability \( \psi_i \) is the random access probability for secondary user \( i \), i.e., the probability that the user with packets to send will send a packet in a slot that is thought to be free from the primary user.

The receiver operating characteristic curves show detection probability as a function of the false alarm probability. This is exactly what we need for optimization purposes. Let us assume the energy detector [10]. In practise, more advanced methods can be used, using knowledge about the signal structure. However, a simple energy detector will suffice for showing the gains from optimizing the detection parameters. When detecting the Gaussian signal in Gaussian noise and using \( N \) real-valued samples, detection probability is

\[ P_D = 1 - \text{chi2cdf} \left( \frac{\text{chi2inv}(1 - P_{FA}, N)}{1 + \gamma}, N \right), \]  

(50)

where

\[ \gamma = \frac{\sigma_N^2}{\sigma_S^2}, \]

is the ratio of signal variance to noise variance, \( \text{chi2cdf} \) is the chi-square cumulative distribution function and \( \text{chi2inv} \) is the inverse of the chi-square cumulative distribution function. It can be shown that for \( N = 2 \)

\[ P_D = P_{FA}^{1/(1+\gamma)}. \]  

(51)

Fig. 32 shows the ROC-curves for various values of \( N \) for \( \gamma = 5 \) dB.
4.3.1 Optimization

Assume first that the $\psi_i = p$, $M = 2$, and $N = 2$. Also assume that detection and false alarm probabilities are equal also for all users, i.e., $P_{D,i} = P_D$ and $P_{FA,i} = P_{FA}$. We get for the primary user delay

$$D_P = \frac{1 - \lambda_1}{p_{FA}^{2(1+\gamma)} + 2 \left(1 - p_{FA}^{1(1+\gamma)}\right)(1-p)p_{FA}^{1(1+\gamma)} + \left(1 - p_{FA}^{1(1+\gamma)}\right)^2 (1-p)^2 - \lambda_1}.$$  

(52)
For the secondary user throughput, we get,

\[
\mu_2 = p(1 - P_{FA})(1 - p(1 - P_{FA})) \times \left(1 - \lambda_1 \left\{ \frac{2(1 - P_{FA}^{(1+\gamma)})}{1 - P_{FA}^{(1+\gamma)}} \left(1 - p \right) + \frac{P_{FA}^{(1+\gamma)}}{2(1 - p)^2 + P_{FA}^{2(1+\gamma)}} \right\} \right) .
\tag{53}
\]

Fig. 33 shows the throughput when \( \lambda_1 = 0.2 \), \( N = 2 \), and \( \gamma = 5 \) dB. Fig. 34 shows the throughput when \( \lambda_1 = 0.8 \), \( N = 2 \), and \( \gamma = 5 \) dB. The optimal operating points are marked with "o". It can be seen that the numerical results indicate that the maximum of the throughput occurs when \( p = 1 \), i.e., \( \psi = 1 \) (with properly selected \( P_{FA} \)). These results do not take the primary user delay constraint into account (although it can be easily handled when \( \psi = 1 \)). Intuitively, \( \psi = 1 \) results in best performance because in that case random access is handled by increasing false alarm probabilities. To increase false alarm probabilities means increasing the detection probability. However, due to the numerical nature of the search, we do not claim that \( \psi = 1 \) always results in the best system design.

Fig. 33. Secondary users’ saturated throughput, \( \lambda_1 = 0.2 \), \( N = 2 \), \( \gamma = 5 \) dB (©2009 IEEE).
Let us now assume $\psi = 1$ so that random access is handled purely by non-zero false alarm probability. We get for arbitrary $M$

$$D_p = \frac{1 - \lambda_1}{P_{FA}^{M/(1+\gamma)} - \lambda_1} \quad (54)$$

Due to the constraint on the primary user delay, we get that the probability of false alarm must satisfy

$$P_{FA} \geq \left(\frac{1 - \lambda_1 + D_{\text{max}}\lambda_1}{D_{\text{max}}}\right)^{(1+\gamma)/M} \quad (55)$$

which requires knowledge (or estimation) of $\lambda_1$, $\gamma$ and $M$. The secondary users’ saturated throughput is,

$$\mu_2 = (1 - P_{FA}) P_{FA}^{M-1} \left(1 - \frac{\lambda_1}{P_{FA}^{M/(1+\gamma)}}\right) \quad (56)$$

Let us assume that $\gamma$ is large. Now we find that the maximum occurs when $P_{FA} \sim (M - 1)/M$. 

---

Fig. 34. Secondary users’ saturated throughput, $\lambda_1 = 0.8$, $N = 2$, $\gamma = 5$ dB (©2009 IEEE).

---

94
This result does not need any other information than the number of secondary users. However, in practise we need to guarantee that (55) is also satisfied. This means that the optimal value cannot always be utilized, especially when the $\gamma$ is small.

Let us assume $\lambda_1 = 0.2$. For $\gamma = 12.18$ dB we get that for $P_{FA} = 0.02$ the detection probability is $P_D = 0.8$, i.e., we get the same operating point as in [101]. Fig. 35 shows the primary user delays when $M = 2$. The vertical lines indicate points where the primary user delays are equal to the maximum value of $D_{max} = 1.3$. It can be observed that the simulation results agree well with the theoretical ones. Fig. 36 shows the corresponding secondary throughputs. It can be seen that when SNR is small, the primary user delay constraint restricts the set of usable values for false alarm probability so that the optimal operating point (regarding only secondary user throughput) cannot be used. Fig. 37 shows the secondary throughputs when $M = 4$.

![Primary user delay as a function of the false alarm, $M = 2$, $\lambda_1 = 0.2$, $D_{max} = 1.3$ (©2009 IEEE).](image)

Fig. 35. Primary user delay as a function of the false alarm, $M = 2$, $\lambda_1 = 0.2$, $D_{max} = 1.3$ (©2009 IEEE).
Fig. 36. Secondary users’ throughput as a function of the false alarm, $M = 2$, $\lambda_1 = 0.2$, $D_{\text{max}} = 1.3$ (©2009 IEEE).

Fig. 37. Secondary users’ throughput as a function of the false alarm, $M = 4$, $\lambda_1 = 0.2$, $D_{\text{max}} = 1.3$ (©2009 IEEE).
4.4 Summary and discussion

Modeling the performance of opportunistic access in CRNs using priority queuing was considered. The waiting time and queue length were theoretically derived using the modified M/D/1 model. The results demonstrated that priority queuing is suitable for studying the feasibility of opportunistic access in slotted systems. PU data traffic characteristics can affect the performance of the SU. Under a high PU arrival rate, the average waiting time and average queueing length of the SU grow, especially when the combined arrival rate approaches the queue utilization factor. Numerical optimization of spectrum sensing parameters was also presented. The false alarm probability was optimized so that SU’s throughput is maximized with the constraint that PU quality of service is maintained. Numerical results indicate a good performance maximizing SU’s throughput can be achieved when the random access probabilities are equal to one and random access is handled instead with false alarm probabilities. It was also demonstrated that when the signal-to-noise ratio is large, there exists a very simple result for the optimal value for the used false alarm probability. Simulation results showed good agreement with the theoretical results and validated the benefits of the optimization.
5 Radio resource allocations in wireless networks

In this chapter, we investigate the problem of resource allocations in wireless networks and study improving wireless networks coverage by using relaying nodes. Wireless networks may cooperate either because they cannot deal with resource allocation demands alone or because they can reduce their cost by working together. One of the basic issues that needs examination is the importance of fair resource sharing among relaying nodes in wireless networks. Sharing of the downlink capacity between multiple nodes using a noncooperative approach is inefficient when the radio resource is scarce. If upstream nodes manipulate their location to exploit the downlink capacity, downstream nodes will suffer disproportionately. The undesirable properties can be avoided by means of a cooperative agreement in which all nodes share the radio resources equally, where downstream nodes are allowed to pay compensation to prevent upstream nodes from exploiting the downlink capacity while encouraging them to cooperate. We describe a novel technique for providing resource allocation mechanisms in a multihop relaying network. The resource allocation problem is formulated as a cooperative game using the Nash bargaining solution (NBS), which allows mobile nodes to fairly share a downlink bandwidth among themselves. We also propose a cooperative game model that addresses the problem of the resource allocation in wireless networks by applying cooperative game theory to identify stable resource allocations, under which wireless networks find it beneficial to cooperate. The proposed model enables a mobile user to split its application between different networks when the applications can not be handled by a single network. The game theory approach considered in this chapter attempted to provide fair resource sharing among competing users (and also how to achieve that). The fairness issue discussed in this chapter can be characterized by how competing users should share the bottleneck resources [111]. The results suggest that an allocation that maximizes revenues will make the cooperation attractive to all networks.

The rest of the chapter is organized as follows: In Section 5.1 we present the related work. Section 5.2 introduces the cooperative multihop relaying problem based on the NBS along with the problem formulation, simulations and numerical results. Resource allocation as a cooperative game among multiple wireless networks is presented.
in Section 5.3. Finally, Section 5.4 concludes the chapter providing summary and discussions on the resource allocation.

5.1 Related work

There has been a significant amount of studies focused on providing solutions to the resource allocation problems in wireless networks. A particular emphasis has been on providing multihop capability to the current networks. Multihop communication allows mobile nodes to maintain network connectivity with base stations using intermediate nodes that act as relays [112–114]. Therefore, the existence of a direct communication link between the base station and the mobile node is not required. Cellular and ad hoc networks integration [115, 116] also facilitates relaying-based communication. Several research studies [117, 118] demonstrated that multihop relaying is a viable solution for extending networks coverage. However, these benefits would vanish if mobile nodes decided not to cooperate in relaying each other’s data. Selfish nodes might pose a serious threat to the fair usage of network resources by not allowing other nodes to use them in data forwarding. Mobile nodes can choose either to cooperate or not to cooperate depending on the perceptions of their relative benefits. In many cases, the main reasons why mobile nodes would prefer not to cooperate include resource and battery level. Mobile nodes which decide to cooperate can be encouraged by introducing incentives for collaboration. They can be compensated in some way for giving up some of their bandwidths and for the loss in their energy resources. The studies presented in [119, 120] suggest that employing incentive mechanisms makes collaboration a rational choice for selfish nodes.

Game theory has been used in modelling wireless communication systems [121]. It can be divided into two broad areas: cooperative games and non-cooperative games. A cooperative game is a game in which the players have the option of planning as a group in advance of choosing their actions. NBS [122] has been used for solving cooperative game models. It has also been used for load balancing [123], bandwidth allocation [124] and fairness in network optimal flow control [125]. In contrast to the significant attention given to end-to-end throughput improvement and resource allocation in multihop communications [126, 127], less attention has been paid to radio resource sharing among the mobile nodes. In this chapter, however, the issue of fair resource distribution between multiple mobile nodes in a multihop relaying environment is investigated.
5.2 Cooperative multihop relaying problem based on the NBS

The NBS attempts to identify a resource division for mobile nodes that is both fair and efficient. It provides an acceptable outcome for all players. Approaches to bargaining fall into two divisions: axiomatic bargaining and strategic bargaining. We will focus on the axiomatic solution, which assumes some desirable properties that must be satisfied by the outcome. In order to describe the cooperative game approach for the resource allocation problem, we will first introduce the notation of "upstream mobile node" and "downstream mobile node," which are defined according to the relative position of mobile nodes in the path from the very first node to the last node. The mobile nodes are numbered 1, 2, ..., L. We assume that mobile nodes are connected via direct point-to-point links to form a linear multihop. Therefore only adjacent mobile nodes can communicate with each other.

Distribution of the downlink capacity among mobile nodes located along the linear multihop link may result in the exploitation of radio resources by upstream nodes, which affect downstream nodes. The obvious solution to the conflicts is cooperation between nodes along the multihop path. Reaching an agreement provides downstream nodes with protection against bandwidth capacity exploitation by upstream neighbors. In this context, the degree of cooperation is influenced by the node’s data rate, selfishness and the relative location of the mobile node. Let $C$ be the downlink radio resource capacity that is to be shared between mobile nodes. Sharing takes place when upstream nodes do not use the entire downlink radio capacity, instead leaving some of it for the downstream nodes. The resource allocation problem can be stated as: Given a downlink radio capacity $C$ and a number of mobile nodes $L$ arranged in a linear multihop, how can we optimally divide the downlink capacity among the mobile nodes.

**Definition 1** A bargaining problem is a pair $(S,d)$ where $S \subset \mathbb{R}^L$ is the set of utility vectors ($S$ is often called the "feasible set" of utilities) that the players can achieve if they cooperate and $d \subset S$ the utility if the bargaining breaks down (i.e., disagreement utility). We assume that $S$ is compact and convex. The set of all bargaining problems is denoted by $B$.

**Definition 2** A bargaining solution is a function $F$ that assigns a non-empty set of feasible utility vectors to all bargaining problems. Formally, a bargaining solution is defined as $F : B \rightarrow \mathbb{R}^L$ such that $F(S,d) \subset S$ for each $(S,d)$. 
Definition 3 Pareto-optimal: Pareto optimality is a measure of efficiency. An outcome of a game is Pareto-optimal if none of the players can increase their payoff without decreasing the payoff of at least one of the other players.

We now turn to formalizing the axioms of fairness. Nash [122] gave four axioms (Nash axioms) that any bargaining solution should satisfy. Once we have formalized these axioms, we will define the NBS that is the unique bargaining solution that satisfies these axioms. A solution which satisfies the following four properties is called the Nash Bargaining Solution.

Symmetry: Symmetry means that if the players’ utilities are exactly the same then they should get symmetric payoffs. A bargaining solution \( F \) satisfies symmetry if for all symmetric problems \( (S, d), (s_1, s_2) \subset F(S, d) \iff (s_2, s_1) \subset F(S, d) \).

Pareto Efficiency: A fair bargain \( F(S, d) \) should give the maximum allocation of utility to the players.

Independence of irrelevant alternatives: The outcome should be independent of irrelevant alternatives.

Invariance with respect to utility transformations: A utility function specifies a player’s preferences. However, different utility functions can be used to model the same preferences. The final outcome should not depend on which of these equivalent utility representations is used.

Definition 4 The NBS is the function \( F \) that assigns a non-empty set of feasible utility vectors to all bargaining problems. Formally, a bargaining solution is defined as \( F : (S, d) \to \mathbb{R}^L \) that satisfies the following maximization problem

\[
\max_{\{s \in S : s \geq d\}} \prod_{i \in N} (s_i - d_i) \tag{57}
\]

With this model, each mobile node needs to know the list of upstream nodes so that it can compute its relative location along the multihop path. It should be noted that the linear multihop link described above favoured upstream nodes, which can use their relative location to control the downstream resource flow. For each mobile node, we assign a list of upstream nodes from the mobile node \( i \). We denote the list of mobile nodes that are upstream from noders \( i \) by \( \text{up}(i) \). The first node in the linear multihop arrangement is the one that is adjacent to the base station. Therefore \( \text{up}(1) = \{0\} \), and \( \text{up}(2) = \{1\} \), and the upstream list for the last mobile node \( L \) is \( \text{up}(L) = \{1, 2, \ldots, L-1\} \). Based on this arrangement, the maximum radio resources available to a mobile node \( i \)
equals to the downlink capacity minus radio resources consumed by upstream nodes $u_p(i)$. Let $f(i)$ represent the bandwidth flow from upstream nodes $u_p(i)$ to node $i$. Let $r = (r_1, r_2, ..., r_L)$ be an $L$-dimensional vector representing the radio resources allocated to each mobile node. The mobile node bandwidth flow is given by

$$f_i = C - \sum_{j \in u_p(i)} r_j \quad (58)$$

such that for each mobile nodes, $i$ the resource allocated/consumed $r_i$ must satisfy $r_i \leq f_i$.

### 5.2.1 The non-cooperative solution

The non-cooperative or disagreement solution is defined as the solution which can result if bargaining breaks down. At this solution, each mobile node maximizes its own utility regardless of the results for other mobile nodes. This disagreement point $r_i^0$ represents the minimum guaranteed resource received by each of the mobile nodes if no agreement is reached. Allocating resources according to this approach without considering the bandwidth demand of other nodes will not typically result in a Pareto-optimal solution. With the disagreement, no incentives transfer would be made if mobile nodes cannot reach an agreement on how to distribute the downlink capacity $C$ among themselves. Then we have the following resource allocation constraints $r_i^0 \leq r_i \leq f_i$. We assume that there is at least one point that mobile nodes prefer to the disagreement point.

### 5.2.2 Cooperative solution using incentives

The non-cooperative solution mentioned above is not an efficient solution because upstream mobile nodes may try to manipulate their location to obtain an unfair higher share of the downlink capacity, ignoring the bandwidth demands of downstream nodes. Cooperation cannot be forced upon unwilling mobile nodes, it is a voluntary process. However, the use of any payment or some form of compensation may facilitate the cooperation process. One way of encouraging cooperation is to use incentives to persuade upstream nodes to negotiate/cooperate, on the basis that getting some compensation may be more effective than exploiting the entire downlink capacity. A good example of this is in [128], which uses incentives and compensation to encourage cooperation between mobile nodes. In the absence of any agreement between mobile nodes, it would be impossible to stop upstream nodes from using more radio resources.
Therefore an agreement between mobile nodes located on the multihop path, specifying the required increase in the bandwidth to be released by the upstream node, combined with the downstream node’s offer to pay incentives or compensation for the increased bandwidth, could lead to a Pareto-optimal solution.

### 5.2.3 Cooperative multihop relaying formulation

The NBS described above can be used for representing the network operating point. However, in order to apply this concept, we have to choose utility functions for players (nodes). Since the bandwidth flow for downstream nodes depends on bandwidth shares of upstream nodes, the disagreement solution is usually of the form $d = (d_1, 0, \ldots, 0)$, i.e., all the downstream nodes will receive no utility in this solution. Let $e = (1, \ldots, 1)^T$ denote the unit vector. If we choose the utility for the $i$-th node to be his bandwidth share $r_i$, then (1) becomes equivalent to the following non-linear optimization problem:

$$\max \ln(r_1 - C) + \sum_{i=2}^{L} \ln r_i $$

subject to

$$e^T \cdot r \leq C, \quad r^0 \leq r,$$

which has at most the trivial solution $r^* = d$. Therefore, we have to apply some other concept of the utility function. Namely, we can define $h_i(r_i)$ as the pure utility function of the $i$-th node from consumption of his bandwidth flow share $r_i$, $v_i(r_i)$ as his standard payment for this share, and $w_i(r_i)$ as his incentive (network) payment for this share. Hence, if the $i$-th node receives the bandwidth flow share $r_i$, his value of utility is the following:

$$u_i(r) = h_i(r_i) - v_i(r_i) - w_i(r_i) + \sum_{j>i} \sigma_{ij} w_j(r_j),$$

where $\sigma_{ij}$ is the incentive share of the $i$-th upstream node received from the $j$-th downstream node such that

$$\sum_{i<j} \sigma_{ij} = 1 \quad \text{for} \quad j = 2, \ldots, L.$$

For example, equal shares correspond to the choice $\sigma_{ij} = 1/(j-1)$ for $i < j$. We suppose that $h_i(0) = v_i(0) = w_i(0) = 0$ for $i = 1, \ldots, L$, $w_1(r_1) \equiv 0$, and also that, for each $i$, there exists $r'_i > r^0_i$ such that $h_i(r_i) - v_i(r_i) - w_i(r_i) > 0$ if $0 < r_i \leq r'_i$. Then we obtain the same disagreement solution $d = (d_1, 0, \ldots, 0)$, where $d_1 = h_1(C) - v_1(C)$. It means that the
downstream nodes receive nothing if the first node takes the whole bandwidth flow $C$. We conclude that the NBS solution can be derived from the following non-linear optimization problem:

$$\max \ln(u_1(r) - d_1) + \sum_{i=2}^{L} \ln u_i(r)$$
subject to
$$e^T r \leq C, \ r^0 \leq r.$$  

(61)

5.2.4 Multihop relaying problem solution

The choice of the solution method for problem (61) depends on the properties of the utility functions $u_i$. More precisely, if we can guarantee the concavity of the cost function in (61) over the feasible set, then it becomes a concave minimization problem and there exists a number of effective iterative methods to find its solution. Note that the cost function involves the implicit constraints $u_1(r) \geq d_1$ and $u_i(r) \geq 0$ for $i = 2, \ldots, L$. It is easy to see that, under these constraints, the concavity of utility functions $u_i$ implies concavity of the cost function. From (60) it follows that this is the case if, for example, $h_i$ is concave, $v_i$ is convex and $w_i$ is affine. One of the well-known utility functions used in communication networks is given by $h_i(r_i) = \omega_i \ln \varphi(r_i)$ [129, 130], where $r_i$ is the allocated bandwidth to node $i$ and $\omega_i$ is node $i$’s budget (or bandwidth sensitivity), and $\varphi : \mathbb{R} \to \mathbb{R}$ is an increasing function.

The most popular and simple choice is $\varphi(r_i) = r_i$, but if we intend to ensure continuity and positivity properties, it can be somewhat modified. In particular, we can set $\varphi(r_i) = 1 + r_i/r_i^0$, thus providing a proportionally fair resource allocation. Then the above properties of $h_i$ is satisfied and we can find a solution of problem (61) with the help of first or second order descent methods. In order to illustrate the approach described, we have carried out numerical experiments with the simplest gradient projection method with exact line search; see e.g., [131, Chapter III, Section 2]. We recall that this method, being applied to the optimization problem

$$\max \ f(x)$$
$$\text{subject to}$$
$$x \in X,$$
can be written as follows:

\[ x^{k+1} = x^k + \mu_k d^k, d^k = y^k - x^k, \]
\[ \mu_k = \arg\max \{ f(x^k + \mu d^k) \mid 0 \leq \mu \leq 1 \}; \]
\[ y^k = \pi_X \{ x^k + \lambda f'(x^k) \}, \]

where \( x^0 \) is an arbitrary point of \( X \), \( \lambda > 0 \). The functions were chosen as follows:

\[ h_i(r_i) = \omega_i \ln(\phi(1 + r_i/r_{0i})), \]
\[ v_i(r_i) = \alpha_i r_i, w_i(r_i) = \beta_i(r_i), \]

where \( \omega_i > 0, \alpha_i > 0, \beta_i \geq 0 \) for \( i = 1, \ldots, N \) and \( \beta_1 = 1 \).

### 5.2.5 Numerical results

The above gradient projection method was implemented using one-dimensional minimization by the dichotomy method. The values of \( \alpha_i \) and \( \beta_i \) were generated randomly and \( \omega_i \) was chosen to ensure the consistency of constraints and the domain of the cost function. Next, we used the standard accuracy measure \( ||d_k|| = ||y_k - x_k|| \). In all the experiments, we set the incentive weights and the minimal resources to be equal, i.e., \( \sigma_{ij} = 1/(j-1) \) for \( i < j \) and \( j = 2, \ldots, N \), \( r_0^i = C/(4L) \) for \( i = 1, \ldots, L \), and the total bandwidth \( C = 1000 \). Accuracy of the one-dimensional optimization (segment length) was set to be 0.0001. The experiments with different data showed a rather stable behavior of the method and weak dependence of its convergence from the starting point. For instance, we first took \( \alpha \) from \([1, 5]\) and \( \beta \) from \([300, 500]\). Then, in case \( L = 10 \) the method attained the accuracy \( \varepsilon = 0.001 \) in 2–4 iterations, in case \( L = 20 \) it attained the accuracy \( \varepsilon = 0.017 \) in 20–25 iterations.

In tables 3, 4, and 5 we list some other detailed results of the numerical experiments with this gradient projection method. In all the cases, we chose the limit accuracy \( \varepsilon = 0.001 \) and \( \alpha \) from \([0.1, 1]\); \( \beta \) from \([300, 500]\). The last column in all tables is the value of the objective function at the minimal possible share point \( r_0^i \) and it is given only for comparison. The number of steps in all the shown results is 50. The total time including on-line printing is indicated, so that the main estimate is the attained accuracy for the given number of steps. Numerical results indicate that the objective function is sensitive towards parameters \( \alpha \) and \( \beta \). The larger the values of \( \beta \) the wider the cost function domain. These properties also indicate the interesting problem of revealing simple criteria of choosing the coefficients ensuring existence of non-trivial solutions.
Calculation time nearly linearly depends on the number of steps. These preliminary results seem satisfactory. It was also found that the best point to start with is always the equal share point, $r_i = C/L, i = 1,...,L$; while the worst one is where all bandwidth goes to the first node $(C,0,0,...,0)$. As to the distribution of bandwidth at the solution, it has been noticed that it is always far from the disagreement point $d$. It is also found that the sum of the allocated radio resources is always equal to the downlink capacity $C$ (no loss of bandwidth), i.e., the NBS solution enables all the nodes to receive positive shares. Since these shares depend on the random coefficients $α$ and $β$, there are some fluctuations of coordinates of the solution found, but usually they are near to those of the equal share point.

### Table 3. Initial Point: Equal Share (©2005 IEEE).

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>Attained $|\mathbf{r}|$</th>
<th>Time (seconds)</th>
<th>Attained $f$</th>
<th>$f(r^0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0180219</td>
<td>6</td>
<td>132.909</td>
<td>123.102</td>
</tr>
<tr>
<td>20</td>
<td>0.0115992</td>
<td>16</td>
<td>266.247</td>
<td>248.866</td>
</tr>
<tr>
<td>30</td>
<td>0.0170272</td>
<td>32</td>
<td>398.693</td>
<td>372.844</td>
</tr>
<tr>
<td>40</td>
<td>0.0218519</td>
<td>52</td>
<td>531.288</td>
<td>496.999</td>
</tr>
<tr>
<td>50</td>
<td>0.0267262</td>
<td>78</td>
<td>663.982</td>
<td>621.164</td>
</tr>
</tbody>
</table>

### Table 4. Initial Point: Minimal share point (©2005 IEEE).

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>Attained $|\mathbf{r}|$</th>
<th>Time (seconds)</th>
<th>Attained $f$</th>
<th>$f(r^0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0264153</td>
<td>6</td>
<td>132.088</td>
<td>123.148</td>
</tr>
<tr>
<td>20</td>
<td>0.0116516</td>
<td>16</td>
<td>265.823</td>
<td>248.446</td>
</tr>
<tr>
<td>30</td>
<td>0.0167311</td>
<td>29</td>
<td>398.665</td>
<td>372.732</td>
</tr>
<tr>
<td>50</td>
<td>0.0258989</td>
<td>74</td>
<td>665.074</td>
<td>665.074</td>
</tr>
</tbody>
</table>
Table 5. Initial Point: All bandwidth to the first node (© 2005 IEEE).

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>Attained |d_k|</th>
<th>Time (seconds)</th>
<th>Attained f</th>
<th>( f(\nu) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0298663</td>
<td>5</td>
<td>123.06</td>
<td>123.03</td>
</tr>
<tr>
<td>20</td>
<td>0.0587363</td>
<td>24</td>
<td>247.738</td>
<td>247.601</td>
</tr>
<tr>
<td>30</td>
<td>0.087606</td>
<td>31</td>
<td>372.42</td>
<td>372.223</td>
</tr>
<tr>
<td>40</td>
<td>0.0975479</td>
<td>54</td>
<td>498.404</td>
<td>498.19</td>
</tr>
<tr>
<td>50</td>
<td>0.0989543</td>
<td>87</td>
<td>622.251</td>
<td>621.778</td>
</tr>
</tbody>
</table>

5.3 Resource allocation as a cooperative game among multiple wireless networks

We start this section by briefly summarizing the basic concepts of cooperative game theory. A Cooperative game is a game in which the players have the option of planning as a group in advance of choosing their actions. In this game, players are able to make enforceable contracts. There are no restrictions on the agreement that may be reached among the players. A cooperative game consists of two elements: a set of players, and a characteristic function specifying the value created by different subsets (coalitions) of the players in the game.

**Coalitions:** Let \( K = \{1, \ldots, k\} \) be a set of \( k \) players. Non-empty subsets of \( K, S, T \subseteq K \) are called a coalition. The coalition form of a \( k \) player game is given by the pair \((K, \nu)\), where \( \nu \) is the characteristic function.

**Grand Coalition:** A grand coalition is a coalition that includes all of the players.

**Characteristic Function:** The characteristic function is a function, denoted \( \nu \), that assigns each coalition \( S \) its maximum gain, the expected total income of the coalition, denoted \( \nu(S) \).

**The Core:** The core is the set of all feasible outcomes that no player or coalition can improve upon by acting for themselves. It was developed as a solution concept for cooperative games. It consists of all undominated allocations in the game. An allocation in the core of a game will always be an efficient allocation.

**Pareto-optimal:** Pareto optimality is a measure of efficiency. An outcome of a game is Pareto-optimal if there is no other outcome that makes every player at least as well off and at least one player strictly better off. This means that a Pareto-optimal outcome cannot be improved upon without hurting at least one player.
**Imputation:** The division of the payoff which can be achieved by all players cooperating is called an imputation of the game. The elements $x = (x_1, x_2, ..., x_k)$ represent the payoff to each player (the imputation) $i$ under coalition structure $P$. The $(x, P)$ is called solution configuration.

### 5.3.1 Cooperative game formulation

We use cooperative Stackelberg [132] games to investigate the resource allocation problem in a heterogeneous network environment. The Stackelberg game is a multi-level game wherein a dominant player (Stackelberg leader) chooses a strategy in the first stage which takes into account the likely reaction of the followers. In the second stage, the Stackelberg followers choose their own strategies, having observed the Stackelberg leader’s decision i.e., they react to the leader's strategy. Fig. 38 shows a Stackelberg game on which network/player 1 assumes the role of the leader and the remaining networks assume the role of the followers. The leader will begin the game by announcing his policy and the followers will take the decision of the leader and optimize their objective functions. In [133], a Stackelberg routing strategy has been used to improve the overall system performance. A cooperative approach to resource allocation has been treated in [134] where the authors investigated a fair and efficient resource allocation scheme on ATM networks for two traffic types contending for a shared network resource.

![Fig. 38. Stackelberg Multilevel Decision.]

Formulating the problem as a cooperative Stackelberg game allows individual networks to cooperate with each other by forming coalitions. Therefore the objective of each network is to maximize the overall objective of the heterogeneous network and fulfill the resource allocation requests from users. Every member of the coalition provides some of the requested resources according to its own operation constraints. Resource allocation in heterogeneous network has been previously addressed in [135]. Therein, services were delivered via the network that is most efficient for that service.
In [136], the authors introduced a fault tolerance architecture to provide continuous QoS support in case of network failures. In [137] a resource reservation strategy that enables scheduling and allocation of resources at an early stage in time has been studied. However, these approaches to resource allocation rely on a single network to handle resource demands and may result in an inefficient resource allocation; it may be that a single network alone cannot fulfill user requests.

We consider a resource allocation problem for a mobile user having access to a heterogeneous wireless network (equipped with multiple network interfaces). Let \( Z = 1, \ldots, z \) denote the set of possible wireless access networks available for the mobile user to request resources from, i.e., the finite player set. Let \( R \) denote the radio resource required by the mobile user to run its multiple independent applications. The basic coalition problem can be described as: Given a set of wireless networks \( Z \) and a resource allocation demand \( R \) they have to satisfy, if the resource demand can’t be satisfied by a single network or when a single network handles the request inefficiently, it is necessary for the wireless networks to cooperate with each other to fulfill the resource demand. With cooperation between networks (by forming coalitions among themselves), the resource can be allocated by splitting the applications over the \( Z \) networks.

We assume that each player \( Z_i \) has resource capacity \( c_i \). The capacity configuration of wireless networks is given by \( c = (c_1, c_2, \ldots, c_k) \), and the total resource of the wireless networks is given by \( C = \sum_{i=1}^{n} c_i \). Each network can use its available resource \( c_i \) for contribution \( r_i \geq 0 \) to the resource allocation request \( R \), such that \( r_i \leq c_i \). In the original Stackelberg games, each player attempts to optimize its own objective with respect to decisions made by other players in the game. The objective of each player can be given by the optimization problem below

\[
\max_{r_i} u_i(r_i) \quad \text{such that} \quad r_i \leq c_i, \quad r_i \geq 0, \quad (62)
\]

where \( u_i \) is the utility function of player \( i \) and \( r_i \) is the resource allocated by player \( i \). Allocating resources according to individual network capabilities without considering the overall heterogeneous network conditions will not typically result in a Pareto-optimal solution. Achieving Pareto-optimal allocation may require cooperation among individual networks by forming a coalition among themselves. Let \( P = P_1, P_2, \ldots, P_j \) denote a coalition structure, where \( P_i \cap P_j = \emptyset \) for all \( i \neq j \). As a result of coalition formation, individual players now act for the benefit of the coalition. Therefore the objective
of each member in the coalition becomes the optimization of the coalition objective subjected to operating constraints. The coalition objective function can be formulated as follows

\[ u'_p(R) = \sum_{i \in P_j} u'_i(r_i) \] (63)

where \( u'_i \) is the coalition utility function that maximizes the total payoff for the coalition for allocating \( R \) resource. Then the overall resource allocated to the mobile user in the heterogenous network is given by: \( R = \sum_{i=1}^{m} r_i \). We assume that there is a transferable utility (means that each coalition can achieve a certain total amount of utility with no restriction on how the total payoff may be divided among coalition members).

It is of the interest for the coalition to take into account the goals of resource allocation and payoff maximization simultaneously with load balancing on the networks and to minimize the delays experienced by user data over the multiple networks. Load balancing in distributed systems has been formulated [123, 138] using game-theoretic approaches. In [123], with the cooperative approach, the authors have shown that the Nash Bargaining Solution provides a Pareto-optimal allocation for the distributed system and is also a fair solution. However, in [138], the authors follow the non-cooperative approach for obtaining a user-optimal load balancing scheme in heterogeneous distributed systems. It is shown that the scheme guarantees the optimality of allocation for each user in the distributed system. We employ an exponential cost function to assign congestion factor (which is a measure of the utilization of a network) to the networks. Using the exponential cost function, the congestion factor is computed for each network with

\[ c_f = e^{c_i} \] (64)

where \( l_i \) is the network load and \( c_i \) is the network capacity. Based on the congestion factor, networks decide whether to allocate resources or not. We use the congestion factor to optimize network performance, e.g., to minimize maximum load or network delay. Fig. 39 shows the plot of the congestion factor. With the congestion factor, the payoff for resource allocation will be different on each network based on how much resource is already used. We allocate resource on networks where the sum of the payoff is maximized. If the request cannot be satisfied on any network, then the request is rejected.
5.3.2 Illustrative example

We consider a cooperative game with three networks/players. Let \( \lambda_i = \frac{1}{cf_i} \), then the payoff for allocating resource on network \( i \) is given by \( x_i = \lambda_i r_i \). The coalition is characterized by the maximum total payoff denoted by vector \( x = (x_1, x_2, ..., x_z) \), where \( x_i \) is the payoff allocated to player \( i \). Let \( r = (r_1, r_2, r_3) \) be a 3–dimensional vector representing the amount of resources allocated to the user by the three players. The objective is to allocate the resources so that the total utility of the coalition is maximized, subjected to resource capacity constraints in each network. We assume that the user bandwidth demand equals 100 Mbits/s. Therefore \( r_1 + r_2 + r_3 = 100 \). Let \( \Lambda = (\Lambda_1, \Lambda_2, \Lambda_3) \) be a 3–dimensional vector representing \( \frac{1}{cf} \) for the multiple networks. Vector \( x = (x_1, x_2, x_3) \) represents the total payoff which is given by \( x = \Lambda r^T \). Table 6 presents the characteristic function of the game.

The imputations are the points \( (r_1, r_2, r_3) \) such that \( r_1 + r_2 + r_3 = 100 \) and \( r_1 \geq 10, r_2 \geq 10, r_3 \geq 20 \). The set of the imputations for the game is represented graphically as shown in Figure 40. The figure shows the core and various other solution points for the game. The core consists of all imputations in the trapezoidal area.
Table 6. Coalition structure.

<table>
<thead>
<tr>
<th>Coalition</th>
<th>Coalition value</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>0</td>
</tr>
<tr>
<td>{1}</td>
<td>10</td>
</tr>
<tr>
<td>{2}</td>
<td>10</td>
</tr>
<tr>
<td>{3}</td>
<td>20</td>
</tr>
<tr>
<td>{1,2}</td>
<td>50</td>
</tr>
<tr>
<td>{1,3}</td>
<td>40</td>
</tr>
<tr>
<td>{2,3}</td>
<td>70</td>
</tr>
<tr>
<td>{1,2,3}</td>
<td>100</td>
</tr>
</tbody>
</table>

Fig. 40. The core of the three-player cooperative game.

5.4 Summary and discussion

The problem of radio resource allocation and fair sharing of radio resources in wireless networks was examined. A cooperative game framework was proposed to provide fair distribution of a downlink capacity among mobile nodes in a multihop radio environment. NBS was used to demonstrate that the cooperative game can enable mobile nodes to receive radio resources fairly. Numerical experiments were performed using the gradient projection method with exact line search to demonstrate the effectiveness of the proposed approach. Experiments with different data showed rather stable behavior.
of the method and weak dependence of its convergence from the starting point. The results also suggested that the solution was far from the disagreement point and the sum of the allocated bandwidth is always equal to the available downlink capacity. The issue of resource allocation where mobile nodes are able to split data traffic over multiple networks was presented. A cooperative game approach was applied to model the resource allocation problem. A coalition structure that enables multiple networks to cooperate so as to fulfill resource allocation demands was presented. A core allocation has been shown to exist for wireless networks. An exponential cost function was used to balance loads on the networks. Results showed that such a resource allocation approach results in higher utilization of network resources and maximizes the outcome while potentially satisfying user performance requirements.
6 Device-to-device (D2D) communication

In infrastructure-based networks, traffic usually goes through a centralized controller such as a base station (BS) even if the source and destination are close to each other. However, direct communication between terminals is also possible. Device-to-device is also called direct terminal-to-terminal (DTT), peer-to-peer (P2P), handset-to-handset, or mobile-to-mobile (m2m). D2D has been an active research area receiving a lot of consideration and has now been included in the LTE in 3GPP Release 12. Advantages of exploiting direct communication between nearby mobile devices include improving spectrum utilization, overall throughput and energy consumption, while enabling new peer-to-peer and location-based applications and services [139]. The author of [139] mentioned that D2D-enabled LTE devices can also become competitive for fallback public safety networks. Recently the authors of [140] presented a literature review on D2D communications and examined the problem of transmission order (TO) optimization D2D communications underlaying the cellular network.

In this chapter, we study D2D communication with localized distribution, so that users tend to gather around some areas (clusters/hot-spots) within the cell such as buildings. Previous analysis about clustering has focused on one-dimensional scenarios. Here we present theoretical analysis of D2D communication with two-dimensional clustering. Additional analysis is also presented for D2D communication with correlated clusters. In addition to the exact results, we also suggest using the point-based approximation to rapidly and easily obtain results. The numerical results show that the gains from direct communication, in terms of blocking probability and carried traffic, depend on the offered traffic. Correlation in cluster selection is shown to significantly improve performance. Point-based approximation is shown to be very useful when the number of clusters is large. We further present simulation results of the performance of direct communication in cellular systems with resource reuse. The presented results include the blocking probabilities and carried traffic.

The rest of the chapter is organized as follows. The related work is presented in Section 6.1. In Section 6.2, the system model is presented. In Section 6.3, results from teletraffic theory are shown for finding the blocking probability and carried traffic as a function of the direct communication probability $p$. In Section 6.4, we find $p$ for situations where there are several circular clusters inside the cell.
based approximation (assuming radius zero for the clusters) is studied in Section 6.5. Numerical results are presented in Section 6.6 for blocking probability and carried traffic. The theoretical results are compared with simulation results. Section 6.7 presents the teletraffic simulation model and simulation results. Finally, summary and discussions are provided in Section 6.8.

6.1 Related work

Direct communication has been proposed for the IMT advanced networks [141]. Standards which already have D2D communication support include the Personal Handyphone System (PHS), the Digital Enhanced Cordless Telecommunications (DECT), HiperLAN/2, Terrestrial Trunked Radio (TETRA) and 802.11e. The advantages of D2D include reduced power requirements which leads to longer battery life [142–144]. More efficient resource use is an additional advantage because direct communication typically requires only half the amount of resources compared to communication through the BS [142–147]. The amount of interference can be reduced if terminals that are close to each other, but in different cells, communicate directly [143]. Due to enhanced channel characteristics, it may also be possible to use different modulation methods or less robust physical layers to increase the bit rate compared to the cellular link [142, 146]. Moreover, increased network efficiency supports more services. Improved performance in shadow fading is also one possible benefit [143, 148]. Direct communication can also improve performance of delay-sensitive applications [146] and enable communication when the BS is malfunctioning [144]. Although direct mode of operation is often free of air time charges, the operator could be involved in establishing direct links with the possibility of some revenue [149, 150].

The practical impact of the advantages of direct communication depends on the number of calls where calling and called terminals are within the direct communication range. Examples where direct communication may be possible include, for example, student communication inside a college campus, communication between colleagues (especially when there are no wired phones available), room to room intercom, communication amongst a hunting group and communication amongst cooperating operational robots [143]. Teletraffic analysis is important for determining the benefits of direct communication. It has been partly performed in [145] by Tanaka et al. By using results presented therein, it is possible to theoretically calculate the blocking probability and carried traffic assuming that direct communication probability \( p \) is known. The direct
communication probability is the probability that the terminals are within direct communication range $D$. In [145], $p$ was found for circular cell where users are uniformly distributed. As in [145], typically when studying cellular systems a uniform distribution of users is assumed. However, to model the fact that people tend to gather around some areas such as buildings using non-uniform localized distribution is necessary [151]. In [152] localized distribution was assumed for direct communication by Kabasawa et al. In their model, users gather in some points. Therein, one-dimensional scenario (street) was assumed. However, two-dimensional approach is necessary to more accurately model cellular and usual wireless communication systems. Two-dimensional localized distribution with resource reuse was simulated in [153]. The results showed significant gains in throughput and power consumption.

6.2 System model

Let us assume a circular cell with radius $R$. To model localized distribution of users we assume that users are concentrated in clusters, as shown in Fig. 41. Denote the number of clusters with $N_C$ and the radius of the clusters with radius $R_C$. Assume that the cell has $n$ channels or, more generally, resource units [145]. It is further assumed that two-hop communication inside the cell through the BS (for example communication between clusters 1 and 3 in Fig. 41) requires two channels and a direct communication inside the cell (for example internal communication in cluster 3 in Fig. 41) consumes one channel [145].

Priority policy is not used so that a direct call is always accepted if there is at least one channel available, and call through the BS is accepted if there are at least two available channels. For simplicity, all calls are assumed to be internal to the cell [152]. In practice, this assumption is not realistic (depends on the scenario), but it can be used to show the upper bound on gains from direct communication. The direct communication range is $D$ and the cell radius is $R$. If distance between nodes is $\leq D$, then direct communication is possible. Otherwise, communication through the BS is always assumed to be available. The call arrivals follow the Poisson process with intensity $\lambda$. The holding times are assumed to be exponential random variables with mean $h$. The offered traffic is $\theta = \lambda h$. We assume that within clusters, the users are uniformly distributed. This model is similar to that used in [154]. Therein, nodes are uniformly distributed in a disk of radius $r_c$ centred at the clump centres. Instead of
uniform distribution, it would also be possible to use the Gaussian distribution to model distribution of the nodes in the clusters [153, 155].

6.3 Teletraffic theory

Blocking probability is the probability that a call is blocked, i.e., there are not enough free channels. Assume that the probability that the nodes are within direct communication range is $p$. Define $a_1 = p \theta$ and $a_2 = (1 - p) \theta$. The blocking probability of calls where direct communication would be possible is $B_1$ and the blocking probability of calls going through the BS is $B_2$. It is well known that [145, 152]

$$B_1 = P_{0,0} \sum_{j=0}^{[n/2]} \frac{a_1^{n-2j}}{(n-2j)!} \frac{a_2^j}{j!},$$  \hspace{1cm} (65)

where $\lfloor x \rfloor$ is the largest integer not exceeding $x$ and

$$B_2 = P_{0,0} \sum_{j=0}^{[n/2]-1} \sum_{i=-2(j+1)+1}^{n-2j} \frac{a_1^{n-2j}}{\pi^i} \frac{a_2^j}{\pi^i} + P_{0,0} \frac{a_1^{[n/2]}}{\pi^i} \sum_{j=0}^{k} \frac{a_2^j}{\pi^i},$$  \hspace{1cm} (66)

where $k = n \mod 2$ and
\[ P_{0,0}^{-1} = \sum_{j=0}^{\frac{n}{2} - 2j} \sum_{i=0}^{\frac{n}{2} - j} \frac{a_i^j a_i^{j}}{l^i j^j} . \] (67)

The carried traffic is
\[ \vartheta_c = a_1 (1 - B_1) + a_2 (1 - B_2) \] (68)

and the overall blocking probability is
\[ B = \frac{a_1 B_1 + a_2 B_2}{a_1 + a_2} . \] (69)

### 6.4 Direct communication probability

The direct communication probability \( p \) plays a critical role in determining the system performance. In [145], \( p \) was found for circular cell where users are uniformly distributed. For localized distribution, where users tend to gather around some points, \( p \) was found in [152] with one-dimensional scenario (such as a street). To study two-dimensional scenarios, let us assume that the first user is in cluster \( i \) and the second user is in cluster \( j \). Assume that users are randomly and uniformly distributed in the clusters so that user locations in the source and destination clusters are random. Let us denote the probability that the distance between the users (two random points in clusters \( i \) and \( j \)) is less than or equal to the direct communication range \( D \) with \( p_{ij} \). It is conditional probability of direct communication given the source and destination clusters. It can be found by using results presented in [156] for random distance between two circles

\[ p_{ij} = \int_{R_C}^{D} \frac{2r}{\pi R_C} \int_{C_1}^{C_2} \int r \arccos \left[ \frac{d^2 + r^2 - 2drcos \phi + d^2 - R_C^2}{2d \sqrt{d^2 + r^2 - 2drcos \phi}} \right] dr d\phi dl , \] (70)

where \( d \) is the distance between the cluster centres, \( p_1 = (r \cos \phi, r \sin \phi) \) is a point in the first cluster which is centered at \((0,0)\) without loss of generality and \( C_1 \) is the region of the first cluster given by

\[ \cos \phi \geq \frac{d^2 + r^2 - (l + R_C)^2}{2dr} \]

for \( d - 2R_C \leq l < d \) and

\[ \cos \phi \leq \frac{d^2 + r^2 - (l - R_C)^2}{2dr} \]

for \( d \leq l < d + 2R_C. \)
In [145] direct communication probability is presented for two uniformly selected points in a cell. By direct application we get, when users are in the same cluster, the direct communication probability

\[ p_{ii} = \frac{2D^2}{2\pi R^2} \arccos \frac{D}{2R} + \frac{2}{2\pi R^2} \arcsin \frac{D}{2R} - \frac{2R^2D}{2\pi R^2} \sqrt{1 - \left( \frac{D}{2R} \right)^2} \]  

(71)

if \( D \leq 2R \) and 1 otherwise. The \( p_{ij} \) is conditional probability of direct communication given that the source is in cluster \( i \) and the destination is in cluster \( j \). The \( \eta_{ij} \) is the probability that the source is in cluster \( i \) and the destination is in cluster \( j \). These probabilities are illustrated in Fig. 42. Now, by using the total probability theorem [157], we find the overall probability of direct communication \( p \) with

\[ p = \sum_{i=1}^{N_C} \sum_{j=1}^{N_C} p_{ij} \eta_{ij} = \sum_{i=1}^{N_C} \sum_{j=1}^{N_C} p_{ij} \eta_{ij} + \sum_{i=1}^{N_C} \sum_{j=1}^{N_C} p_{ij} (\eta_{ij} + \eta_{ji}) + \sum_{i=1}^{N_C} \eta_{ii}, \]

(72)

where we have used the fact that \( p_{ji} = p_{ij} \) and \( p_{ii} = p_{11} \) (same radius is assumed for all clusters).

\[ \eta_{i,j} \quad \text{Probability of source being in cluster } i \text{ and destination being in cluster } j \]

\[ p_{i,j} \quad \text{Probability of direct communication between users given that source is in } i \text{ and destination in } j \]

Fig. 42. Probabilities \( \eta_{ij} \) (source in cluster \( i \) and destination in cluster \( j \)) and \( p_{ij} \) (direct communication probability given that source is in cluster \( i \) and destination in cluster \( j \)) (c)2009 IEICE).

The probability matrix \( \eta \) containing the values of \( \eta_{ij} \) has an infinite number of possible combinations. For example, we can have the following \( N_C \times N_C \) probability
matrix

\[
\eta = \begin{bmatrix}
\sigma & 1 & 1 & \cdots & 1 \\
1 & \sigma & 1 & \cdots & 1 \\
1 & \cdots & \ddots & \cdots & 1 \\
1 & 1 & \cdots & \sigma & 1 \\
1 & \cdots & 1 & \cdots & \sigma,
\end{bmatrix} / \xi \quad (73)
\]

where \( \xi = [N_C(\sigma + N_C - 1)] \). This case models a situation where the intra-cluster communication and inter-cluster communication have different probabilities. Note that if \( \sigma \) is high, it means that intra-cluster communication has increased probability. On the other hand, if \( \sigma \) is small, it means that inter-cluster communication happens with increased probability. In the case without correlation (\( \sigma = 1 \)) every combination of source and destination is equally probable. We get,

\[
p = p_{11} \sum_{i=1}^{N_C} \frac{\sigma_i}{(\sigma_i + N_C - 1)} + \frac{2}{N_C(\sigma_i + N_C - 1)} \sum_{i=1}^{N_C} \sum_{j=i+1}^{N_C} p_{ij}. \quad (74)
\]

It also possible to use a somewhat more general probability matrix given by:

\[
\eta = \begin{bmatrix}
\sigma_1 & 1 & 1 & \cdots & 1 \\
1 & \sigma_2 & 1 & \cdots & 1 \\
1 & \cdots & \ddots & \cdots & 1 \\
1 & 1 & \cdots & \sigma_{N_C-1} & 1 \\
1 & \cdots & 1 & \cdots & \sigma_{N_C},
\end{bmatrix} / \xi \quad (75)
\]

where \( \xi = \sum_{i=1}^{N_C} \sigma_i + N_C(N_C - 1) \). Now, we get that direct communication probability is

\[
p = \frac{p_{11} \sum_{i=1}^{N_C} \sigma_i}{\sum_{i=1}^{N_C} \sigma_i + N_C(N_C - 1)} + \frac{2}{\sum_{i=1}^{N_C} \sigma_i + N_C(N_C - 1)} \sum_{i=1}^{N_C} \sum_{j=i+1}^{N_C} p_{ij}. \quad (76)
\]

which depends on \( \sigma_i \) only through the sum \( \sum_{i=1}^{N_C} \sigma_i \), i.e., the total amount of intra-cluster communication is what matters and not how it is distributed among clusters. Note that when \( \sum_{i=1}^{N_C} \sigma_i \) approaches infinity (high level correlation), then \( p = p_{ii} \), i.e., finding the overall direct communication probability reduced to finding the direct communication probability for intra-cluster communication.

121
6.5 Point-based approximation

6.5.1 Theory

Assume that there are \( N_C \) points with radius zero where users gather and that \( \sigma = 1 \). Now, by using equations (72)–(73), we get

\[
p = \frac{N_C + \sum_{i=1}^{N_C} \sum_{j \neq i}^N p_{ij}}{N_C^2} = \frac{N_C + \sum_{i=1}^{N_C} |G_i|}{N_C^2},
\]

(77)

where, because in point-based situation \( p_{ij} \) is either zero (no direct communication) or 1 (always direct communication),

\[
|G_i| = \sum_{j=1,j \neq i}^{N_C} p_{ij} \in \{0, 1, \cdots, N_C - 1\}
\]

(78)

is the number of points that a node can directly communicate with, not including itself. It can be determined given the locations of \( N_C \) points and the direct communication range \( D \). Results similar to (77) have been given in [152] (used therein for one-dimensional case). In addition to situations where the clusters are points, the point-clusters can be used to rapidly and easily get approximate results even when the clusters are not actually points, especially when the number of clusters is large. For these situations, we propose to improve the above equation by taking into account the easily found probability of direct communication within a cluster as given in eq. (71) (calculated with the real \( R_C \) instead of \( R_C = 0 \)),

\[
p = \frac{1}{N_C} p_{ii} + \frac{\sum_{k=1}^{N_C} |G_k|}{N_C^2}.
\]

(79)

6.5.2 Binomial approximation

Let us now study cases where cluster locations are random. Then \( |G_k| \) is a random variable with some distribution. However, distribution of \( p \) given by (77) is also interesting. It is a random variable with a finite number of possible values. The values it can get are

\[
p_k = \frac{N_C + 2k}{N_C^2}, k = 0, 1, \cdots, M
\]

(80)
where the number of combinations of selecting two clusters from \(N_C\) (the binomial coefficient) is

\[
M = \binom{N_C}{2} = \frac{N_C(N_C - 1)}{2}.
\]  

(81)

The possible values are as given above because intra-cluster communication in each cluster is always assumed to be possible (distance zero). Also, the factor two appears because if direct communication is possible to cluster \(j\) from cluster \(i\), then it is also possible to cluster \(i\) from cluster \(j\). By making the approximation that all the links are independent from each other, we get the corresponding probabilities

\[
n_k = \binom{M}{k} \phi^k (1 - \phi)^{M-k}, \quad k = 0, 1, \ldots, M
\]  

(82)

where \(\phi\) is the probability that two random points (uniformly distributed in the cell) are within direct communication range and it is given by (71) (with \(R\) instead of \(R_C\)). Now theoretical averaged results over different realizations can be found by calculating the desired results (such as carried traffic) for each possible value of \(p_k\) and summing them with weights \(n_k\). The above approach results in approximation because the links have correlation between each other. For example, if direct communication is possible from node \(i\) to nodes \(j\) and \(k\), then the probability that direct communication is possible between nodes \(j\) and \(k\) is increased. Numerical results are shown later to show the accuracy of the proposed approximation.

### 6.6 Numerical results

Let us assume that the clusters locations are as listed in Table 7. Users are uniformly distributed within the clusters. The resulting blocking probabilities and carried traffic are shown in Fig. 43 and Fig. 44, respectively. The results were obtained by using numerical integration of (70). It can be seen that when the offered traffic is 12.45, direct communication gives significant advantage in terms of blocking probabilities compared to a situation without direct communication \((D = 0)\). However, the gain in carried traffic is not that large. On the other hand, when the offered traffic is 18.675, direct communication significantly increases the carried traffic. But the gains in terms of blocking probabilities are not that large. Additionally, it can be seen that the theoretical and simulation results agree very well.
The point-based approximation is also shown in Fig. 43 and Fig. 44. It can be seen that, as expected, due to the small number of clusters it is heavily quantized. Due to this, exact theoretical results are still useful especially when the number of clusters is small. Figs 45 and 46 show blocking probabilities in the case of correlated clusters with the correlation model given by (73). It can be seen that large correlation significantly improves system performance. In the limit when correlation is infinite, results correspond to $p_{ii}$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-49.5385</td>
<td>-52.4414</td>
</tr>
<tr>
<td>-4.7266</td>
<td>25.9856</td>
</tr>
<tr>
<td>43.4574</td>
<td>27.3656</td>
</tr>
</tbody>
</table>

**Table 7. Cluster \((x, y)\)-coordinates (©2009 IEICE).**
Fig. 44. Carried traffic with $n = 30$ channels, cell radius $R = 100$, direct communication range $D$ between 0 and 200, number of clusters $N_C = 3$ and cluster radius $R_C = 20$. Solid lines are theoretical results (exact and point based) and markers denote simulation results (©2009 IEICE).

Fig. 45. Blocking probabilities with $n = 30$ channels, cell radius $R = 100$, direct communication range $D$ between 0 and 200, number of clusters $N_C = 3$ and cluster radius $R_C = 20$. Offered traffic is 12.45 (©2009 IEICE).
Let us next assume that there are $N_C = 8$ clusters with cluster centres as listed in Table 8. Simulation results for that situation are compared with point clusters-based approximation (theoretical results for $R_C = 0$) in Figs 47 and 48. It can be seen that the theoretical approximation matches the simulation results rather well, thus showing the usefulness of the point-based approximation for situations with large numbers of clusters. The results shown were calculated using the proposed equation (79). Let us continue to use the point-based approximation. Fig. 49 shows the carried traffic averaged over different realizations of the cluster locations. The simulation results are based on 500 realizations of the cluster locations with various numbers of clusters. The cluster locations were selected randomly and uniformly inside the cell [158]. The presented theoretical results are based on the binomial approximation presented in Section 6.5.2. Although the results here are presented for $D$ ranging from zero to two times the cell radius, in a practical system the direct communication range is probably less than the cell radius to unfavourable locations of the antennas, better sensitivity of the base station and limited transmit power of the terminals. If we limit our discussion to cases where $D \leq R$, then we can see that the gains from direct communication decrease
(but they still exist, especially with correlated clusters). An additional factor to keep in mind is that when $D$ is large interference levels may rise above those allowed in the cellular frequency reuse plan.

Table 8. Clusters $(x,y)$-coordinates, \( N_C = 8 \) (©2009 IEICE).

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.1735</td>
<td>13.1008</td>
</tr>
<tr>
<td>-8.1547</td>
<td>33.1499</td>
</tr>
<tr>
<td>20.1022</td>
<td>-70.2141</td>
</tr>
<tr>
<td>40.8354</td>
<td>53.4385</td>
</tr>
<tr>
<td>-20.0690</td>
<td>71.4665</td>
</tr>
<tr>
<td>-16.1098</td>
<td>-41.6565</td>
</tr>
<tr>
<td>53.2527</td>
<td>-41.3839</td>
</tr>
<tr>
<td>-48.4780</td>
<td>31.4366</td>
</tr>
</tbody>
</table>

Fig. 47. Blocking probabilities with $n = 30$ channels, cell radius $R = 100$, direct communication range $D$ between 0 and 200, number of clusters $N_C = 8$ and cluster radius $R_C = 20$. Offered traffic is 18.6750. Solid lines denote the point-based theoretical approximation (©2009 IEICE).
Fig. 48. Carried traffic with \( n = 30 \) channels, cell radius \( R = 100 \), direct communication range \( D \) between 0 and 200, number of clusters \( N_C = 8 \) and cluster radius \( R_C = 20 \). Offered traffic is 18.6750. Solid lines denote the point-based theoretical approximation (©2009 IEICE).

Fig. 49. Carried traffic with \( n = 30 \) channels, cell radius \( R = 100 \), direct communication range \( D \) between 0 and 100, number of clusters is varied. Offered traffic is 18.6750. Simulated results are averaged over 500 realizations of cluster locations (©2009 IEICE).
6.7 Teletraffic simulation of D2D with resource reuse

In this section, we perform simulation study of D2D communication focusing on estimating the blocking probability when radio resource reuse (clustering, direct communication range) is taken into consideration. In [153] the localized distribution with resource reuse was investigated and the results demonstrate that throughput and power consumption could be improved. We consider a cellular network where device-to-device communication between nodes is allowed. We consider that cells are homogeneous and that calls are originated in a cell (i.e., a single cell represents a network). We assume \( N \) channels or, more generally, resource units are available in the cell. It is further assumed that direct communication calls require one channel, while conventional calls that go through the base station consume two channels. The direct communication range is \( D \) and the cell radius is \( R \). If the distance between terminals is less than \( D \), then direct communication is assumed to be possible. The nodes are randomly and independently distributed in the cell. Here, for simplicity, all calls are assumed to be internal to the cell and the internal calls have the Poisson process intensity \( \lambda \). The holding times are assumed to be exponential random variables. Analysis using this model has been performed in [145]. However, the results therein do not take into account some factors such as mobility, the fact that terminals may tend to gather around some places (clustering) depending on the scenario, as well as the resource reuse.

Several models have been proposed in [152] for studying the effects of localized distribution. Therein, one-dimensional scenario (street) was assumed. However, spatial reuse of channels was not studied. According to [152] and [159], significant gains cannot always be expected with mobility. We assume that interference range of a link is \( q = (1 + \Delta) \) times its communication distance [160]. If possible, we try to use resources already used somewhere in the cell, so that there is room for new arrivals. It should be noted that channel re-use is only allowed for D2D communications. Note that we do not claim that this approach results in jointly optimal scheduling. The flowchart shown in Fig. 50 illustrates how newly arrived calls are handled. Depending on the type of the call, the new call is accepted as follows:

- D2D calls are admitted if there is one channel available. Depending on the interference range, it is possible to re-use radio resources or channels for making D2D calls. Resource reuse will be attempted first before attempting to allocate new channels.
- BS calls are admitted when there are at least two channels available. No resource re-use is allowed for cellular calls.
The number of channels allocated for BS calls increases by 2 when a BS call arrives. Similarly, when a D2D call is completed, a used channel will be released, or the number of reused channels will be decreased if the call utilizing a reused channel finishes. When a BS call finishes, channels allocated for BS calls decrease by 2. Accepted calls hold the channel until the calls are completed, after which time the resources/channels are released and become available for new calls. Theoretically, resource reuse could be modeled as follows: an arrival of a D2D call will cause the system to allocate a used channel with some probability $\alpha$, where $\alpha$ is the probability of resource reuse. With probability $(1 - \alpha)$, the system will allocate a new channel for the D2D connection. Note that the probability of reuse depends on the number of reuses already occurring in the cell. One of the major challenges with resource reuse is the interference. Mobile terminals have to be able to efficiently use the same resource simultaneously without causing harmful interference to each other. To address this issue, several studies have investigated this problem. In [160], a node $X_{R(i)}$ can successfully receive transmission from node $X_{(i)}$ if it does not lie within $(1 + \Delta)$ times the range of any other concurrent
transmitter as explained in equation (83) [160]

\[ |X_k - X_{R(i)}| \geq (1 + \Delta)|X_k - X_{R(k)}|, \]  \hspace{1cm} (83)

where \( X_i \) is the original transmitter, \( X_{R(i)} \) is the intended receiver, \( X_k \) is the interfering transmitter and \( X_{R(k)} \) is the interfering receiver.

To efficiently utilize the radio resources, we use the fact that users tend to gather around buildings, forming clusters. When studying scenarios with clustering we assume that within the cell there are \( N_c \) circular clusters, within which the terminals are uniformly distributed. The source and destination clusters are selected independently, as in [152], except in the case of correlated clusters where the probability of choosing a destination from the same cluster as the source is increased; we assumed that selecting the same cluster is \( \Theta \) (correlation factor) times more probable than any other cluster. Clustering allows for D2D to use spatial resource reuse more effectively. Therefore, radio resource reuse can play an important role on the direct communication improvement in terms of blocking probability and carried traffic. As illustrated in Fig. 41, three clusters are formed within a cell. These clusters could reuse the resources used by other clusters.

### 6.7.1 Simulation results

In this section we present a comparison of the resource reuse and no resource reuse effect on the direct communication between terminals. Simulation parameters are given in Table 9. Fig. 51 shows that the blocking probability increases with the offered load and decreases with the direct communication range. For the carried traffic case, Fig. 52 indicates that the carried traffic increases with increase in both the offered load and the direct communication range. Fig. 53 shows the blocking probability against the offered load and direct communication range while Fig. 54 presents the carried traffic versus offered load and direct communication range. The results are obtained for the cases with radio resource reuse and correlated clusters. Similar to the results obtained for the without resource reuse case, the blocking probability increases with the offered load and decreases with the direct communication range as illustrated in Fig. 53.
Table 9. Simulation parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offered load</td>
<td>0-40</td>
</tr>
<tr>
<td>Mean service time</td>
<td>1.5</td>
</tr>
<tr>
<td>Direct communication range $D$</td>
<td>$0 \leq D \leq 200$</td>
</tr>
<tr>
<td>Number of channels</td>
<td>30</td>
</tr>
<tr>
<td>Cell radius</td>
<td>100</td>
</tr>
<tr>
<td>Number of clusters</td>
<td>3</td>
</tr>
<tr>
<td>Radius of clusters</td>
<td>20</td>
</tr>
<tr>
<td>Correlation factor $\Theta$</td>
<td>2</td>
</tr>
<tr>
<td>Interference distance multiplier $q$</td>
<td>2</td>
</tr>
</tbody>
</table>

![Fig. 51. Blocking probabilities versus offered load and D2D range (with correlated clusters and without resource reuse).](image-url)
Fig. 52. Carried traffic with correlated clusters and without resource reuse.

Fig. 53. Blocking probabilities with resource reuse and correlated clusters.
Fig. 54. Carried traffic with resource reuse and correlated clusters.

Fig. 55. Blocking probabilities versus D2D range with different interference multipliers.
Furthermore, as shown in Fig. 54, the carried traffic increases with increases in both the offered load and the direct communication range. Comparing the results, there is a difference in the performance between resource reuse and no resource reuse in terms of reduced blocking probabilities and increased carried traffic when resource reuse is utilized. Fig. 55 shows the effect of the interference distance multiplier $q$ to the blocking probabilities. It can be seen that with $q = 2$, blocking probabilities are significantly smaller than without resource reuse. Even with $q = 6$ there are some gains compared to a situation without resource reuse.

### 6.8 Summary and discussion

Performance analysis for blocking probabilities and carried traffic for D2D communication was presented. Theoretical analysis for D2D communication with clustering in two-dimensional scenarios with and without correlated clustering were considered. Gains from D2D were found to be dependent on the direct communication range and also on the offered traffic level. With large offered traffic, direct communication showed significant gains in the carried traffic. However, when the offered traffic was small, direct communication showed significant gains in the blocking probability. Clusters correlation was found to increase the D2D gain. In addition to the exact results, a point-based approximation approach to rapidly and easily obtain results was also proposed. The usefulness of the point-based approximation when the number of clusters is large was shown. However, the proposed exact theoretical approach is still useful when the number of clusters is small. Teletraffic simulations for D2D communications have also been performed where blocking probabilities and carried traffic were evaluated. Simulation results show that D2D communication performance could be improved largely by reusing radio resources within the same cell. Interference distance multiplier was found to have a significant impact on the blocking probabilities.
Chapter 7 Conclusions and future work

The thesis focused on developing analytical framework to evaluate the performance of CRNs. Resource allocation including D2D communication in wireless networks was also investigated. The performance of CRNs under an imperfect sensing environment was studied in Chapter 2 using the CTMC-based analytical model. The analytical framework incorporated key performance evaluation parameters such as full state-dependent transitions, multi-channel support and handoff capability to enable accurate theoretical analysis. A recursive approach was employed to allow the CTMC state transition matrix to be found for an arbitrary number of channels. Numerical and simulation results showed that the imperfect spectrum sensing has a significant effect on the performance of the CRNs and thus sensing errors need to be considered in theoretical modeling. The analysis was further extended in Chapter 3 to study the effect of the $\lambda_{FAR}$ on the performance of CRNs. The CTMC-based analytical model was improved to more accurately evaluate the performance of CRNs by taking $\lambda_{FAR}$ into account. The proposed model not only includes sensing errors by incoming SUs, but also takes into account the misdetection and false alarm probabilities by ongoing SUs. The proposed model is also capable of examining other performance evaluation parameters such as the effect of interference tolerance $\tilde{T}_{TOL}$ among PUs and SUs as well as the effect of SUs’ residual self-interference.

Formulas for performance evaluation metrics such as primary forced termination probability, secondary forced termination probability, primary blocking probability, secondary successful probability and radio resources utilization were derived. Extensive simulations were performed to validate the accuracy of the analytical model. Simulation results are in excellent agreement with the analytical results. The results showed that high $\lambda_{FAR}$ can severely degrade PU performance and reduce the overall system resource utilization. Results also showed that decreasing the interference tolerance $\tilde{T}_{TOL}$ has a negative effect on the performance of PUs as it reduces primary successful probability and increases their forced termination probability. A similar effect was also observed with the increase in the SU residual interference distortion factor. Large amounts of residual interference deteriorates the detection probability and leads to a reduced PU performance. The incorporation of $\lambda_{FAR}$ into the CTMC model allows for obtaining exact and accurate state transition probabilities that improves calculation of
the performance evaluation measures. The results from the proposed analytical model provide a new insight into the operation of CRNs and can be used to develop practical and more accurate CRN performance evaluation models.

Detection of the occupancy status of the time slots in CRNs is characterized by false alarm and detection probabilities. The $P_D$ needs to be large enough to protect the primary user. However, too large $P_D$ means increased $P_{FA}$, thus reducing the secondary user’s transmit opportunities. Chapter 4 proposed a numerical optimization of the $P_{FA}$ so that SU throughput is maximized with the constraint that PU quality of service is maintained. Numerical results showed that SUs’ maximum throughput is obtained when the random access probabilities are equal to one and random access is handled instead with false alarm probabilities. The results also indicated that when the signal-to-noise ratio is large, there exists a very simple result for the optimal value for the used $P_{FA}$. The results are important for studying the feasibility of opportunistic access in slotted systems. Simulation results validate the benefits of the optimization. Additionally, a priority queuing approach was used to model opportunistic access in CRNs. Waiting time and queue length were theoretically derived using the modified M/D/1 model. Simulation results showed a good match with the theoretical results. The results indicated that the performance of the SU depends on the data traffic characteristics of the PU, and under a high PU arrival rate the average waiting time and average queueing length of the SU grow especially when the combined arrival rates approach the queue utilization factor.

Resource allocation in wireless networks was addressed in Chapter 5. A game-theoretic approach to radio resource allocation for the downlink capacity in multihop wireless networks is introduced. The importance of fair resource sharing among mobile nodes located along a multihop link was examined. A novel technique has been proposed for providing fair distribution of a downlink capacity among mobile nodes in a multihop relaying radio environment. The resource allocation problem was formulated as a cooperative game using the Nash Bargaining Solution (NBS), which enables mobile nodes to fairly share a downlink bandwidth among themselves. Numerical experiments for the NBS model have been performed using the gradient projection method with exact line search to demonstrate the effectiveness of this approach. Experiments with different data showed rather stable behavior of the method and weak dependence of its convergence from the starting point. The results also suggested that the solution was far from the disagreement point and the sum of the allocated bandwidth is always equal to the available downlink capacity. It has also been shown that the undesirable properties (such as the ability of upstream nodes to manipulate their locations to exploit...
the downlink capacity) can be avoided by means of a cooperative agreement in which all nodes share the radio resources equally. Another approach for resource allocation in wireless networks where users are able to split parallel applications over multiple networks was proposed using cooperative game theory. The proposed approach enabled multiple networks to cooperate so as to fulfill resource allocation demands. The results showed that the resource allocation approach results in higher utilization of network resources and maximizes the outcome while potentially satisfying user performance requirements. An exponential cost function was used to balance the load on the networks.

Chapter 6 focused on studying D2D communications in the cellular system. The aim was to study direct communication with localized distribution, so that users tend to gather around some areas (clusters) within the cell such as buildings. Analysis for blocking probabilities and carried traffic for direct communication with clustering in two-dimensional scenarios with and without correlated clustering was presented. Furthermore, teletraffic simulations studies for D2D in the cellular system have also been performed where the effect of resource reuse, clustering and direct communication range have been taken into consideration. D2D communications brings several advantages such as reduced power requirements, efficient resource use, reduced interference, improved performance in shadow fading and increased network efficiency. However, the practical impact of the advantages of D2D communication depends on the number of calls where calling and called terminals are within the direct communication range. In addition to the exact results, point-based approximation was employed so that results can be easily obtained. The numerical results showed that gains from direct communication depend on the direct communication range and also on the offered traffic level. With large offered traffic, direct communication showed significant gains in the carried traffic. On the other hand, when the offered traffic was smaller, direct communication showed significant gains in the blocking probability. When clusters are correlated, further gains can be obtained. Correlation in cluster selection is shown to significantly improve performance. The usefulness of point-based cluster approximation when the number of clusters is large was shown. However, the proposed exact theoretical approach is still useful when the number of clusters is small. Simulation results showed that D2D communication performance could be improved largely by reusing radio resources within the same cell. It has also been demonstrated that interference distance multiplier has a significant impact on the blocking probabilities.

Although several research problems have been addressed by the thesis, there are still some limitations and shortcomings. One of the limitations of the thesis is that the SU
spectrum access is mainly based on the spectrum detection. However, several research work point out to the importance of taking into account the issue of the aggregate interference produced by multiple secondary users. Another shortcoming of the thesis is that it did not take into account the issues of the robust and reliable spectrum detection especially in the weak SNR environments. Although these issues were partially taken care of with the modelling of residual interference and guard bands, modelling of those (aggregate interference and robust detection) would have enabled the proposed models to be used in a more practical system. Thus there is a room for improvement in future works to handle these shortcomings. As already stated previously, measurements have shown that network traffic models utilized in the thesis can lead to poor fitting to several PU types. However, the use of the Poisson modeling (which has been widely used in the literature [39, 44, 45]) made it possible to develop the analytical model. To overcome this shortcoming future work could apply network traffic models that match modern packet data traffic. While developing the analytical models, several assumptions and simplifications were made in order to get tractable models (for both analysis and computation). Future research may consider eliminating these assumptions and simplifications for more realistic models while keeping the model tractability in order.

There are still many interesting problems that require further studies. For example, in future works, it would be useful to investigate the case where only SU calls will be terminated in collisions between PU and SU in CRNs. Cooperative spectrum sensing can be considered for improving detection performance and/or for mitigating the effects of fading. Additionally, adaptive sensing parameters based on PU channel utilization can also be studied. Further study is also needed to optimize sensing duration. The analysis could also be extended to model the time-slotted system under shadowing and fading conditions. More work can be done to improve resource allocation results by studying coalition formation strategies. Open D2D problems that require further study include theoretical results for the case with resource reuse for sufficiently spatially separated links, taking into account shadowing, generalized non-uniform traffic distributions and using network coding (NC) to reduce resources needed for communication through the central controller [161].
References


558. Aula, Matti (2016) Optical emission from electric arc furnaces
559. Ferdinand, Nuwan Suresh (2016) Low complexity lattice codes for communication networks
560. Xue, Qiang (2016) Analysis of near-optimal relaying schemes for wireless tandem and multicast relay networks
563. Huusko, Jarkko (2016) Communication performance prediction and link adaptation based on a statistical radio channel model
564. Nguyen, Vu Thuy Dan (2016) Transmission strategies for full-duplex multiuser MIMO communications systems
565. Keränen, Pekka (2016) High precision time-to-digital converters for applications requiring a wide measurement range
566. Koivuranta, Elisa (2016) Optical monitoring of flocs and filaments in the activated sludge process
567. Lohikoski, Päivi (2016) Information processing in global virtual NPD projects
568. Kauppila, Osmo (2016) Integrated quality evaluation in higher education
569. Kisko, Anna (2016) Microstructure and properties of reversion treated low-Ni high-Mn austenitic stainless steels
570. Postila, Heini (2016) Peat extraction runoff water purification in treatment wetlands constructed on drained peatlands in a cold climate
571. Happonen, Tuomas (2016) Reliability studies on printed conductors on flexible substrates under cyclic bending
572. Soderi, Simone (2016) Evaluation of industrial wireless communications systems’ security
574. Tolonen, Arto (2016) Product portfolio management over horizontal and vertical portfolios

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Isameldin Mohammed Suliman

PERFORMANCE ANALYSIS OF COGNITIVE RADIO NETWORKS AND RADIO RESOURCE ALLOCATION