Coexistence and Energy Efficiency in Wireless Networks

IOANNIS GLAROPOULOS

Doctoral Thesis in Telecommunications
Stockholm, Sweden, 2015

TRITA-EE 2015:009
ISSN 1653-5146
ISBN 978-91-7595-520-9

School of Electrical Engineering
KTH, Stockholm, Sweden

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i telekommunikation tisdag den 12 Maj 2015 klockan 13:15 i sal F3, Lindstedtsvägen 26, KTH.

Tryck: Universitetsservice US AB

Abstract

Dynamic spectrum access has been recently proposed to increase the utilization of the licensed spectrum bands, and support the constantly growing volumes of mobile traffic in the modern society. At the same time, the increasing demand for wireless connectivity, as a result of the rapid emergence of innovative wireless and mobile services, has led to the deployment of various wireless technologies in the open ISM bands. This thesis addresses the effective coexistence among the diverse wireless technologies in the above scenarios, and the energy efficiency of the deployed wireless systems, both listed among the key challenges that wireless networking is facing today.

We discuss cooperative sensing, a fundamental mechanism for allowing unlicensed users perform opportunistic access in the licensed spectrum. Considering the scenario where the users perform both sensing and unlicensed spectrum access, we evaluate the efficiency of multi-channel cooperative sensing schemes with respect to the per user achievable capacity. We conclude that a careful optimization of both the number of sensed channels, and the allocation of sensing duties to the network users is necessary to achieve high capacity gains in large-scale networks of unlicensed users.

We address a number of energy efficient design issues for sensor networks and wireless LANs. We study how to improve the energy efficiency of low-power sensor networks operating under the interference from a coexisting WLAN. We propose a cognitive, cross-layer access control mechanism that minimizes the energy cost for multi-hop WSN communication, by deriving energy-optimal packet lengths and single-hop transmission distances, based on the knowledge of the stochastic channel activity patterns of the interfering WLAN. We show that the proposed mechanism leads to significant performance improvements on both energy efficiency, as well as end-to-end latency in multi-hop WSN communication, under different levels of interference. Additionally, we develop and validate the considered WLAN channel activity model and implement efficient, lightweight, real-time parameter estimation methods.

We investigate how to enhance the multi-hop communication performance in ad hoc WLANs, when 802.11 stations operate under a power saving duty-cycle scheme. We extend the traffic announcement scheme of the 802.11 power saving mode, allowing the stations to propagate pending frame notifications to all nodes in the end-to-end forwarding path of a network flow. We study the performance of the proposed scheme with respect to end-to-end packet delay and signaling overhead, while we investigate the impact on the achievable duty-cycle ratios of the wireless stations. For the purpose of the evaluation, and for the comparison with the standard 802.11 power saving mechanism, we implement the protocol extension in a development platform.

Finally, we study how the combination of the objectives for energy efficiency and a high quality of service impacts the topology stability of self-organized ad hoc networks comprised of individual agents. Based on a non-cooperative game theoretic model for topology formation, we identify key extensions in the nodes' strategy profile space that guarantees a stable network formation under multi-objective player utility functions.

Sammanfattning

Dynamic spectrum access har nyligen föreslagits som ett sätt att öka utnyttjandet av licensierade frekvensband, och på så vis stödja det moderna samhällets ständigt växande volym av mobiltrafik. Samtidigt har den ökade efterfrågan på trådlös anslutning, till följd av snabbt framväxande, innovativa trådlösa och mobila tjänster, lett till utbyggnaden av diverse trådlösa tekniker i de öppna ISM-banden. Denna avhandling behandlar effektiv samexistens bland de olika trådlösa teknikerna i ovanstående scenarier och energieffektiviteten hos de utplacerade trådlösa systemen, två av de nyckelutmaningar som trådlösa nätverk står inför idag.

Vi diskuterar kooperativ avkänning, en grundläggande mekanism för att olicensierade användare opportunistiskt ska kunna få åtkomst till licensierade spektrum. Utifrån scenariot där användarna utför både avkänning och olicensierad spektrumtillgång utvärderar vi effektiviteten med avseende på varje användares uppnåeliga kapacitet. Vi drar slutsatsen att en noggrann optimering av både antalet avkända kanaler och tilldelningen av avkänningsuppgifter till nätanvändare är nödvändiga för att uppnå höga kapacitetsvinster i storskaliga nätverk av olicensierade användare.

Vi tar upp ett antal frågor om energieffektiv design för trådlös sensornätverk (WSN) och WLAN. Vi studerar hur man kan förbättra energieffektiviteten hos ett sensornätverk som verkar under störningar från ett samexisterande WLAN. Vi föreslår en kognitiv, lageröverskridande mekanism för åtkomstkontroll som minimerar energikostnaden för multi-hop kommunikation i WSN. Framtagningen av åtkomstkontrollen sker genom härledning av energioptimerade paketlängder och överföringsavstånd, baserat på kunskap om stokastiska kanalaktivitetsmönster i störande WLAN. Vi visar att den föreslagna mekanismen leder till betydande prestandaförbättringar både avseende energieffektivitet, och end-to-end latens i multi-hop WSN kommunikation under olika nivåer av störningar. Dessutom utvecklar vi och validera den föreslagna kanalaktivitetsmodellen för WLAN och implementerar effektiva och lätta realtidsmetoder för skattning av parametrar.

Vi undersöker hur man kan förbättra prestandan för multi-hop kommunikation i ad hoc WLANs då 802.11 stationer verkar enligt energisparande duty-cycle system. Vi utvidgar tekniken för trafikmeddelande hos 802.11 i energisparläge, och studerar prestanda i det föreslagna systemet med avseende på end-to-end fördröjning och behovet av ytterligare signalering. Samtidigt undersöker vi effekten av de uppnåeliga duty-cycle förhållandena hos de trådlösa stationerna. För utvärdering och jämförelse med standardmekanismen för energisparande i 802.11 implementerar vi det utvidgade protokollet i en utvecklingsplattform.

Slutligen studerar vi hur kombinationen av mål för energieffektivitet och hög kvalitet av tjänster påverkar stabiliteten i topologin hos självorganiserade ad hoc nätverk bestående av enskilda aktörer. Baserat på en modell för icke-kooperativa spel vid topologi-bildning, identifierar vi viktiga tillägg till nodernas strategiska profil som garanterar en stabil nätverksbildning enligt spelarnas multi-objektiv nyttofunktioner.

Acknowledgments

First, I would like to express my sincere gratitude to my thesis advisor, Assoc. Professor Viktoria Fodor, for her continuous guidance. I am grateful to her for the support, encouragement, patience, as well as for all our fruitful discussions that provided me with invaluable feedback on the research problems I have worked over the past years. I thank her for offering me – several times – the opportunity to leave KTH for short-term research internships, which allowed me to expand my expertise in research fields originally outside the scope of my thesis work. I am thankful to Professor Gunnar Karlsson, director of the Laboratory for Communication Networks, for offering me the opportunity to become a member of LCN, as well as for guaranteeing a creative and relaxing working environment for all doctoral students in our group.

Within the six years of my PhD studies I had the pleasure and the honor to visit several research institutes and collaborate on exciting projects with highly motivated researchers. I want to thank Professor Chiara Petrioli for hosting me in "La Sapienza" in 2010, and for all her constructive feedback for my work. Her influence is present within a major part of this thesis. I am grateful to Dr. Stefan Mangold for hosting me in Disney Research Zurich in 2013, for introducing me to the world of the Internet of Things, and for giving me the opportunity to develop technical skills, highly needed for my future career plans. Finally, I would like to thank Professor Thiemo Voigt for accepting me in his group in the Swedish Institute of Computer Science, and for giving me the opportunity to contribute on several exciting projects within the world of embedded systems. To all of my colleagues in La Sapienza, Disney and SICS: thanks so much for a stimulating working atmosphere! Furtermore, I want to thank the co-authors of my papers, Maria Papadopouli, Loreto Pescosolido, Vladimir Vukadinovic, Carlo Fiscione, and my students Alex Vizcaino Luna and Marcello Lagana, for the great work we accomplished together. It has been a pleasure to work with all of you!

In 2008 I joined LCN, a small laboratory in KTH with six PhD students, and saw it growing today into a large group with more than twenty researchers. I am grateful to the "old" LCNers, Rolf, György, Fetahi, Ian, Vladimir, Ognjen, Sylvia and Ljubica for an enjoyable working environment, great social events and our inspiring Friday-"fika" sessions! Special thanks to my long-term office-mates, Olafur, Ilias, Liping and Valentino, for breaking the routine of the everyday life at the office with unforgettable scientific (and non-scientific) discussions. To all the younger members of LCN: I wish you good luck with your PhD studies!

I would like to thank my friends in Stockholm, for being here for me and for sharing the experience of living abroad. Big thanks to Maria for her support throughout all the steps of my PhD years, constantly reminding me that there is life outside wireless communication research. Finally, I would like to thank my parents for their eternal support, love and care. Alex, my younger brother deserves the final mention in this acknowledgements section: Thanks for all your love, and big congrats for defending your PhD dissertation before me!

Contents

Chapter 1

Introduction

1.1 Motivation

In the recent decades our society has witnessed a dramatic increase in the demand for wireless connectivity in industrial and residential areas, as well as an exponential growth in the volumes of data traffic as a result of the proliferation of mobile broadband services, such as video telephony, personal communication, and mobile multimedia streaming services. At the same time, rapidly emerging application scenarios in the context of Wireless Sensor Networks (WSN) and the Internet of Things (IoT), such as smart homes, building automation, surveillance, and complex industrial control systems, increase the need for wireless connectivity, in both machine-to-machine and machine-to-cloud communication scenarios.

Having to rely on limited spectrum, allocated by regulatory bodies, mobile operators have addressed the exponentially increasing demand for mobile data traffic by, both, expanding the coverage and the deployment density of mobile networks, as well as by investigating ways to increase the efficiency of the allocated licensed spectrum. Dynamic spectrum access, based on the innovative concept of software-defined radio, constitutes an *hierarchical* spectrum sharing paradigm, enabling a more efficient use of the radio spectrum by allowing the co-deployment of wireless systems that can exploit the burstiness of mobile traffic and, thus, make use of temporarily non-utilized licensed spectrum.

Wireless Local Area Networks (WLAN) have addressed the need for wireless connectivity by promoting a *flat*, un-coordinated, unlicensed deployment of WLAN access points in the open industrial, scientific and medical (ISM) spectrum bands, offering cheap, broadband, wireless internet access to machines and individuals. Beside WLANs, mesh radio technologies, such as 802.15.4-based 6LoWPAN, as well as ultra low power wireless Personal Area Network (WPAN) solutions, such as Bluetooth and ZigBee, make use of the unlicensed ISM bands, in an effort to provide cost-efficient machine-to-machine (M2M) communication, for both consumer and large-scale industry quality IoT applications.

The *coexistence* of diverse network technologies in the same spectral bands introduces two significant challenges. First, it requires interference management mechanisms that will effectively restrict the interference to licensed networks in scenarios of hierarchical coexistence. This advance may allow for spectrum regulation changes, which will permit unlicensed access within D-TV, UMTS and LTE spectral resources. Second, it requires access protocol mechanisms that will ensure a fair sharing of spectral resources in case of *flat*, or *heterogeneous* coexistence, that is the co-deployment of *secondary* wireless systems with diverse characteristics in terms of transmission power, coverage and data rates. Instead of being optimized for standalone operation, wireless protocols need to be designed in a way that guarantees efficient access in the shared spectrum bands.

The tremendous expansion in the deployment of wireless systems, in an effort to satisfy the increasing demands for wireless connectivity, has turned energy efficiency into one of the most important considerations in wireless networking. Energy efficient communication can lower the operational costs of wireless systems, allowing for large-scale infrastructure deployments, or permit the realization of environmentally sustainable solutions, such as energy harvesting. Being energy efficient, battery-operating wireless devices with finite power supplies can maximize their operational lifetime, which is a desired feature in scenarios where mobility and portability are crucial application requirements. Lifetime maximization can, additionally, lower the required frequency of human intervention for network re-configuration, and, in general, decrease network maintenance costs. At the same time, energy efficiency should not be guaranteed at the cost of low network performance. Therefore, energy efficient design comprises of mechanisms – spanning, possibly, multiple layers of the wireless protocol stack – that ensure both a low-power operation for the wireless devices, and a high quality of performance.

1.2 Scope and Outline of this Thesis

This thesis focuses on a number of design issues related to efficient wireless coexistence and low-power wireless network operation. The first part of the thesis concentrates on performance modeling and analysis of cognitive access control mechanisms that can guarantee an efficient coexistence between heterogeneous wireless networks. The thesis contributes to the following topics:

- Hierarchical coexistence: we investigate the efficiency of cooperative spectum sensing schemes in cognitive radio networks, with respect to the achievable capacity of the unlicensed users. We study the case of dense ad hoc congnitive networks, evaluating the fundamental limits of secondary capacity under constraints on the interference to the coexisting primary network.

- Flat heterogeneous coexistence: we design a cognitive access control scheme for wireless sensor networks that operate under WLAN interference. The scheme is based on a stochastic characterization of the WLAN channel activ-

ity and employs cross-layer optimizations to increase the energy efficiency in WSN communication.

- Stochastic WLAN modeling: we introduce and analyze stochastic models for WLAN channel activity and develop efficient methods for real-time model parameterization to support interference-aware cognitive access control.

The second part of the thesis addresses issues related to energy efficiency in wireless networks. We focus on the following topics:

- We address the challenge of optimizing the WLAN power saving mechanism to alleviate the negative effects of radio duty-cycling on the communication performance in multi-hop 802.11 ad hoc networks.
- We study topology control in energy-constrained self-organized wireless sensor networks under a game-theoretic formulation with multi-objective player utility functions, reflecting both lifetime and QoS performance objectives.

The thesis is structured as follows: In Chapter 2 we discuss challenges and solution approaches regarding efficient heterogeneous wireless coexistence. Chapter 3 surveys network design approaches towards enhancing the energy efficiency in wireless networks. In Chapter 4 we give a more detailed description of the main analytic and simulation tools that were used in this thesis. Chapter 5 includes a summary of the original contibutions, while Chapter 6 presents the main conclusions derived in this thesis, along with possible directions for future research.

Chapter 2

Wireless Coexistence

Wireless coexistence defines the scenario when various communication networks – often operating on different radio technologies – coexist in the same geographical area and spectrum space. Wireless coexistence can be the result of the deployment of unlicensed, dynamic spectrum access-based networks operating within a licensed spectrum space [1]. Alternatively, it can be the natural outcome of the uncoordinated deployment of several networks inside the same open spectrum band [2]. In both scenarios, however, the spectrum resources must be shared among multiple networks.

The increasing number of wireless and mobile applications and services emerging in the modern society, and the inherent problem of spectrum scarcity make wireless coexistence the ruling scenario, rather than the exception, and, therefore, demand for a rethinking of the mechanisms that regulate shared spectrum access.

Under wireless coexistence the spectrum access mechanisms should be designed for addressing two fundamental issues. In general, they should ensure that the available spectrum is shared, among the different network entities, as efficiently as possible. This implies that the coexisting networks should effectively discover opportunities to utilize their spectrum resources in a way that maximizes their performance. In the particular scenarios involving dynamic spectrum access, the access mechanisms should guarantee that the unlicensed networks are able to adapt their transmission schemes in a way that the resulting interference to the co-deployed licensed networks is controlled.

Efficient spectrum access design should, therefore, be *cognitive*, i.e. aware of the activity of the coexisting networks. In this Chapter we look into the key components of cognitive access mechanisms (Fig. 2.1) that enable an efficient wireless coexistence. We then introduce the most common performance metrics, with respect to which the efficiency of these access mechanisms is evaluated. Finally, we discuss the design and optimization of cognitive access mechanisms under both the aforementioned scenarios of wireless coexistence, focusing on the challenges and the solutions for regulating effectively the utilization of the shared radio spectrum.

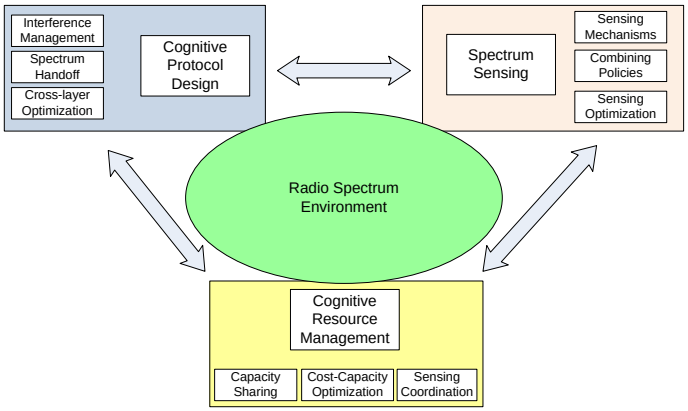


Figure 2.1: Interactions among the key components of a cognitive access scheme.

2.1 Cognitive Spectrum Access

Spectrum sensing

The first challenge in the case of wireless coexistence is how to effectively detect the presence of the co-deployed networks. *Spectrum* or *channel sensing* refers to the mechanism of detecting the presence of transmitted signals within a particular frequency band by listening to the channel. Spectrum sensing offers instantaneous spatio-temporal information about the status of the sensed channel (or spectrum band). Wireless terminals utilize this information to assess both the opportunity of performing a successful transmission within the particular band, as well as the probability of causing harmful interference to a coexisting wireless transmission [3]. In addition to that, spectrum sensing – performed over longer periods – can be used to characterize the statistical properties of spectrum occupancy in the neighborhood of a wireless user [4]. Based on this statistical information that user can adapt its long-term channel access behavior in order to avoid communication impairments due to the coexisting networks and, thus, maximize its communication performance.

Wireless terminals may perform spectrum sensing based on energy detection schemes [5][6] when the nature and the format of the transmitted signals are unknown. Alternatively, they utilize more sophisticated schemes, like match-filter, or cyclo-stationarity-based detectors [7], when a-priori knowledge of the particular signal characteristics is available.

Due to channel noise and signal attenuation phenomena, spectrum sensing is in general imperfect, leading to frequent erroneous channel activity assessments by the sensing devices. The performance of spectrum sensing degrades rapidly with the distance between the transmitter and the sensing device, which decreases

the *signal-to-noise-ratio* over the sensing link. In addition, channel fading and shadowing on the sensing *link* limit the reliability of spectrum sensing mechanisms; this reliability can be increased by enforcing *cooperation* among several sensing devices [8], exploiting the spatial diversity over the sensing links [9][10][11][12]. The cooperative decision can be either *hard*, that is, based on combining individual decisions at each sensing device [13], or *soft* when it combines raw channel sensing measurements at each device [14][15]. Optimal soft decision combining [14] is shown to outperform hard combining schemes as the decision is made exploiting all the knowledge obtained through spectrum sensing.

The *cost* of sensing reflects the *resources* allocated to spectrum sensing, namely the sensing *time* or the sensing *energy* that are spent by the sensing devices, or the signalling and processing *overhead* of exchanging sensing results, in order to perform the collaborative decision. Sensing *optimization* aims at maximizing the achievable sensing performance, subject to certain constrains on the sensing cost [16].

Cognitive network protocol design

Cognitive network control refers to the design of wireless medium access, link-layer and routing control schemes aiming at achieving an efficient utilization of the transmission opportunities within the shared spectrum, discovered via sensing. Cognitive network control addresses two fundamental issues. It enables *interference management*, that is, it regulates the interference among the coexisting networks, and optimizes MAC and routing schemes for communication performance enhancement in coexistence scenarios.

Interference management builds on the information provided by channel sensing. To control the interference to a licensed network, an unlicensed user may need to immediately evacuate a spectrum band on which a signal originating from the licensed system has been detected. Alternatively, the user may apply an effective *power control* scheme, that is adapt its transmission power at a level that it does not cause harmful interference to the ongoing detected transmission [17]. Interference management may additionally involve *channel hopping* [18][19] mechanisms, where wireless users migrate to a different channel in order to mitigate the interference with the detected signals, thus, protecting both their own and the detected wireless transmissions.

Spectrum sensing and frequency hopping can be combined into efficient *spectrum sensing and handoff* schemes [20][21][22], where users dynamically modify their sensing and channel access policies based on the obtained sensing results, in order to limit the interference to and from the coexisting networks.

In addition to the instantaneous information provided by spectrum sensing, a cognitive network control scheme may utilize a-priori statistical knowledge of the transmission patterns of the users of the coexisting networks. Such schemes involve the optimization of a set of *cross-layer* transmission parameters. As far as *Medium Access Control* (MAC) is concerned, cognitive access schemes optimize the

frame transmissions lengths to avoid collisions with the users of the co-deployed networks [23]. Cognitive *routing* schemes involve routing traffic dynamically, avoiding network nodes with limited spectrum resources. Under multi-hop communication cognitive access control may optimize the next-hop selection, with the objective of maximizing the performance of the end-to-end communication under the interference of the coexisting networks [24]. For such solutions it is crucial that the a-priori knowledge of the aforementioned transmission behavior is sufficiently accurate, while, at the same time, it can be obtained at minimal cost.

Finally, cognitive network control may employ medium access protocol techniques that enhance the robustness of single-hop communication, such as enforcing enhanced link-layer transmission handshake mechanisms, thus, improving collision detection and interference mitigation. Alternatively, it may involve mechanisms for smooth inter-operation between the coexisting networks, for example, by a-priori assuming [25], or by identifying the tranmission patterns of the co-deployed networks – decoding link-layer management transmissions [26] – to enable in this way a more efficient spectrum sharing.

Cognitive resource management

In wireless coexistence scenarios *cognitive resource management* refers, to the process of determining the amount of network resources that needs to be spent for discovering transmission opportunities. In addition, it manages the allocation of the resulting transmission opportunities to the network users.

Spectrum resource management models the inherent tradeoff between the resources allocated for spectrum sensing and the resulting sensing performance, that reflects the *cognitive capacity*, that is the amount of spectrum resources available for the network users. This modeling enables the derivation of the sensing parameters that result in a target *cost-capacity* operational point for the cognitive system. As a representative example of cognitive resource management, [27] addresses the problem of sensing efficiency maximization in cognitive radio networks. Considering that the time spent for sensing reflects a capacity loss for the users, the work aims at optimizing the lengths of the spectrum sensing periods.

In the context of collaborative sensing, and since discovering spectrum opportunities requires effort from a set of cooperating users, these users need to decide how large part of the spectrum space they intend to sense and utilize. On one side, a large space may increase the number of channels to sense, so that there are more transmission opportunities to share. On the other side, this requires more sensing efforts from the users, revealing that there is an optimal spectrum space to be sensed that depends, additionally, on the *capacity* requirements of the existing users [28].

An important challenge is how the discovered transmission opportunities will be allocated among the existing wireless users. Optimally, a fair spectrum resource sharing scheme is desired, which implies that the sensing cost of each wireless user quantitatively reflects its achievable transmission capacity [29]. In addition

to that, wireless users may, in general, have different capacity requirements; this diversity among the individual user requirements or objectives needs to be taken into consideration when distributing the cost of spectrum sensing so as to provide strong inscentives for cooperation to the wireless users [30].

2.2 Performance Metrics in Cognitive Coexistence

Sensing and interference control

The *cross-network interference*, defined as the interference between the coexisting networks, can be viewed from two different perspectives: from the transmitter's, or the interferer's, and from the receiver's perspective. From the interferer's point of view we aim at evaluating the ability of a network to detect and effectively *avoid* to cause interference to the co-existing systems. From the perspective of a receiver, we aim at quantifying the ability of a system to efficiently operate in the presence of interfering networks.

Interference avoidance

The ability of wireless system to effectively detect and avoid interfering with a co-deployed network is quantitatively captured by the probabilities of *missed detection*, p_{MD}, and *false alarm*, p_{FA}. p_{MD} denotes the probability that a transmitted signal at an arbitrary point in time is not detected by the users of a coexisting network who aim at simultaneously utilizing the same transmission band in the neighborhood of the transmitted signal. On the other side, p_{FA} defines the probability that channel sensing results in a false detection of signal presence due to channel noise. *Local* missed detection refers to the sensing performance at individual sensing devices, while *global* or *cooperative* missed detection refers to the collaborative detection process by a set of devices. Regardless of the exact spectrum sensing model that is applied,

$$p_{\mathrm{MD}} \triangleq p_{\mathrm{MD}}\left(SNR(d), T_s\right)$$

is a decreasing function of the instantaneous *signal to noise ratio* at the sensing device, while it decreases with the duration of the sensing time allocated for sensing, T_s. As SNR is a decreasing function of the distance separation, d, missed detection probability increases with the length of the sensing link.

Missed detection events, however, do not necessarily result in cross-network interference, unless multiple users from different networks simultaneously attempt to utilize the same channel in the neighborhood of each other. Therefore, a network that intends to operate without causing harmful interference to a coexisting wireless system calculates the *probability of interference*, P_I, on a channel as the joint probability of two events: *i)* a missed detection of an ongoing transmission from a user of the coexisting network in the particular channel, and *ii)* a channel access attempt by a network user that collides with the ongoing transmission, resulting in

transmission error:

$$P_I \triangleq \Pr\{\text{missed detection}, \text{collision}\}.$$

Under wireless coexistence interference can not be completely avoided, due to the imperfections in spectrum sensing and the stochastic nature of the channel access. Instead, coexistence is regulated based on practical non-zero interference *constraints*, i.e. $P_I \leq P_I^{\max}$, which, if met, guarantee an acceptable system performance.

Surviving cross-network interference

From the receiver's point of view we are interested in assessing the ability of a wireless device to communicate successfully under the interference of the co-deployed networks [31]. We quantitatively capture the efficiency of coexistence by evaluating for a transmitter-receiver pair the probability of *successful communication*,

$$\Pr\{\text{success}|d_{\text{t-r}}\},$$

in the presence of cross-network interference. Communication success decreases with the transmitter-receiver spatial separation, $d_{\text{t-r}}$, [24], since a higher distance decreases the receiver signal power, and, consequently, exposes the transmission to potential interference from a larger area,

$$\frac{\partial \Pr\{\text{success}|d_{\text{t-r}}\}}{\partial d_{\text{t-r}}} \leq 0, \quad d_{\text{t-r}} > 0.$$

In addition to that, communication success depends heavily on the transmission properties of the coexisting networks, which, in turn, depend predominantly on the traffic patterns of their users. In general, the duration of the communication, t, decreases $\Pr\{\text{success}|d_{\text{t-r}}, t\}$, since it increases the time interval within which this transmission is exposed to cross-network interference,

$$\frac{\partial \Pr\{\text{success}|d_{\text{t-r}}, t\}}{\partial t} \leq 0, \quad t > 0.$$

Cross-network interference estimation

Efficient wireless coexistence is facilitated if the networks configure their communication mechanisms based on the knowledge of the stochastic spatio-temporal channel access patterns of the co-deployed systems [23]. An accurate modeling and parameter estimation of the channel usage is, therefore, desired under wireless coexistence.

Channel usage patterns – including the durations and the autocorrelation properties of the *active* and *idle* channel periods – depend on the traffic workload of the network users, on the network topology, and on the underlying medium access

mechanisms [32][33]. These factors must be considered when introducing a tractable wireless *channel occupancy* modeling [34]. The *applicability* of the channel occupancy model is assessed applying a *goodness-of-fit* tests, of a set of measurements or observations, against the expected observations under the model in question.

Following the model validation, an efficient parameter estimation algorithm must be designed. The estimation efficiency is assessed by the resulting *accuracy* of the estimated parameters, evaluated by the *parameter estimation errors* as a function of the resources spent for channel occupancy estimation. As the channel occupancy parameterization is performed by the users collecting active and idle period duration samples, with the help of their own channel sensing infrastructure, we evaluate the efficiency of the parameter estimation as the minimum required number of collected samples that guarantee that the parameter estimation error drops below a predefined threshold.

Communication performance

Achievable capacity

Under wireless coexistence, we define a network's *achievable capacity* [35] as the total amount of the shared spectrum resources available for communication. The achievable capacity, C, is a function of the spectrum sensing performance of a network, quantified through the missed detection and false alarm probabilities, the total number of sensed bands, M, as well as the aggregate cross-network channel *load*, ρ, within the sensed spectrum space.

$$C \triangleq C\left(M, \rho, p_{\mathrm{MD}}, p_{\mathrm{FA}}\right). \tag{2.1}$$

The network achievable capacity is then shared among the users, N, of the network, leading to the *per-user* average achievable capacity,

$$C(N) = \frac{C\left(M, \rho, p_{\mathrm{MD}}, p_{\mathrm{FA}}\right)}{N}. \tag{2.2}$$

QoS-related metrics

$C(N)$ indicates the per-user spectrum resources that are available for communication, reflecting nominal user communication performance. Additionally, we may want to evaluate the impact of wireless coexistence on the practically experienced communication quality. For that we introduce a set of user QoS-related performance metrics.

We introduce the *end-to-end transmission delay*, to evaluate the communication delays in multi-hop wireless networks as a result of cross-network interference. The end-to-end delay depends on the experienced interference along the multi-hop transmission paths, which affects the expected number of retransmissions, ETX_r, on each link of the path, where ETX_r is inversely proportional to the probability

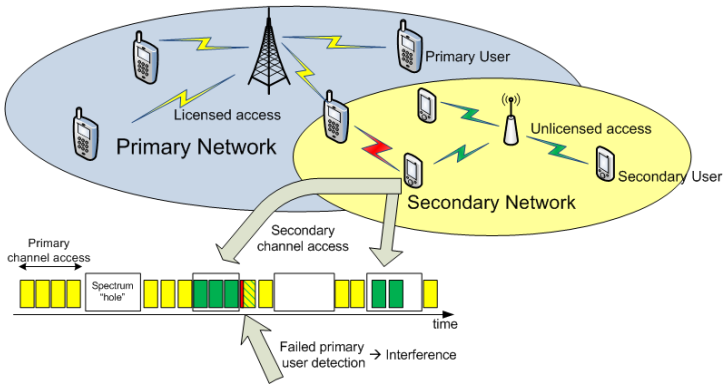


Figure 2.2: Hirearchical primary-secondary network coexistence with the secondary network performing dynamic, unlicensed spectrum access.

of successful transmission,

$$ETX_r = \frac{1}{\Pr\{\text{success}|r\}}.$$

Similarly, in multi-hop wireless networks *end-to-end throughput* defines the information delivery rate – in bits per time unit – between a source and the respective destination node under cross-network interference. Multi-hop paths experiencing high cross-network interference should normally be avoided, in order to maintain high throughput, and to limit the experienced end-to-end delays [36].

Energy efficiency is a commonly set objective for communication networks formed by energy-constrained wireless devices. Designing an energy efficient protocol stack is a fundamental prerequisite, in order to guarantee a sufficiently long network lifetime. Protocol design is energy efficient, when it minimizes the energy cost per transmitted unit of information. Considering, in general, multihop communication scenarios, we quantify energy efficiency by defining the *normalized energy cost* metric [24], which gives the total energy required for transmitting a unit of information over a unit of distance towards the final destination node.

2.3 Design Challenges for Coexistence Scenarios

Hierarchical coexistence: The case of primary-secondary network coexistence

Traditional regulatory access mechanisms in cellular networks, such as exclusive spectrum licensing and spatial frequency reuse often fail to guarantee an efficient usage of the available spectrum [37]. Spectrum may remain highly underutilized

as a result of low instantaneous demand for wireless traffic exchange within the licensed networks [38][3], caused by high spatio-temporal burstiness in user traffic demand. Licensed spectrum underutilization has been experimentally proven in a broad set of scenarios [39], and in particular for cellular – UMTS and LTE – communication networks [37][40].

Parallel to this, we have witnessed the emergence of broadband wireless internet services with lower requirements in terms of user-experience QoS, including data delivery delay, jittering, or packet loss rates. Such services can be supported by unlicenced, low-priority, *dynamic spectrum access*-based networks [41] [1] [42] that coexist with the licensed (or primary) networks and make use of the temporarily non-utilized licensed spectrum (Fig. 2.3).

Wireless coexistence, however, introduces the need for interference control between the licensed, and the unlicensed – or *secondary* – users (SU), since licensed users should not experience any communication performance degradation due to the operation of the unlicensed network. In other words, *interference management*, based on spectrum sensing [43], is the key component behind the deployment of unlicensed (or secondary) communication networks.

Spectrum sensing & capacity maximization

Spectrum sensing is the fundamental mechanism for identifying appropriate transmission opportunities and for protecting the licensed or *primary* user operation. The efficient design of spectrum sensing involves optimizations at both local and global (cooperative) level.

Local sensing optimization: At local level, cognitive users must first optimize the length of their sensing measurements [20][44]. Short-period sensing measurements increase the probability of missed-detecting an active primary user, while longer sensing periods reduce the time available for secondary communication and increase the energy consumption of sensing. As typically there are more than one channels available for secondary access, sensing is often perform sequentially over a set of multiple channels. An important challenge here is how to optimize the order in which sensing is carried out in each of the bands. The work in [45] optimizes the sensing order taking the long term occupancy statistics of the respective channels and minimizes the required sensing energy cost while maintaining a target missed detection probability at each sensed band. To increase energy efficiency sensing order optimization can be combined with dynamically adjusting the sensing time duration [46], upon achieving a target performance. Spectrum sensing can, additionally, employ learning techniques for deriving the optimal sensing order [47], to maximize reliability. Optimal sensing policies may be applied in order to select a particular subset of channels to sense, for example, based on long-term channel availability [48] or short-term band occupancy along with channel quality statistics [49].

Sensing resource allocation: At cooperative level, sensing performance increases with optimal combining of individual sensing measurements, based on the experienced SNR levels at the sensing devices [14], the individual measurement reporting reliability [50], or the correlation among sensing results [51].

In addition, efficient cooperative sensing involves the optimization of the total sensed bandwidth [52] and the extent of cooperation among sensing devices. As discovering spectrum opportunities requires effort from the cognitive users, the users need to decide, first, how large part of the spectrum space, dedicated for unlicensed operation, they want to utilize, and, second, how many of them should cooperate for sensing each band in the spectrum space. On one side, the users may increase the number of channels to sense, so that there are more transmission opportunities to share. On the other side, this requires more sensing efforts from each SU. Similarly, increasing the number of cooperative users lowers the resulting missed detection probability [53], at the expense of linearly increased sensing resource requirement for detecting channel availability. In Paper A, we address the above joint optimization aiming at maximizing the achievable per-user cognitive capacity, as it was defined in Section 2.2 and show how the density of the secondary network, and the desired coexisting licensed network interference constraint are important design factors.

Sensing coordination: After determining the number of users to participate in the cooperative decisions, a remaining issue is how to decide on the exact sensing duties to be allocated to the existing secondary users. This problem is often defined as *sensing coordination* [54]. Correlation-aware sensing cordination schemes [55] aim at guaranteeing that the users sensing the same bands experience uncorrelated channel gains on the sensed links. Sensing coordination may rely on a centralized mechanism that distributes sensing coordination information to the secondary users, ensuring a similar missed detection rate over each of the sensed bands. Alternatively, a distributed approach lets the existing secondary users individually select a set of bands to sense. Clearly the first approach achieves a higher capacity due to balanced detection performance in each sensed band, at the expense of a significant signalling overhead that is required to distribute the coordination information to the users. Such overhead may be prohibited in scenarios where energy efficiency is desired or in cases where time constraints require fast cooperative sensing decisions. In Paper A we define and analyze sensing allocation mechanisms, spanning from fully randomized to fully centralized sensing coordination schemes, and conclude that there exists a constant performance gap between the centralized and distributed approaches that is independent of the network density and the remaining design factors. We achieve this by analytically deriving the asymptotic performance limits for the aforementioned sensing coordination schemes.

Heterogeneous flat coexistence: The case of WSN and WiFi

Flat wireless coexistence is the result of uncoordinated co-deployment of networks operating in overlapping subsets of the open spectrum ISM bands. As opposed to

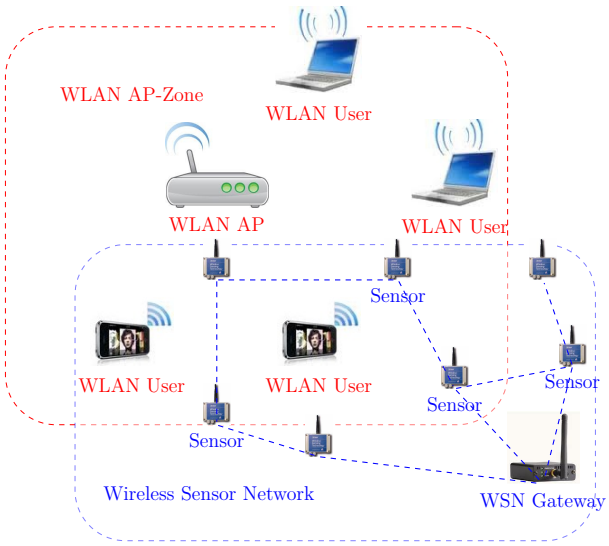


Figure 2.3: Heterogeneous coexistence of 802.11 and 802.15.4 networks in the 2.4GHz ISM band.

the case of hierarchical coexistence, where exclusive spectrum ownership demands efficient interference avoidance mechanisms, flat coexistence focuses on developing protocols that, instead, guarantee an efficient operation for all systems.

In recent years we have witnessed a rapid increase in the technologies operating in the 2.4GHz ISM band, with the common characteristics of being license-free networks, employing random medium access schemes, and supporting error and delay-tolerant communication services. Among the most popular systems we list the wireless sensor networks with customized communication standards, IEEE 802.15.4-based personal area networks (WPAN), IEEE 802.11-based wireless LANs, as well as Bluetooth networks, cordless phones and RFID communication systems.

Due to the different transmission characteristics of the aforementioned systems, flat coexistence is, defined as heterogeneous [24], and imposes different challenges in the design of the different network players. Systems with relatively high transmission power levels, combining, additionally, efficient broadband physical layer, enhanced radio hardware and moderate communication ranges, often do not experience any performance degradation due to the operation of coexisting networks. The protocol stack of such systems can, therefore, be designed and optimized considering standalone operation.

On the opposite side, the performance of systems operating within narrow-band channels and with relatively low transmission power may be severely affected by the presence of high-powered systems. For such networks, the performance of the channel access control mechanisms can be significantly improved, if their design is

cognitive, i.e. aware of the radio environment, including the presence and channel occupancy patterns of the coexisting networks.

In this thesis we focus on the popular scenario of a low-power WSN that operates under the interference of a coexisting WLAN (Fig. 2.3). Heterogeneous coexistence is justified by the relatively high difference in the transmission power of the two network technologies. Due to this difference, WLAN terminals are *blind* towards the WSN transmissions [4], and do not back off when a transmission is initiated that overlaps with that of a WSN packet. As a result of such packet collisions, WSN communication performance degrades, while WLAN througput is hardly affected by WSN interference, a scenario that is often defined as *asymmetric* interference.

The negative impact of the cross-network WLAN interference on the WSN performance has been underlined in a plethora of experimental studies [56], while similar studies have been conducted for Bluetooth systems [57] [58]. In order to survive the WLAN interference and, thus, guarantee a high communication performance, WSNs must employ smart channel access mechanisms, i.e. avoid using the wireless channel simultaneously with the WLAN terminals. We review here the basic principles of cognitive coexistence in the case of flat-hierarchy, asymmetric interference scenarios.

WLAN white space characterization

Model design and validation: Identifying and capturing the statistical properties of the spatio-temporal WLAN channel occupancy enables the WSN users to assess accurately the transmission opportunities under WLAN coexistence [34]. The first step towards this direction is the adoption of an appropriate stochastic model that can describe WLAN occupancy in a broad range of WLAN networking scenarios. To be attractive for analytic performance studies and cognitive access control design, a good model candidate must be relatively simple. It must, additionally, bare the structure and the required degrees of freedom that ensure a good potential of capturing the behavior of WLAN channel occupancy at a *microscopic* level [59], that is, modeling directly the short term temporal behavior of the channel status in WLAN networks.

Related work in this area includes the seminal approach in [33] that derives an analytic model for the impact of IEEE 802.11 MAC protocol on channel occupancy assuming saturated traffic. WLAN traffic, however, is far from saturated; consequently, channel usage models are usually developed based on a-priori considered traffic generation patterns [60] [61], or workload models derived from measurement studies [62] [63] [2]. In this thesis we adopt the interesting approach introduced in [23], where an ON-FF semi-Markovian model is employed to characterize the WLAN channel usage. A significant challenge in WLAN activity characterization is to assess the generality of the proposed model; this may be conducted based on real traces of WLAN channel usage collected from public WLAN hotspot measurements [64], or generated in testbed experiments [2]. Instead, Paper D validates the model applicability over a broaded range of traffic workload scenarios, generated

based on experimentally driven high-layer 802.11 traffic statistics [65], in an effort to close the gap between *macroscopic* WLAN traffic workload modeling [65]–[73] and microscopic channel usage models. Focusing primarily on modeling the idle channel periods, we show that the proposed model exhibits excellent fitting under diverse WLAN scenarios, due to its inherent *mixture* distribution for the idle period lengths, consisting of a right truncated term that models the short 802.11 DCF back-off periods, and a heavy-tailed [74] term for the longer periods of WLAN terminals' inactivity.

Model parameterization: WSN terminals rely on channel sensing, in order to collect a sequence of channel occupancy samples – active and idle period lengths – and to parameterize the WLAN channel usage model [2]. The challenge rises due to the sensing limitations of the WSN terminals, which may only partially detect the WLAN channel activity. Thus, in [75] we enhance the adopted WLAN model considering the WSN limited sensing range, and prove the existence of a closed-form expression for the model stochastic distribution functions on the Laplace transform domain [76].

Estimation algorithms are required to be computationally efficient, in order to be able to run on constrained-resource devices, such as sensor nodes. CPU constraints impose limits on the complexity of the estimation algorithms, while memory constrains require *on-the-fly* computation of the model parameters, without the need for storing the collected WLAN *empirical* channel occupancy traceset. In [75] we describe a estimation algorithm based on maximum-likelihood maximization and show that for a target estimation accuracy, as defined in Section 2.2, the convergence speed – in number of samples – depends on the percentage of the observable WLAN activity. In an attempt to satisfy potential memory limitations, in Paper C [77] we develop an estimation algorithm that allows WSN terminals to dynamically re-compute the model parameters based on a real-time sample collection mechanism. The algorithm structure is based on a modified version of an iterative *discrete stochastic optimization* scheme [78]. In Paper C we prove the algorithm convergence stability based on the properties of the WLAN channel occupancy functions.

Interference-aware protocol design

Under WLAN coexistence WSN terminals need to control channel access in a way that it alleviates the harmful WLAN interference and ensure an effective use of the shared ISM spectrum band. Traditional interference mitigation schemes include *channel hopping* mechanisms, where WSN nodes measure and tune to the best available band for communication [79] [80] [81]. However, the effectiveness of these schemes is debatable, particularly in cases where all considered bands exhibit similar statistical interference. Alternative approaches focus on mitigating the cross-network interference by adding information redundancy [31][82] or by partial intervention with the WLAN MAC operation [26]. The efficiency of these

approaches is accompanied by either significant transmission overhead, or hardware extensions in WSN design.

Effort has therefore been put on exloiting the knowledge of 802.11 channel activity patterns leading to cognitive access control, alternatively denoted as *interference-aware* MAC design. Approaches similar to the seminal work in [83] attempt to jointly optimize polices for channel access and discovery of transmission opportunities, based on a-priori known traffic statistics of the interfering network. A requirement for a wide system-optimization approach is to efficiently couple the cognitive access mechanism with the WLAN channel occupancy model derivation [4] [23]. Our work in Paper B addresses the challenges of model estimation, and cognitive access optimization over partially observable WLAN activity. It shows that the WLAN occupancy statistics serve as input for both the design of the channel sensing scheme, as well as for the optimization of the WSN transmission policies and can, therefore, maximize the probability of transmission success, as defined in Section 2.2 under cross-network interference.

Chapter 3

Energy Efficiency

Energy efficiency is perceived as one of the most important concerns in wireless networking. In a wide range of network applications involving wireless devices with finite energy supplies, energy efficient operation is the key factor behind extending the lifetime of the devices to reasonable times. Typical examples of such application scenarios are battery-powered, radio-capable consumer electronics, such as wearable sport gadgets, health monitoring, or entertaintment electronics, where the requirement for energy efficiency is enforced by battery size limitations, driven by the consumers' demand for portability and minimal device design. In rapidly emerging networking applications within the context of the Internet of Things, such as smart home appliances, building automation, or smart cities, energy efficiency is, additionally, required for scaling up the deployed network infrastructures, while guaranteeing environmentally sustainable operation. Finally, the rapid proliferation of applications for wireless sensor networks, such as monitoring environmental conditions, or targeting surveillance, actuation and automation on complex industrial control systems [84], demands for energy efficient design in an effort to maintain low operational costs, thus, alleviate the concerns about the profitability of smart automation and monitoring solutions in large-scale industrial production.

Energy efficient design in wireless networking refers to two fundamental engineering tasks. The first task is to define appropriate metrics, based on which the energy efficiency of a network can be quantitatively evaluated. The second task is to come up with the required architectural changes in network design, and to engineer novel communication protocols, which will allow the wireless devices to utilize their energy resources as effectively as possible, while maintaining a high quality of service for the applications that use the networking infrastructure.

3.1 Evaluating the Communication Energy Efficiency

Metrics for energy efficiency

Transmission cost: As the major source of energy consumption of low-power wireless devices is associated with their radio operations, the primary mechanism for achieving energy efficiency is the minimization of the nodes' communication energy cost per unit of transmitted information. In general, communication protocol operations involve an inevitable transmission *overhead* – in the form of frame header extensions, link layer packet retransmissions to increase reliability, as well as medium access and routing protocol signalling – which may significantly increase the communication energy cost. We can quantify the cost of transmission overhead by normalizing the energy consumption with the amount of information transmitted by the wireless devices. In multi-hop networking scenarios, the protocol energy efficiency must account for the end-to-end energy cost of information delivery. Based on the above considerations, in this thesis we quantify the energy efficiency of communication protocols by defining the *normalized energy cost* metric [24], which gives the total energy required for transmitting a unit of information over a unit of distance towards the final destination node for a multi-hop end-to-end transmission.

Lifetime: One of the key directions towards energy efficient design is the maximization of the network lifetime, which is the amount of time when the wireless network can sustain the target operational performance. Network lifetime depends not only on the communication energy efficiency at the individual wireless devices, but, additionally, on the distribution of energy consumption among the nodes in the network. In multi-hop networks, exessive traffic relaying increases a node's energy consumption, and may lead to node failures, which impose a severe threat on the connectivity and the stability properties of the network topology. As a wireless node depends on relaying nodes so as to transmit and receive traffic over multiple hops, its lifetime is strongly correlated with the lifetime of the relaying nodes. Therefore, in this thesis we express the *lifetime* of a wireless device as a function of the lifetime of the nodes, on which this device relies, in order to achieve target connectivity properties. 'Energy efficient design' refers to efficient network formation, traffic routing, and topology control that increase the lifetime of the wireless devices.

Duty-cycle ratio: In addition to the energy consumption when transmitting or receiving data, the wireless devices may spend a significant amount of energy resources when they remain *idle*, that is, when they listen to the radio channel waiting to receive information. *Radio duty-cycling* is proposed as the straightforward approach towards mitigating the energy cost of idle listening [85]. Duty-cycling mechanisms are implemented on the *medium access control* (MAC) level [86]. Duty-cycling demands the wireless devices activate their radios only when

they need to participate in data exchange; in the absence of relevant traffic, devices can transit to *sleep* [87] or *doze* state [88] to save energy. Energy savings are high when the devices remain in the doze state for long periods. Therefore, in this thesis we evaluate the efficiency of duty-cycling by the achievable *sleep ratio*, that is the percentage of time a wireless device can operate with its radio de-activated. 'Energy efficient' design concerns the optimization of duty-cycling parameters that maximize the sleep ratio of the devices in a network, under a given traffic workload.

Delay overhead: While it effectively decreases the cost of idle listening, duty-cycling may introduce significant delays in traffic exchange, as the devices are not able to receive data when they are in sleep state. Thus, data transmission needs to be buffered until the receiver wakes-up. In a multi-hop transmission, buffering delays may occur at each intermediate node introducing significant latency in traffic delivery, which, in turn, imposes concerns about the applicability of duty-cycling. A duty cycling-based protocol is efficient when the energy cost savings are achieved at the expence of low traffic exchange delays. Therefore, in this thesis we introduce the *delay overhead* metric to quantify the impact of duty-cycling on data delivery delays. In multi-hop networking scenarios the delay overhead denotes the end-to-end traffic exchange delays as a result of duty-cycling. Here, 'energy efficient design' refers to developing wireless duty cycle-based protocols that maintain a low multi-hop delay overhead.

Energy efficient protocol design in wireless networks

Energy efficiency in wireless networks can be viewed from two opposite perspectives: as a performance objective, or as a built-in constraint in network design. When constituting an objective, energy efficiency refers to the architecture and the optimization of network protocols, that decrease the energy cosumption of resource-constrained wireless devices. From the perspective of a constraint, energy efficient design refers to network architectural changes that allow network protocols to maintain high performance standards, while operating with limited energy resources. This chapter surveys recent developments related to energy efficient design, with the emphasis put on *cross-layer* approaches, where different protocol modules and building blocks, e.g. medium access, routing, or topology control are jointly designed for performance improvements, with respect to the aforementioned performance metrics.

We begin with energy efficient design approaches in wireless ad hoc and sensor networks, and close the discussion with novel advances and contributions towards energy efficiency in 802.11 (WLAN) networks.

3.2 Energy Efficient Design for Wireless Ad hoc and Sensor Networks

Cross-layer protocol optimizations for energy efficiency

Wireless sensors employ *power control* as a means of regulating their energy consumption level [89]. Power control covers a broad area of power conservation techniques that aim at increasing energy efficiency, while satisfying performance requirements, such as througphout, data-rates, and link reliability. Reducing transmission power at the devices decreases, in general, communication range and data-rates. Therefore, on network scope, power control is often coupled with routing and link scheduling optimization, in order to minimize the normalized communication energy cost of the wireless nodes subject to connectivity and data delivery requirements. The seminal works in [90][91] propose algorithmic solutions that jointly optimize routing selection and power allocation in a wireless network, so as to mimimize the normalized energy cost for a given traffic workload that describes the rates, at which traffic is to be delivered between specific source-destination pairs. The optimal route selection takes into account the interference between simultaneous transmissions, thus, schedules neighboring link transmissions in different time slots.

Wireless sensors may, additionally, control the transmission overhead, and, consequently, the communication energy efficiency, by applying *packet length optimization* techniques [92]. Large packet sizes decrease the framing overhead, but lead to higher packet error rates, and, therefeore, frequent retransmissions, since the packets are exposed for a longer period to channel noise and interference from simultaneous tranmsissions. Consequently, large packet sizes may increase the retransmission overhead at the MAC layer. Packet length optimization is, therefore, a crucial design factor for energy efficient communications in wireless networks [93]. Intuitively, packet size and routing may also be jointly optimized for energy efficiency in multi-hop wirless networks [94]. To decrease the normalized end-to-end energy cost of communication, nodes may, for example, select to route their traffic via a larger number of intermediate hops, choosing shorter single-hop links, where large packets can be transmitted with high reliability.

Under coexistence with high-power wireless networks, wireless devices may perform packet length optimization with the help of efficient statistical characterization of the cross-network interference [23]. If the channel activity model of the high-power interferer is known, or can be determined, low-power wireless devices can trade-off larger framing overhead with larger retransmission overhead, to determine the optimal packet size that mimimizes the normalized energy cost [4]. In Paper D we jointly optimize WSN packet size and next-hop transmission distance to maximize energy efficiency under known WLAN interference patterns.

In the context of energy efficient design *topology control* mechanisms are employed in an effort to prolong the lifetime of resource-constrained nodes in the wireless network. A plethora of approaches propose *load-balanced* network topologies, where traffic flows are directed in a way that avoids significant irregularities in

the local energy consumption of the nodes [95][96]. In several approaches, topology control is coupled with power control, so that the wireless devices can select the optimal set of neighbor links and then determine the optimal transmission power, based on the experienced interference on each link [97].

Energy efficiency under QoS considerations

While energy efficiency is an important design factor, the vast majority of networking application scenarios introduces equally important QoS considerations, such as the traffic delivery delay and the reliability of the routing paths. In several cases there exists an inherent conflict between these two categories of design goals. Aiming at achieving higher energy savings and increased lifetime, wireless devices might need to compromise the quality of service. Therefore, significant effort is devoted to network optimization oriented towards both energy and performance efficiency.

The trade-off between energy efficiency and network performance may be analyzed under a multi-objective, system-wide optimization perspective. The analysis requires, first, a cost model that quantitatively reflects both classes of design goals. The wireless devices may, then, employ routing selection, power and topology control, to minimize system-wide cost functions, determined by the considered cost model [98]. In other approaches, the desired trade-off between load-balancing and reliability of traffic delivery is reflected in the routing and MAC protocol parameterization [99]. This allows WSNs to dynamically – and in a distributed fashion – adapt to temporal changes in the traffic workload and the link quality in the network.

Several application scenarios lead to dynamic and self-organized sensor networks, where the assumption of system-wide optimization can not be easily justified [100]. In certain cases, wireless devices might be owned by different entities, with an objective of maximizing their own performance. Scenarios with nodes having self-optimizing, or *selfish* goals, are, in principle, studied applying game-theoretic tools. A key question in such cases, is whether the wireless devices can coverge to stable network operation points, namely *Nash equilibria* (NE), where nodes maximize their individual performance that reflects both a high node lifetime, and a low delay overhead [101][102]. Certain contributions demonstrate that Nash equilibria do not always exist under cost models reflecting contradictory objectives [98]. Several other studies employ constructive methodology, to prove that such NE exist, by formulating iterative games that lead to stable energy efficient topologies [103][100][104].

In an attempt to guarantee stable network formations, a few interesting approaches re-formulate the game-theoretic model of the topology control to extend the space of strategies for the wireless devices by introducing bilateral negotiations between wireless nodes that wish to form communication links [105][106]. We follow a similar approach in Paper F, where the wireless nodes negotiate the quality of traffic relaying they offer, in addition to selecting routing paths for their own traffic. We show that such a strategy space expansion leads to stable Nash equilibrium

topologies, even when the wireless devices aim for both node lifetime, as well as QoS maximization.

Duty-cycling in wireless sensor networks

Duty-cycling in wireless sensor networks is implemented based on periodic radio *wake-up* [85] and transition to sleep state, unless traffic needs to be exchanged by the sensor device. Such a scheme achieves high sleep ratio, in case of low traffic demand, where nodes sleep most of the time. Key design aspects, which characterize the duty-cycling efficiency are the sleep and wake-up scheduling at the sensors [107]. Schemes with fixed wake-up and doze time lengths, and, eventually, sleep ratios [85] perform well when the traffic demand remains fairly stable over time. In case of temporal variations in traffic workload, agile duty-cycling schemes [107] adapt the durations of wake-up and sleep time for higher energy savings. Further efficiency can be achieved through *dynamic* duty-cycling schemes [108], where *idle listening* is restricted to a short time interval in the begining of the wake-up period, or where appropriate signaling enables nodes to dynamically end the wake-up time upon completion of traffic exchange.

WSNs achieve further energy saving gains when employing interference-aware sleep transition policies [109]. In Paper D, WSN devices sense the channel in the beginning of a wake-up period and transit immediately to sleep mode, if channel activity is detected, so as not to risk a possible frame collision due to overlapping transmissions with interfering radio activity.

Despite the undeniable energy savings, duty-cycling can introduce a significant delay overhead in multi-hop traffic delivery, particularly in case sensors sleep for long periods. *Opportunistic routing* under duty-cycling aims at minimizing the delay overhead, by dynamically relaying sensor traffic via any neighboring nodes that are awake and geographically closer to the destination node [110]. From the system-optimization perspective, the duty-cycling ratio may be jointly optimized along with multi-path routing and duty-cycle scheduling, under target delivery delays, and connectivity requirements, the energy efficiency of the sensor network can be minimized by determining optimal values for the sleep ratio and the wake-up schedule of the sensors [111].

A major challenge in WSN duty-cycling is how to achieve schedule-synchronization between the communicating sensor nodes. Exessive signaling overhead, large node populations, node mobility, or churn do not allow for controlling the duty-cycle schedules of the network nodes in a centralized fashion. Transmitter-receiver *rendezvous* in time is, therefore, achieved, on the basis of transmitter-initiated duty-cycle MAC protocols [112], where potential transmitters use frame *preambles* to wake-up the intended receivers. Approaches like [113] rely on periodic channel sampling checking for preambles that indicate upcoming packet transmissions. In the same context [114] reduces excessive preambling applying short *strobbing*, which additionally embedds target receiver addressing to wake-up only the intended receiver. Instead of using preambling, [115] employs *opportunistic schedule learning*

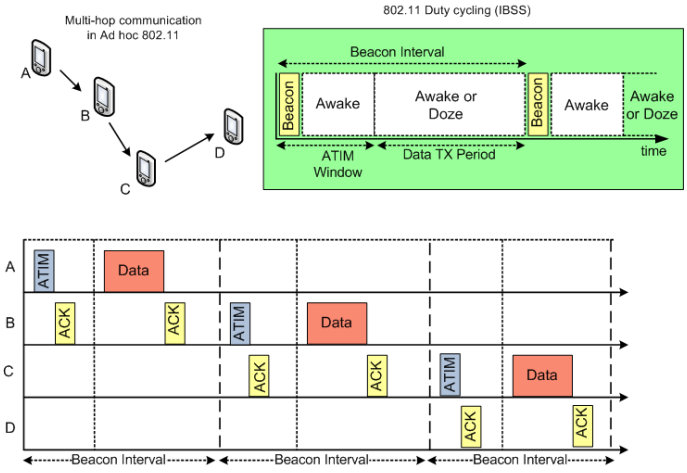


Figure 3.1: Power-saving mode in IEEE 802.11 ad hoc networks.

where WSN devices obtain the scheduling information of nodes within their physical neighborhood by overhearing data transmissions, while in [116] synchronization is achieved through periodically broadcasting scheduling information.

Duty-cycle synchronization is, however, not perfect due to hardware imperfections leading to CPU clock drifts at the WSN devices, and thus, synchronization *gaps* need to be taken into consideration when designing duty cycling-based protocols for WSNs. A common approach to mitigate the effect of synchronization offsets is by slightly increasing the duration of either preamble transmissions or idle listening [116] [113]; a similar approach is implemented in Paper D where sensors wait for a period equal to the maximum synchronization offset before attempting frame transmissions, to guarantee that the intented receiver is awake.

3.3 Duty-cycling in WLAN Ad hoc Networks

The energy consumption of the 802.11-based radio operation is significantly higher compared to 802.15.4-based radio link layers – commonly used in WSN devices – offering, in exchange, higher communication ranges and transmission rates. Such communication characteristics may be beneficial in sensor applications involving generation of a large amount of data traffic. In addition, 802.11 enables sensor devices to interact directly with consumer electronics, such as smartphones, tablets or laptops, that is, market segments where WLAN connectivity dominates over alternative wireless technologies.

However, due to the large transmission energy consumption, 802.11 radio operations can quickly drain the batteries of energy-constrained WLAN-enabled devices.

Therefore, 802.11 duty-cycling, as a means of power-saving, is crucial for enabling WLAN connectivity in WSN and IoT applications [117]. For this reason radio duty-cycling, has been standardized for IEEE 802.11 for both infrastructure (BSS) and ad hoc (IBSS) network mode, and is denoted as *Power Saving Mode* (PSM) [88]. In 802.11 infrastructure networks, stations in power-save mode remain in doze state if they have no traffic to transmit, and wake-up periodically in order to check whether they need to receive data from the access point. Traffic indications for each associated station are embedded in the beacon frame sent periodically by the access point.

In 802.11 ad hoc networks, all stations have synchronized duty-cycle schedules, and compete for a beacon transmission at the beginning of a duty-cycle period (Fig. 3.3), defined as *beacon interval*. Stations in power saving mode must transit to the *awake* state at the beginning of the duty-cycle. The stations announce pending data traffic to other stations by transmitting unicast, short traffic *announcement* (ATIM) frames. ATIM frames are transmitted and acknowledged during the ATIM *window*, during which all stations are awake. After the expiration of the ATIM window, stations transit to doze state for the remaining of the duty-cycle, if they have not been involved in an ATIM/ACK message exchange.

802.11 PSM defines a non-adaptive duty-cycle scheme for decreasing the idle listening at WLAN stations. Interesting research directions in the literature propose protocol enhancements for the 802.11 PSM operation in infrastructure WLANs [118]–[125], involving load driven wake-up scheduling, and joint time division and power control optimization, aiming for further energy preservations through higher achievable sleep ratios. In ad hoc WLANs enhancing the standard PSM involves the introduction of mechanisms for early transition to the doze state [123], [126]–[129], by including additional information in the ATIM frames [126], or by delaying a beacon transmission attempt [123], indicating, implicitely, the lack of pending traffic.

A key approach towards higher sleep ratios is the optimization of the ATIM window length. [127] maximizes the percentage of time the stations remain in doze state by optimizing the duration of the ATIM window subject to throughput constraints for a given traffic workload at the stations. A major challenge in the ad hoc 802.11 PSM is the signalling overhead, imposed by the unicast ATIM frames at each beacon cycle. PSM signaling may be significant in case of multiple concurrent source-destination traffic flows in a WLAN. Exploiting ATIM frame overhearing [130], [131] and deferring from ATIM transmissions is a promising solution towards lower PSM signaling overhead.

Despite the achieved energy savings, PSM results in significant frame delivery delays, as data may need to be buffered at intermediate stations, and delayed until the following beacon interval, when a new ATIM exchange process will notify the next-hop station towards the final destination to remain active and receive the packet (Fig. 3.3). To limit the *multi-hop end-to-end latency* solutions target MAC and routing *cross-layer* approaches, classified as *static* [128] – where the ad hoc networks are organized hierarchically forming a back-bone with PSM-disabled

stations – or *on-demand* [132], where stations temporarily disable duty-cycling, if they are notified from on-demand routing protocols that they may need to relay traffic. Inversely, the routing protocol may on-demand select the packet forwarding paths based on knowledge of the current power saving status of the stations [133].

The challenge of decreasing the end-to-end latency in multi-hop 802.11 ad hoc networks can be addressed effectively, if stations that implement duty-cycling can dynamically adapt to instantaneous end-to-end traffic demand and defer from transitions to doze state, so that packets can be forwarded within the same beacon interval. This can be realized through cognitive path prediction mechanisms, where stations infer whether they need to remain awake, based on overheard traffic announcements associated with past end-to-end transmission flows [134]. Such approaches achieve high energy and delay efficiency, paticularly in networks with traffic bursts that are long enough so that stations learn the end-to-end path flows and remain awake until the whole data stream is forwarded to the destination station. In scenarios with sporadic, or very dynamic traffic workload, the performance of path learning mechanisms is limited. In such cases, the objective for joint energy and delay efficiency can be addressed effectively, if stations can immediately infer the final destination of an 802.11 frame and wake-up all involved stations on the routing path, at the event of frame generation or reception. In Paper E we address the above issue by proposing a cross-layer approach, where WLAN stations embed the MAC address of the final destination of a WLAN packet inside the ATIM message. Upon receiving ATIM frames, the stations can determine the destination node of a pending 802.11 packet, and can, therefore, forward the received notification to the next-hop station in the routing path, in the current ATIM window. Under such scheme all stations involved in the end-to-end delivery will be notified to remain awake in the current beacon interval. Therefore, the WLAN packet may arrive at the final destination with minimum delay overhead.

Chapter 4

Analytic Models, Methods & Evaluation Tools

This Chapter presents an overview of the main analytic scientific methods and evaluation tools that are used in the context of this thesis. We begin with the main theoretic background for spectrum sensing techniques and for interference modeling in wireless networks. Spectrum sensing has been studied in Paper A, while interference modeling has been considered for the cognitive MAC design in Paper B. We give a brief introduction on stochastic modeling in wireless networks that has been considered extensively in this work. We continue, by presenting the basic theoretic tools for distribution fitting, parameter estimation, and stochastic model validation, which we later employ in Papers C and D for WLAN spectrum occupancy characterization. We close the Chapter by introducing, briefly, our simulation platforms and implementation tools, based on which we evaluated our protocol design proposals in Papers B and E.

4.1 Modeling of the Physical Interference

As discussed in Section 2.2 the physical interference model aim at describing the success of a frame transmission in the presence of temporary overlapping transmissions in the neighborhood of the receiving device. Therefore, the interference model relies on the underlying signal propagation model.

Under a *path-loss*-based signal attenuation model [135], the received signal power, $P_{Rx}(r)$, degrades with the distance, $r > 0$ between the transmitting and receiving device:

$$P_{Rx}(r) = P_0 \cdot P_{L_0} r^{-\eta} \tag{4.1}$$

where P_{L_0}, η denote the signal attenuation at a reference (one meter) distance, and the path-loss *exponent*, respectively. In order to correctly decode a received packet, a terminal needs to receive it with a *Signal to Noise plus Interference Ratio* (SINR)

greater than a given threshold, ζ_{SINR}. Assuming the existence of a single interfering node at distance R_{INT}, the SINR under a path-loss model becomes:

$$\text{SINR} = \frac{P_0 P_{L_0} r^{-\eta}}{P_{\text{INT}} P_{L_0} R_{\text{INT}}^{-\eta} + \sigma_N^2} \tag{4.2}$$

where $P_{\text{INT}}, \sigma_N^2$ denote the power of the interfering signal and AWGN, respectively. The considered SINR threshold combined with the path-loss channel model result in a *disk* interference model, that is a circular interference zone around the receiving terminal, with *interference radius*, $R_{\mathcal{I}}$:

$$R_I(r, \zeta_{\text{SINR}}, P_{\text{INT}}, P_0) \triangleq \sqrt[\eta]{\frac{\zeta_{\text{SINR}} P_{\text{INT}} P_{L_0}}{P_0 P_{L_0} r^{-\eta} - \zeta_{\text{SINR}} \sigma_N^2}}. \tag{4.3}$$

In the event of a temporal overlap between a frame reception and a transmission within the interference zone defined by (4.3), the outcome is a frame collision, resulting in a packet loss event. It is clear that under a fixed SINR threshold the interference radius can be decreased by either increasing the transmission power, P_0, or by decresing the transmission distance, r.

The disk interference model is ideal, as it presents a clear geographic boundary, between the area where frame collision occurs with probability 1, and the area where interference from overlapping transmissions is not harmful. In the presence of shadow fading on the channel, the disk interference model is no longer valid, as the collision events depend on the instantaneous shadowing gains on the transmission and on the interfering link.

Under log-normal shadowing [135] the shadowing *gain* on a transmission link is modelled by a log-normal random variable, thus, (4.1) is extended as:

$$P_{Rx}(d, \zeta) = P_0 \cdot P_{L_0} \cdot r^{-\eta} \cdot 10^{\zeta/10}, \tag{4.4}$$

where ζ is an instance of a zero-mean Gaussian variable, Z with standard deviation, σ_{sh}:

$$f_Z(\zeta) = \frac{1}{\sigma_{\text{sh}} \sqrt{2\pi}} e^{-\frac{\zeta^2}{2\sigma_{\text{sh}}^2}}.$$

Consider that Z_0, Z_{INT} denote the shadowing gains on the transmission and on the interference link. Assuming identical and independent distributions, the instantaneous interference radius $R_{\mathcal{I}}$ is a function of the shadowing realizations:

$$R_{\mathcal{I}}(r, \zeta_{\text{INT}}, \zeta_0, P_0, P_{\text{INT}}) = \sqrt[\eta]{\frac{\zeta_{\text{SINR}} P_{\text{INT}} P_{L_0} 10^{\zeta_{\text{INT}}/10}}{P_0 P_{L_0} r^{-\eta} 10^{\zeta_0/10} - \zeta_{\text{SINR}} \sigma_N^2}}. \tag{4.5}$$

In addition, channel shadowing has an impact on the performance of spectrum sensing. The missed detection probability, p_{MD}, defined in Section 2.2, depends on the received signal power at the sensing device, that is a function of the shadowing

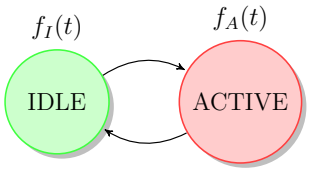


Figure 4.1: A generic two-state stochastic channel occupancy model

gain. Assuming that the gains, ζ_{INT}, is constant within a spectrum sensing period, t_s, the missed detection probability must be averaged over the distribution of the shadowing gain:

$$p_{\text{MD}}(t_s, R_{\text{INT}}) \triangleq \int_0^\infty p_{\text{MD}}(t_s, R_{\text{INT}}, \zeta_{\text{INT}}) f_Z(\zeta) d\zeta, \tag{4.6}$$

while the channel shadowing, clearly, has no effect on the false alarm probability, p_{FA}.

As we show in Paper B channel shadowing introduces uncertainty in the spatial distribution of the interfering sources resulting in lower transmission efficiency in the flat coexistence scenario.

4.2 Stochastic Models for Channel Activity in Wireless Networks

Background

The *channel activity*, or channel usage, in wireless networks is, in general, a stochastic process that reflects the status of the wireless medium, whether it is *active* or *idle*. The channel activity is strongly correlated with the traffic arrival process at the network nodes [136][137]. Its stochastic properties depend, additionally, on the medium access protocol that involves, in general, randomized channel access operations [33][60].

Based on the above consideration, we can use, in the simplest scenario, a *two-state* model to describe the temporal evolution of the channel status in a wireless network. This model is shown in Fig 4.1. The generic functions, $f_{\text{A}}(t), f_{\text{I}}(t)$, denote the probability distributions of the active, and idle channel durations, respectively. As the channel status constantly alters between active and idle state, we can define the stochastic processes, P_{T_A}, P_{T_I}, that describe the sequences of active and idle channel period durations. Thus, $T_I(k)$, denotes the duration of the k-th idle period. $T_I(k)$ is randomly distributed with density function $f_{\text{I}}(t)$.

In general, the random processes, P_{T_A}, P_{T_I}, are not *white*, that is, the process sample values at different sequence indexes are correlated. The correlation is quantitatively evaluated by the *auto-correlation functions*, $R_{T_A}(\tau), R_{T_I}(\tau)$, which are

defined as:

$$R_{T_A}(\tau) = \frac{\mathrm{E}[(T_A(k) - \mu_{T_A})(T_A(k+\tau) - \mu_{T_A})]}{\sigma^2_{T_A}}. \tag{4.7}$$

$$R_{T_I}(\tau) = \frac{\mathrm{E}[(T_I(k) - \mu_{T_I})(T_I(k+\tau) - \mu_{T_I})]}{\sigma^2_{T_I}}. \tag{4.8}$$

Under the assumption of the absence of correlation, the channel activity model is classified as *semi-Markovian*[61]. The model becomes a Markovian one, if, additionally, $f_{\mathrm{A}}(t), f_{\mathrm{I}}(t)$ are exponential distributions [23].

To capture more accurately the behavior of the channel status, we can extend the two-state model to a finite state model, where, in each state, the channel is either active or idle, however, the durations of the active or idle periods are not drawn from the same distributions. Multi-state models can better capture the behavior of the channel status, as a function of traffic arrival and protocol dynamics, at the expense of higher complexity.

Under wireless coexistence scenarios, discussed in Chapter 2, a wide range of network protocol mechanisms may be optimized based on the knowledge of the stochastic patterns of the channel activity. The accuracy of the applied models is critical for the performance of protocol optimizations. In the remainder of this Section we present a brief overview of the analytic tools and methodologies for model parameter estimation and verification, that have used in the context of this thesis.

Parameter estimation techniques

Parameter estimation refers to the procedure of determining the parameters of the functions that constitute the model components, based on a finite set of samples collected from the actual random process that is the subject of our modeling. The estimation of parameters is conducted using distribution fitting techniques, i.e. we determine the appropriate parameter values, for which the analytic distribution matches closely with the *empirical* one that is generated using the collected samples.

Maximum likelihood estimation

Consider a random variable T probability density function, $f_T(t|\boldsymbol{\theta}), t \in \mathbb{R}$, where, $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_K\}$ denotes the vector of the parameters of the density function. The estimation of the parameter set relies on a set of M samples, (or realization) of the random variable, T: $\{t_1, \ldots, t_M\}$.

Based on the collected samples, the *maximum likelihood estimator* (MLE), determines the set of parameters as the solution of the following maximization problem:

$$\boldsymbol{\theta}^* = \{\theta_1^*, \ldots, \theta_K^*\} = \arg \max_{\theta_1,\ldots,\theta_K} f_{T_1,\ldots,T_M}(t_1, \ldots, t_M|\theta_1, \ldots, \theta_K) \tag{4.9}$$

where $f_{T_1,\ldots,T_M}(t_1, \ldots, t_M|\theta_1, \ldots, \theta_K)$ denotes the joint distribution density considering all the collected samples. Assuming an uncorrelated sequence of distribution

realizations, (4.9) reduces to

$$\boldsymbol{\theta}^* = \{\theta_1^*, \ldots, \theta_M^*\} = \arg \max_{\theta_1,\ldots,\theta_K} \prod_{m=1}^{M} f_{T_m}\left(t_m|\theta_1, \ldots, \theta_K\right). \tag{4.10}$$

or, if considering the *log*-likelihood:

$$\boldsymbol{\theta}^* = \{\theta_1^*, \ldots, \theta_M^*\} = \arg \max_{\theta_1,\ldots,\theta_K} \sum_{m=1}^{M} \log\left[f_{T_m}\left(t_m|\theta_1, \ldots, \theta_K\right)\right]. \tag{4.11}$$

We derive the numerical solution of (4.10) by forcing the partial derivatives to zero:

$$\boldsymbol{\theta}^* = \{\theta_1^*, \ldots, \theta_M^*\} = \arg_{\theta_1,\ldots,\theta_K} \left\{ \frac{\sum_{m=1}^{M} \partial \log\left[f_{T_m}\left(t_m|\theta_1, \ldots, \theta_K\right)\right]}{\partial \theta_K} = 0, \forall k \right\}. \tag{4.12}$$

In Paper B we apply a MLE estimator for deriving the parameters of the generalized Pareto distribution [138] that is employed for modeling the heavy tailed behavior of the 802.11 white spaces, based on a modified estimator developed in accordance with [139], to account for a left-truncated nature of the white space distribution.

Estimation in the Laplace domain

The MLE-based estimation can be applied when the closed-form expression of the probability density function, $f_T(t|\boldsymbol{\theta})$, exists. There are, however, cases of distributions that lack a closed form expression in the probability domain. Such cases include, most commonly, composite variables that are the result of the superposition of individual, random variables.

In some particular scenarios, where this superposition comprises summations of uncorrelated variables, it is possible to derive a closed-form expresion in the *Laplace domain* of a variable,

$$f^*_{T|\boldsymbol{\theta}}(s) = \int_0^\infty f_T(t;\boldsymbol{\theta}) e^{-st} dt. \tag{4.13}$$

by expoiting the Laplace transform (LT) property that the sum of independent random variables leads to a joint density function, whose LT is the product of the individual transforms of each variable:

In case of finite discrete random summations, the derivation of the LT expression requires the *generating function* of the discrete distribution that models the random sum. Consider, for instance, that we need to calculate the LT of the following variable:

$$\mathcal{T}_{\mathcal{N}} = \sum_{i=1}^{\mathcal{N}} T^{(i)}, \tag{4.14}$$

where $T^{(i)} \propto T$ and, $\mathcal{N}$ is a discrete variable with probability mass function $p_{\mathcal{N}}$. The generating function of the random variable $\mathcal{N}$ is:

$$\mathcal{G}_{\mathcal{N}}(z) = \sum_{k=0}^{\infty} p_{\mathcal{N}}(k) z^k. \tag{4.15}$$

The, the LT of $\mathcal{T}_{\mathcal{N}}$ is given by:

$$f^*_{\mathcal{T}_{\mathcal{N}}}(s) = \mathcal{G}_{\mathcal{N}}\left(f^*_T(s)\right). \tag{4.16}$$

The above property is applied in [75] and in Paper C, where we derive the Laplace transform of the partially-observed WLAN idle time distribution, as a geometrically distributed sum of WLAN cycles, consisting of consequtive idle an active WLAN periods. Due to this complex combination, the idle time distribution lacks a closed-form expression. Therefore, as MLE can not be applied for deriving the optimal distribution parameters, we develop a heuristic estimation method that relies on the one-to-one correspondence between the Laplace transform and the probability density function [76]. We determine the optimal values for the parameter set as the solution to the following minimization problem:

$$\boldsymbol{\theta}^* \triangleq \{\theta_1, \ldots, \theta_K\} = \arg \min_{\theta_1,\ldots,\theta_K} \frac{1}{S} \sum_{k=0}^{S} \left[f^*_T(s_k|\boldsymbol{\theta}) - f^*_{Te}(s_k;\boldsymbol{\tau})\right]^2 \tag{4.17}$$

where $\mathcal{S} = \{s_0, \ldots, s_S\}$ is a finite discrete subset of the $s-$domain, and $f^*_{Te}(s_k;\boldsymbol{\tau})$ is the *empirical* LT with respect to the parameter set, $\boldsymbol{\theta}$, calculated from the collected distribution sample sequence, $\boldsymbol{\tau} = \{t_1, \ldots, t_M\}$. In Paper C we show that the empirical LT of a random distribution can be calculated on the fly, from the sample sequence as follows:

$$f^*_{Te}(s_k;\boldsymbol{\tau}) = \frac{1}{M} \sum_{m=1}^{M} e^{-s_k t_m} \tag{4.18}$$

Discussion We stress that the outcome of the minimization of (4.17) does not necessarily correspond to the MLE-based estimation of (4.9). The developed heuristic method is, instead, based on the uniqueness of the Laplace transformation, and on the a-priori considered assumption that random distributions with similar Laplace transforms exhibit similar stochastic behaviors. The performance of the proposed heuristic method is assessed in Paper C and in [75], where we evaluate the parameter estimation errors when the heuristic is performed on sample sequences with given parameter sets. In [75] we study the effect of the sample seuqence lengths on the parameter estimation accuracy. We show that the proposed algorithm leads to efficient parameter estimation under sequence lengths in the order of 10^2 to 10^3, while the estimation errors are practicaly eliminated under input sequences with a length of 10^4 to 10^5 samples. The following Section introduces the reader to stochastic optimization-based methodologies for solving the optimization problem in (4.17).

Discrete stochastic optimization

Optimization problems of the form

$$\mathcal{K}_n^* := \arg \min_{\mathcal{K}_n \in \boldsymbol{\mathcal{K}}} \{c(n)\} \tag{4.19}$$

can be solved by applying traditional optimization tools [140], when the objective function, $c(n)$, has a deterministic and closed analytic form. Instead, when the objective function includes stochastic components, optimization is, generally, a harder problem, as the exact relation between the function value with respect to the variables under optimization is obscured. *Stochastic optimization* constitues a general category of methodologies for solving optimization problems, whose objective functions are functions of random variables. Consider the simplest setting of a discrete stochastic optimization problem:

$$\boldsymbol{\mathcal{K}}^* = \arg \min_{\mathcal{K}_n \in \boldsymbol{\mathcal{K}}} \{c(n) = E[X_{\mathcal{K}_n}]\} \tag{4.20}$$

where $c(n) = E[X_{\mathcal{K}_n}]$, is the expectation of a random variable X, whose distribution depends on the variables, $\mathcal{K}_n$, under optimization. *Simulation-based* techniques are employed, when the expectation $E[X_{\mathcal{K}_n}]$ can be determined based on a sequence, $\{X_{\mathcal{K}_n}\}$, of observations obtained via simulations. The optimization problem is *discrete*, when the sample spaces of the variables under optimization are countable sets. Brute force methods, including an exhaustive calculations over all sample space points are, clearly, inefficient for large sample spaces. Approaches such as [78], address the problem by constructing a discrete Markov model over the sample space and study the conditions for convergence to global minima of the objective function. [141] proposes a sample average approximation method for solving discrete stochastic optimization problems with uncostrained objective functions with finite variance. The work assumes a white sampling process over the random objective function and shows that the proposed methods converges to the optimal values under un-constrained sample sizes. In [142] simulation-based stochastic programming is extended to cover constrained optimization problems. A typical consideration in discrete stochastic optimization theory is the covergence rate of the proposed techniques [78], [141], [143], [144]. This consideration is critical in case the stochastic objective function includes time-variant components, and, therefore, time-variant global minimizers. [145] proposes a random-search, Markov-based algorithm that exhibits fast-convergence properties, thus, performing well under stochastic objective functions with high temporal dynamics.

In the context of our work the derivation of the optimal values of (4.17) shall be combined with the process of collecting the sample sequence, $\{t_1, \ldots, t_M\}$, on the basis of an iterative algorithm, where the new samples refine the output of the optimal values. This approach is similar to the ones presented in [78], [141] and has two important advantages. First, the collected samples do not need to be stored in advance before the estimation process begins, limiting the required algorithm

memory. Second, by enforcing smart stopping rules, the execution time – with respect to the number of iterations – can be reduced. In the following we detail the desciption of the adopted discrete stochastic optimization methodology.

The iterative stochastic optimization algorithm [78] aiming at solving the problem of Eq. (4.20), has a discrete and finite set of *states*,

$$\boldsymbol{\mathcal{K}} \triangleq \{\mathcal{K}_1, \ldots, \mathcal{K}_K\}$$

that correspond to the set of possible outcomes of the algorithm. Denote by $\boldsymbol{\mathcal{L}} = \{\mathcal{L}_1, \ldots, \mathcal{L}_L\} \subset \boldsymbol{\mathcal{K}}$ the set of global minimizers of the function c, i.e.

$$\forall \mathcal{L}_i \in \boldsymbol{\mathcal{L}}, \mathcal{K}_n \in \boldsymbol{\mathcal{K}} \setminus \boldsymbol{\mathcal{L}},\ c\left(\mathcal{L}_i\right) < c(\mathcal{K}_n) \tag{4.21}$$

$$\forall i, j = 1, 2, ..., L,\ c(\mathcal{L}_i) = c(\mathcal{L}_j). \tag{4.22}$$

Stochastic optimization algorithms take $\{X_{\mathcal{K}_n}\}$ as the input and outputs an element $\mathcal{L}_i \in \boldsymbol{\mathcal{L}}$. The algorithm is *iterative*, that is, it involves a *search* process that repeats itself as more samples of the random sequence are obtained from the sampling process.

Searching initiates from an arbitrary state, $\mathcal{K}_i \in \boldsymbol{\mathcal{K}}$. In each iteration step, m, the process selects a new state, $\mathcal{K}_j \neq \mathcal{K}_i$, uniformly at random, and obtains the observation of a random variable $Z_{l_m}^{\mathcal{K}_i \to \mathcal{K}_j}$, which is a function of the random variables $\{X_{\mathcal{K}_i}\}_{l_m}$, $\{X_{\mathcal{K}_j}\}_{l_m}$. In general, l_m is a function of the iteration step, m. In most of the cases, however, it is convenient to define l_m as the total number of random observations obtained until iteration m. $\{X_{\mathcal{K}_j}\}_{l_m}, \{X_{\mathcal{K}_i}\}_{l_m}$ denote the current estimation of $c(j), c(i)$, respectively, given the collected samples. The stochastic optimization algorithm moves to the new state $\mathcal{K}_j$, if $Z_{l_m}^{\mathcal{K}_i \to \mathcal{K}_j} > 0$.

We denote by $\mathcal{K}_m$ the algorithm state after iteration m and with $Q_m(\mathcal{K}_n)$ the *popularity* of state $\mathcal{K}_n$, that is the total number of times the algorithm has visited (or remained at) state $\mathcal{K}_n \in \boldsymbol{\mathcal{K}}$ until iteration m. The output of the algorithm is chosen as the most popular state.

Discussion Due to the random selection of the next candidate state at each iteration step, the algorithm corresponds to a discrete time, discrete space Markov process, where the state space is the set $\boldsymbol{\mathcal{K}}$. The transition probabilities, however, are time-variant as they depend on the current number of collected samples that refine the empirical distribution of the sampled process.

In [78] it is shown that the algorithm converges almost surely to a minimizer of $c(n)$, that is, a member of $\boldsymbol{\mathcal{L}}$, after sufficiently large number of iterations, if the following conditions hold:

Condition 1. *For each $\mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}$ and $l \in \mathbb{N}$, there exists a random variable $Z_l^{(\mathcal{K}_i \to \mathcal{K}_j)}$ such that the limit $\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i \to \mathcal{K}_j)} > 0\}$ exists for all $\mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}$ and for all $\mathcal{K}_i \in \boldsymbol{\mathcal{L}}, \mathcal{K}_j \notin \boldsymbol{\mathcal{L}}, \mathcal{K}_n \neq \mathcal{K}_i, \mathcal{K}_j$, and $l \in \mathbb{N}$,*

$$\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_j \to \mathcal{K}_i)} > 0\} > \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i \to \mathcal{K}_j)} > 0\}, \tag{4.23}$$

$$\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_n\to\mathcal{K}_i)} > 0\} \geq \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_n\to\mathcal{K}_j)} > 0\}, \tag{4.24}$$

$$\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i\to\mathcal{K}_n)} \leq 0\} \geq \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_j\to\mathcal{K}_n)} \leq 0\}. \tag{4.25}$$

Condition 2. *$\{l_m\}$ is a sequence of positive integers such that $l_m \to \infty$ as $m \to \infty$.*

Condition 3. *The Markov matrix $\mathcal{P}$ defined in the following equations is irreducible.*

$$\mathcal{P}(\mathcal{K}_i, \mathcal{K}_j) = \frac{1}{K-1} \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i\to\mathcal{K}_j)} > 0\} \quad \forall \mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}, \mathcal{K}_i \neq \mathcal{K}_j, \tag{4.26}$$

$$\mathcal{P}(\mathcal{K}_i, \mathcal{K}_i) = \frac{1}{K-1} \sum_{\mathcal{K}_j \in \boldsymbol{\mathcal{K}}\setminus\{\mathcal{K}_i\}} \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i\to\mathcal{K}_j)} \leq 0\} \quad \forall \mathcal{K}_i \in \boldsymbol{\mathcal{K}}. \tag{4.27}$$

The above result is asymptotic, i.e. convergence is guaranteed after an infinite number of iterations. In practical cases, however, a stopping rule is required to limit the algorithm execution time. In [75] we define both the maximum number of iteration steps, and a stopping rule based on the number of consequent iteration steps that the algorithm remains in the same state. The maximum number of iterations is determined by the length of the sample sequence and the number of samples integrated into the algorithm at each iteration step.

In Paper C we apply the described stochastic optimization algorithm for the estimation of the distribution parameters of the 802.11 idle period duration, which presents a closed form expression in the Laplace domain, and show that the aforementioned conditions are satisfied ensuring the convergence of the algorithm.

Model validation tools

Stochastic model validation provides us with an analytic framework for verifying whether a physical random process can be accurately described by a stochastic model. In the context of this thesis, model validation is performed by analytic techniques, discussed briefly in this Section.

Goodness-of-fit

First, we aim at evaluating how well a derived analytic stochastic model fits a set of real observations originating from a considered random process; a procedure described as a *goodness-of-fit* evaluation.

D-value: Goodness-of-fit is quantitatively assessed based on the *D-value* of the *Kolmogorof-Smirnoff* test, which is defined as the supremum of the differences between the estimated analytic distribution model and the empirical distribution generated by the set of collected samples:

$$D = \sup_{t_m \in \boldsymbol{\tau}} \left| F_T(t_m; \hat{\boldsymbol{\theta}}) - F_{T_{\text{emp}}}(t_m; \boldsymbol{\tau}) \right|. \tag{4.28}$$

In (4.28) $\boldsymbol{\tau}$ is a sequence of real samples collected from the considered random process, $\hat{\boldsymbol{\theta}}$ is the vector of estimated parameters, while $F_T, F_{T_{\text{emp}}}$ denote the analytic and empirical cumulative distribution functions, respectively, which are evaluated on the values of the collected samples. A low D-value indicates the good fitting performance of the analytic model. We underline, however, that the D-value is a conservative goodness-of-fit metric, as it considers the supremum of the point-wise difference between the two functions, instead of the average.

Kolmogorof-Smirnoff test: The D-value measures the fitting offset between the empirical distribution and its candidate analytic fit. A goodness-of-fit *test* assesses the probability that the collected samples of a given random process do originate from the fitted analytic distribution. In this thesis we employ the *two-sample* Kolmogorof-Smirnoff (K-S) test [74]. In particular, we evaluate the K-S *null hypothesis*, i.e. the probability that a sequence of samples generated from the candidate analytic fit can originate from the same distribution as the sequence of the real collected samples. The evaluation is done considering the two-sample K-S statistic:

$$K_n = \sqrt{\frac{n}{2}} \sup_{t_m \in \boldsymbol{\tau}, \hat{t}_m \in \hat{\boldsymbol{\tau}}} \left| F_{T_{\text{emp}}}(t_m; \boldsymbol{\tau}) - F_T(\hat{t}_m; \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}) \right|, \tag{4.29}$$

where $\hat{\boldsymbol{\tau}}$ denotes the sample sequence from the fitted analytic distribution, and n is the length of both sequences. The null hypothesis is assessed by calculating the $p-$value of the test, that is the probability of obtaining a test statistic, K_n, at least as extreme as the observed one. The null hypothesis is rejected at a *significance level* $\alpha \in (0, 1)$, if $K_n > K_\alpha$, where K_α is the *critical* value [74] defined as $K_\alpha = k : \Pr\{K_n > k\} < \alpha$. Typical values of the significance level are $\alpha = 0.1$ or $\alpha = 0.05$.

In Paper B we apply the two-sample K-S test in order to verify or reject the stochastic model that aims at capturing the random process of the 802.11 idle channel durations.

Whiteness property validation

White random processes have the fundamental property that the generated samples are uncorrelated random variables. Such a property is often desired, as it simplifies the stochastic analysis of complex systems. It is, however, not always possible to either justify or verify the assumption of a white random process based on the system functional properties. Therefore, before being introduced as an assumption in the system model, the whiteness property of a random process needs to be experimentally validated.

The whiteness of a stochastic process is, traditionally, verified by inspecting the *autocorrelation* of the sample series generated by the process:

$$R_T(i) = \sum_{m=0}^{n-i-1} t_{m+i} \cdot t_m, \quad t_m \in \mathbb{R}, \ \forall m \tag{4.30}$$

A non-zero *lag-i* autocorrelation, $i > 0$, implies lack of independence among the generated process samples. In most of the cases, howerver, we need to decide on the whiteness of a sample sequence based on a limited number of input samples, which, in general, results in $R_T(i) \neq 0$, for $i > 0$. One solution to this challenge is to compare the statistical behavior of the lag-1 autocorrelation of the input sequence against the lag-i autocorrelation of a *white reference*, that is, a sequence of random samples assumed to be uncorrelated. In case the autocorrelation outputs of the compared processes have significantly different statistical properties, we can safely assume that the input sequence exhibits correlation among the generated samples, and, therefore, the whiteness is not validated. In Paper B we design a test for independence based on this principle, and use it for the characterization of the 802.11 channel usage process in terms of whiteness.

4.3 Simulation Tools

Network simulation tools enable the evaluation of networking scenarios where the complexity limits the applicability of analytically-based performance studies. The particular selection of the appropriate simulation tools is based on the following set of criteria.

An appropriate network simulation tool must provide accurate mathematical models for the real phenomena that are expected to affect the network performance, such as signal propagation and interference models or packet error rate models. A *modular*, or *component-based* structure of a network simulation is desirable, as it facilitates the protocol design and evaluation, allowing for direct testing of a specific network component (e.g. a protocol) by plugging it in the appropriate position in the protocol stack. Simulators should also be *extensible* to enable the design of additional features or components without major modifications in the rest of the simulator platform.

NS-Miracle: The NS-Miracle framework [146], which is based on the popular NS-2 simulation platform, fulfills the aforemention design criteria, and has, therefore, been selected as the platform for the simulation-based evaluation within the context of this thesis. In addition, NS-Miracle offers a broad set of wireless network protocols already implemented and extensively tested, allowing for rapid implementation and evaluation of customized network stacks. Our work in Paper D and in [24], [75], [109] benefits from the feature-rich implementations of the IEEE 802.11 and 802.15.4 protocol variations, while the detailed NS-Miracle physical layer and channel modeling makes it higly attractive for the wireless coexistence scenarios discussed in Chapter 2. The simulation-based study in Paper D is applied to validate the numerical evaluation of the proposed cognitive access mechanism, which is conducted based on a simplifying set of model assumptions and approximations for analytic tractability.

MiXiM framework: The MiXiM framework is an extension of the Omnet++ simulation platform featuring similar libraries for wireless network protocols as NS-Miracle. MiXiM facilitates the debugging of the implementation code, mainly due to its graphical accessories that offer a direct, real-time illustration of network protocol operations. In addition, its statistic toolboxes facilitate the collection of measurement data and its aggregation into network performance statistics. In [59] we have used MiXiM's detailed traffic generation libraries for implementing the various 802.11 traffic workload scenarios for the purpose of WLAN channel occupancy characterization.

Jemula: The *Java Emulator (Jemula)* framework [147] has been used for the simulation development in the context of our work in [148] [149]. Jemula facilitates protocol development and debugging owing to its Java-based implementation, and offers an advanced, real-time graphical interface for protocol operations. Its 802.11 protocol library makes Jemula suitable for the evaluation of the proposed power-save 802.11 MAC enhancements; the lack of detailed physical layer modeling is, however, a major challenge not only for wireless coexistence scenarios, but also for standalone, dense ad-hoc 802.11 network deployments with frequent frame collisions.

Chapter 5

Summary of Original Work

Paper A: Spectrum sharing with low power primary networks
Ioannis Glaropoulos, and Viktoria Fodor
Published in *Proc. of IEEE Dynamic Spectrum Access Networks (DySPAN)*, 2014.

Summary: Access to unused spectrum bands of primary networks requires a careful optimization of the secondary cooperative spectrum sensing, if the transmission powers in the two networks are comparable. In this case the reliability of the sensing depends significantly on the spatial distribution of the cooperating nodes. In this paper we study the efficiency of cooperative sensing over multiple bands, sensed and shared by a large number of secondary users, which form an ad-hoc cognitive network. We show that the per-user cognitive capacity is maximized, if both the number of bands sensed by the secondary network as a whole, and the subsets of these bands sensed by the individual nodes are optimized. We derive the fundamental limits under different sensing duty allocation schemes. We show that with some coordination the per user cognitive capacity can be kept nearly independent from the network density.

The author of this thesis performed the work presented in this paper under the supervision of the second author.

Paper B: Energy efficient COGnitive MAC for sensor networks under WLAN co-existence
Ioannis Glaropoulos, Marcello Laganà, Viktoria Fodor, and Chiara Petrioli
To appear in *IEEE Transactions on Wireless Communications*, 2015.

Summary: Energy efficiency has been the driving force behind the design of communication protocols for battery-constrained wireless sensor networks (WSNs). The energy efficiency and the performance of the proposed protocol stacks, however, degrade dramatically in case the low-powered WSNs are subject to interference from high-power wireless systems such as WLANs. In this paper we propose COG-

MAC, a novel cognitive medium access control scheme (MAC) for IEEE 802.15.4-compliant WSNs that minimizes the energy cost for multihop communications, by deriving energy-optimal packet lengths and single-hop transmission distances based on the experienced interference from IEEE 802.11 WLANs. We evaluate COG-MAC by deriving a detailed analytic model for its performance and by comparing it with previous access control schemes. Numerical and simulation results show that a significant decrease in packet transmission energy cost, up to 66%, can be achieved in a wide range of scenarios, particularly under severe WLAN interference. COG-MAC is, also, lightweight and shows high robustness against WLAN model estimation errors and is, therefore, an effective, implementable solution to reduce the WSN performance impairment when coexisting with WLANs. *The author of this thesis performed the major part of the work presented in this paper, including the analytic modeling and optimization of COG-MAC, the numerical performance evaluation and the design of the simulation experiments. The second author has contributed in the design of the NS simulator, upon which the simulation-based evaluation was conducted. The author of this thesis wrote and revised this paper together with the third author, based on the feedback offered by the fourth author.*

Paper C: Discrete Stochastic Optimization Based Parameter Estimation for Modeling Partially Observed WLAN Spectrum Activity
Ioannis Glaropoulos, and Viktoria Fodor
Published in *Infocommunications Journal*, 2012.

Summary: Modeling and parameter estimation of spectrum usage in the ISM band would allow the competing networking technologies to adjust their medium access control accordingly, leading to the more efficient use of the shared spectrum. In this paper we address the problem of WLAN spectrum activity model parameter estimation. We propose a solution based on discrete stochastic optimization, that allows accurate spectrum activity modeling and can be implemented even in wireless sensor nodes with limited computational and energy resources.

The author of this thesis performed the work presented in this paper under the supervision of the second author.

Paper D: Closing the gap between traffic workload and channel occupancy models for 802.11 networks
Ioannis Glaropoulos, Alexandre Vizcaino Luna, Viktoria Fodor, and Maria Papadopouli
Published in the *Elsevier Journal of Adhoc Networks*, 2014.

Summary: The modeling of wireless network traffic is necessary to evaluate the possible gains of spectrum sharing and to support the design of new cognitive protocols that can use spectrum efficiently in network environments where diverse technologies coexist. In this paper we focus on IEEE 802.11 wireless local area net-

works and close the gap between two popular levels of modeling, macroscopic traffic workload modeling and microscopic channel occupancy modeling. We consider traffic streams generated by established traffic workload models and characterize the networking scenarios where a simple, semi-Markovian channel occupancy model accurately predicts the wireless channel usage. Our results demonstrate that the proposed channel occupancy model can capture the channel idle time distribution in most of the scenarios, while the Markovian assumption can not be validated in all cases.

The author of this thesis performed the major part of the work presented in this paper, under the supervision of the third author, and based on the suggestions and feedback provided by the fourth author. The second author of the paper performed a significant part of the simulation experiments in Section 5. The paper was written by the author of this thesis in collaboration with the third author.

Paper E: Enhanced power saving mode for low-latency communication in multi-hop 802.11 networks
Vladimir Vukadinovic, Ioannis Glaropoulos, and Stefan Mangold
Published in the *Elsevier Journal of Adhoc Networks*, 2014.

Summary: The Future Internet of Things (IoT) will connect billions of battery-powered radio-enabled devices. Some of them may need to communicate with each other and with Internet gateways (border routers) over multi-hop links. While most IoT scenarios assume that for this purpose devices use energy-efficient IEEE 802.15.4 radios, there are use cases where IEEE 802.11 is preferred despite its potentially higher energy consumption. We extend the IEEE 802.11 power saving mode (PSM), which allows WLAN devices to enter a low-power doze state to save energy, with a traffic announcement scheme that facilitates multi-hop communication. The scheme propagates traffic announcements along multi-hop paths to ensure that all intermediate nodes remain awake to receive and forward the pending data frames with minimum latency. Our simulation results show that the proposed Multi-Hop PSM (MH-PSM) improves both end-to-end delay and doze time compared to the standard PSM; therefore, it may optimize WLAN to meet the networking requirements of IoT devices. MH-PSM is practical and software-implementable since it does not require changes to the parts of the IEEE 802.11 medium access control that are typically implemented on-chip. We implemented MH-PSM as a part of a WLAN driver for Contiki OS, which is an operating system for resource-constrained IoT devices, and we demonstrated its efficiency experimentally.

The protocol design proposed in this paper has been the joint effort of the author of this thesis and the first author of the paper. The author of this thesis carried out the protocol simulation development, the implementation, and the simulation evaluation, while the field experiments were performed in collaboration with the first author. All authors collaborated in writing the paper.

Paper F: The Stability of Multiple Objective RPL Tree Formation
Ioannis Glaropoulos, and Viktoria Fodor
Submitted to *IFIP Med-Hoc-Net*, 2015.

Summary: We address the problem of RPL tree formation in self-organized, multi-hop, wireless sensor networks, where resource-constrained nodes may independently select their routing paths that maximize their performance. We study the result of the tree formation applying a non-cooperative game-theoretic model, and show that multiple objectives may lead to unstable Nash graphs with unwanted traffic cycling. To ensure stability we propose an extension of the node's strategy space, denoted as selective routing, that efficiently eliminates non-acyclic formations from the set of Nash equilibria, while the resulting routing decisions comply standard RPL.

The author of this thesis performed the work presented in this paper under the supervision of the second author.

List of publications not included in this thesis

- Ioannis Glaropoulos, "Energy efficient COGnitive MAC for sensor networks under WLAN co-existence – Revised technical report", KTH, 2015.
- Ioannis Glaropoulos, Vladimir Vukadinovic, and Stefan Mangold, "Contiki80211: an IEEE 802.11 radio link-layer for the Contiki OS", in Proc. of *IEEE International Conference on Embedded Software and Systems (ICESS)*, 2014.
- Hossein Shokri, Ioannis Glaropoulos, Viktoria Fodor, Carlo Fiscione, and Antony Ephremides, "Green Sensing and Access: Energy-Throughput Tradeoffs in Cognitive Networking", submitted to *IEEE Communication Magazine* (major revision), 2014.
- Ioannis Glaropoulos, Stefan Mangold, and Vladimir Vukadinovic "Enhanced IEEE 802.11 power saving for multi-hop toy-to-toy communication", in Proc. of *Green Computing and Communications (GreenCom), IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, 2014.
- Ioannis Glaropoulos, Alexandre Vizcaino Luna, Viktoria Fodor, and Maria Papadopouli, "WLAN channel occupancy modeling and validation", in Proc. of *Swedish Communication Technologies Workshop* (Swe-CTW), 2013.
- Marcello Lagana, Ioannis Glaropoulos, Viktoria Fodor, and Chiara Petrioli, "Modeling and estimation of partially observed WLAN activity for cognitive WSNs", in Proc. of *IEEE Wireless Communications and Networking Conference (WCNC)*, 2012.
- Ioannis Glaropoulos, and Viktoria Fodor, "Cognitive WSN MAC for WLAN co-existence", in Proc. of *Scandinavian Workshop of Wireless Adhoc Networks*, 2011.
- Ioannis Glaropoulos, Viktoria Fodor, Loreto Pescosolido, and Chiara Petrioli, "Cognitive WSN transmission control for energy efficiency under WLAN coexistence", in Proc. of *ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, 2011.
- Amin Nahvi, Viktoria Fodor, and Ioannis Glaropoulos, "Performance of deterministic local sensing aggregation under interference", in Proc. of *ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, 2010.
- Ioannis Glaropoulos, and Viktoria Fodor, "On the efficiency of distributed spectrum sensing in ad-hoc cognitive radio networks", in Proc. of *ACM Mobicom, Cognitive Radio Networks Workshop*, 2009.

- Viktoria Fodor, Ioannis Glaropoulos, and Loreto Pescosolido, "Detecting low-power primary signals via distributed sensing to support opportunistic spectrum access", in Proc. of *IEEE International Conference of Communications (ICC)*, 2009.
- Ioannis Glaropoulos, and Viktoria Fodor, "Distributed spectrum sensing for opportunistic and cognitive spectrum access", in Proc. of *Swedish National Computer Networking Workshop (SNCNW)*, 2009.
- Viktoria Fodor, and Ioannis Glaropoulos, "On the gains of deterministic placement and coordinated activation in sensor networks", in Proc. of *IEEE Global Telecommunications Conference (Globecom)*, 2008.

Chapter 6

Conclusions and Future Work

This thesis presents cognitive control and cross-layer optimization techniques in wireless systems, aiming at addressing two important challenges in wireless networking: heterogeneous coexistence and energy efficiency. In this chapter we summarize the main contributions of our work, emphasizing on the most significant conclusions that we derived.

We studied the efficiency of cooperative sensing in ad hoc cognitive networks, where the primary and secondary systems bare similar transmission characteristics. We focused on the particular scenario where the secondary users performing spectrum sensing also aim at utilizing the discovered spectrum opportunities, while satisfying, additionally, interference avoidance constraints for primary users. Sensing efficiency was therefore evaluated with respect to the achievable cognitive capacity for the secondary network. We showed that the cognitive capacity approaches zero in dense networks, if cooperative sensing is performed on a limited spectrum bandwidth. We therefore defined and evaluated various sensing allocation schemes, where the cognitive users sense a limited subset of bands. We developed appropriate analytic models for the sensing efficiency under random, coordinated, and optimal allocation schemes, evaluating the performance gaps due to the different levels of coodination among the secondary users. We studied the fundamental limits of the cognitive capacity in highly dense cognitive networks and showed that the achievable capacity converges to a limit that depends on the transmission characteristics of the primary and the secondary users, on the parameters of local sensing, as well as on the primary interference constraints. Using numerical evaluation we concluded that while the performance gap between random and coordinated sensing is significant, optimizing the sensing allocation based on the actual number of secondary nodes leads to little additional gain and does not compensate for the increased coordination overhead in dense secondary networks.

We addressed the challenge of increasing the energy efficiency in low-power, 802.15.4-based wireless sensor networks, operating under the interference of coexisting 802.11 networks, through the introduction of cross-layer control mechanisms,

that are cognitive of the radio environment, as imposed by the WLAN activity. We proposed COG-MAC, a new cognitive MAC protocol for wireless sensor networks, that aims at minimizing the energy loss due to unsuccessful WSN communication as a result of WLAN interference on the transmitted channel. COG-MAC builds upon known stochastic models for the WLAN channel activity, uses a smart clear channel assessment mechanism and performs channel access, with optimal packet length and transmission distance, to increase the probability of successful packet transmission. We performed a detailed evaluation of COG-MAC based on an accurate analytic performance model and showed that it significantly outperforms benchmark solutions, particularly under severe WLAN interference. We stressed that all the building blocks of COG-MAC are essential for achieving the objective of energy efficient communication. A detailed simulation study of COG-MAC revealed that the protocol achieves significant gains even in multihop WSN topologies.

We addressed the challenge of developing an easy-to-use, yet, accurate stochastic model for WLAN channel activity, as an essential component of congitive access control. We proposed an iterative, simulation-based, discrete stochastic optimization algorithm, to efficiently estimate the model parameters from a set of observations of the channel activity. We showed that the developed algorithm asymptotically converges to the actual model parameters and evaluated the required number of samples with respect to the target estimation accuracy. In addition, we addressed the question, whether the proposed stochastic WLAN channel activity model is a realistic modeling approach for capturing the channel usage patterns in practical 802.11 networks. In particular, we considered traffic streams, generated by established traffic workload models and, additionally, from real WLAN tracesets, and compared the simulated WLAN activity process with the ones predicted by the stochastic model. Our results showed that in a wide range of scenarios the proposed WLAN model can sufficiently capture the behavioral patterns of real WLAN channel activity. We finally identified the aggregate WLAN channel load, and the particular mixture of application-layer traffic as the dominate factors with significant impact on the accuracy of the proposed channel activity model.

Believing that emerging Internet of Things applications will require low-power communication between IoT radio devices and consumer electronics, typically using Wi-Fi for network connectivity, we addressed the challenge of optimizing the 802.11 duty-cycling mechanism, so as to achieve a high communication performance without compromising its enery efficiency. We proposed a multi-hop extension of the standard IEEE 802.11 power saving mehcanism, that enables low-latency communication in multi-hop ad hoc 802.11 networks. We implemented our solution on an embedded open-source platform, demonstrating its effectiveness via an extensive simulation and experimental evaluation. We showed that the enhanced power saving mechanism increases the sleep ratio of the 802.11 stations, to further extend their lifetime under finite power supply. We stressed two important features of our approach: first, that it is software implementable and, additionally, backward-compatible with the standardized 802.11 power saving mechanism, which guarantees interoperability with legacy WLAN devices.

Finally, we studied the problem of network formation in self-organized, multi-hop, wireless sensor networks, where the wireless devices independently select their routing paths towards a gateway node, in order to maximize their individual performance. In particular, we focused on the case where battery-constrained nodes have multi-objective utility functions, aiming at maximing both the experienced QoS and their lifetime. Considering RPL as the routing protocol in the wireless network, we studied the directed acyclic graph, or tree, formation using tools from non-cooperative game theory, investigating the existence and the quality of mixed-strategy Nash equilibiria in the formulated tree-formation game. Our main results showed that under generic node multi-objective utility functions, multi-parent selection strategies lead to unstable Nash directed graphs, that may contain undesired traffic cycling, strictly prohibited by RPL. We therefore proposed an extension of the strategy profile space of the game, where parent nodes may adopt different forwarding polices for their children. We demonstrated that this policy extension can efficiently eliminate non-acyclic graphs from the set of Nash equilibria of the tree formation game.

Future research directions

Our work on cognitive transmission control demonstrated the importance of both the stochastic characterization of the wireless environment, as well as the cross-layer optimization for the design of efficient solutions that will allow for a smooth coexistence between heterogeneous technologies in the open spectrum bands. Yet, it is clear that including, additionally, power and topology control on a system-wide optimization scheme would lead to further performance gains. The exact way under which all the aforementioned building blocks would be combined in a multi-dimensional optimization scheme remains, however, an open issue.

Cognitive transmission control schemes should be agile, thus, include efficient runtime control mechanisms to adapt to a dynamically changing cross-network interference. Such mechanisms would require a quick, yet, accurate assessment of the impact of the protocol operations and of the current interference on the high-level network performance. Existing runtime protocol adaptation frameworks, such as [150], could be extended, providing a detailed mapping of the stochastic characteristics of the cross-network interference on the system performance.

In-line with a plethora of research contributions, this thesis tackles the problem of energy efficient design in wireless sensor networks of resource constrained devices, as the key factor for realizing green, large-scale infrastructures for monitoring and actuation applications. What remains to be assessed thoroughly is the impact of energy efficient protocol stacks on the performance of upper-level networking, including transport and application layer protocols. The lack of deep experimentation on the inter-operability between energy-efficient design and high-level networking impedes the introduction of such large-scale infrastuctures. Early approaches, such as [110], recognize the challenge of the interplay between energy-efficient medium access control and RPL, and showed, instead, that opportunistic RPL fits well with

traditional MAC-layer duty-cycling. Future reseach could focus on analyzing the performance of opportunistic RPL on top of congitive access control schemes, like COG-MAC, under heterogeneous coexistence.

If energy efficient WLAN infrastructures are to be used in large-scale IoT systems, an important question is how the increase in the network scale impacts the interplay between the mechanisms for energy-efficiency and the remaining layers of the wireless protocol stack. The performance of several protocol operations in duty-cycling ad hoc 802.11 networks relies heavily on how accurately the network terminals can synchronize their beacon intervals. Beacon synchronization in large-scale, multi-hop ad hoc WLANs is, however, highly inefficient. The applicability of duty-cycling in multi-hop WLANs is, therefore, an important direction for future investigation. However, network scale raises serious concerns on the performance of cross-layer operations, which constitute essential building blocks of several proposed energy efficiency optimization schemes, but are still under experimentation, or have not yet been fully standardized. A large-scale evaluation of such cross-layer mechanisms is, therefore, crucial for an accurate performance assessment of the various energy efficient design approaches for ad hoc WLANs.

Finally, in the context of the Internet of Things application scenarios, a fundamental research question remains open: Which radio link-layer technology is most suitable for energy-efficient IoT network infrastructures? Several studies, including [151], question whether the combination of IEEE 802.15.4 duty cycle-based link-layers and RPL-based routing over 6LoWPAN results in a network stack with highest preformance with respect to energy efficiency, claiming that under certain application scenarios resource-constrained devices may achieve higher energy savings while operating with an IEEE 802.11-based network stack. Research could, therefore, aim at identifying the key parameters of IoT application scenarios with a decisive impact on the selection of the underlying network stack.

Bibliography

[1] S. Haykin, "Cognitive radio: brain-empowered wireless communications", *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201 –220, 2005.

[2] S. Geirhofer, J. Z. Sun, L. Tong, and B. M. Sadler, "Cognitive frequency hopping based on interference prediction: theory and experimental results", *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 13, pp. 49–61, 2 2009. [Online]. Available: `http://doi.acm.org/10.1145/1621076.1621082`.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey", *Computer Networks*, vol. 50, no. 13, pp. 2127 –2159, 2006. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1389128606001009`.

[4] J. Huang, G. Xing, G. Zhou, and R. Zhou, "Beyond Co-existence; Exploiting WiFi White Space for ZigBee Performance Assurance", in *Proceedings of IEEE International Conference on Netw. Protocols*, 2010.

[5] H. Urkowitz, "Energy detection of unknown deterministic signals", in *Proc. IEEE, vol. 55, pp. 523-531*, 1967.

[6] F. Digham, M. Alouini, and M. Simon, "On the energy detection of unknown signals over fading channels", in *Proceedings of the IEEE International Conference on Communication (ICC)*, 2003.

[7] J. Lunden, V. Koivunen, A. Huttunen, and H. Poor, "Collaborative cyclostationary spectrum sensing for cognitive radio systems", *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4182–4195, 2009.

[8] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: a survey", *Phys. Commun.*, vol. 4, no. 1, pp. 40–62, Mar. 2011. [Online]. Available: `http://dx.doi.org/10.1016/j.phycom.2010.12.003`.

[9] G. Ganesan and Y. G. Li, "Agility improvement through cooperative diversity in cognitive radio", in *Proc. IEEE Globecom, St. Louise, Missouri, USA,*, 2005, pp. 2505–2509.

[10] D. Duan, L. Yang, and J. Principe, "Cooperative diversity of spectrum sensing for cognitive radio systems", *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3218–3227, 2010.

[11] S. M. Mishra, A. Sahai, and R. W. Broderson, "Cooperative sensing among cognitive radios", in *ICC 2005, Istanbul, Turkey*, 2006.

[12] A. Sahai, R. Tandra, and N. Hoven, "Opportunistic spectrum use for sensor networks: the need for local cooperation", *Berkeley Wireless Research Center*, 2006.

[13] V. Fodor, I. Glaropoulos, and L. Pescosolido, "Detecting low-power primary signals via distributed sensing to support opportunistic spectrum access", in *IEEE International Conference on Communications*, 2009, pp. 1 –6.

[14] Z. Quan, S. Cui, and A. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks", *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, 2008.

[15] D. Sun, T. Song, M. Wu, J. Hu, J. Guo, and B. Gu, "Optimal sensing time of soft decision cooperative spectrum sensing in cognitive radio networks", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 4124–4128.

[16] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications", *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[17] X. Kang, Y.-C. Liang, A. Nallanathan, H. Garg, and R. Zhang, "Optimal power allocation for fading channels in cognitive radio networks: ergodic capacity and outage capacity", *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 940–950, 2009.

[18] J. Li and X. Liu, "A frequency diversity technique for interference mitigation in coexisting bluetooth and wlan", in *Communications, 2007. ICC '07. IEEE International Conference on*, 2007, pp. 5490–5495.

[19] H. Su and X. Zhang, "Channel-hopping based single transceiver mac for cognitive radio networks", in *42nd Annual Conference on Information Sciences and Systems, (CISS'08)*, 2008, pp. 197–202.

[20] H. Shokri-Ghadikolaei, F. Sheikholeslami, and M. Nasiri-Kenari, "Distributed multiuser sequential channel sensing schemes in multichannel cognitive radio networks", *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2055–2067, 2013.

[21] H.-B. Chang and K.-C. Chen, "Auction-based spectrum management of cognitive radio networks", *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1923–1935, 2010.

[22] C.-W. Wang and L.-C. Wang, "Modeling and analysis for proactive-decision spectrum handoff in cognitive radio networks", in *Communications, 2009. ICC '09. IEEE International Conference on*, 2009, pp. 1–6.

[23] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive medium access: constraining interference based on experimental models", *IEEE Selected Areas in Communications*, vol. 26, no. 1, 2008.

[24] I. Glaropoulos, M. Lagana, V. Fodor, and C. Petrioli, "Energy Efficient COGnitive MAC for Sensor Networks under WLAN co-existence", *IEEE Transactions on Wireless Communications (To appear)*, 2015.

[25] B. H. Jung, J. W. Chong, C. Jung, S. M. Kim, and D. K. Sung, "Interference mediation for coexistence of wlan and zigbee networks", in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, 2008, pp. 1–5.

[26] K. Chowdhury and I. Akyildiz, "Interferer Classification, Channel Selection and Transmission Adaptation for Wireless Sensor Networks", in *Proceedings of IEEE International Conference on Communications*, Dresden, Germany, Jun. 2009, pp. 1–5.

[27] E. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, "Optimization of cooperative sensing in cognitive radio networks: a sensing-throughput tradeoff view", *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 5294–5299, 2009.

[28] I. Glaropoulos and V. Fodor, "Spectrum sharing with low power primary networks", in *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, 2014, pp. 315–326.

[29] I. Malanchini, M. Cesana, and N. Gatti, "On spectrum selection games in cognitive radio networks", in *IEEE Global Telecommunications Conference (GLOBECOM'09)*, 2009, pp. 1–7.

[30] J. Rajasekharan, J. Eriksson, and V. Koivunen, "Cooperative game theory and auctioning for spectrum allocation in cognitive radios", in *2011 IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2011, pp. 656–660.

[31] C. Liang, N. Priyantha, J. Liu, and A. Terzis, "Surviving Wi-Fi interference in low power ZigBee networks", in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, 2010, pp. 309–322.

[32] D. Kotz and K. Essien, "Analysis of a campus-wide wireless network", in *Proceedings of the 8th annual international conference on Mobile computing and networking*, ser. MobiCom '02, Atlanta, Georgia, USA: ACM, 2002, pp. 107–118. [Online]. Available: `http://doi.acm.org/10.1145/570645.570659`.

[33] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535 –547, 2000.

[34] S. Geirhofer, L. Tong, and B. Sadler, "Dynamic spectrum access in the time domain: modeling and exploiting white space", *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66 –72, 2007.

[35] I. Glaropoulos and V. Fodor, "On the efficiency of distributed spectrum sensing in ad-hoc cognitive radio networks", in *Proceedings of the 2009 ACM Workshop on Cognitive Radio Networks*, ser. CoRoNet '09, Beijing, China: ACM, 2009, pp. 7–12. [Online]. Available: `http://doi.acm.org/10.1145/1614235.1614238`.

[36] M. Vuran and I. Akyildiz, "XLP: a Cross-Layer Protocol for Efficient Communication in Wireless Sensor Networks", *IEEE Transactions on Mobile Computing*, vol. 9, no. 11, pp. 1578–1591, 2010.

[37] M. A. McHenry, "NSF spectrum occupancy measurements, project summary", Shared Spectrum Company, Tech. Rep., 2005.

[38] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access", *IEEE Communications Magazine*, vol. 47, no. 3, pp. 88–95, 2009.

[39] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal", in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum*, ser. TAPAS '06, Boston, Massachusetts: ACM, 2006. [Online]. Available: `http://doi.acm.org/10.1145/1234388.1234389`.

[40] A. Palaios, J. Riihijarvi, and P. Mahonen, "From paris to london: comparative analysis of licensed spectrum use in two european metropolises", in *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, 2014, pp. 48–59.

[41] J. Mitola, "Software radios: survey, critical evaluation and future directions", *IEEE Aerosp. Electro. Syst. Mag.*, vol. 8, pp. 25 –36, 2005.

[42] I. F. Akyildiz, W.-Y. Lee, and K. R. Chowdhury, "CRAHNs: cognitive radio ad hoc networks", *Ad Hoc Networks*, vol. 7, no. 5, 2009.

[43] M. Timmers, S. Pollin, A. Dejonghe, A. Bahai, L. V. der Perre, and F. Catthoor, "Accumulative interference modeling for distributed cognitive radio networks", *Journal of Communications*, vol. 4, no. 3, 2009.

[44] P. Wang, L. Xiao, S. Zhou, and J. Wang, "Optimization of detection time for channel efficiency in cognitive radio systems", in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, 2007.

[45] W.-Y. Lee and I. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks", *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845–3857, 2008.

[46] Y. Pei, Y.-C. Liang, K. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: optimal sensing strategy, power allocation, and sensing order", *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1648–1659, 2011.

[47] A. Mendes, C. Augusto, M. da Silva, R. Guedes, and J. de Rezende, "Channel sensing order for cognitive radio networks using reinforcement learning", in *IEEE 36th Conference on Local Computer Networks (LCN)*, 2011, pp. 546–553.

[48] H. Jiang, L. Lai, R. Fan, and H. Poor, "Optimal selection of channel sensing order in cognitive radio", *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 297–307, 2009.

[49] T. V. Nguyen, H. Shin, T. Quek, and M. Win, "Sensing and probing cardinalities for active cognitive radios", *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1833–1848, 2012.

[50] S. Chaudhari, J. Lunden, V. Koivunen, and H. Poor, "Cooperative sensing with imperfect reporting channels: hard decisions or soft decisions?", *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 18–28, 2012.

[51] L. Khalid and A. Anpalagan, "Cooperative sensing with correlated local decisions in cognitive radio networks", *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 843–849, 2012.

[52] R. Fan and H. Jiang, "Optimal multi-channel cooperative sensing in cognitive radio networks", *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, 2010.

[53] K. Koufos, K. Ruttik, and R. Jantti, "Distributed sensing in multiband cognitive networks", *IEEE Transactions on Wireless Communications*, vol. 10, no. 5, pp. 1667–1677, 2011.

[54] C. Song and Q. Zhang, "Cooperative spectrum sensing with multi-channel coordination in cognitive radio networks", in *IEEE International Conference on Communications*, 2010.

[55] A. Cacciapuoti, I. Akyildiz, and L. Paura, "Correlation-aware user selection for cooperative spectrum sensing in cognitive radio ad hoc networks", *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 2, pp. 297–306, 2012.

[56] L. Lo Bello and E. Toscano, "Coexistence Issues of Multiple Co-located IEEE 802.15.4/ZigBee Networks Running on Adjacent Radio Channels in Industrial Environments", *IEEE Transactions on Industrial Informatics*, vol. 5, no. 2, pp. 157–167, 2009.

[57] L. Angrisani, M. Bertocco, D. Fortin, and A. Sona, "Experimental study of coexistence issues between IEEE 802.11b and IEEE 802.15.4 Wireless Networks", *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 8, pp. 1514–1523, 2008.

[58] D. G. Yoon, S. Y. Shin, W. H. Kwon, and H. S. Park, "Packet error rate analysis of ieee 802.11b under ieee 802.15.4 interference", in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, vol. 3, 2006, pp. 1186 – 1190.

[59] I. Glaropoulos, A. V. Luna, V. Fodor, and M. Papadopouli, "Closing the gap between traffic workload and channel occupancy models for 802.11 networks", *Ad Hoc Networks*, no. 0, pp. –, 2014. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1570870514000821`.

[60] P. Rathod, O. Dabeer, A. Karandikar, and A. Sahoo, "Characterizing the exit process of a non-saturated ieee 802.11 wireless network", in *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, ser. MobiHoc '09, New Orleans, LA, USA: ACM, 2009, pp. 249–258. [Online]. Available: `http://doi.acm.org/10.1145/1530748.1530783`.

[61] J. Misic and V. Misic, "Characterization of idle periods in IEEE 802.11e networks", in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, 2011, pp. 1004 –1009.

[62] L. Stabellini, "Quantifying and modeling spectrum opportunities in a real wireless environment", in *Proceedings of Wireless Communications and Networking Conference (WCNC)*, 2010, pp. 1 –6.

[63] C. Ghosh, S. Pagadarai, D. Agrawal, and A. Wyglinski, "A framework for statistical wireless spectrum occupancy modeling", *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 38 –44, 2010.

[64] C. Phillips and S. Singh, "Analysis of wlan traffic in the wild", in *Proceedings of the 6th International IFIP-TC6 Conference on Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet*, ser. NETWORKING'07, Atlanta, GA, USA: Springer-Verlag, 2007, pp. 1173–1178.

[65] F. Hernández-Campos, M. Karaliopoulos, M. Papadopouli, and H. Shen, "Spatio-temporal modeling of traffic workload in a campus WLAN", in *Proceedings of the 2nd annual international workshop on Wireless internet*, ser. WICON '06, Boston, Massachusetts: ACM, 2006. [Online]. Available: `http://doi.acm.org/10.1145/1234161.1234162`.

[66] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Usage patterns in an urban wifi network", *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1359–1372, Oct. 2010. [Online]. Available: `http://dx.doi.org/10.1109/TNET.2010.2040087`.

[67] A. Ghosh, R. Jana, V. Ramaswami, J. Rowland, and N. Shankaranarayanan, "Modeling and characterization of large-scale wi-fi traffic in public hot-spots", in *INFOCOM, 2011 Proceedings IEEE*, 2011, pp. 2921 –2929.

[68] G. He, J. C. Hou, W.-P. Chen, and T. Hamada, "Characterizing individual user behaviors in wlans", in *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*, ser. MSWiM '07, Chania, Crete Island, Greece: ACM, 2007, pp. 132–137. [Online]. Available: `http://doi.acm.org/10.1145/1298126.1298150`.

[69] X. G. Meng, S. H. Y. Wong, Y. Yuan, and S. Lu, "Characterizing flows in large wireless data networks", in *Proceedings of the 10th annual international conference on Mobile computing and networking*, ser. MobiCom '04, Philadelphia, PA, USA: ACM, 2004, pp. 174–186. [Online]. Available: `http://doi.acm.org/10.1145/1023720.1023738`.

[70] H. Feng, Y. Shu, and O. W. Yang, "Research on characterization of wireless lans traffic", *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 8, pp. 3179 –3187, 2011. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1007570410005526`.

[71] F. Wamser, R. Pries, D. Staehle, K. Heck, and P. Tran-Gia, "Traffic characterization of a residential wireless internet access", *Telecommunications Systems, Springer*, vol. 48, no. 1-2, 2010.

[72] J.-K. Lee and J. C. Hou, "Modeling steady-state and transient behaviors of user mobility: formulation, analysis, and application", in *Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing*, ser. MobiHoc '06, Florence, Italy: ACM, 2006, pp. 85–96. [Online]. Available: `http://doi.acm.org/10.1145/1132905.1132915`.

[73] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network", in *Proceedings of the 1st international conference on Mobile systems, applications and services*, ser. MobiSys '03, San Francisco, California: ACM, 2003, pp. 303–316. [Online]. Available: `http://doi.acm.org/10.1145/1066116.1066127`.

[74] J.-Y. L. Boudec, *Performance evaluation of computer and communication systems*, E. Press, Ed. EPFL Press, 2010.

[75] M. Lagana, I. Glaropoulos, V. Fodor, and C. Petrioli, "Modeling and estimation of partially observed wlan activity for cognitive wsns", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2012, pp. 1526–1531.

[76] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, M. Hill, Ed. McGraw Hill, 2002.

[77] I. Glaropoulos and V. Fodor, "Discrete stochastic optimization based parameter estimation for modeling partially observed WLAN spectrum activity", *Infocommunications Journal*, vol. 4, no. 2, pp. 11–17, 2012.

[78] S. Andradottir, "A global search method for discrete stochastic optimization", *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 513–530, 1996. [Online]. Available: http://dx.doi.org/doi/10.1137/0806027.

[79] J. Ansari and P. Mähönen, "Channel selection in spectrum agile and cognitive mac protocols for wireless sensor networks", in *Proceedings of the 8th ACM international workshop on Mobility management and wireless access*, ser. ACM MobiWac'10, Bodrum, Turkey: ACM, 2010, pp. 83–90. [Online]. Available: http://doi.acm.org/10.1145/1868497.1868511.

[80] K. il Hwang, S.-S. Yeo, and J. H. Park, "Adaptive multi-channel utilization scheme for coexistence of IEEE 802.15.4 LR-WPAN with other interfering systems", in *Proceedings of the 11th IEEE International Conference on High Performance Computing and Communications, 2009. HPCC'09*, 2009, pp. 297–304.

[81] M. Hanninen, J. Suhonen, T. Hamalainen, and M. Hannikainen, "Link quality-based channel selection for resource constrained WSNs", *Springer Advances in Grid and Pervasive Computing*, vol. 6646, pp. 254 –263, 2011.

[82] Y. Wu, G. Zhou, and J. Stankovic, "ACR: Active Collision Recovery in Dense Wireless Sensor Networks", in *Proceedings of IEEE International Conference on Computer Communications*, 2010, pp. 1–9.

[83] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework", *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589 –600, 2007.

[84] *EU FP7 HYDROBIONETS Deliverable D2.2 - Scenario descriptions and system requirements (v2)*, http://www.hydrobionets.eu/index.php/deliverables, 2013.

[85] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient mac protocol for wireless sensor networks", in *Proceedings of Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002).*, vol. 3, 2002, 1567–1576 vol.3.

[86] A. Dunkels, "The ContikiMAC radio duty cycling protocol", Swedish Institute of Computer Science, Tech. Rep. T2011:13, Dec. 2011. [Online]. Available: http://dunkels.com/adam/dunkels11contikimac.pdf.

[87] *IEEE standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks. Specific requirements. Part 15.4: Wireless Medium Access Control and Physical Layer Specifications for Low-Rate Wireless Personal Area Networks*, 2006.

[88] *IEEE standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks. Specific requirements. Part 11: Wireless LAN Medium Access Control and Physical Layer specifications*, 2012.

[89] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen, "A survey of energy efficient network protocols for wireless networks", *Wirel. Netw.*, vol. 7, no. 4, pp. 343–358, Sep. 2001. [Online]. Available: `http://dx.doi.org/10.1023/A:1016627727877`.

[90] R. Cruz and A. Santhanam, "Optimal routing, link scheduling and power control in multihop wireless networks", in *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, vol. 1, 2003, 702–711 vol.1.

[91] R. Bhatia and M. Kodialam, "On power efficient communication over multihop wireless networks: joint routing, scheduling and power control", in *Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 2, 2004, 1457–1466 vol.2.

[92] W. Dong, C. Chen, X. Liu, Y. He, Y. Liu, J. Bu, and X. Xu, "Dynamic Packet Length Control in Wireless Sensor Networks", *IEEE Transactions on Wireless Communications,*, vol. 13, no. 3, pp. 1172–1181, 2014.

[93] Y. Sankarasubramaniam, I. Akyildiz, and S. McLaughlin, "Energy efficiency based packet size optimization in wireless sensor networks", in *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications*, 2003, pp. 1–8.

[94] W. Dong, Y. Liu, C. Wang, X. Liu, C. Chen, and J. Bu, "Link quality aware code dissemination in wireless sensor networks", in *19th IEEE International Conference on Network Protocols (ICNP)*, 2011, pp. 89–98.

[95] Y.-F. Huang, T.-H. Tan, Y.-D. Wang, J.-J. Liaw, and D. Yin, "Performance of an Energy Efficient Routing Design for Wireless Sensor Networks", in *10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN)*, 2009, pp. 196–201.

[96] C. Patra, M. Chattopadhyay, P. Bhaumik, and A. Roy, "Using self organizing map in wireless sensor network for designing energy efficient topologies", in *2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE)*, 2011, pp. 1–6.

[97] Y. Zhu, M. Huang, S. Chen, and Y. Wang, "Energy-Efficient Topology Control in Cooperative Ad Hoc Networks", *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 8, pp. 1480–1491, 2012.

[98] A. Nahir and A. Orda, "The energy-delay tradeoff in wireless networks - System-wide optimization and game-theoretic perspectives", Department of Electrical Engineering, Technion, Israel, Tech. Rep., 2007.

[99] P. Di Marco, C. Fischione, G. Athanasiou, and P.-V. Mekikis, "Harmonizing MAC and routing in low power and lossy networks", in *IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 231–236.

[100] R. Komali, A. MacKenzie, and R. P. Gilles, "Effect of Selfish Node Behavior on Efficient Topology Design", *IEEE Transactions on Mobile Computing*, vol. 7, no. 9, pp. 1057–1070, 2008.

[101] R. Kannan and S. Iyengar, "Game-theoretic models for reliable path-length and energy-constrained routing with data aggregation in wireless sensor networks", *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1141–1150, 2004.

[102] S. Eidenbenz, V. S. A. Kumar, and S. Zust, "Equilibria in topology control games for ad hoc networks", *Mob. Netw. Appl.*, vol. 11, no. 2, pp. 143–159, Apr. 2006. [Online]. Available: `http://dx.doi.org/10.1007/s11036-005-4468-y`.

[103] N. Sadagopan, M. Singh, and B. Krishnamachari, "Decentralized Utility-based Sensor Network Design", English, *Mobile Networks and Applications*, vol. 11, no. 3, pp. 341–350, 2006. [Online]. Available: `http://dx.doi.org/10.1007/s11036-006-5187-8`.

[104] R. Banner and A. Orda, "Bottleneck Routing Games in Communication Networks", *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 6, pp. 1173–1179, 2007.

[105] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjorungnes, "Network Formation Games Among Relay Stations in Next Generation Wireless Networks", *IEEE Transactions on Communications*, vol. 59, no. 9, pp. 2528–2542, 2011.

[106] R. Johari, S. Mannor, and J. N. Tsitsiklis, "A Contract-Based Model for Directed Network Formation", *Games and Economic Behavior*, vol. 56, pp. 56–2006, 2003.

[107] P. Lin, C. Qiao, and X. Wang, "Medium access control with a dynamic duty cycle for sensor networks", in *IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 3, 2004.

[108] T. van Dam and K. Langendoen, "An adaptive energy-efficient mac protocol for wireless sensor networks", in *Proceedings of the 1st international conference on Embedded networked sensor systems*, ser. SenSys '03, Los Angeles, California, USA: ACM, 2003, pp. 171–180. [Online]. Available: `http://doi.acm.org/10.1145/958491.958512`.

[109] I. Glaropoulos, V. Fodor, L. Pescosolido, and C. Petrioli, "Cognitive wsn transmission control for energy efficiency under wlan coexistence", in *Sixth International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, 2011, pp. 261–265.

[110] S. Duquennoy, O. Landsiedel, and T. Voigt, "Let the Tree Bloom: Scalable Opportunistic Routing with ORPL", in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '13, Roma, Italy: ACM, 2013, 2:1–2:14. [Online]. Available: `http://doi.acm.org/10.1145/2517351.2517369`.

[111] O. Yang and W. Heinzelman, "An Adaptive Sensor Sleeping Solution Based on Sleeping Multipath Routing and Duty-cycled MAC Protocols", *ACM Trans. Sen. Netw.*, vol. 10, no. 1, 10:1–10:30, Dec. 2013. [Online]. Available: `http://doi.acm.org/10.1145/2529977`.

[112] T. Kim, I. Kim, Y. Sun, and Z. Jin, "Physical Layer and Medium Access Control Design in Energy Efficient Sensor Networks: An Overview", *IEEE Transactions on Industrial Informatics*, vol. PP, no. 99, pp. 1–1, 2014.

[113] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks", in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, ser. SenSys '04, Baltimore, MD, USA: ACM, 2004, pp. 95–107. [Online]. Available: `http://doi.acm.org/10.1145/1031495.1031508`.

[114] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: a short Preamble MAC Protocol for Duty-Cycled Wireless Sensor Networks", in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems*, ser. SenSys'06, Boulder, Colorado, USA: ACM, 2006, pp. 307–320. [Online]. Available: `http://doi.acm.org/10.1145/1182807.1182838`.

[115] F. Ashraf, R. Crepaldi, and R. Kravets, "Know your neighborhood: a strategy for energy-efficient communication", in *IEEE 7th International Conference on Mobile Adhoc and Sensor Systems (MASS'10)*, 2010, pp. 392–401.

[116] W. Ye, F. Silva, and J. Heidemann, "Ultra-low duty cycle MAC with scheduled channel polling", in *Proceedings of the 4th international conference on Embedded networked sensor systems*, ser. SenSys'06, Boulder, Colorado, USA: ACM, 2006, pp. 321–334. [Online]. Available: `http://doi.acm.org/10.1145/1182807.1182839`.

[117] S. Tozlu, M. Senel, W. Mao, and A. Keshavarzian, "Wi-fi enabled sensors for internet of things: a practical approach", *IEEE Communications Magazine*, vol. 50, no. 6, pp. 134–143, 2012.

[118] D. Qiao and K. Shin, "Smart power-saving mode for ieee 802.11 wireless lans", in *Proc. IEEE Infocom*, 2005.

[119] C.-H. Gan and Y.-B. Lin, "An effective power conservation scheme for ieee 802.11 wireless networks", *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1920–1929, 2009.

[120] Y. He and R. Yuan, "A novel scheduled power saving mechanism for 802.11 wireless lans", *IEEE Transactions on Mobile Computing*, vol. 8, no. 10, pp. 1368–1383, 2009.

[121] Y. Xie, X. Luo, and R. Chang, "Centralized psm: an ap-centric power saving mode for 802.11 infrastructure networks", in *Sarnoff Symposium, 2009. SARNOFF '09. IEEE*, 2009, pp. 1–5.

[122] P. Agrawal, A. Kumar, J. Kuri, M. Panda, V. Navda, and R. Ramjee, "Opsm - opportunistic power save mode for infrastructure ieee 802.11 wlan", in *Communications Workshops (ICC), 2010 IEEE International Conference on*, 2010, pp. 1–6.

[123] J.-M. Choi, Y.-B. Ko, and J.-H. Kim, "Enhanced power saving scheme for ieee 802.11 dcf based wireless networks", in *Personal Wireless Communications*, ser. Lecture Notes in Computer Science, M. Conti, S. Giordano, E. Gregori, and S. Olariu, Eds., vol. 2775, Springer Berlin Heidelberg, 2003, pp. 835–840.

[124] X. Chen, S. Jin, and D. Qiao, "M-psm: mobility-aware power save mode for ieee 802.11 wlans", in *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, 2011, pp. 77–86.

[125] X. Perez-Costa and D. Camps-Mur, "A protocol enhancement for ieee 802.11 distributed power saving mechanisms no data acknowledgement", in *Mobile and Wireless Communications Summit, 2007. 16th IST*, 2007, pp. 1–7.

[126] D.-Y. Kim and C.-H. Choi, "Adaptive power management for IEEE 802.11-based ad hoc networks", in *Proc. 5th World Wireless Congress*, San Francisco, USA, May 2004.

[127] E.-S. Jung and N. Vaidya, "An energy efficient MAC protocol for wireless LANs", in *Proc. IEEE Infocom*, New York, USA, 2002.

[128] S. Yongsheng and T. A. Gulliver, "An energy-efficient MAC protocol for ad hoc networks", *Wireless Sensor Network*, vol. 1, no. 5, pp. 407–416, 2009.

[129] N. Rajangopalan and C. Mala, "Modified power save model for better energy efficiency and reduced packet latency", *American Journal of Engineering and Applied Sciences*, vol. 5, no. 3, pp. 237–242, 2012.

[130] W. Akkari, A. Belghith, and A Ben Mnaouer, "Enhancing power saving mechanisms for ad hoc networks using neighborhood information", in *Proc. Int. Wireless Comm. and Mobile Comp. Conf. (IWCMC)*, Crete, Greece, Aug. 2008, pp. 794–800.

[131] A. Belghith and W. Akkari, "Neighborhood aware power saving mechanisms for ad hoc networks", in *Proc. IEEE Conf. Local Computer Networks (LCN)*, IEEE, Montreal, Canada, Oct. 2008.

[132] R. Zheng and R. Kravets, "On-demand power management for ad hoc networks", in *IEEE INFOCOM*, vol. 1, 2003, pp. 481–491.

[133] R.-H. Hwang, C.-Y. Wang, C.-J. Wu, and G.-N. Chen, "A novel efficient power-saving MAC protocol for multi-hop MANETs", *Int. Journal of Comm. Systems*, vol. 26, no. 1, pp. 34–55, Jan. 2013.

[134] C. Hu and J. Hou, "Lisp: a link-indexed statistical traffic prediction approach to improving ieee 802.11 psm", in *Proc. Int. Conf. Distributed Computing Systems*, 2004.

[135] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[136] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless lan", *SIGMETRICS Perform. Eval. Rev.*, vol. 30, no. 1, pp. 195–205, Jun. 2002. [Online]. Available: http://doi.acm.org/10.1145/511399.511359.

[137] M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, and J. Zahorjan, "Measurement-based characterization of 802.11 in a hotspot setting", in *Proceedings of the 2005 ACM SIGCOMM workshop on Experimental approaches to wireless network design and analysis*, ser. E-WIND '05, Philadelphia, Pennsylvania, USA: ACM, 2005, pp. 5–10. [Online]. Available: http://doi.acm.org/10.1145/1080148.1080150.

[138] J. d. Castillo and J. Daoudi, "Estimation of the generalized pareto distribution", *Statistics & Probability Letters*, vol. 79, no. 5, pp. 684 –688, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167715208004987.

[139] J. Hasler, D. Li, and M. Raschke, "Estimation for the generalized pareto distribution using maximum likelihood and goodness of fit", *Communications in Statistics - Theory and Methods*, vol. 40, no. 14, pp. 2500–2510, 2011. eprint: http://www.tandfonline.com/doi/pdf/10.1080/03610920903324874. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03610920903324874.

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

[141] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, "The Sample Average Approximation Method for Stochastic Discrete Optimization", *SIAM J. on Optimization*, vol. 12, no. 2, pp. 479–502, Feb. 2002. [Online]. Available: http://dx.doi.org/10.1137/S1052623499363220.

[142] Y. Luo and E. Lim, "Simulation-based optimization over discrete sets with noisy constraints", in *Proceedings of the Winter Simulation Conference (WSC)*, 2011, pp. 4008–4020.

[143] W. B. Powell, A. Ruszczynski, and H. Topaloglu, "Learning algorithms for separable approximations of discrete stochastic optimization problems", *Mathematics of Operations Research*, vol. 29, pp. 814–836, 2004.

[144] S. Bhatnagar, V. Mishra, and N. Hemachandra, "Stochastic Algorithms for Discrete Parameter Simulation Optimization", *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 4, pp. 780–793, 2011.

[145] O. N. Gharehshiran, V. Krishnamurthy, and G. Yin, "Adaptive search algorithms for discrete stochastic optimization: A smooth best-response approach", *arXiv preprint arXiv:1402.3354*, 2014.

[146] *Ns-miracle: multi-interface cross-layer extension library for the network simulator*, http://telecom.dei.unipd.it/pages/read/58/. [Online]. Available: `http://telecom.dei.unipd.it/pages/read/58/`.

[147] S. Mangold, *Jemula802*, https://github.com/schmist/Jemula802. [Apr-2013].

[148] I. Glaropoulos, S. Mangold, and V. Vukadinovic, "Enhanced IEEE 802.11 power saving for multi-hop toy-to-toy communication", in *IEEE International Conference on Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom)*, 2013, pp. 603–610.

[149] V. Vukadinovic, I. Glaropoulos, and S. Mangold, "Enhanced power saving mode for low-latency communication in multi-hop 802.11 networks", *Elsevier Journal of Adhoc Networks*, 2014.

[150] M. Zimmerling, F. Ferrari, L. Mottola, T. Voigt, and L. Thiele, "PTunes: Runtime Parameter Adaptation for Low-power MAC Protocols", in *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*, ser. IPSN '12, Beijing, China: ACM, 2012, pp. 173–184. [Online]. Available: `http://doi.acm.org/10.1145/2185677.2185730`.

[151] W. W. Jeff Drake David Najewicz, *Energy Efficiency Comparisons of Wireless Communication Technology Options for Smart Grid Enabled Devices*, General Electric Company, 2010.

Paper **A**

Spectrum sharing with low power primary networks

Ioannis Glaropoulos and Viktoria Fodor

Published in Proc. of IEEE Dynamic Spectrum Access Networks (DySpan), 2014.

Spectrum sharing with low power primary networks

Ioannis Glaropoulos and Viktoria Fodor*

Access Linnaeus Center
KTH, Royal Institute of Thechnology
Stockholm, Sweden

Abstract

Access to unused spectrum bands of primary networks requires a careful optimization of the secondary cooperative spectrum sensing, if the transmission powers in the two networks are comparable. In this case the reliability of the sensing depends significantly on the spatial distribution of the cooperating nodes. In this paper we study the efficiency of cooperative sensing over multiple bands, sensed and shared by a large number of secondary users. We show that the per user cognitive capacity is maximized, if both the number of bands sensed by the secondary network as a whole, and the subsets of these bands sensed by the individual nodes are optimized. We derive the fundamental limits under different sensing duty allocation schemes. We show that with some coordination the per user cognitive capacity can be kept nearly independent from the network density.

1 Introduction

Cognitive radio networks (CRNs), based on sensing the radio environment and adapting the transmission strategies accordingly, may allow for increased utilisation of radio spectrum resources. Due to the cognitive capabilities CRNs can efficiently control the interference among the competing networks [1], or provide secondary access to a spectrum band, ensuring that the incumbent, primary users do not experience severe performance degradation [2].

Cognitive networks gain information on the availability of the radio spectrum through *spectrum sensing* or by accessing some *spectrum database*, decide about their *spectrum access* strategy based on this information, and perform *interference management* to control the interference caused to the other networks in the area. Therefore, the efficiency of the

*This work was supported in part by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216076 FP7 (SENDORA) and by the ICT TNG Strategic Research Area.

cognitive network depends both on the accuracy of the spectrum availability information and the efficiency of the channel access and interference management.

In this paper we consider the specific case of secondary access, and focus of the efficiency of spectrum sensing performed by the nodes of the cognitive network. Spectrum sensing has the significant advantage that the provided information is timely. Unfortunately, spectrum sensing performed locally at the SUs can not give accurate information about the spectrum availability due to the impairments of the wireless channel and the hardware limitations of the sensors [3]. Therefore, to increase sensing performance, *cooperative sensing* is required, where sensing results, exchanged among several sensing nodes, are combined, in order to reliably detect the presence of primary transmissions. The feasibility of cooperative sensing has been shown for the secondary use of Digital TV white space, where the primary transmission is high power and has low time dynamics [4] [5] [6].

In this paper we consider the more challenging scenario, when the both the primary and secondary transmissions use comparable, low transmissions powers. Spectrum sensing and secondary access are challenging in this scenario, since the local sensing performance degrades rapidly as the distance between the primary transmitter and the secondary sensing node increases, and at the same time a large area around the primary receiver needs to be protected, to avoid harmful secondary interference.

As the nodes of the secondary network both perform sensing and aim at utilizing the discovered spectrum bands, we evaluate the effect of the secondary user density on the per user achievable *cognitive capacity*. We show that as the network density increases, the performance of the cooperative sensing of a single band saturates, and the secondary network needs to optimize both the number of utilized bands and the number of bands sensed by a single user, to maximize the cognitive capacity.

The contribution of the paper is as follows:

- We provide an analytic framework to evaluate the efficiency of cooperative sensing in terms of the per user cognitive capacity under interference limitations, when the primary and secondary transmission characteristics are similar.
- We define and analyze sensing allocation mechanisms, spanning from limited to extended spectrum sensing and random to optimal sensing duty allocation.
- We study the fundamental limits of the cognitive capacity in highly dense cognitive networks and show how it is bounded by the constraints of local sensing performance.
- We demonstrate with numerical examples that sensing optimization can achieve significant gain in dense networks.

The paper is organized as follows. Related work is presented in Section 2. In Section 3 we describe the networking scenario, the considered optimization problem and give the local sensing model. Section 4 presents the analytic model of the capacity optimization of limited sensing under primary interference constraints. In Section 5 we introduce and evaluate the different sensing extension schemes. Section 6 concludes the paper.

2 Related Work

The optimization of secondary cognitive access, including sensing and channel access control is extensively studied in the literature. Here we consider the specific case of energy detection based cooperative spectrum sensing over multiple bands. The key issue in the design of the cooperative sensing solutions is the overhead introduced by the sensing itself and by the sensing control.

Considering the control of the cooperative sensing and the fusion of the sensing results, proposed solutions are based on a common control channel, e.g. [7], or distributed consensus protocols [8] [9], demonstrating that the control overhead is not significant, due to the localized nature of the decision processes. Similarly, distributed solutions are proposed to coordinate the access to the cognitive channels [10] [11].

The overhead of sensing depends on several parameters, as the frequency sensing needs to be performed with, the time needed to sense an individual band, the number of bands needed to be sensed by a single user, the granularity of the sensing results to be shared, and the efficiency of the decision combining. The frequency of the spectrum sensing affects the energy consumption overhead and, together with the time spent for sensing, gives the ratio of time that is surely lost for the secondary communication. Therefore, [12] [13] optimize the sensing interval based on the primary channel access statistics. Once the sensing interval is set, the aim is to optimize the time spent for the sensing process, which typically means to find the number of bands to be sensed and the per band sensing time, such that primary interference constraints are met and the secondary sensing performance or the secondary throughput is maximized [14] [15].

Sensing decision combining under cooperative sensing falls in one of two categories, hard decision combining or soft decision combining. Under hard decision combining the local decisions about the band availability are combined, with AND, OR or some k-out-of-N fusion rule, while under soft decision combining the quantized energy measurements are shared for the cooperative decision. Hard decision combining is often considered for its limited transmission overhead. As [14] and [16] show, the OR rule is typically more efficient than the AND one, while optimized decision combining outperforms these two, especially if even the local decision thresholds are carefully tuned. However, considering the effect of the average primary SNR and the number of nodes participating in the cooperative decision, all these schemes have similar behavior. Hard and soft decision combining are compared in [17], which concludes that under transmission errors the gain of soft decision combining is limited.

The allocation of sensing duties to a limited set of sensors is addressed in [18], considering a priori knowledge of the band occupancy probability and of the average SNR. The set of sensing nodes is optimized based on the experienced signal propagation environment in [19]; the proposed solution is further improved in [20], where learning is applied to select the sensing nodes taking even the sensing delay and the control traffic into account. [21] optimizes the number of users sensing a single band in a multichannel environment, recognizing that with increased fading more and more nodes need to sense the same band, which decreases the number of bands accessible for the secondary network.

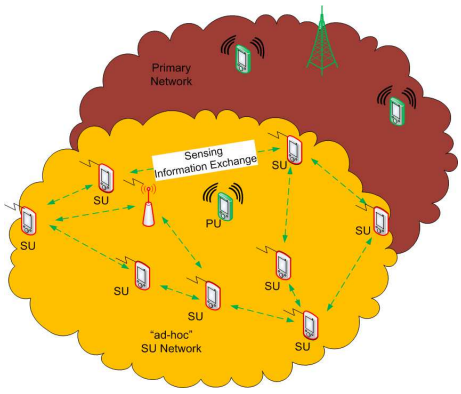


Figure 1: Secondary Cognitive Radio Network, coexisting with the Primary Network in the same area.

Most of the studies on sensing and interference management, however, consider primary users with large transmission power, resulting in equal average SNRs at the cognitive sensing nodes. Few works are available for the scenario considered in this paper, where the cooperating SUs see significantly different average SNR levels. In [22] the authors propose a framework for cooperative sensing where local sensing is distance-dependent. Taking into account the spatial distribution of the cognitive users, the authors model the accumulative interference to a PU, without, however, considering a capacity maximization problem. In [23] a specific case of this scenario is considered, where the primary transmission range is significantly lower than the secondary one. Sensing and channel access is considered jointly in [24], however only for the scenario where the local sensing parameters are not distance dependent and the subsets of channels sensed by the given secondary users are not optimized.

3 System Description

3.1 Sensing and interference management

We consider a primary and a cognitive secondary network in the same geographical area, as shown in Figure 1. The primary and the secondary transmitters have similar characteristics, thus similar transmission range. Secondary users (SUs) are randomly dispersed in the area, with density ρ. SUs perform spectrum sensing over a set of frequency bands with the help of their embedded sensing equipment. SUs exchange information to perform cooperative sensing, derive their relative locations [25] [26], control the individual sensing processes, and coordinate the access to the cognitive bands [7]– [11].

We consider four different cases of *sensing duty allocation*: Under *limited spectrum*

sensing each of the SUs senses the same set $\mathcal{M}$ of narrow frequency bands, $|\mathcal{M}| = M$. We refer to M as the local sensing budget. M is limited by the nodes' hardware constrains, that is, $M < M_{\max}$. The goal of the sensing duty allocation optimization is to find the optimal M value and cooperative sensing parameters for a given SU density.

Under *extended spectrum sensing* the SUs together aim at sensing a set of $\mathcal{W}$ bands, of size $|\mathcal{W}| = W$, defined as the *nominal sensing budget*. Each secondary user i may sense a different subset of bands, $\mathcal{M}_i \subset \mathcal{W}$. As nodes have similar sensing capabilities, we still consider local sensing budgets $|\mathcal{M}_i| = M < M_{\max}$, $\forall i$. If the users sense the same subset, $\mathcal{M}_i = \mathcal{M} = \mathcal{W}$, the scenario reduces to the limited spectrum sensing. We consider three different policies for sensing duty allocation under extended spectrum sensing:

1. *Random sensing*: each of the secondary users selects the subset $\mathcal{M}_i$ of frequency bands, by picking each band of $\mathcal{W}$ with the same probability. W, M and the cooperative sensing parameters are optimized for the given SU density.

2. *Coordinated sensing*: the secondary users coordinate the sensing duty allocation, such that each band of $\mathcal{W}$ is sensed by approximately the same number of users. Again, optimization is performed for the average user density.

3. *Optimal sensing*: the secondary network is aware of the instantaneous number of SUs in the area and performs *dynamic sensing budget adjustment* accordingly. Then, it performs coordinated sensing, considering the actual sensing budget given by the optimized W and M.

SUs operate in a time-slotted, slot-synchronized manner. They conduct spectrum measurements and share spectrum availability information at the beginning of a time-slot, and transmit in the second part of the time-slot, if free bands have been detected. We consider hard decision combining with local energy detection, where each sensor shares only its binary only noise or signal present decision, and OR decision rule, that is, a band is considered as occupied if at least one sensor decides for signal present.

Figure 2 illustrates the main principles of the considered sensing and interference management framework. An arbitrary primary transmitter is surrounded by a *prohibited area*, inside which simultaneous secondary transmissions within the same frequency band would cause interference. The radius of the prohibited area, R_I, is determined by the transmission characteristics of primary and secondary users, associated with the transmission ranges R_p and R_s respectively. Considering the worst case scenario, when the primary receiver of the particular transmitter lies in the border of the primary transmission range, R_I becomes $R_I = R_p + R_s$, as shown in the Figure. In the rest of the analysis we will assume that transmission ranges are fixed and so the radius R_I is fixed and known to the secondary users.

To detect a transmitting PU at a given location, a subset of the SUs inside the related prohibited area performs cooperative sensing. Since the reliability of the local sensing decreases with the distance to the transmitter, we define the *sensing area* as a disk centered at the considered PU location. The size of the sensing area, that is, the extent of the cooperative sensing, is controlled by the cooperation radius $R_c \leq R_{\mathcal{I}}$. As shown in Figure 2 spectrum sensing for the considered PU location is conducted by $N_{\mathcal{A}}$ SUs inside the area

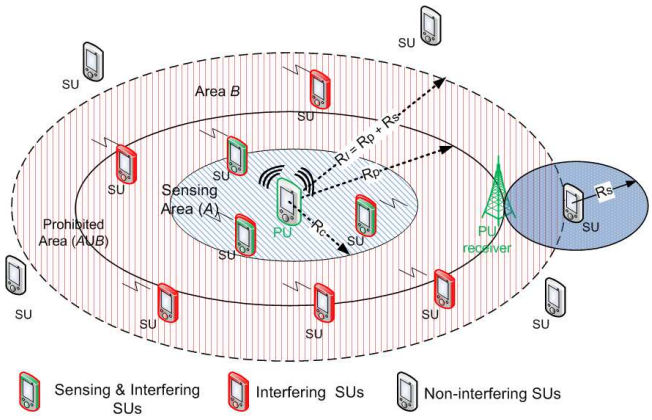


Figure 2: Interference modelling scheme: radius R_I defines the disk that corresponds to the prohibited area of a PU, which is determined by the communication ranges of the primary and secondary users.

$\mathcal{A}$. Area $\mathcal{B}$ is the area where existing SUs do not sense for the primary user but can cause harmful interference. The union $\mathcal{A} \cup \mathcal{B}$ constitutes the prohibited area for the considered primary user location, with $N_{\mathcal{A}} + N_{\mathcal{B}}$ SUs.

As for the cognitive transmission, we consider an idealized channel access scheme in the secondary network, where, after each sensing period, the *available* bands, i.e. the ones, for which cooperative sensing resulted in a correct detection of a free band or a missed detection of an occupied one, are assigned fairly to the SUs within $\mathcal{A} \cup \mathcal{B}$, with at most one band for an SU at a time.

3.2 Sensing optimization

Our aim is to maximize the per user cognitive network capacity, that is, the amount of spectrum resources available for each of the secondary users. The secondary users need to satisfy their bandwidth requirements through their own cooperative spectrum sensing. Their ability to extend sensing reliably to a large portion of radio spectrum enhances, as their population in a certain area increases, but at the same time, the available resources need to be shared among a higher number of SUs.

To capture this trade-off we define the per user *effective cognitive capacity*, C, as the ratio of the spectrum resources that are available for cognitive communication and the sum of resources requested by the secondary users, assuming that even a mis-detected band is useful resource for the cognitive communication. We limit $C \leq 1$, as an SU can transmit on one band only.

The secondary channel access is limited by the *probability of interference* at the primary user from the secondary network. To reflect the fact that channels detected free may remain

unutilized if all SUs are satisfied, we define the probability of interference by the joint probability that the primary transmission is not detected and the band is assigned to a secondary user in the interference area.

We aim at optimizing the sensing duty allocation such that $\mathbf{C}$, the expected per user *effective cognitive capacity* under given SU density ρ is maximized, while respecting $P_{\mathcal{I}}^{(\max)}$, the primary system interference constraint. Formally:

$$\begin{aligned} &\text{maximize } \mathbf{C} \\ &\text{subject to } P(\mathcal{I}) \leq P_{\mathcal{I}}^{(\max)} \\ &\qquad\qquad M \leq M_{\max}, \end{aligned} \tag{1}$$

In addition, under optimal extended sensing the $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$ values are known, and the optimization can be reformulated as:

$$\begin{aligned} \text{maximize} \quad & C(N_{\mathcal{A}}, N_{\mathcal{B}}) \\ \text{subject to} \quad & P(\mathcal{I}|N_{\mathcal{A}}, N_{\mathcal{B}}) \leq P_{\mathcal{I}}^{(\max)}, \\ & 0 < M \leq M_{\max}, \end{aligned} \tag{2}$$

and $\mathbf{C}$ is the average of the achieved $C(N_{\mathcal{A}}, N_{\mathcal{B}})$ over the distributions of $N_{\mathcal{A}}, N_{\mathcal{B}}$.

There are numerous input parameters of this optimization problem and numerous system variables to be optimized. Specifically, as input parameter, we consider $R_{\mathcal{I}}$, the radius of the prohibited area; $P_{\mathcal{I}}^{(\max)}$, the primary interference constraint; the sensing parameters as $M_{\max}$, the maximum number of sensed bands, and T_s, the total sensing time; and finally, ρ, the secondary network density. The optimized system variables are γ, the threshold value of the local sensing; R_c, the radius of the sensing area; W, the number of sensed bands; and M, the number of bands sensed by a single SU.

Note, that we do not consider the optimization of the sensing time T_s, the length of the cognitive time slot, the fusion rule, and the overhead of cooperation, to keep the problem tractable. As discussed in Section 2, related results can be found e.g., in [12]– [17].

3.3 Local sensing framework

During the sensing period an arbitrary secondary user measures the energy that is received within each of the frequency bands in its local sensing budget. Then it makes a binary decision, regarding the existence of an active primary transmission, by comparing the measured signal energy with a predefined energy decision threshold. We define the two complementary hypotheses as follows:

$$\begin{cases} \mathcal{H}_0 : x[k] = \upsilon[k] \ \text{ only noise} \\ \mathcal{H}_1 : x[k] = h \cdot s[k] + \upsilon[k] \ \text{ signal present,} \end{cases} \tag{3}$$

where x is the received primary signal, h is the channel coefficient of the link between the SU and the hypothetical primary transmitter and s is the transmitted signal, which is

assumed to be constant for the sake of simplicity. The local test variable $\mathbf{y}$ is formed by squaring and integrating K samples of the received signal during the per band sensing time, t_s:

$$\mathbf{y} = \frac{1}{K}\sum_{k=1}^{K} |x[k]|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_i, \tag{4}$$

where γ is the selected decision threshold. For Rayleigh flat-fading channel with lognormal shadowing, the channel coefficient obtains the following form:

$$h = \frac{1}{(r/r_0)^{\eta/2}} e^{j\phi} \alpha \, 10^{\zeta/20}. \tag{5}$$

In the above expression r is the distance between the hypothetic primary transmitter location and the sensing device, r_0 is a close-in reference, η is the path-loss exponent, α, ζ are a unit - variance Rayleigh and a zero-mean Gaussian random variable respectively representing small scale fading and shadowing and ϕ is a random phase shift uniformly distributed in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Based on (4) and (5) the probability density function of the test variable $\mathbf{y}$ under the two hypotheses is given in the following formulas:

$$p_{\mathbf{y}}(y;\mathcal{H}_0) = \frac{1}{(\sigma/K)^{2K}\Gamma(K)} y^{K-1} e^{y/(\sigma/K)^2},$$

$$p_{\mathbf{y}}(y;\mathcal{H}_1/\alpha,\zeta) = \frac{1}{(\sigma/K)^2}\left(\frac{y}{q^2}\right)^{\frac{(K-1)}{2}} e^{-\frac{\left(q^2+y\right)}{(\sigma/K)^2}} I_{K-1}\left(\sqrt{y}\frac{2q}{(\sigma/K)^2}\right),$$

where $q^2 \triangleq P_W K \frac{10^{\zeta/10}\alpha^2}{(r/r_0)^{\eta}}$. P_W is the transmitted signal power in the considered band, i.e. $P_W = |s|^2$, σ^2 is the noise power and $I_K(\cdot)$ is the K-th order Bessel modified function of the first kind.

Local missed detection and false alarm probabilities with respect to the energy threshold γ are given as in [24] – by averaging over the random variables α and ζ:

$$p_{md}(r,\gamma) = \int_0^{\gamma}\int_0^{\infty}\int_{-\infty}^{\infty} p_{\mathbf{y}}(y;\mathcal{H}_1/\alpha,\zeta)\, p_\zeta p_\alpha d\zeta d\alpha dy, \tag{6}$$

$$p_{fa}(r,\gamma) = \int_{\gamma}^{+\infty} p_{\mathbf{y}}(y;\mathcal{H}_0)\, dy. \tag{7}$$

Since the number of signal samples integrated at the energy detector can be consider large enough, we approximate the above density functions with Gaussian, obtaining the following simplified expressions:

$$\hat{p}_{md}(r,\gamma) = Pr(\text{no detection}|\mathcal{H}_1) = \\ = 1 - \mathbf{E}_{\alpha^2}[\mathbf{E}_{\zeta}[Q(\frac{\gamma-\sigma^2-\alpha^2 10^{\zeta/10} P_W (r/r_0)^{-\eta}}{\sqrt{2\sigma^4/K}})]], \tag{8}$$

$$\hat{p}_{fa}(r,\gamma) = Pr(\text{false detection}|\mathcal{H}_0) = Q(\frac{\gamma-\sigma^2}{\sqrt{2\sigma^4/K}}),\ \forall i. \tag{9}$$

As the received signal level depends significantly on r, we define the detection threshold γ as the linear combination of the received noise and the *expected* signal power at the SU, given distance r to the hypothetic transmitter: $\gamma : \mathbb{R}_+ \to (\sigma^2, \infty)$, that is:

$$\gamma = \gamma(r) = \gamma_0 \cdot \sigma^2 + \gamma_1 \cdot P_W r^{-\eta}, \tag{10}$$

Consequently, secondary users relatively close to the considered primary transmitter location apply higher decision thresholds and decrease their local false alarm probability.

To calculate the expected local false alarm and missed detection probabilities, consider the situation in Figure 2. Sensing collaboration is extended to a circular sensing area around the tested PU location, determined by the radius R_c. Since SUs are uniformly and independently distributed, the expected local missed detection probability of any SU within the sensing area is given by:

$$\mathbf{p}_{md}(R_c,\gamma) = \int_0^{R_c} \hat{p}_{md}(r,\gamma)\mathbf{p}(r|R_c)dr = \int_0^{R_c} \frac{2}{R_c^2} r \hat{p}_{md}(r,\gamma)dr, \tag{11}$$

where $\mathbf{p}(r|R_c)$ denotes the probability that a secondary user lies inside the infinitely thin ring at distance r from the PU (index i is omitted for the sake of simplicity):

$$\mathbf{p}(r|R_c) = Pr\{\text{user lies in the ring at distance } r\} = \frac{2rdr}{R_c^2}.$$

Similarly, secondary users cooperating within a sensing area with radius R_c where no primary transmitters are active generate false alarm with expected local false alarm probability:

$$\mathbf{p}_{fa}(R_c,\gamma) = \int_0^{R_c} \hat{p}_{fa}(r,\gamma)\mathbf{p}(r|R_c)dr = \int_0^{R_c} \frac{2}{R_c^2} r \hat{p}_{fa}(r,\gamma)dr. \tag{12}$$

4 Maximizing the Cognitive Capacity under Limited Sensing

First we evaluate the efficiency of limited sensing, that is, when all SUs in the sensing area of a PU location sense the same set of narrow frequency bands. For the analysis we assume that there is only one active PU, and derive how large part of the remaining free capacity can be used by the SUs in the interference area, such that the interference limit towards the PU is respected.

4.1 Analytic model for interference and cognitive capacity with primary user operating in a single band

Consider that the primary user in Figure 2 transmits in band $V \in \mathcal{M}$, while all the other bands of $\mathcal{M}$ are not utilized by the primary network. The PU encounters interference, $\mathcal{I}$, on band V, if both *i)* spectrum sensing on band V results in a missed detection, and *ii)* during the following time interval this band is assigned – based on the access scheme – to a secondary user that lies inside its prohibited area.

SUs in area $\mathcal{A}$ use the total sensing time T_s to sense sequentially $M \leq M_{\max}$ bands. The available sensing time for any band is thus $t_s = T_s/M$.

To model the random location of secondary users, we consider the SUs' population in areas $\mathcal{A}$ and $\mathcal{B}$ be independent Poisson variables with the same expected density, ρ.

Following the definition in Section 3.2, we express $P(\mathcal{I}|N_\mathcal{A}, N_\mathcal{B})$, the probability of interference to the primary user, conditioned on the number of the existing secondary users in areas $\mathcal{A}$ and $\mathcal{B}$, as:

$$P(\mathcal{I}|N_\mathcal{A}, N_\mathcal{B}) = Pr\{\text{miss. det.}|N_\mathcal{A}\} \cdot Pr\{\text{use}|N_\mathcal{A}, N_\mathcal{B}\}. \tag{13}$$

In the following we derive the expressions for both factors in (13). Based on the *OR* decision rule, we obtain the missed detection probability of the cooperative sensing:

$$\begin{aligned} &Pr\{\text{miss. det.}|N_\mathcal{A}\} \triangleq p_{MD}(N_\mathcal{A}) = \\ &= \textstyle\prod_{i=1}^{N_\mathcal{A}} Pr\{\text{miss. det. node } i\} = (\mathbf{p}_{md}(R_c, \gamma))^{N_\mathcal{A}}. \end{aligned} \tag{14}$$

Notice that (14) assumes uncorrelated local measurements. The assumption has been justified in [27]. Furthermore, the probability that the band V will be assigned for cognitive operation to a secondary user in the prohibited area is given by:

$$Pr\{\text{use}|N_\mathcal{A}, N_\mathcal{B}\} = \sum_{j=0}^{M-1} \min\{1, \frac{N_\mathcal{A} + N_\mathcal{B}}{j+1}\} \cdot P_{fr}(j), \tag{15}$$

where $P_{fr}(j)$ defines the probability that j out of the unused $M-1$ bands are available for cognitive operation after sensing. A band may not be available for cognitive use if spectrum sensing in this band resulted in a false alarm. Since the false alarm probability is independent for each sensed band,

$$\begin{aligned} P_{fr}(j) &\triangleq \Pr\{j \text{ bands detected free}\} \\ &= \tbinom{M-1}{j} (p_{FA}(N_\mathcal{A}))^{M-1-j} (1 - p_{FA}(N_\mathcal{A}))^j, \end{aligned} \tag{16}$$

where $p_{FA}(N_\mathcal{A})$ is the false alarm probability of cooperative sensing:

$$\begin{aligned} p_{FA}(N_\mathcal{A}) &\triangleq \Pr\{\text{false alarm in a single band}|N_\mathcal{A}\} \\ &= 1 - (1 - \mathbf{p}_{fa}(R_c, \gamma))^{N_\mathcal{A}}. \end{aligned} \tag{17}$$

Notice in (14) and (17) that $\mathbf{p}_{md}(R_c,\gamma)$ and $\mathbf{p}_{fa}(R_c,\gamma)$ – given in (11) and (12) respectively – also depend on the available sensing time for each band, which is inversely proportional to the number of sensed bands, M. Finally, the expected interference at the primary user from the secondary network is given by the following expression:

$$\begin{aligned} P(\mathcal{I}) &= \textstyle\sum_{N_\mathcal{A}=0}^{\infty}\sum_{N_\mathcal{B}=0}^{\infty} P(\mathcal{I}|N_\mathcal{A},N_\mathcal{B})\cdot\mathbf{p}_{N_\mathcal{A}}\mathbf{p}_{N_\mathcal{B}} = \\ &= \textstyle\sum_{N_\mathcal{A}=0}^{\infty}\sum_{N_\mathcal{B}=0}^{\infty}[P(\mathcal{I}|N_\mathcal{A},N_\mathcal{B})\cdot\frac{(|\mathcal{A}|\rho)^{N_\mathcal{A}}}{N_\mathcal{A}!}e^{-(|\mathcal{A}|\rho)}\frac{(|\mathcal{B}|\rho)^{N_\mathcal{A}}}{N_\mathcal{B}!}e^{-(|\mathcal{B}|\rho)}], \end{aligned} \tag{18}$$

where $P(\mathcal{I}|N_\mathcal{A},N_\mathcal{B})$ is given by (13) based on the derivations in (14), (15) and (16) and $\mathbf{p}_{N_\mathcal{A}}$ and $\mathbf{p}_{N_\mathcal{B}}$ define the probabilities of having $N_\mathcal{A}$ and $N_\mathcal{B}$ in areas $\mathcal{A}$ and $\mathcal{B}$ respectively. These probabilities depend on the sizes of the areas, denoted as $|\mathcal{A}|$ and $|\mathcal{B}|$, corresponding to radii R_c and $R_\mathcal{I}$ respectively. Note, that they do not depend on the components of $R_\mathcal{I}$, that is, R_P and R_S.

The *effective cognitive capacity*, as defined in Section 3.2, is a function of the number of SUs in areas $\mathcal{A}$ and $\mathcal{B}$, and depends on the number of bands detected free in $\mathcal{M}\setminus V$, and on the probability that the primary transmission on band V is not detected:

$$\begin{aligned} C(N_\mathcal{A},N_\mathcal{B}) \triangleq \textstyle\sum_{j=0}^{M-1}[&\min\{1,\tfrac{j}{N_\mathcal{A}+N_\mathcal{B}}\}(1-p_{MD}(N_\mathcal{A}))+ \\ &+\min\{1,\tfrac{j+1}{N_\mathcal{A}+N_\mathcal{B}}\}p_{MD}(N_\mathcal{A})]\cdot P_{fr}(j). \end{aligned} \tag{19}$$

The expected effective cognitive capacity depends on the SU distribution, that is:

$$\mathbf{C} = \sum_{N_\mathcal{A}=0}^{\infty}\sum_{N_\mathcal{B}=0}^{\infty} C(N_\mathcal{A},N_\mathcal{B})\cdot\mathbf{p}_{N_\mathcal{A}}\mathbf{p}_{N_\mathcal{B}}. \tag{20}$$

Although it is not shown in (19), (20), $\mathbf{C}$ is a function of the decision threshold γ and the cooperation radius R_c, so it is related to all of the system design parameters that we wish to optimize.

Given (18) and (19), the solution of the optimization problem (1) is not straightforward, since the objective and constraint functions are not convex, and therefore extensive search over the system variables γ, R_c and M, is required. The search process is impeded by the fact that the feasible set of (1) is generally not compact, since $P(\mathcal{I})$ is not even monotonic with respect to the system variables. It is though possible to reduce the searching borders of R_c, within which extensive search is, however, still necessary.

First we solve a modified version of (1), with the interference constraint function, $P(\mathcal{I})$, reduced to the missed detection probability,

$$\tilde{P}(\mathcal{I}) = \sum_{N_\mathcal{A}=0}^{\infty} p_{MD}(N_\mathcal{A}) = e^{-(1-\mathbf{p}_{md}(R_c,\gamma))|\mathcal{A}|\rho}. \tag{21}$$

Now $\tilde{P}(\mathcal{I})$ is increasing with γ and M, while it is decreasing with R_c. Since the cognitive capacity, $\mathbf{C}$ is decreasing with R_c as well, the optimal radius, $\tilde{R}_c^*$, for the modified problem is given by:

$$\tilde{R}_c^*(M,\gamma) = \tilde{P}(\mathcal{I})^{-1}(P_\mathcal{I}^{(\max)}),\ \forall M,\gamma, \tag{22}$$

where the inverse function is defined with respect to R_c. Since $R_c \leq R_{\mathcal{I}}$, γ is restricted by the set of maximal values $\gamma^{(\max)}$, which are the solutions of:

$$\gamma^{(\max)} = P(\mathcal{I})^{-1}(P_{\mathcal{I}}^{(\max)}, R_{\mathcal{I}}),$$

where the inverse function is taken with respect to γ.
Since $Pr\{\text{use}|N_{\mathcal{A}}, N_{\mathcal{B}}\} \leq 1$ and for low values of $N_{\mathcal{A}}, N_{\mathcal{B}}$ it holds $Pr\{\text{use}|N_{\mathcal{A}}, N_{\mathcal{B}}\} < 1$ we have:

$$P(\mathcal{I}) \mid_{R_c = \tilde{R}_c^*(M,\gamma)} < P_{\mathcal{I}}^{(\max)}.$$

A a result, the optimal radius, $R_c^*(M,\gamma)$ for any M, γ in (1) is lower than $\tilde{R}_c^*(M,\gamma)$.

Similarly, a lower bound for $R_c^*(M,\gamma)$ can be derived by considering a lower bound for $Pr\{\text{use}\}$, assuming zero false alarm probability:

$$\hat{Pr}\{\text{use}\} = \min\{1, \frac{N_A + N_B}{M}\}.$$

This leads to:

$$\hat{P}(\mathcal{I}) = \sum_{N_{\mathcal{A}}} \sum_{N_{\mathcal{B}}} p_{MD}(N_{\mathcal{A}}) \min\{1, \frac{N_A + N_B}{M}\} \mathbf{p}_{N_{\mathcal{A}}} \mathbf{p}_{N_{\mathcal{B}}} \tag{23}$$

and

$$\hat{R}_c^*(M,\gamma) = \hat{P}(\mathcal{I})^{-1}(P_{\mathcal{I}}^{(\max)}),\ \forall M, \gamma. \tag{24}$$

Finally, the optimal $R_c^*(M,\gamma)$ is the minimum solution of $P(\mathcal{I}) \mid_{M,\gamma} = P_{\mathcal{I}}^{(\max)}$ found by extensive search in a discretized version of the interval $(\hat{R}_c^*, \tilde{R}_c^*)$.

Since $Pr\{N_{\mathcal{A}} + N_{\mathcal{B}} \leq M\}$ decreases with ρ, $\hat{P}(\mathcal{I})$ approaches $\tilde{P}(\mathcal{I})$, as ρ increases. As a result, $\hat{R}_c^*$ approaches $\tilde{R}_c^*$, which significantly reduces the size of the search interval for the original problem in (1) for the considered dense networks.

4.2 Cognitive capacity and interference modelling with primary users operating in multiple bands

Let us now consider the scenario when the PU in Figure 2 utilizes a set of frequency bands. We allow this set to be random in each time slot, assuming a simple ON-OFF model for primary user activity, parameterized by the per band PU load, w:

$$\Pr\{\text{band } V_i \text{ is occupied}\} = w \in (0,1),\ \forall V_i \in \mathcal{M}. \tag{25}$$

Following the derivation of $P(\mathcal{I})$ in Section 4.1, only (16) has to be reformulated, since it gives $P_{fr}(j)$, the probability that j bands are detected free, which now additionally depends on the number of bands occupied by the PU.

According to the activity model of (25), $v(k)$, the probability that k bands are occupied in addition to band V is:

$$v(k) = \binom{M-1}{k} w^k (1-w)^{M-1-k},\ 0 \leq k \leq M-1, \tag{26}$$

and the probability that j bands are detected free is given as:

$$P_{fr}(j) = \sum_{k=0}^{M-1} P_{fr}(j|k)v(k). \tag{27}$$

Let us denote by n the number of frequency bands that have been detected free, while occupied by the PU, that is detection has failed, and by m the number of bands that have been successfully detected free by the SUs. The random variable $j = n+m$ gives the total number of the available bands. As the bands experience independent load, we get:

$$\begin{aligned} P_{fr}(j|k) &= \Pr\{j \text{ bands detected free} \mid k \text{ are used}\} = \\ &= \quad f_{n|k} * f_{m|k},\ 0 \le n \le M-1, n+m = j, \end{aligned}$$

where $*$ denotes convolution, and $f(n|k)$ and $f(m|k)$ represent the distributions of variables n and m respectively ($n \le k$, $m \le M-1-k$):

$$\begin{aligned} f_{n|k} &= \binom{k}{n} (p_{MD}(N_{\mathcal{A}}))^n (1-p_{MD}(N_{\mathcal{A}}))^{k-n} \text{ and} \\ f_{m|k} &= \binom{M-k-1}{m} (p_{FA}(N_{\mathcal{A}}))^{M-k-1-m} (1-p_{FA}(N_{\mathcal{A}}))^m, \end{aligned}$$

with $p_{MD}(N_{\mathcal{A}})$ and $p_{FA}(N_{\mathcal{A}})$ given in (14) and in (17).

The effective cognitive capacity **C** is then given by (19), with $P_{fr}(j)$ defined in (27).

4.3 Performance evaluation

Table 1: Parameter Setting for WLAN Case Study

Parameter	**Value**
Primary Signal Bandwidth	22MHz (1 WLAN Band)
Primary Signal Power	15dBm
Path Loss (η)	4.5
Shadowing (μ, σ^2)	0dBm, 10dB
AWGN Power (σ^2)	-96dBm
Interference Limit ($P_{\mathcal{I}}^{(\max)}$)	10^{-3}
Total Sensing Time (T_s)	2.5msec
Sensed Band Size (B_W)	200kHz
Max. Number of Sensed Bands ($M_{\max}$)	100
Signal Power in Sensed Band (P_W)	-5dBm
Prohibited Area Radius ($R_{\mathcal{I}}$)	300m

Let us now evaluate the efficiency of limited sensing, that is, the achievable effective cognitive capacity, as a function of the network density. The capacity is derived through the numerical solution of the optimization problem in (1).

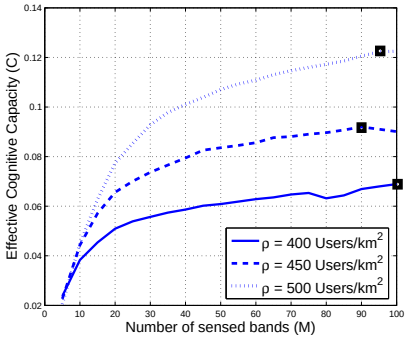


Figure 3: Effective cognitive capacity with respect to the total number of sensed frequency bands, M, for various cognitive network densities. Interference limit: $P_{\mathcal{I}}^{(\max)} = 10^{-3}$. Optimal values are marked.

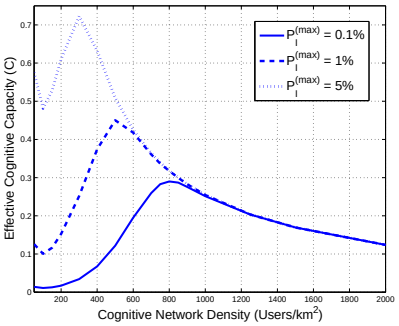


Figure 4: Effective cognitive capacity with respect to average secondary network density and for various interference limits.

4.3.1 Parameter setting

As an example, we consider local primary and secondary networks, with transmission parameters comparable to IEEE 802.11x WLANs. Table 1 lists the input parameters for our numerical analysis, unless otherwise stated.

The particular value for $R_{\mathcal{I}}$, which defines the size of the prohibited area, is selected based on practical transmission ranges in wireless local area networks. $M_{\max}$ is chosen equal to 100 bands. Each of those narrow bands has a bandwidth of $B_W = 200$kHz, as proposed in [28].

4.3.2 Numerical analysis

First we consider the case when the PU only occupies a single band and evaluate how M – the number of bands sensed – affects the effective cognitive capacity. That is, M is now input parameter of (1). Figure 3 shows, that M is indeed a parameter to be optimized, since depending on the SU density ρ there can be a local optimum $M < M_{\max}$. An expansion of M behind this value increases the probability of false alarms, which leads to decreased per user cognitive capacity. After a local minimum, the capacity is expected to increase again, since in the unrealistic case of $M \to \infty$, the probability of interference tends to zero even without sensing, and consequently, the cognitive capacity tends to 1.

The effective cognitive capacity, as a function of the secondary network density and for different interference limits is depicted in Figure 4. The capacity is determined via the numerical solution of (1). Under a given interference constraint, the cognitive network capacity reaches a highest value as network density increases, due to improved sensing

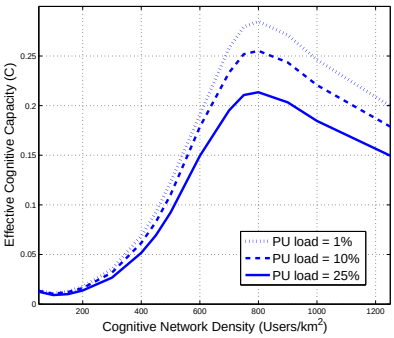


Figure 5: Effective cognitive capacity with respect to average secondary network density, for different primary user activity levels, w. Interference limit: $P_{\mathcal{I}}^{(\max)} = 10^{-3}$.

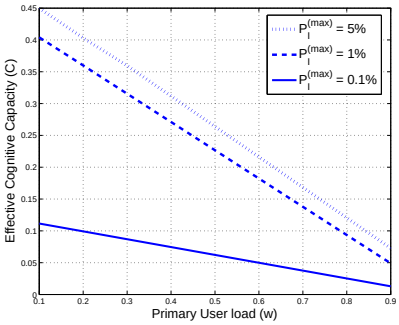


Figure 6: Effective cognitive capacity as a function of primary user activity load, w, for various interference constraint values. User density: ρ=500 Users/Km^2

efficiency. Above this "optimal" network density the capacity falls, as the now marginal improvement of sensing efficiency is not sufficient to accommodate the increasing need for bandwidth. For high user densities the cognitive capacity does not depend much on the interference limit, as a result of the high spectrum sensing performance. We observe a local minimum of cognitive capacity at low network densities. For densities below this value the sensing performance is weak, but as only few SUs are in the protected area, only few bands are accessed, and thus the probability of interference remains low.

Figure 5 shows the cognitive network capacity when the primary user operates in multiple bands. While the capacity decreases with the PU load, the the optimal network density, that is, where the capacity is maximized, does not seem to be affected. At the same time Figure 6 shows that the cognitive capacity decreases nearly linearly with the PU load, for the considered strict interference limits $P_{\mathcal{I}}^{(\max)}$.

Due to this independence, and for the sake of simplicity, we consider primary users operating in a single band in the rest of the paper. The extension for variable PU load is straightforward.

Our evaluation proves that the set of frequencies that can be sensed reliably by the SUs limits the achievable effective capacity in dense secondary networks. To overcome this limitation, the set of frequency bands available for cognitive transmission have to be extended, without the decrease of the per band sensing time of the SUs. This can be achieved by extended sensing, that is, by allowing the SUs to sense different subsets of the primary bands.

5 Extended Spectrum Sensing

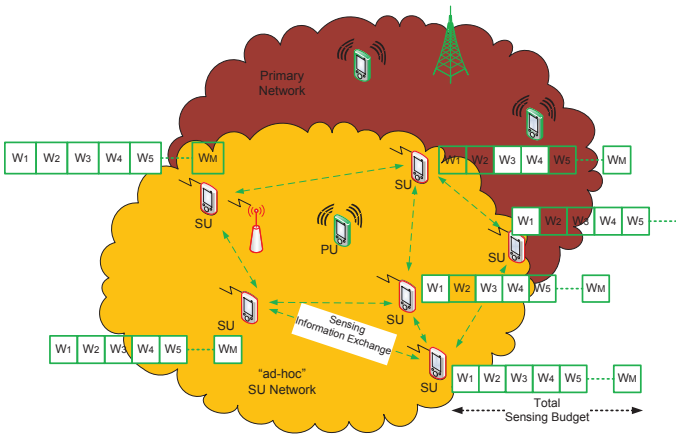


Figure 7: The distribution of sensing duties in a dense cognitive network.

Let us now evaluate the performance of extended sensing. As defined in Section 3.1, the SUs together sense a set of $\mathcal{W}$ primary bands, of size $|\mathcal{W}| = W$, giving the nominal sensing budget. This nominal sensing budget can be larger than the local sensing budget of the nodes, $\mathcal{M}_i \subset \mathcal{W}$ and $|\mathcal{M}_i| = M \leq M_{\max} < W$. This situation is depicted in Figure 7. If all the users sense the same subset $\mathcal{M}$, the scenario is reduced to the one of Section 4 and $\mathcal{M}_i = \mathcal{W}, \forall i$.

5.1 Analytic model for extended spectrum sensing with different sensing policies

5.1.1 Random sensing

In this case the secondary users randomly select the subset $\mathcal{M}_i$ of frequency bands to sense, independently from each other. An SU selects each of the spectrum bands with equal probability, that is, the probability that a band is sensed by an arbitrary SU i is:

$$p_{W,i} = \frac{M}{W} = p_W, \ \forall i = 1, ..., N_{\mathcal{A}}.$$

Consider the band $V \in \mathcal{W}$ used by the primary system. With probability p_W an arbitrary SU belongs to the $N_{\mathcal{A}}^V$ SUs that sense band V and additional $M-1$ out of the remaining $W-1$ spectrum bands, while with probability $(1-p_W)$ it belongs to the rest $N_{\mathcal{A}} - N_{\mathcal{A}}^V$ SUs which sense M out of the $W-1$ bands.

Let us introduce $p(j,k|N_{\mathcal{A}})$, the probability that when $N_{\mathcal{A}}$ SUs perform sensing, k of them sense band V and j bands in $\mathcal{W}\setminus V$ are available for transmission, that is, are correctly detected as free.

With $p(j,k|N_{\mathcal{A}})$ we can express $P(\mathcal{I}|N_{\mathcal{A}},N_{\mathcal{B}})$, the probability of interference on band V, similarly to (13) – (15), as:

$$\begin{aligned} &P(\mathcal{I}|N_{\mathcal{A}},N_{\mathcal{B}}) = \\ &= \textstyle\sum_{j=0}^{W-1}\sum_{k=0}^{N_{\mathcal{A}}} Pr\{\text{miss. det.}|k\}\cdot Pr\{\text{use}|N_{\mathcal{A}},N_{\mathcal{B}},j\}p(j,k|N_{\mathcal{A}}) = \\ &= \textstyle\sum_{j=0}^{W-1}\sum_{k=0}^{N_{\mathcal{A}}} p_{MD}(k)\min\{1,\frac{N_{\mathcal{A}}+N_{\mathcal{B}}}{j+1}\}p(j,k|N_{\mathcal{A}}). \end{aligned} \tag{28}$$

Similarly, the variable $p(j,k|N_{\mathcal{A}})$ helps us to express the cognitive capacity. We calculate $C(N_{\mathcal{A}},N_{\mathcal{B}})$ according to (19), with some changes. First, the missed detection probability $p_{MD}(N_{\mathcal{A}})$ depends now on the number of SUs sensing channel V, $p_{MD}(N_{\mathcal{A}}) = \sum_{k=0}^{N_{\mathcal{A}}} p_{MD}(k)p(k)$, where k is the number of SUs sensing a band, which follows a binomial distribution with parameters $N_{\mathcal{A}}$ and p_W. Similarly, we replace $P_{fr}(j)$ with $\sum_{k=0}^{N_{\mathcal{A}}} p(j,k|N_{\mathcal{A}})$.

The direct calculation of $p(j,k|N_{\mathcal{A}})$ suffers from combinatorial complexity due to the random band selection at the SUs. Therefore, we propose a recursive algorithm to calculate these probabilities, where we integrate the individual sensing results of each of the $N_{\mathcal{A}}$ SUs sequentially, exploiting that they are independent and stochastically identical.

We define the two-dimensional stochastic process, where state $\mathbb{S}_{j,k}^{(s)}$ defines the event that j spectrum bands are available and k users choose to sense band V, after the results of the first s SUs are integrated. $\mathrm{S}_{j,k}^{(s)}$ denotes the probability of state $\mathbb{S}_{j,k}^{(s)}$, with $\sum_{n=0}^{W} S_{n,k}^{(s)} = 1$. The vector $S_k^{(s)} \in \mathbb{R}^W$ denotes the stochastic vector of the available spectrum bands, given k, after incorporating the sensing processes of the first s SUs.

In each iteration, the process can move the from a state $\mathbb{S}_{j,k}^{(s)}$ to states $\mathbb{S}_{j-n,k}^{(s+1)}$ and $\mathbb{S}_{j-n,k+1}^{(s+1)}$, based on whether the $s+1$-th user senses band V and given the false alarm events that it may generate. The value of n, which indicates the additional number of bands that are "infected" by a false alarm event after incorporating the sensing from the $s+1$-th SU, is bounded by:

$$\max\{0, l-(W-1-j)\} \leq n \leq \min\{l,j\}, \tag{29}$$

with l denoting the number of the generated false alarm events by user $s+1$. Given that the false alarm probability is the same for any spectrum band, and considering that SUs select to sense any spectrum band with the same probability, we can express the conditional probability of n new bands getting infected by false alarm as:

$$\pi_{j\to j-n|l} = \binom{l}{n}\prod_{u=0}^{n-1}\frac{j-u}{W-1-u}\cdot\prod_{u=0}^{l-n-1}(1-\frac{j-n}{W-1-n-u}),$$

which indicates that n out of l false alarms infect new bands, while the rest infect already infected bands. Clearly, for values of n outside the bounds given in (29) the conditional transition probabilities are zero. The unconditioned transition probabilities are computed

by averaging based on the probability mass function of the number of the false alarm events l, $p(l|L) = \binom{L}{l}\mathbf{p}_{fa}^{l}(1-\mathbf{p}_{fa})^{L-l}$, where $L = M$ if SU $s+1$ does not sense band V and $L = M-1$ otherwise:

$$\pi_{j,k\to j-n,k} = \sum_{l=0}^{M} \pi_{j\to j-n|l} \cdot p(l|M) p_W,$$

$$\pi_{j,k-1\to j-n,k} = \sum_{l=0}^{M-1} \pi_{j\to j-n|l} \cdot p(l|M-1)(1-p_W).$$

The initial state vector of the recursion is: $S_0^{(0)} = (0,0,....,1)$. The state vector in step $s+1$ is calculated as:

$$S_{j,k}^{(s+1)} = \sum_{i=0}^{W-1} \pi_{i,k\to j,k} \cdot S_{i,k}^{(s)} + \sum_{i=0}^{W-1} \pi_{i,k-1\to j,k} \cdot S_{i,k-1}^{(s)}, \tag{30}$$

or in matrix form:

$$S_k^{(s+1)} = {S_k^{(s)}}^T \cdot \Pi + {S_{k-1}^{(s)}}^T \cdot \Pi^V, \tag{31}$$

where $\Pi, \Pi^V \in \mathbb{R}^{W\times W}$ are the *transition matrices*:

$$\Pi = [\pi_{i,k\to j,k}], \quad \Pi^V = [\pi_{i,k-1\to j,k}],$$

$\forall i, j = 0,1,...,W-1$ and $\forall k = 0,1,...,N_{\mathcal{A}}$.

With the above recursive process we calculate $S_{j,k}^{(N_{\mathcal{A}})}$, and set $p(j,k|N_{\mathcal{A}}) = S_{j,k}^{(N_{\mathcal{A}})}$.

Considering the recursive method described above, the computation complexity of calculating $p(j,k|N_{\mathcal{A}})$ is of the order of $O(N_{\mathcal{A}} \cdot M \cdot W)$. Moreover, due to the recursive nature of the method, we can calculate the state distribution for $N_{\mathcal{A}} = i+1$ SUs based on the already computed state distribution for i SUs with a complexity of $O(M \cdot W)$. As a result, the iterative algorithm makes it possible to calculate the capacity and interference values with polynomial complexity.

5.1.2 Coordinated sensing

Under coordinated sensing the $N_{\mathcal{A}}$ SUs in the sensing area cooperate to select the individual local sensing budgets $\mathcal{M}_i, i = 1,...,N_{\mathcal{A}}$, such that each spectrum band is sensed by approximately the same number of secondary users, $N_{\mathcal{A}d} \approx N_{\mathcal{A}}/(W/M) = N_{\mathcal{A}}M/W$. The approximation is accurate for $N_{\mathcal{A}} \gg W/M$, which is expected in dense networks.

To calculate the probability of interference, $P(\mathcal{I})$, we can follow (13) – (18), but considering cooperative sensing by $N_{\mathcal{A}d}$ SUs for each spectrum band. Consequently, the missed detection probability becomes:

$$p_{MD}(N_{\mathcal{A}d}) = (\mathbf{p}_{md}(R_c,\gamma))^{N_{\mathcal{A}d}},$$

while the band utilization probability is:

$$\Pr\{\text{use}|(N_{\mathcal{A}},N_{\mathcal{B}})\} = \sum_{j=0}^{W-1} \min\{1, \frac{N_{\mathcal{A}}+N_{\mathcal{B}}}{j+1}\} P_{fr}(j), \tag{32}$$

with

$$P_{fr}(j) = \binom{W-1}{j}(p_{FA}(N_{\mathcal{A}d}))^{W-1-j}(1-p_{FA}(N_{\mathcal{A}d}))^{j}. \tag{33}$$

The effective cognitive capacity is again calculated according to (19), using $p_{MD}(N_{\mathcal{A}d})$ and $P_{fr}(j)$ from the equations above.

5.1.3 Optimal sensing with dynamic sensing budget adjustment

Now the secondary network maintains information on the actual number of SUs in the area at each point in time, optimizes the system variables W, M, R_c and γ according to (2), and then allocates sensing duties as in the case of coordinated sensing. The effective cognitive capacity in this case is expected to outperform the ones of random and coordinated sensing at the cost of a higher control traffic overhead. $P(\mathcal{I}|\mathcal{N}_{\mathcal{A}},\mathcal{N}_{\mathcal{B}})$ and $C(N_{\mathcal{A}},N_{\mathcal{B}})$ can be calculated as for coordinated sensing.

5.2 Capacity limits in highly dense cognitive networks

Let us now investigate the asymptotic behavior of the effective cognitive capacity at very high SU densities. We detail the evaluation of the optimal sensing extension scheme, and summarize the results for the random and coordinated schemes.

In highly dense networks we can approximate the capacity, defined in (19), with respect to the total number of SUs in the prohibited area, $N = N_{\mathcal{A}} + N_{\mathcal{B}}$. Let us introduce $P_{fr}(j|N)$ and $P_{fr}(j|N_{\mathcal{A}})$ as the probability of j channels detected free given N SUs in the prohibited area, and given $N_{\mathcal{A}}$ SUs in the sensing area, respectively. Given N, $N_{\mathcal{A}}$ has binomial distribution, with $\mathbf{q}_N(N_{\mathcal{A}}) = \binom{N}{N_{\mathcal{A}}}[(\frac{R_c}{R_{\mathcal{I}}})^2]^{N_{\mathcal{A}}}[1-(\frac{R_c}{R_{\mathcal{I}}})^2]^{N-N_{\mathcal{A}}}$.

Considering that for large N $p_{MD}(N_{\mathcal{A}}) \approx 0$ and approximating $\min\{1, \frac{j}{N}\}$ with $\frac{j}{N}$, we get:

$$\begin{aligned}
\mathbf{C}(N) &\approx \textstyle\sum_{j=0}^{W-1} \min\{1, \frac{j}{N}\} P_{fr}(j|N) \\
&\approx \textstyle\sum_{j=0}^{W-1} j/N \cdot P_{fr}(j|N) \\
&= \textstyle\sum_{N_{\mathcal{A}}=0}^{N} \sum_{j=0}^{W-1} j/N \cdot P_{fr}(j|N_{\mathcal{A}}) \mathbf{q}_N(N_{\mathcal{A}}) \\
&= \hat{\mathbf{C}}(N).
\end{aligned} \tag{34}$$

In the optimal sensing extension case $N_{\mathcal{A}d} = N_{\mathcal{A}}M/W$ SUs sense each of the bands, and thus $\hat{\mathbf{C}}(N)$ can be further approximated as:

$$\begin{aligned}
&\hat{\mathbf{C}}_o(N) = \\
&= \tfrac{W-1}{N} \textstyle\sum_{N_{\mathcal{A}}=0}^{N} (1-\mathbf{p}_{fa})^{N_{\mathcal{A}d}} \mathbf{q}_N(N_{\mathcal{A}}) \\
&\approx \tfrac{W-1}{N} \exp\{[(1-\mathbf{p}_{fa})^{M/W} - 1] N (\tfrac{R_c}{R_{\mathcal{I}}})^2\} \\
&\approx \tfrac{W}{N} \exp\{[(1-\mathbf{p}_{fa})^{M/W} - 1] N (\tfrac{R_c}{R_{\mathcal{I}}})^2\}.
\end{aligned} \tag{35}$$

In dense networks, the probability that a band, that is detected free, is indeed allocated to an SU is close to one, that is, $Pr\{\text{use}|N\} \approx 1$. Based on this, we can approximate the probability of interference, defined in (13) – (15), as:

$$\begin{aligned} P_o(\mathcal{I}|N) &\approx \textstyle\sum_{N_\mathcal{A}=0}^{N} \mathbf{p}_{md}^{N_\mathcal{A}\cdot M/W} \mathbf{q}_N(N_\mathcal{A}) \approx \\ &\approx \exp\{[\mathbf{p}_{md}^{M/W} - 1]N(\tfrac{R_c}{R_\mathcal{I}})^2\} = \hat{P}_o(\mathcal{I}|N). \end{aligned} \tag{36}$$

For $\hat{P}_o(\mathcal{I}|N) = P_\mathcal{I}^{(\max)}$ we can then express W as:

$$W = M\ln(\mathbf{p}_{md})/\ln\left(1 + \frac{\ln(P_\mathcal{I}^{(\max)})}{N(R_c/R_\mathcal{I})^2}\right), \tag{37}$$

which already shows that in highly dense networks the size of the nominal sensing budget $\mathcal{W}$ has to be increased with N, to achieve optimal performance.

Replacing (37) in (35) we get the cognitive capacity with respect to the local sensing parameters, M, R_c and γ:

$$\begin{aligned} \hat{\mathbf{C}}_o(N) = &\frac{M\ln(\mathbf{p}_{md})}{N\ln(1+\frac{\ln(P_\mathcal{I}^{(\max)})}{N(\frac{R_c}{R_\mathcal{I}})^2})} \cdot \\ &\cdot \exp\left\{[(1-p_{fa})^{\ln(1+\ln(P_\mathcal{I}^{(\max)})/N(\frac{R_c}{R_\mathcal{I}})^2)/\ln(\mathbf{p}_{md})} - 1]N(\tfrac{R_c}{R_\mathcal{I}})^2\right\}. \end{aligned} \tag{38}$$

Proposition 1: In highly dense networks and with optimal sensing extension, the effective cognitive capacity asymptotically tends to the following limit:

$$\lim_{N\to\infty} \hat{\mathbf{C}}_o(N) = M\frac{R_c^2/R_\mathcal{I}^2 \ln \mathbf{p}_{md}}{\ln P_\mathcal{I}^{(\max)}} (P_\mathcal{I}^{(\max)})^{\frac{\ln(1-\mathbf{p}_{fa})}{\ln \mathbf{p}_{md}}}. \tag{39}$$

Proof: The limit can be derived analytically from (38).

□

Proposition 2: The limit of the effective cognitive capacity given in (39) is an increasing function of M.

Proof: The gradient of (39), with respect to M is always positive.

□

According to Proposition 2, the optimum M is $M_{\max}$, and W is given by (37). Therefore, the optimization problem to maximize the limit of the cognitive capacity under optimal sensing extension becomes two dimensional:

$$\mathbf{C}_o = \max_{R_c,\gamma}\{M_{\max}\frac{R_c^2/R_\mathcal{I}^2 \ln \mathbf{p}_{md}}{\ln P_\mathcal{I}^{(\max)}} P_\mathcal{I}^{(\max)\frac{\ln(1-\mathbf{p}_{fa})}{\ln \mathbf{p}_{md}}}\}. \tag{40}$$

Furthermore, as this is now an unconstrained optimization problem, maximization can be performed sequentially for the variables of the system, thus reducing the complexity significantly:

$$\mathbf{C}_o = M_{\max}/[R_{\mathcal{I}}^2 \ln P_{\mathcal{I}}^{(\max)}] \cdot \max_{R_c} \left\{ \max_{\gamma} \{ R_c^2 \ln \mathbf{p}_{md} P_{\mathcal{I}}^{(\max)} \frac{\ln(1-\mathbf{p}_{fa})}{\ln \mathbf{p}_{md}} \} \right\}. \tag{41}$$

Corollary: The asymptotic cognitive capacity limit of optimal extended spectrum sensing depends on the input parameters such as the interference probability constraint and the size of the prohibited area, and on the performance of the local sensing, which in turn depends on the primary transmission power and the sensing time, but even on the optimal sensing area.

The process of deriving the approximate capacity and interference, and finally the capacity limits of random and coordinated sensing extension is similar. Specifically, for the coordinated scheme:

$$\hat{\mathbf{C}}_c(N) \approx \frac{W-1}{\alpha_c(M)\pi\rho R_{\mathcal{I}}^2} e^{-\pi\rho R_{\mathcal{I}}^2} [e^{\alpha_c(M)\pi\rho R_{\mathcal{I}}^2} - 1], \tag{42}$$

where $\alpha_c(M) = e^{[(1-\mathbf{p}_{fa})^{M/|\mathcal{W}|}-1](\frac{R_c}{R_{\mathcal{I}}})^2}$, and

$$\hat{P}_c(\mathcal{I}|N) = \exp\{-\pi\rho R_{\mathcal{I}}^2 (1 - e^{\beta_c(M)})\}, \tag{43}$$

where $\beta_c(M) = [\mathbf{p}_{md}^{M/W} - 1] \cdot (\frac{R_c}{R_{\mathcal{I}}})^2$.

For the random sensing scheme:

$$\hat{\mathbf{C}}_r(N) \approx \frac{W-1}{\alpha_r(M)\pi\rho R_{\mathcal{I}}^2} e^{-\pi\rho R_{\mathcal{I}}^2} [e^{\alpha_d(M)\pi\rho R_{\mathcal{I}}^2} - 1], \tag{44}$$

where $\alpha_r(M) = \exp\{[e^{-\mathbf{p}_{fa}M/W} - 1](\frac{R_c}{R_{\mathcal{I}}})^2\}$, and

$$\hat{P}_r(\mathcal{I}|N) = \exp\{-\pi\rho R_{\mathcal{I}}^2 (1 - e^{\beta_r})\}, \tag{45}$$

where $\beta_r = [\exp\{-(1-\mathbf{p}_{md})M/W\} - 1](R_c/R_{\mathcal{I}})^2$.

5.3 Performance evaluation

In this section we evaluate the performance of extended sensing, using the parameters from Table 1. The presented results are based on the numerical evaluation of (1) for random and cooperative sensing extension, (2) for optimal sensing extension and (41), (42) and (44) for the capacity limits of the different schemes.

We consider first the relationship between ρ, the average density of the cognitive network and the size of the nominal sensing budget W that is required to maximize the effective

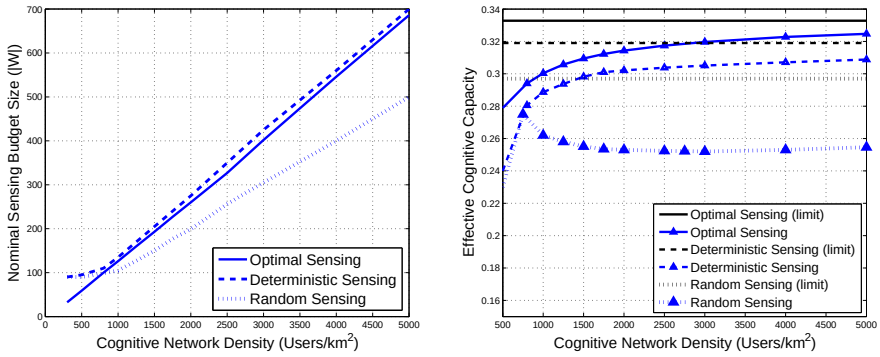
cognitive capacity. This is an important design factor, as it determines the spectrum that should be "reserved" for the particular cognitive network to maximize spectrum efficiency.

Figure 8(a) depicts the optimal value of W as a function of the density of the cognitive network, for the random and coordinated sensing extension schemes and the average of the optimal values for the optimal sensing scheme. For the optimal scheme W increases linearly with the number of SUs. When we consider networks with high density the nominal budget required by the coordinated and random schemes has a similar linear behavior. For low or average densities, however, an increased frequency budget is required, to compensate for the reduced quality of cooperative sensing. The optimal sensing budget in the case of random sensing is lower compared to the required ones under coordinated and optimal sensing. This is explained by the need to keep the number of SUs sensing a band reasonably high, even for the bands that happen to be selected by relatively few SUs.

Figure 8(b) shows the effective cognitive capacity with respect to the average density of the cognitive network, together with the asymptotic limits at $\rho \to \infty$. We recall from Figure 4 that under the considered $P_{\mathcal{I}}^{(max)} = 10^{-3}$, limited sensing with $W = M$ achieved maximum performance of $\mathbf{C} \approx 0.28$ at a density of ca. 800 users/km^2. Considering sensing extension with random sensing we observe the same peak. The capacity then falls, since, as reflected in Figure 8(b), in this density region the nominal sensing budget can not be extended significantly. Finally, the capacity increases with the network density, approaching the limit very slowly. Under coordinated and optimal sensing, the achievable capacity increases monotonically, approaching the limit with decreasing rate.

Clearly, the optimal sensing scheme outperforms the other discussed policies. The random sensing is the worst of the three, where the lack of coordination of the individual sensing processes at the SUs leads to insufficient sensing at some of the bands. . The gain of the optimal sensing over the coordinated one is however around 5%, which shows that coordinated sensing with balanced allocation of sensing resources is an efficient way of ensuring reliable sensing on all bands. As a result, the small additional gain of optimal sensing may not justify the additional control and computational complexity.

Finally, we evaluate how the capacity limit is affected by the input parameters, that is, the interference probability constraint, $P_{\mathcal{I}}^{(\max)}$, and prohibited area radius, $R_{\mathcal{I}}$. We consider the optimal sensing extension scheme, but the behavior of the other schemes is similar. As shown in Figure 9, the capacity limit becomes larger than 1 under loose interference constraint and small prohibited area, which shows that our approximation in (34), $\min\{1, \frac{j}{N}\} \approx \frac{j}{N}$ is not tight in this region. The capacity limit is very sensitive to the size of the prohibited area; as seen from (41) it decreases with $1/R_{\mathcal{I}}^2$, a result of the quadratic increase of interfering SUs that have to share the spectrum resources. The capacity limit with respect to $P_{\mathcal{I}}^{(\max)}$, however, increases slowly, especially for large prohibited areas. Increasing the probability constraint with two orders of magnitude leads to less than doubled cognitive capacity limit.



(a) (b)

Figure 8: (a): Nominal Sensing Budget for highly dense cognitive networks, for the different sensing coordination schemes, and (b): Effective cognitive capacity for highly dense cognitive networks, for the different sensing coordination schemes.

6 Conclusion

In this paper we investigated the efficiency of spectrum sensing in dense cognitive networks, when the transmission characteristics of the primary and secondary systems are similar, and consequently the local spectrum sensing performance is highly distance dependent. We introduced the per user effective cognitive capacity as the performance metric, to reflect that the SUs performing the sensing also aim at utilizing the spectrum. We showed that the per user cognitive capacity decreases and approaches zero in dense networks if limited sensing is used, that is, all SUs sense the same set of bands.

We defined and evaluated different solutions to extend the set of sensed bands by letting each SU to sense only a subset of them. We have shown that extended sensing with coordination among the users leads to per user cognitive capacity that increases together with the SU density, converging to a limit which depends on the primary and secondary transmission characteristics, on the primary interference constraints and on the parameters of the cooperative sensing. Under random sensing allocation the capacity limit first falls as in the case of limited sensing, but then stabilizes and increases slightly with the SU density. Based on the numerical results we concluded that the performance gap of random and coordinated sensing is significant, while optimizing sensing based on the actual number of secondary nodes leads to little additional gain.

References

[1] S. Geirhofer, Lang Tong, and B.M. Sadler. Dynamic spectrum access in the time domain: Modeling and exploiting white space. *IEEE Communications Magazine*,

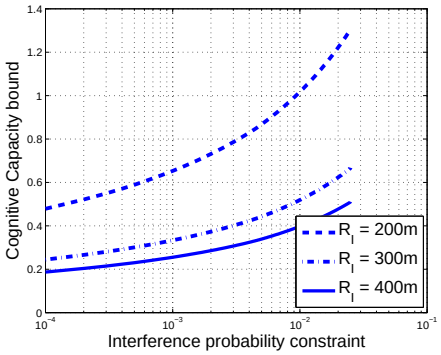


Figure 9: Effective cognitive capacity limit with respect to the interference constraint imposed by the primary system and for different values of interference range.

45(5):66–72, 2007.

[2] M. Song, C. Xin, Y. Zhao, and X. Cheng. Dynamic spectrum access: from cognitive radio to network radio. *IEEE Wireless Communications*, 19(1):23–29, 2012.

[3] A. Sahai, N. Hoven, S. M. Mishra, and R. Tandra. Fundamental tradeoffs in robust spectrum sensing for opportunistic frequency reuse. Technical report, Berkeley, 2006.

[4] E. Visotsky, S. Kuffner, and R. Peterson. On Collaborative Detection of TV Transmissions in Support of Dynamic Spectrum Sharing. In *IEEE Dynamic Spectrum Access Networks*, 2005.

[5] A. Sahai, R. Tandra, and N. Hoven. Opportunistic spectrum use for sensor networks: the need for local cooperation. *Berkeley Wireless Research Center*, 2006.

[6] D. Duan, L. Yang, and J.C. Principe. Cooperative diversity of spectrum sensing for cognitive radio systems. *IEEE Transactions on Signal Processing*, 58(6):3218–3227, 2010.

[7] Ian F. Akyildiz, Won-Yeol Lee, and Kaushik R. Chowdhury. CRAHNs: Cognitive radio ad hoc networks. *Ad Hoc Networks*, 7(5):810 – 836, 2009.

[8] Z. Li, F.R. Yu, and M. Huang. A distributed consensus-based cooperative spectrum-sensing scheme in cognitive radios. *IEEE Transactions on Vehicular Technology*, 59(1):383–393, 2010.

[9] P. Di Lorenzo, S. Barbarossa, and A.H. Sayed. Decentralized resource assignment in cognitive networks based on swarming mechanisms over random graphs. *IEEE Transactions on Signal Processing*, 60(7):3755–3769, 2012.

[10] C. Cormio and K. R. Chowdhury. A survey on {MAC} protocols for cognitive radio networks. *Ad Hoc Networks*, 7(7):1315 – 1329, 2009.

[11] J. Xiang, Y. Zhang, and T. Skeie. Medium access control protocols in cognitive radio networks. *Wireless Communications and Mobile Computing*, 10(1):31–49, 2010.

[12] Y. Pei, A. T. Hoang, and Y.-C. Liang. Sensing-Throughput Tradeoff in Cognitive Radio Networks: How Frequently Should Spectrum Sensing be Carried Out? In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications,(PIMRC'07)*, 2007.

[13] X. Xing, T. Jing, H. Li, Y. Huo, X. Cheng, and T. Znati. Optimal Spectrum Sensing Interval in Cognitive Radio Networks. *IEEE Transactions on Parallel and Distributed Systems*, 99(PrePrints), 2013.

[14] E.C.Y. Peh, Y.-C. Liang, Y. Liang Guan, and Y. Zeng. Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view. *IEEE Transactions on Vehicular Technology*, 58(9):5294–5299, 2009.

[15] R. Fan and H. Jiang. Optimal multi-channel cooperative sensing in cognitive radio networks. *IEEE Transactions on Wireless Communications*, 9(3), 2010.

[16] W. Han, J. Li, Z. Tian, and Y. Zhang. Efficient cooperative spectrum sensing with minimum overhead in cognitive radio. *IEEE Transactions on Wireless Communications*, 9(10):3006–3011, 2010.

[17] S. Chaudhari, J. Lunden, V. Koivunen, and H.V. Poor. Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions? *IEEE Transactions on Signal Process.*, 60(1):18–28, 2012.

[18] C. Song and Q. Zhang. Cooperative spectrum sensing with multi-channel coordination in cognitive radio networks. In *IEEE International Conference on Communications*, 2010.

[19] A.S. Cacciapuoti, I.F. Akyildiz, and L. Paura. Correlation-aware user selection for cooperative spectrum sensing in cognitive radio ad hoc networks. *IEEE Journal on Selelected Topics of Signal Processing*, 30(2):297–306, 2012.

[20] B. F. Lo and I. F. Akyildiz. Reinforcement learning for cooperative sensing gain in cognitive radio ad hoc networks. *Wireless Networks*, 19(6):1237–1250, 2013.

[21] K. Koufos, K. Ruttik, and R. Jantti. Distributed sensing in multiband cognitive networks. *IEEE Transactions on Wireless Communications*, 10(5):1667–1677, 2011.

[22] M.l Timmers, S. Pollin, A. Dejonghe, A. Bahai, L. Van der Perre, and F. Catthoor. Accumulative interference modeling for distributed cognitive radio networks. *Journal of Communications*, 4(3), 2009.

[23] Jihoon Park, P. Pawelczak, P. Gronsund, and D. Cabric. Analysis framework for opportunistic spectrum ofdma and its application to the ieee 802.22 standard. *IEEE Transactions on Vehicular Technology*, 61(5):2271–2293, 2012.

[24] V. Fodor, I. Glaropoulos, and L. Pescosolido. Detecting low-power primary signals via distributed sensing to support opportunistic spectrum access. In *IEEE International Conference on Communications*, 2009.

[25] C. Taylor, A. Rahimi, J. Bachrach, H. Shrobe, and A. Grue. Simultaneous localization, calibration, and tracking in an ad hoc sensor network. In *The Fifth International Conference on Information Processing in Sensor Networks, (IPSN'06)*, 2006.

[26] J. Gribben, A. Boukerche, and R. W. Nelem Pazzi. Scheduling for scalable energy-efficient localization in mobile ad hoc networks. In *IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON)*, 2010.

[27] I. Glaropoulos and V. Fodor. On the efficiency of spectrum sensing in adhoc cognitive radio networks. In *ACM Mobicom Workshop on Cognitive Wireless Networking (CoRoNet)*, 2009.

[28] B. Mercier et al. Sensor networks for cognitive radio: Theory and system design. In *ICT Mobile Summit*, 2008.

Paper B

Energy efficient COGnitive MAC for sensor networks under WLAN coexistence

Ioannis Glaropoulos, Marcello Laganá, Viktoria Fodor, and Chiara Petrioli

To appear in IEEE Transactions on Wireless Communications, 2015.

Energy Efficient COGnitive MAC for Sensor Networks under WLAN Co-existence

Ioannis Glaropoulos, Marcello Laganà, Viktoria Fodor, and Chiara Petrioli*

Abstract

Energy efficiency has been the driving force behind the design of communication protocols for battery-constrained wireless sensor networks (WSNs). The energy efficiency and the performance of the proposed protocol stacks, however, degrade dramatically in case the low-powered WSNs are subject to interference from high-power wireless systems such as WLANs. In this paper we propose COG-MAC, a novel cognitive medium access control scheme (MAC) for IEEE 802.15.4-compliant WSNs that minimizes the energy cost for multihop communications, by deriving energy-optimal packet lengths and single-hop transmission distances based on the experienced interference from IEEE 802.11 WLANs. We evaluate COG-MAC by deriving a detailed analytic model for its performance and by comparing it with previous access control schemes. Numerical and simulation results show that a significant decrease in packet transmission energy cost, up to 66%, can be achieved in a wide range of scenarios, particularly under severe WLAN interference. COG-MAC is, also, lightweight and shows high robustness against WLAN model estimation errors and is, therefore, an effective, implementable solution to reduce the WSN performance impairment when coexisting with WLANs.

Index terms— WSN, energy efficiency, cognitive networks, coexistence, IEEE 802.11, IEEE 802.15.4.

1 Introduction

The increasing number of different wireless technologies sharing the open spectrum bands, such as the 2.4GHz ISM band, demands for a rethinking of the protocols regulating the spectrum access. As the medium access control (MAC) schemes are carefully designed for one given technology, they are not anymore able to achieve

*I. Glaropoulos, M. Laganà, and V. Fodor are with the KTH, Royal Institute of Technology, Sweden. C. Petrioli is with the University of Roma "La Sapienza", Italy.

the objective of efficient and "fair" sharing of the wireless resources when operating under interference from heterogeneous technologies.

In this paper we consider the specific case of the coexistence of IEEE 802.11 wireless local area networks (WLANs) and IEEE 802.15.4-compliant wireless sensor networks (WSNs). Both technologies apply carrier sensing-based medium access control with collision avoidance. In addition, WSNs try to locate the narrow frequency band with less harmful interference for their operations. Unfortunately, all these techniques do not avoid high interference and frequent packet losses in the WSN, which are mainly caused by the significantly different transmission bandwidths and powers of the two technologies competing for the same resource.

As shown in [1], the WLAN terminals operate in a relatively broad channel and at a higher transmission power than WSNs. Therefore, they are *blind* to the narrow-band, low-powered WSN transmissions, and do not defer channel access, due to the overlapping WSN packet transmission.

In all this, the WLAN transmissions remain basically unaffected by the low WSN interference, while WSN packets are lost. Fortunately, measurement results show that the WLAN traffic is rather bursty with long *white spaces*, when the channel is idle because all WLAN users are inactive [2]. Therefore, in order to maximize its performance, the WSN should be able to transmit in these long interference-free times, thus, being cognitive of the radio environment as imposed by the WLAN activity.

In this paper we propose and evaluate a new COGnitive MAC (COG-MAC) protocol for wireless sensor networks, which extends the carrier sense-based MAC and aims at minimizing the energy loss due to unsuccessful transmissions over the interfered channel. Our paper provides the following contributions. 1) We give a characterization of the WLAN channel usage patterns as seen by the sensor nodes, taking into account the nodes' limited channel estimation capabilities, and propose techniques for distributed WLAN usage pattern estimation. 2) Based on these resulting WLAN channel usage characterization we design COG-MAC, that optimizes the packet length and the transmission distance, and performs WLAN activity-aware channel access to ensure that WSN nodes transmit in the long WLAN white space periods. 3) We provide an accurate analytical model that describes the probability of COG-MAC packet transmission success. We use the model to optimize the WSN packet size and the single-hop WSN transmission distance to minimize the *normalized energy cost* metric, which we define as the energy required to successfully transmit and receive a unit of information over a unit of distance. 4) We show that all the basic components of COG-MAC are essential for achieving the objective of energy efficient communication, and COG-MAC, compared to previous access schemes, reduces the normalized energy cost up to 66%, and can significantly decrease the end-to-end energy cost in a multihop WSN without increased delay.

The rest of the paper is organized as follows. Related work is presented in Section 2. Section 3 describes the networking scenario and the interference and sensing models and Section 4 gives the WLAN channel activity model. In Section 5

we describe the proposed protocol stack, followed by its mathematical analysis in Section 6. In Section 7 we present a numerical evaluation of COG-MAC along with a comparison with traditional WSN MAC schemes, while a simulation study is presented in Section 8. We conclude the paper in Section 9.

2 Related Work

Energy efficient communications have been extensively studied for stand alone WSNs [3][4]. The key idea for energy efficiency in sensor networks is to minimize idle listening, by letting the sensors turn off their radios whenever idle, controlled by duty-cycling [5][6][7][8][9], or by wake-up radios [10].

It is recognized, however, that cross-network interference can have significant effect on the network performance, as it is shown for coexisting WSNs in [11] and for WLAN and Bluetooth interference in [12][13]. WSN multi-channel operation aims at avoiding this cross-network interference by tuning to the best available band for communication [14][15][16]. These solutions are efficient as long as there exist channels with no or low interference, but lose effectiveness when all considered channels suffer from interference with similar statistical behavior.

Therefore, as wireless channels are getting densely populated, it is important to design protocols that can work efficiently even in the presence of cross-network interference. Many of the proposed solutions build on the known characteristics of the interfering networks. [17] employs narrow-band sensing, with additional HW cost, to identify and utilize the channels, where the wide-band device can effectively coexist with narrow-band transmissions, while in [18], the sensors force the WLAN to back off by sending frequent (one per DIFS), high power jamming signals during their packet transmission, which needs complex PHY layer and leads to increased energy consumption in the WSN. Instead, the effect of interference is minimized without changing the WLAN behaviour in [19] and [20] introducing WSN packet header and payload redundancy.

Recent works investigate how to avoid WLAN interference by employing channel availability predictions. The case of a non-saturated single WLAN AP is studied in [21], modeling the packet arrivals at the users as a Bernoulli process. In [22] a Poisson arrival process is considered, and WLAN output buffers are modeled as M/G/1 queues, resulting in sub-geometric idle period distribution. While these models capture the effect of the WLAN MAC, their generality is limited, since they are based on simple, rather unrealistic traffic models.

To capture the effect of realistic network load, [2][23][24][25] use traffic traces to find the distribution of WLAN idle periods. These results show that idle periods can be short contention periods, in the range of hundreds of microseconds, or heavy-tailed white spaces, where WLAN users are inactive. As it is demonstrated in [25], the average white space length depends on the WLAN load and the traffic characteristics, and is in the milliseconds range. In [26] similar results are derived

based on the self-similar nature of WLAN traffic.

Considering that, due to the low bitrate, the transmission times in the WSN are comparable to the average WLAN idle period length, it is important to capture and utilize the heavy-tail characteristics of the WLAN channel usage. Therefore, in our work we apply the model of [2][27] where a mixture distribution is proposed to model the idle periods, capturing the two basic sources of inactivity, the long heavy-tailed white space periods, when the WLAN users are inactive, and the short contention windows. Given this WLAN channel usage model we claim that the WSN on one hand needs to avoid channel access in the contention windows and on the other hand it needs to optimize transmissions in the long white space periods [28], which are the key functions in the proposed COG-MAC.

3 Networking Scenario, Interference and Sensing Models

We consider a WLAN Access Point (AP) zone that covers an area where a WSN is deployed. The WSN nodes are battery-powered and operate on a single 5MHz channel inside the 2.4GHz ISM band. They transmit information over multiple hops, and are able to estimate the distance to their neighboring nodes [29].

WLAN users are distributed inside the AP zone and operate on a 802.11 22MHz channel, covering the WSN channel. The WLAN transmission power is in the order of 15-20dBm. The WLAN terminals are blind to the WSN nodes [26], that is, the WLAN carrier-sense mechanism does not detect the low-power WSN signals, which results in collisions, and hence packet losses in the WSN. On the opposite side, the WSN nodes transmit with a signal power that is in the order of 0-3dBm [30] and, thus, their impact on WLAN operation is negligible [23]. Therefore, to ensure efficient WSN communication, sensor nodes need to consider the WLAN activity when transmitting. In the remainder of this section we clarify our assumptions on the interference and sensing models used throughout the paper. The signal propagation is assumed to be adequately described by a simple path-loss model. In order to correctly receive a packet, a WSN node needs to receive it with Signal to Interference plus Noise Ratio (SINR) above a given threshold, denoted as ζ_{SINR}, where the interference is assumed to be caused by a single active WLAN transmitter. Then, the path loss-based propagation model results in circular interference zones around receiving sensors, with radius R_I [31]:

$$R_I(r, \zeta_{\text{SINR}}, \eta, P_{\text{WSN}}, P_{\text{WLAN}}) = \sqrt[\eta]{\frac{\zeta_{\text{SINR}} P_{\text{WLAN}} P_{L_0}}{P_{\text{WSN}} P_{L_0} r^{-\eta} - \zeta_{\text{SINR}} \sigma_N^2}}, \tag{1}$$

where r is the distance between the transmitting and receiving sensor, η is the channel path-loss exponent, P_{WSN} is the WSN transmission power, P_{WLAN} is the fraction of WLAN transmission power inside the narrow WSN band, and P_{L_0} and

σ_N^2 denote the attenuation at 1m reference distance and the noise power, respectively. Whenever an overlap occurs between a WSN packet transmission and a WLAN transmission within the receiving sensor node interference zone of radius R_I, we assume that the WSN packet is lost.

The WSN nodes perform channel sensing based on energy detection through their build-in Receiver Signal Strength Indicator (RSSI) [30]. In the proposed system two kinds of sensing are performed. Repeated sensing over long periods of time for WLAN activity model estimation, and short-time sensing for channel access control. The performance of both kinds of sensing is bounded by the *maximum sensitivity level* ψ_0 of the sensor, stating the minimum signal level that can be detected [30]. Short sensing time leads to probabilistic energy detection, characterised by the probability of *missed detection*, p_{MD}, when a signal is not detected, and the probability of *false alarm*, p_{FA}, when the sensing results in "signal detected" decision, even when the channel is idle [32]. The false alarm probability p_{FA} is a function of the sensing time t_s and of the energy decision threshold γ, $p_{\mathrm{FA}}(t_s;\gamma)=\mathcal{Q}((\gamma-\sigma_N^2)/(\sigma_N^2\sqrt{2/(f_s t_s)}))$, where f_s denotes the sampling frequency. In this paper we consider a target p_{FA} which gives γ as [32]:

$$\gamma(p_{\mathrm{FA}}) = \max\left\{\psi_0, \sigma_N^2\left[1+\sqrt{2/(f_s t_s)}\mathcal{Q}^{-1}(p_{\mathrm{FA}})\right]\right\}. \tag{2}$$

The missed detection probability, p_{MD}, depends on the received signal power, $P_{Rx}(d)$, given as a function of the distance, d, to the transmitter, $P_{Rx}(d) = P_{\mathrm{WLAN}} P_{L_0} d^{-\eta}$. It also depends on the decision threshold, γ, which in turn is determined, using (2), by the target p_{FA}:

$$p_{\mathrm{MD}}(t_s, d; \gamma(p_{\mathrm{FA}})) = 1 - \mathcal{Q}\left(\frac{\gamma(p_{\mathrm{FA}}) - (\sigma_N^2 + P_{Rx}(d))}{\sigma_N^2\sqrt{2/(f_s t_s)}}\right). \tag{3}$$

During the long-period sensing the sensors keep measuring the channel to collect samples of active and idle period durations. Due to the longer sensing time, p_{FA} approaches zero. The missed detection probability p_{MD} also approaches zero inside the sensors $\mathcal{A}_{\mathrm{CCA}}$, the CCA area, where all transmissions are detected, and approaches 1 outside the $\mathcal{A}_{\mathrm{CCA}}$.

Under the path-loss propagation model the CCA area is circular; its radius depends on the WLAN transmission power and can be controlled by tuning the CCA threshold $\psi \geq \psi_0$ [30]:

$$R_{\mathrm{CCA}} \triangleq R_{\mathrm{CCA}}(\psi) = \left(\frac{\psi - \sigma_N^2}{P_{\mathrm{WLAN}} P_{L_0}}\right)^{-1/\eta}. \tag{4}$$

We derive the COG-MAC performance model considering path-loss-based channel attenuation. However, the model can be extended for more generic signal attenuation models, at the expense of increased analytic complexity. In [33] we give the extended model, based on channel attenuation enhanced with log-normal shadowing, and evaluate the effect of shadowing on the protocol performance in Section 7.

4 The WLAN Channel Activity Model

We model the WLAN channel activity as a semi-Markovian system of active and idle periods as originally proposed in [23] and validated in [25]. We call this model the *Global View*, since it captures the global WLAN activity. Fig. 1(a) depicts the states of the Global View model and their merging into a two-state semi-Markovian chain. The states of Data, SIFS and ACK transmissions are grouped together into a single *Active* state and the states that represent the WLAN Contention Window period (CW) and the WLAN *white space* (WS) due to user inactivity are merged into a single *Idle* state. The distributions of the active and idle states, $f_A(t)$ and $f_I(t)$, respectively, define how long the WLAN channel remains in either state. As proposed in [23], a *uniform* distribution in a range $[\alpha_{\mathrm{ON}}, \beta_{\mathrm{ON}}]$ sufficiently models the active channel periods. The idle distribution is modeled as a *mixture* of uniformly distributed idle periods within $[0, \alpha_{\mathrm{BK}}]$, corresponding to the WLAN contention periods, and long, zero-location *generalized Pareto*-distributed idle periods with parameters (ξ, σ) that capture the heavy-tailed behavior of the white spaces. The *percentage of contention periods* $p \in (0, 1)$ determines the shape of the mixed idle distribution, which obtains the following form [23]:

$$f_I(t) \triangleq p f_I^{(\mathrm{CW})}(t) + (1-p) f_I^{(\mathrm{WS})}(t) = \\ = \begin{cases} p \cdot \frac{1}{\alpha_{\mathrm{BK}}} + (1-p) \cdot \frac{1}{\sigma}(1 + \xi \frac{t}{\sigma})^{-\frac{1+\xi}{\xi}}, & t \le a_{\mathrm{BK}}, \\ (1-p) \cdot \frac{1}{\sigma}(1 + \xi \frac{t}{\sigma})^{-\frac{1+\xi}{\xi}}, & t > a_{\mathrm{BK}}, \end{cases} \tag{5}$$

while the active distribution is given as:

$$f_A(t) = \frac{1}{\beta_{\mathrm{ON}} - \alpha_{\mathrm{ON}}}, \quad t \in (\alpha_{\mathrm{ON}}, \beta_{\mathrm{ON}}). \tag{6}$$

We define the WLAN *load* as the percentage of time the channel is active due to WLAN operation:

$$\rho \triangleq \frac{E[T_{\mathrm{ON}}]}{E[T_{\mathrm{ON}}] + E[T_{\mathrm{OFF}}]} = \frac{\frac{\alpha_{\mathrm{ON}} + \beta_{\mathrm{ON}}}{2}}{p\frac{\alpha_{\mathrm{BK}}}{2} + (1-p)\frac{\sigma}{1-\xi}}. \tag{7}$$

Additionally, this provides the probabilities of active and idle channel at an arbitrary point in time, $p_A = \rho$ and $p_I = (1 - \rho)$, respectively.

Our objective is that of estimating the parameters of the model by means of sensor node observations. To this purpose, we define a *Local View* model by "extracting" from the Global View the WLAN channel activity as seen by a single sensor node. Due to sensitivity limitations the sensors detect WLAN transmitters only within the CCA area, that is, they observe the WLAN activity only partially, with some probability, given by the observable load parameter p_{CCA}. Assuming uncorrelated consequent WLAN transmissions the channel activity pattern seen

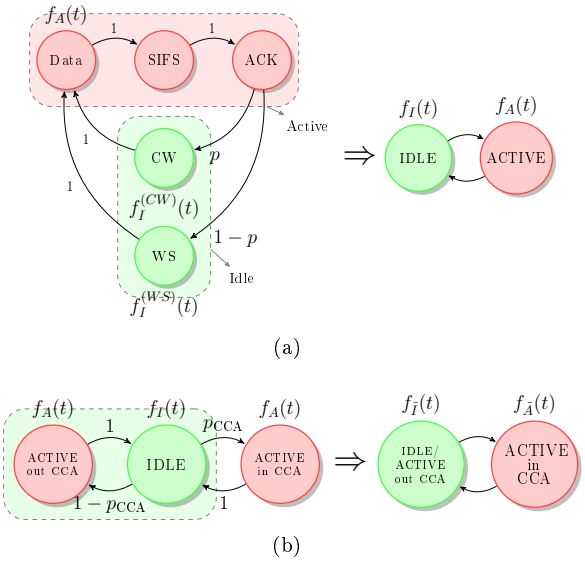


Figure 1: (a): The Global View model with all channel states and the reduced two-state semi-Markovian model. (b): The 3-state semi-Markovian model and its 2-state equivalent for the Local View channel usage modeling.

by a sensor can be described with a 3-state semi-Markovian system, as shown in Fig. 1(b), distinguishing between detected and non-detected WLAN activity that occur with probabilities, p_{CCA} and $1-p_{\mathrm{CCA}}$, respectively. Finally, by merging the states at which the sensor detects an idle channel we obtain the *Local View*, as a 2-state semi-Markovian system with the *observable* sojourn time distributions $f_{\tilde{A}}(t)$ and $f_{\tilde{I}}(t)$. It holds that $f_{\tilde{A}}(t) = f_A(t)$, but $f_{\tilde{I}}(t) \neq f_I(t), \quad \forall p_{\mathrm{CCA}} < 1$. The observable idle channel period consists of a random number of WLAN *cycles*, that is, consecutive idle and un-detected active periods, followed by an additional idle period. Its distribution, $f_{\tilde{I}}(t)$, is, therefore, a random-term convolution-based function of the idle and active time distributions, $f_I(t)$ and $f_A(t)$, and of the observable load p_{CCA}, and can be expressed in closed form only in the Laplace transform (LT) domain, as shown in [34]:

$$f^*_{\tilde{I}}(s) = f^*_I(s) \cdot p_{\mathrm{CCA}} \;/\; \left[1 - (1 - p_{\mathrm{CCA}}) f^*_I(s) f^*_A(s)\right], \tag{8}$$

where $f^*(s)$ denotes the Laplace transform of function f. We discuss the feasibility of parameter estimation in [34][35], where we propose an estimation algorithm that integrates dynamically the collected samples, and therefore runs efficiently on memory-constrained sensor devices.

5 The Cognitive WSN

We propose a WSN COGnitive Medium Access Control (COG-MAC) that employs WLAN usage prediction and channel sensing so as to minimize the energy cost for *unicast* WSN communication under WLAN interference. In particular, it aims at minimizing the transmission energy spent by sensors for transmitting and receiving data packets. COG-MAC can be combined with some duty-cycling or wake-up enabled solution that is responsible for minimizing the energy spent due to idle listening. Therefore, the COG-MAC design does not consider the idle listening energy costs.

The operation of COG-MAC is divided into two main phases. The first one is the *estimation and optimization phase*, when a sensor listens to the channel and gathers samples of the active and idle times, estimates the Local View parameters and selects the optimal one-hop transmission distance and the optimal packet size. The second one is the *transmission phase*, when the sensor transmits and receives data packets. The sensor moves back to the first phase either periodically, or when it experiences a performance drop, suggesting that the estimated WLAN activity parameters are no longer valid (i.e., WLAN activity has significantly changed).

5.1 Estimation and optimization

During the estimation phase potential transmitter (TR) and receiver (RR) sensors listen to the channel and gather active and idle times for estimating the WLAN channel activity. As shown in Fig. 2(a), they perform the measurements for the maximized CCA area $\mathcal{A}_{\mathrm{CCA}}$ (denoted by $\mathcal{A}^T_{\mathrm{CCA}}$ and $\mathcal{A}^R_{\mathrm{CCA}}$ for TR and RR respectively) by using the maximum sensitivity level ψ_0, leading to $R_{\mathrm{CCA}} = R_{\mathrm{CCA}}(\psi_0)$. Based on these measurements they derive the Local View parameters, that is, the parameters of the functions $f_A(t), f_I(t)$, and p_{CCA}. The required number of the samples and thus the length of the estimation phase depends on the target estimation accuracy, which in turn is determined by the sensitivity of COG-MAC. Therefore, we discuss this issue in Section 7.

In addition, for a better estimate of the spatial distribution of the active WLAN users, each sensor also evaluates $\hat{p}^T_{\mathrm{CCA}}$, the common load it can observe within the overlap of the CCA areas. Specifically, it measures the load in the disk area $\hat{\mathcal{A}}^T_{\mathrm{CCA}}$ by filtering the measurements with a changed sensitivity level ψ, such that for a TR–RR distance r, $\hat{R}_{\mathrm{CCA}} = R_{\mathrm{CCA}} - r$. At the end of the estimation phase the sensors receive the observable load values from the potential receivers, denoted by p^R_{CCA}. Based on the locally estimated and received WLAN channel activity model parameters, the sensors select the transmission parameters that are expected to result in minimum energy consumption per bit and meter, according to the model and the implementation given in Section 6. Specifically, they optimize the packet size, to trade-off the probability of interference with a new WLAN transmission and the useful information transmitted per packet. They optimize the transmission

distance, to trade-off the probability that a new WLAN transmission does not cause harmful interference and WSN packet transmission can continue even after the white space period, and the progression towards the multihop destination.

5.2 Transmissions with COG-MAC

The estimation and optimization phase is followed by the *transmission phase* when actual network operation occurs. We assume that the WSN operates under a duty-cycling or wake-up radio based protocol, to limit the energy that is spent in idle listening [5][7][10]. In case of duty-cycling, the WSN nodes are synchronized. Synchronization gaps are, however, expected, as a result of CPU clock drift, and have to be accounted for. Their maximum value $t_{\mathrm{SYNC}}^{\max}$ is determined by the frequency of synchronization data exchange.

Fig. 2(b) shows the COG-MAC operation within a duty-cycle for potential transmitters (TR) and receivers (RR). The duty-cycle of the TR nodes starts with a guard time (denoted as SYNC in the Figure) equal to $t_{\mathrm{SYNC}}^{\max}$, ensuring that channel sensing and transmission do not overlap due to the lack of perfect synchronization. The medium access control is a modified CSMA/CA with the key component of dual channel sensing. As it is shown in Fig. 2(b), the *on* time of the duty-cycle begins with two short channel sensing measurements with a duration of t_s, separated by a time *gap* of t_{gap}, where $\alpha_{\mathrm{ON}} \geq t_{\mathrm{gap}} \geq \alpha_{\mathrm{BK}}$. The sensor's RF circuit can be powered-off between the short channel measurements. If the channel state is correctly detected as idle at both measurements, the sensors can safely assume that the spectrum was idle in the entire time and characterize the idle period as a white space. The operation in the rest of the cycle is determined by the sensing result. If any of the measurements have indicated an active state, the sensor immediately transits to *sleep* mode to save energy. Instead, sensors with idle measurements stay awake and follow a CSMA/CA-based channel access with RTS/CTS exchange. RTS/CTS has been shown to be beneficial in [28], as it allows TR and RR to share their view on the channel status, and increases the probability that the current period is indeed a long white-space (and not the case when a WLAN station is transmitting outside the CCA area), as the total observable load includes the percentage seen by the RR node. We evaluate the usefulness of RTS/CTS under fixed packet size in Section 7. Sensors are assumed to transit to sleep mode after packet transmission, and retransmit packets, if required, in consecutive duty cycle periods.

6 COG-MAC Optimization

In this Section we define the COG-MAC energy consumption model, and formulate the packet size and transmission distance optimization problem. This formulation allows us to trade-off the probability of successful packet transmission with its use-

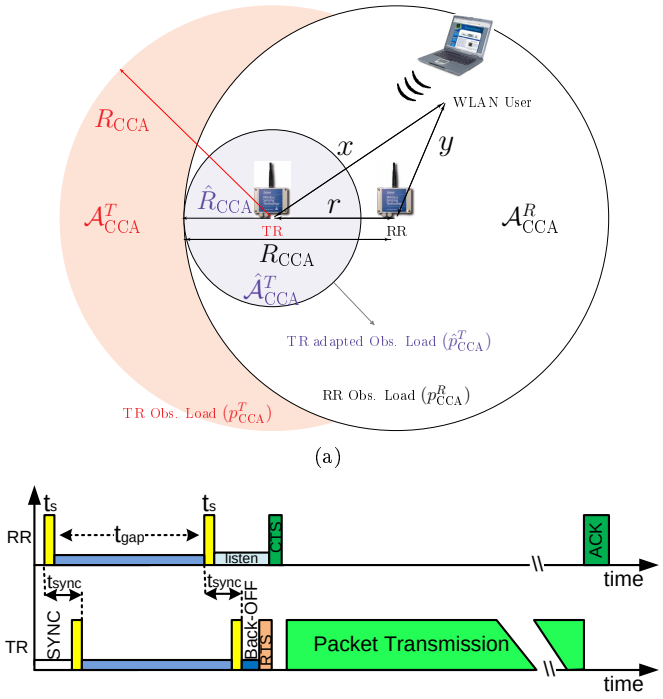


Figure 2: (a):TR and RR CCA areas, $\mathcal{A}^T_{\mathrm{CCA}}$ and $\mathcal{A}^R_{\mathrm{CCA}}$, and the adapted CCA area $\hat{\mathcal{A}}^T_{\mathrm{CCA}}$. (b): Time diagram of COG-MAC operations.

fulness, considering the amount of information transmitted and the distance travelled. Then we present the detailed analytic model of COG-MAC that is required for the optimization. Due to space limitation the presentation is restricted to the case of perfect synchronization, RTS/CTS-based access and simple path-loss model. Interested readers can find the respective model extensions in [28][33]. Moreover, in order to focus on the effect of WLAN interference, we consider the case of low WSN load, when the probability of sensors competing for the channel is low. The model can be extended for the high load case, including expected delays of channel access due to contention resolution.

6.1 Energy efficiency optimization

COG-MAC consumes energy for computing and storing the optimal transmission parameters, for packet transmission, and for sensing, listening and packet reception. Below we focus on the energy spent for radio operations, as their energy consumption in typical sensor nodes is at least two orders of magnitude higher than that of computations. We consider the WSN communication to be energy optimal when the

energy cost of transmitting and receiving a unit of information at unit distance is minimized. Therefore, we define the main performance measure as the energy consumption at the TR and RR nodes until successful packet delivery, that is, through the sequence of possibly unsuccessful and eventually successful RTS/CTS *handshake* and packet transmission attempts, normalized by the amount of information transmitted and the distance covered.

We consider a fixed power cost, $P_{\mathrm{ON}}^{\mathrm{WSN}}$, for channel sensing, transmitting, receiving, and idle listening during the RTS/CTS handshake. Consequently, the expected TR energy cost of attempting a handshake is $e_{\mathrm{hs}}^{\mathrm{TR}} = P_{\mathrm{ON}}^{\mathrm{WSN}}\,(t_{\mathrm{hs}} + 2t_{\mathrm{s}})$, where $2t_{\mathrm{s}}$, t_{hs} are, respectively, the durations of the repeated sensing and the handshake. The latter includes the synchronization gap and the RTS/CTS process. Let $\mathcal{T}$ denote the event of successful handshake. Assuming that the WLAN channel state is uncorrelated at the consecutive handshake attempts, the number of unsuccessful handshakes has geometric distribution with parameter $P\{\mathcal{T}\}$, and the expected energy cost until handshake success becomes:

$$E_{\mathrm{hs}}^{\mathrm{TR}}(r) = e_{\mathrm{hs}}^{\mathrm{TR}} \;/\; P\{\mathcal{T}\}. \tag{9}$$

At the receiver, the handshake energy cost, $e_{\mathrm{hs}}^{\mathrm{RR}}$, depends on whether the node has participated in the handshake. Under perfect synchronization and considering $t_{\mathrm{hs}} >> t_s$ the expected cost can be approximated as:

$$e_{\mathrm{hs}}^{\mathrm{RR}} \approx \begin{cases} e_{\mathrm{hs}}^{\mathrm{TR}}, & \text{RR participates in HS} \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Similar to (9), the expected handshake cost will be:

$$E_{\mathrm{hs}}^{\mathrm{RR}}(r) = e_{\mathrm{hs}}^{\mathrm{TR}}\left[1 + (1/P\{\mathcal{T}\} - 1)\cdot P\{\mathcal{T}_P|\overline{\mathcal{T}}\}\right], \tag{11}$$

with $P\{\mathcal{T}_P|\overline{\mathcal{T}}\}$ denoting the probability that RR has participated on an, otherwise, failed handshake attempt.

The energy cost for transmitting and for receiving a packet with transmission time t is, similarly, $P_{\mathrm{ON}}^{\mathrm{WSN}}t$. If a packet transmission attempt fails, a new handshake must be established before attempting a new transmission. Consequently, the expected energy cost of successful packet delivery with packet transmission time t becomes:

$$E_{\mathrm{trans}}(r,t) = \left[E_{\mathrm{hs}}^{\mathrm{TR}}(r) + E_{\mathrm{hs}}^{\mathrm{RR}}(r) + 2\cdot P_{\mathrm{ON}}^{\mathrm{WSN}}\cdot t\right] \;/\; P\{\text{transmission success}|\mathcal{T}\}. \tag{12}$$

The theoretical optimal packet length and receiver distance, t^*, r^*, as a function of the parameters, $(\xi, \sigma, p, p_{\mathrm{CCA}}, \hat{p}_{\mathrm{CCA}}^{T})$ is given by:

$$(t^*, r^*) = \arg\min_{t,r}\left\{E_{\mathrm{trans}}(r,t)\left[r\cdot(R_{\mathrm{WSN}}t - L_0)\right]\right\}, \tag{13}$$

where R_{WSN} and L_0 denote the WSN transmission rate and the packet overhead, respectively. (13) can be easily modified to consider only t or r WSNs with known node distance or packet size respectively.

We derive the optimal values numerically by solving the above optimization problems applying the bisection method. For practical implementation the optimization problem can be solved a-priori, and the optimal packet size and next-hop distance pairs for a set of WLAN load parameter vectors can be stored in the sensor.

6.2 COG-MAC probability of successful handshake and transmission

In this Section we derive analytically the probabilities of successful handshake and packet transmission, required in (9), (12), respectively.

The TR starts a handshake by transmitting an RTS packet, if its dual sensing process gave idle channel status. Let $\hat{I}_T^{(i)}, \hat{I}_R^{(i)}$ denote the events that the i-th channel measurement is idle at the TR and RR nodes, respectively, with $i = 1, 2$. The handshake attempt is successful if the RR node is awake, as a result of a pair of idle measurements, and, additionally, if the communication is not disturbed by an ongoing, miss-detected WLAN transmission, or by a WLAN transmission that starts during the period of the handshake within the TR or RR interference regions. After a successful handshake the packet transmission itself will be successful, if all WLAN sources within the interference region of the RR remain silent during the whole packet transmission time. We derive the probability of successful handshake and transmission in five steps.
1) We define the spatial distribution of WLAN sources, as seen by the TR node, based on the a priori measured observable load values and the TR-RR distance r.
2) Using Bayesian inference we derive the probability of idle and active channel status at the TR and RR nodes, given the observed idle state measurements at the TR.
3) We derive the distribution of the interference-free time that remains after the dual sensing process.
4) Based on the previous steps, we express the probability $P\{\mathcal{T}\} \triangleq P\{\mathcal{T}|\hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}$ of a successful handshake between the TR and the RR node.
5) Finally, conditioned on the successful handshake, the probability of successful packet transmission $P\{\text{transmission success}|\mathcal{T}\}$ is expressed as a function of transmission distance and packet length.

Let $S \in \mathcal{S} = \{I\} \cup \{A_{\text{XY}}(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ be the channel status. The status is either idle, I, or active, $A_{\text{XY}}(x, y)$, with a WLAN *source* at distances $(X, Y) = (x, y) \in \mathcal{X} \times \mathcal{Y}$ from the TR and RR nodes, respectively (see Fig. 2(a)). $\mathcal{X} \times \mathcal{Y}$ denotes the set of all possible WLAN source positions. $S^{(i)} \in \mathcal{S}$ denotes the channel status during the i-th sensing measurement, where $i = 1, 2$.

6.2.1 Spatial distribution of WLAN interfering sources

The spatial distribution of WLAN sources around the TR and RR nodes affects their miss-detection probabilities, as well as the probability that such a source within the TR/RR interference region starts to transmit during the WSN packet transmission. As shown in Fig. 2(a), the TR can estimate the joint distribution of the distances X,Y of a possible active WLAN source based on the a-priori known observable load values, p_{CCA}^R, received from RR, and $\hat{p}_{\text{CCA}}^T$, measured by the TR itself. Since $\hat{\mathcal{A}}_{\text{CCA}}^T \subseteq \mathcal{A}_{\text{CCA}}^R$, an arbitrary WLAN source lies in the area $\hat{\mathcal{A}}_{\text{CCA}}^T$ with probability

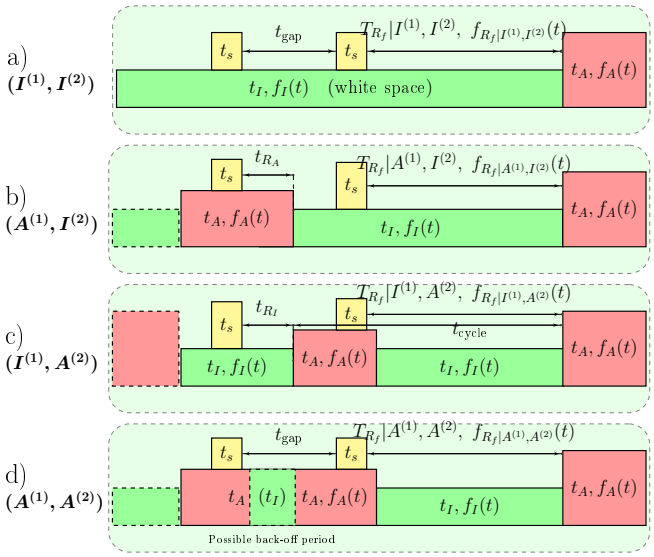


Figure 3: Diagram for the calculation of remaining time densities.

$\hat{p}^T_{\text{CCA}}$, in the area $\mathcal{A}^R_{\text{CCA}} \setminus \hat{\mathcal{A}}^T_{\text{CCA}}$ with probability $p^R_{\text{CCA}} - \hat{p}^T_{\text{CCA}}$, and in the AP area outside $\mathcal{A}^R_{\text{CCA}}$ otherwise. Since there is no additional a-priori information available about the WLAN source locations, the TR assumes that these sources are located uniformly at random inside the respective areas. In addition, we approximate the AP area as a disc around TR with radius $R_{\max}$. This approximation does not affect the model accuracy significantly, unless the TR happens to be very close to the border of the AP area. Let $f_X(x)$ denote the unconditional probability density function of distance X, for uniformly random WLAN source locations in an disk area around TR with radius $R_{\max}$. Similarly, $f_{Y|X}(x, y)$ denotes the density of the distance Y from the RR node, given $X{=}x$, and $f_{XY}(x, y) = f_X(x) f_{Y|X}(x, y)$ denotes the unconditional joint distance density. These functions depend on the distance r between the two sensors and can be derived through basic geometry. The distribution of distance X, conditioned on the reported observable load values $\hat{p}^T_{\text{CCA}}$ and p^R_{CCA} can be expressed as:

$$p_{A_X}(x) = f_X(x) \cdot \begin{cases} \hat{p}^T_{\text{CCA}} \ / \ \nu_{\hat{R}_{\text{CCA}}}, & x \le \hat{R}_{\text{CCA}}, \\ \frac{F_{Y|X}(x, R_{\text{CCA}})(p^R_{\text{CCA}} - \hat{p}^T_{\text{CCA}})}{\nu_{R_{\text{CCA}}} - \nu_{\hat{R}_{\text{CCA}}}} + \frac{\overline{F}_{Y|X}(x, R_{\text{CCA}})(1 - p^R_{\text{CCA}})}{1 - \nu_{R_{\text{CCA}}}}, & \text{otherwise}, \end{cases} \tag{14}$$

where $\nu_{\hat{R}_{\text{CCA}}}, \nu_{R_{\text{CCA}}}$ denote the ratio of the observable areas $\hat{\mathcal{A}}^T_{\text{CCA}}$ and $\mathcal{A}^R_{\text{CCA}}$, respectively, over the total WLAN AP area with radius $R_{\max}$. Similarly, the conditional distance density $p_{A_{Y|X}}(x, y)$, can be expressed as:

$$p_{A_{Y|X}}(x,y) = f_{Y|X}(x,y)\cdot \cdot \begin{cases} \hat{p}^T_{\mathrm{CCA}} \;/\; \nu_{\hat{R}_{\mathrm{CCA}}}, & x \le \hat{R}_{\mathrm{CCA}} \\ \left(p^R_{\mathrm{CCA}} - \hat{p}^T_{\mathrm{CCA}}\right) \;/\; \left(\nu_{R_{\mathrm{CCA}}} - \nu_{\hat{R}_{\mathrm{CCA}}}\right), & x > \hat{R}_{\mathrm{CCA}}, y \le R_{\mathrm{CCA}} \\ \left(1 - p^R_{\mathrm{CCA}}\right) \;/\; \left(1 - \nu_{R_{\mathrm{CCA}}}\right), & \text{otherwise.} \end{cases} \tag{15}$$

Finally, we have $p_{A_{\mathrm{XY}}}(x,y) = f_X(x) \cdot p_{A_{Y|X}}(x,y)$ and $p_{A_Y}(y) = \int_{\mathcal{X}} p_{A_{\mathrm{XY}}}(x,y)dx$.

6.2.2 Bayesian inference of channel status

We derive now the *posterior* distribution of channel status, $\left(S^{(1)}, S^{(2)}\right)$, given the observed TR idle measurements, applying Bayesian formulation. To calculate $P\{\left(S^{(1)}, S^{(2)}\right)|\hat{I}^{(1)}_T, \hat{I}^{(2)}_T\}$ we use the following decomposition:

$$\begin{aligned} &P\{\left(S^{(1)}, S^{(2)}\right)|\hat{I}^{(1)}_T, \hat{I}^{(2)}_T\} = \\ &= P\{S^{(1)}|\hat{I}^{(1)}_T, \hat{I}^{(2)}_T\} \cdot P\{S^{(2)}|S^{(1)}, \hat{I}^{(1)}_T, \hat{I}^{(2)}_T\} = \\ &= P\{S^{(1)}|\hat{I}^{(1)}_T\} \cdot P\{S^{(2)}|S^{(1)}, \hat{I}^{(2)}_T\}. \end{aligned} \tag{16}$$

Conditioned on the first idle measurement the channel status is either idle or active with the following probabilities:

$$P\{I^{(1)}|\hat{I}^{(1)}_T\} = (1 - p_{\mathrm{FA}})p_I \left[\int_{\mathcal{X}} p_{\mathrm{MD}}(u)p_{A^{(1)}_X}(u)du + (1 - p_{\mathrm{FA}})p_I\right]^{-1}, \tag{17}$$

$$P\{A^{(1)}_X|\hat{I}^{(1)}_T\} = p_{\mathrm{MD}}(x)p_{A^{(1)}_X}(x) \left[\int_{\mathcal{X}} p_{\mathrm{MD}}(u)p_{A^{(1)}_X}(u)du + (1 - p_{\mathrm{FA}})p_I\right]^{-1}, \tag{18}$$

$$P\{A^{(1)}_{\mathrm{XY}}|\hat{I}^{(1)}_T\} = p_{\mathrm{MD}}(x)p_{A^{(1)}_{\mathrm{XY}}}(x,y) \left[\int\int_{\mathcal{XY}} p_{\mathrm{MD}}(u)p_{A^{(1)}_{\mathrm{XY}}}(u,v)dudv + (1-p_{\mathrm{FA}})p_I\right]^{-1} \tag{19}$$

where $p_{A^{(1)}_X}(x) = p_A \cdot p_{A_X}(x)$ is the probability that a WLAN source is active at a distance x from the TR at the time of the first measurement, and $p_{A^{(1)}_{\mathrm{XY}}}(x,y) = p_A \cdot p_{A_{\mathrm{XY}}}(x,y)$ is the probability that it is active at distances x, y from TR and RR respectively.

To derive the second term of (16), we first express the status transition probabilities, $P\{S^{(2)}|S^{(1)}\}$, following Fig. 3. For $I^{(1)}$ (Fig. 3(a),(c)):

$$P\{I^{(2)}|I^{(1)}\} = \overline{F}_{R_I}(t_{\mathrm{gap}}), \tag{20}$$

$$P\{A^{(2)}_{\mathrm{XY}}|I^{(1)}\} = \left(1 - P\{I^{(2)}|I^{(1)}\}\right) p_{A_{\mathrm{XY}}}(x,y). \tag{21}$$

In the above, $F_{R_I}(t)$ denotes the distribution function of the remaining idle time, T_{R_I}, As the time of the first measurement is uniformly distributed within the WLAN

idle period: $F_{R_I}(t) = \int_t^\infty (1 - \frac{t}{z}) f_I(z)dz$. Similarly, for $S^{(1)} = A^{(1)}_{\rm XY}$, $\forall x, y$ we obtain (Fig. 3(b)):

$$P\{I^{(2)}|A^{(1)}_{\rm XY}\} = \int_0^{t_{\rm gap}} \overline{F}_I(t_{\rm gap} - z) f_{R_A}(z)dz, \tag{22}$$

with $f_{R_A}(t) = \int_t^\infty \frac{1}{z} f_A(z)dz$ denoting the density of the remaining active WLAN period. For $S^{(2)} = A^{(2)}_{\rm XY}(x, y)$ we need to distinguish between the cases of a continuous active period, with the same WLAN source being active, or that of a short idle period between the measurements (Fig. 3(d)):

$$\begin{aligned} &P\{A^{(2)}_{\rm XY}|A^{(1)}_{\rm XY}(x_1, y_1)\} = \\ &= \overline{F}_{R_A}(t_{\rm gap})\delta(x_1, y_1) + [1 - \overline{F}_{R_A}(t_{\rm gap}) - P\{I^{(2)}|A^{(1)}_{\rm XY}\}]p_{A_{\rm XY}}(x, y). \end{aligned} \tag{23}$$

Finally, we define the channel status probabilities conditioned on the second idle measurement and based on the a-priori status transition probabilities calculated in (20)-(23):

$$P\{I^{(2)}|I^{(1)}, \hat{I}^{(2)}_T\} = \frac{(1-p_{\rm FA})P\{I^{(2)}|I^{(1)}\}}{(1-p_{\rm FA})P\{I^{(2)}|I^{(1)}\} + \int\int_{\mathcal{XY}} p_{\rm MD}(u)P\{A^{(2)}_{\rm XY}|I^{(1)}\}dudv}, \tag{24}$$

$$P\{A^{(2)}_{XY}|I^{(1)}, \hat{I}^{(2)}_T\} = \left(1 - P\{I^{(2)}|I^{(1)}, \hat{I}^{(2)}_T\}\right) p_{A_{\rm XY}}(x, y), \tag{25}$$

$$P\{I^{(2)}|A^{(1)}_{XY}, \hat{I}^{(2)}_T\} = \frac{(1-p_{\rm FA})P\{I^{(2)}|A^{(1)}_{XY}\}}{(1-p_{\rm FA})P\{I^{(2)}|A^{(1)}_{XY}\} + \int\int_{\mathcal{XY}} p_{\rm MD}(u)P\{A^{(2)}_{XY}|A^{(1)}_{XY}\}dudv}, \tag{26}$$

$$P\{A^{(2)}_{XY}|A^{(1)}_{XY}(x_1, y_1), \hat{I}^{(2)}_T\} = \frac{p_{\rm MD}(x)P\{A^{(2)}_{XY}|A^{(1)}_{XY}(x_1,y_1)\}}{(1-p_{\rm FA})P\{I^{(2)}|A^{(1)}_{XY}\} + \int\int_{\mathcal{XY}} p_{\rm MD}(u)P\{A^{(2)}_{XY}|A^{(1)}_{XY}\}dudv}. \tag{27}$$

6.2.3 Conditional remaining interference-free time

We define T_{R_F}, $T^{(\rm TR,RR)}_{R_F}$ as the total *interference-free time* remaining at the RR sensor and at both the TR and RR sensors, respectively, after the sensing process at the nodes, and derive the densities $f_{R_F|S^{(1)},S^{(2)}}(t)$ and $f^{(\rm TR,RR)}_{R_F|S^{(1)},S^{(2)}}(t)$, given the channel status at the time of the TR measurements. T_{R_F} includes the interval between the end of the second sensing measurement and the start of the following active WLAN period, and a *geometric* number of successive WLAN *cycles*, i.e. pairs of successive active and idle WLAN periods with density $f_C(t) = f_I(t) * f_A(t)$, representing consecutive WLAN transmissions outside the interference area. The distribution of T_{R_F} can be numerically calculated with the help of Laplace transform, similar to (8) and for all $S^{(1)}, S^{(2)} \in \mathcal{S}$:

$$f_{R_F|S^{(1)},S^{(2)}}(t) = \mathcal{L}^{-1}\left\{\frac{p_{\mathcal{IN}} f^*_{R_f|S^{(1)},S^{(2)}}(s)}{1 - (1 - p_{\mathcal{IN}})f^*_I(s)f^*_A(s)}\right\}, \tag{28}$$

For additionally undisturbed TR, we obtain the density:

$$f^{(\mathrm{TR,RR})}_{R_F|S^{(1)},S^{(2)}}(t) = \mathcal{L}^{-1}\left\{\frac{p^{(\mathrm{TR,RR})}_{\mathcal{IN}} f^*_{R_f|S^{(1)},S^{(2)}}(s)}{1-(1-p^{(\mathrm{TR,RR})}_{\mathcal{IN}})f^*_I(s)f^*_A(s)}\right\}, \tag{29}$$

where $p_{\mathcal{IN}}$, the probability that an activated WLAN source interferes with the RR reception is:

$$p_{\mathcal{IN}} = \begin{cases} p^R_{\mathrm{CCA}}\left(R_{\mathcal{I}}(r)/R_{\mathrm{CCA}}\right)^2, & \text{if } R_{\mathcal{I}}(r) \le R_{\mathrm{CCA}} \\ p^R_{\mathrm{CCA}} + (1-p^R_{\mathrm{CCA}})\left(R^2_{\mathcal{I}}(r) - R^2_{\mathrm{CCA}}\right)/\left(R^2_{\max} - R^2_{\mathrm{CCA}}\right), & \text{otherwise,} \end{cases} \tag{30}$$

and the probability that the source lies, additionally, inside the TR interference area is given as:

$$p^{(\mathrm{TR,RR})}_{\mathcal{IN}} = \int_{y\le R_{\mathcal{I}}(r)}\int_{x\le R_{\mathcal{I}}(r)} p_{A_{\mathrm{XY}}}(x,y)dxdy. \tag{31}$$

The interval between the end of the second sensing measurement and the start of the first active WLAN period is denoted by T_{R_f}, and its density depends on the channel status. In the derivations we approximate the sensing period as $T_s = 2t_s + t_{\mathrm{gap}} \approx t_{\mathrm{gap}}$, since $t_{\mathrm{gap}} \gg t_s$.

For $\left(I^{(1)}, I^{(2)}\right)$ (Fig. 3(a)) we safely classify the idle period as white space, and consequently: $$f_{R_f|I^{(1)},I^{(2)}}(t) = f_{R_I}(t+t_{\mathrm{gap}})/\overline{F}_{R_I}(t_{\mathrm{gap}}). \tag{32}$$

In the case $(A^{(1)}_{\mathrm{XY}}, I^{(2)})$ (Fig. 3(b)), a transition from active to idle status occurs sometime $z \le t_{\mathrm{gap}}$ after the first TR measurement, and the idle period may also be a back-off, which gives:

$$f_{R_f|A^{(1)}_{\mathrm{XY}},I^{(2)}}(t) = \frac{1}{\overline{F}_{R_A}(t_{\mathrm{gap}})}\int_0^{t_{\mathrm{gap}}} f_{R_A}(z)\frac{f_I(t+t_{\mathrm{gap}}-z)}{\overline{F}_I(t_{\mathrm{gap}}-z)}dz. \tag{33}$$

For $S^{(2)} = A^{(2)}_{\mathrm{XY}}$ (Fig. 3(c)) the channel is active at the second measurement, and the remaining time, T_{R_f}, is given by the remaining active and the following idle period:

$$f_{R_f|I^{(1)},A^{(2)}_{\mathrm{XY}}}(t) = \frac{1}{\overline{F}_{R_I}(t_{\mathrm{gap}})}\int_0^{t_{\mathrm{gap}}} f_{R_I}(z)\frac{f_C(t+t_{\mathrm{gap}}-z)}{\overline{F}_A(t_{\mathrm{gap}}-z)}dz. \tag{34}$$

Finally under $(A^{(1)}_{XY}, A^{(2)}_{XY})$ (Fig. 3(d)), the active period may be continuous between the two measurements, or interrupted by a short idle time. In the case of continuous active period:

$$f_{R_f|A^{(1)}_{XY}(x_1,y_1),A^{(2)}_{XY}(x_1,y_1)}(t) = \frac{1}{\overline{F}_{R_A}(t_{\mathrm{gap}})}\int_{t_{\mathrm{gap}}}^{\infty} f_{R_A}(z)f_I(t+t_{\mathrm{gap}}-z)dz, \tag{35}$$

while in the case of a short idle period between the measurements, $(x_1,y_1) \neq (x_2,y_2)$:

$$\begin{aligned} &f_{R_f|A^{(1)}_{XY},A^{(2)}_{XY}}(t) = \\ &= \frac{1}{\overline{F}_{R_A}(t_{\mathrm{gap}})}\cdot\int_0^{t_{\mathrm{gap}}} f_{R_A}(z_1)\frac{\int_0^{t_{\mathrm{gap}}-z_1} f_I(z_2)}{F_I(t_{\mathrm{gap}}-z_1)}\cdot\frac{f_{\mathrm{cycle}}(t+t_{\mathrm{gap}}-z_1-z_2)}{\overline{F}_A(t_{\mathrm{gap}}-z_1-z_2)}dz_1dz_2. \end{aligned} \tag{36}$$

6.2.4 TR-RR handshake success

For a TR node aiming at communicating with an RR at distance r we calculate the probability of successful handshake, conditioned on the idle TR measurements. The event of handshake success, $\mathcal{T}$, requires, first, idle measurements at the receiver, $\hat{I}_R^{(1)}, \hat{I}_R^{(2)}$. Second, it requires that no WLAN transmission interferes with the RTS/CTS handshake. Since the duration of the handshake is expected to be significantly lower than the WLAN activity dynamics, we approximate the second constraint as the requirement that all the active WLAN sources lie outside the interference regions of both the TR and the RR for the entire handshake period t_{hs}:

$$P\{\mathcal{T}|S^{(1)},S^{(2)}\} \approx P\{\hat{I}_R^{(1)},\hat{I}_R^{(2)}|S^{(1)},S^{(2)}\}\overline{F}^{(\mathrm{TR,RR})}_{R_F|S^{(1)},S^{(2)}}(t_{\mathrm{hs}}). \tag{37}$$

If the channel status is indeed idle during both of the TR measurements, i.e., $(S^{(1)},S^{(2)}) = (I^{(1)},I^{(2)})$, the handshake is successful if there is no false alarm at the RR, and the remaining interference-free time at both the RR and TR is longer than the total duration of the handshake t_{hs}. That is:

$$P\{\mathcal{T}|I^{(1)},I^{(2)}\} = (1-p_{\mathrm{FA}})^2\overline{F}^{(\mathrm{TR,RR})}_{R_F|I^{(1)},I^{(2)}}(t_{\mathrm{hs}}). \tag{38}$$

With a similar reasoning, the conditional handshake success probability for the remaining channel status cases becomes:

$$P\{\mathcal{T}|A^{(1)}_{\mathrm{XY}}(x,y),I^{(2)}\} = (1-p_{\mathrm{FA}})p_{\mathrm{MD}}(y)\overline{F}^{(\mathrm{TR,RR})}_{R_F|A^{(1)}_{\mathrm{XY}},I^{(2)}}(t_{\mathrm{hs}}),\ x\in\mathcal{X}, y\in\mathcal{Y}, \tag{39}$$

$$P\{\mathcal{T}|I^{(1)},A^{(2)}_{XY}(x,y)\} = (1-p_{FA})p_{\mathrm{MD}}(y)\overline{F}^{(\mathrm{TR,RR})}_{R_F|I^{(1)},A^{(2)}_{\mathrm{XY}}}(t_{\mathrm{hs}}),\ x,y\geq R_I(r), \tag{40}$$

$$\begin{aligned} &P\{\mathcal{T}|A^{(1)}_{\mathrm{XY}}(x_1,y_1),A^{(2)}_{\mathrm{XY}}(x_2,y_2)\} = \\ &= p_{\mathrm{MD}}(y_1)p_{\mathrm{MD}}(y_2)\overline{F}^{(\mathrm{TR,RR})}_{R_F|A^{(1)}_{\mathrm{XY}},A^{(2)}_{\mathrm{XY}}}(t_{\mathrm{hs}}), x_2,y_2\geq R_I(r). \end{aligned} \tag{41}$$

The probability of handshake success is then calculated by averaging over all possible cases (Eq. (42)).

$$\begin{aligned} P\{\mathcal{T}|\hat{I}_T^{(1)},\hat{I}_T^{(2)}\} &= P\{\mathcal{T}|I^{(1)},I^{(2)}\}P\{I^{(1)},I^{(2)}|\hat{I}_T^{(1)},\hat{I}_T^{(2)}\}+ \\ &+\iint_{\mathcal{XY}} P\{\mathcal{T}|A^{(1)}_{\mathrm{XY}}(x,y),I^{(2)}\}P\{A^{(1)}_{\mathrm{XY}}(x,y),I^{(2)}|\hat{I}_T^{(1)},\hat{I}_T^{(2)}\}dxdy \\ &+\iint_{x,y\geq R_I(r)} P\{\mathcal{T}|I^{(1)},A^{(2)}_{\mathrm{XY}}(x,y)\}P\{I^{(1)},A^{(2)}_{\mathrm{XY}}(x,y)|\hat{I}_T^{(1)},\hat{I}_T^{(2)}\}dxdy \\ &+\iint_{\mathcal{XY}}\iint_{x_2,y_2\geq R_I(r)} P\{\mathcal{T}|A^{(1)}_{\mathrm{XY}}(x_1,y_1),A^{(2)}_{\mathrm{XY}}(x_2,y_2)\}\cdot \\ &\quad\cdot P\{A^{(1)}_{\mathrm{XY}}(x_1,y_1),A^{(2)}_{\mathrm{XY}}(x_2,y_2)|\hat{I}_T^{(1)},\hat{I}_T^{(2)}\}dx_1dy_1dx_2dy_2. \end{aligned} \tag{42}$$

Finally, we calculate the conditional probability $P\{\mathcal{T}_P|\overline{\mathcal{T}}\}$ needed in (11) applying Bayesian inference:

$$
\begin{aligned}
&P\{\mathcal{T}_P|\overline{\mathcal{T}}\} \triangleq P\{\text{RR attempting HS}|\text{HS failed}\} = \\
&= 1 - P\{\text{RR not attempt. HS}|\text{HS failed}\} = \\
&= 1 - P\{\text{RR not attempt. HS}\} / \left[1 - P\{\mathcal{T}|\hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}\right]
\end{aligned} \tag{43}
$$

where $P\{\text{RR not attempt. HS}\} = 1 - P\{\hat{I}_R^{(1)}, \hat{I}_R^{(2)}|\hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}$. To calculate the last factor, we compute $P\{\hat{I}_R^{(1)}, \hat{I}_R^{(2)}|S^{(1)}, S^{(2)}\}$, $\forall (S^{(1)}, S^{(2)})$ from (38)-(41) by substituting $t_{\text{hs}} = 0$, and average over all possible cases as in (42).

6.2.5 Successful packet transmission

Finally, let us express the probability of successful packet transmission, now conditioned on the success of the handshake. We update all $(S^{(1)}, S^{(2)})$, through Bayesian inference:

$$
P\{S^{(1)}, S^{(2)}|\mathcal{T}\} = P\{S^{(1)}, S^{(2)}|\mathcal{T}, \hat{I}_T^{(1)}, \hat{I}_T^{(2)}\} = \frac{P\{\mathcal{T}|S^{(1)}, S^{(2)}\} P\{S^{(1)} S^{(2)}|\hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}}{P\{\mathcal{T}|\hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}}, \tag{44}
$$

where $P\{\mathcal{T}|\hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}$ is defined in (42), the terms $P\{\mathcal{T}|S^{(1)}, S^{(2)}\}$ are derived in the previous section and $P\{S^{(1)}, S^{(2)}|\hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}$ is derived from (16).

Similarly, we update the total remaining interference-free time, T_{R_F}, with respect to the total length of the handshake time, including the synchronization delay, $t_{\text{hs}}^T = t_{\text{hs}} + t_{\text{SYNC}}$, as it is measured by the TR node:

$$
\begin{aligned}
&\overline{F}_{R_F|(S^{(1)}, S^{(2)}), \mathcal{T}}(t) = P\{T_{R_F} \geq t + t_{\text{hs}}^T | T_{R_F}^{(\text{TR,RR})} \geq t_{\text{hs}}^T\} = \\
&= P\{T_{R_F} \geq t + t_{\text{hs}}^T, T_{R_F}^{(\text{TR,RR})} \geq t_{\text{hs}}^T\} / P\{T_{R_F}^{(\text{TR,RR})} \geq t_{\text{hs}}^T\} \approx \\
&\approx \overline{F}_{R_F|(S^{(1)}, S^{(2)})}(t + t_{\text{hs}}^T) / \overline{F}_{R_F|(S^{(1)}, S^{(2)})}^{(\text{TR,RR})}(t_{\text{hs}}^T),
\end{aligned} \tag{45}
$$

where the respective density functions are given in (28) and (29), and the approximation is valid due to the relatively short WSN handshake time with respect to the average WLAN cycle duration. Finally, from (42) and (45) we express the probability that a packet of transmission duration t will be successfully transmitted as:

$$
P\{\text{transmission success}|\mathcal{T}\} = \sum \overline{F}_{R_F|(S^{(1)}, S^{(2)}), \mathcal{T}}(t) P\{S^{(1)}, S^{(2)}|\mathcal{T}, \hat{I}_T^{(1)}, \hat{I}_T^{(2)}\}, \tag{46}
$$

where the summation is over all possible channel status $S^{(1)}, S^{(2)}$.

7 Numerical Performance Evaluation

We evaluate the performance of COG-MAC, based on the analytic model in Section 6, by comparing it to non-cognitive WSN MAC schemes that – similar to COG-MAC – are controlled by duty-cycling. In particular, we consider an ALOHA-type

Table 1: Parameter setup for the performance evaluation

Channel Model	
Path-Loss Exponent (η)	3.0
Ref. Distance attenuation (P_{L_0})	Isotropic $(\lambda/(4\pi)))^2$
Noise Power (σ_N^2)	-174 dB/Hz
WLAN Properties & Modeling	
Bandwidth (B_{WLAN})	22MHz (802.11x Channel)
Tx-Power	18dBm (22MHz Channel)
Tx-Power inside WSN band (P_{WLAN})	12dB
Max. back-off period (α_{BK})	700μsec
Active period interval ($\alpha_{\mathrm{ON}}, \beta_{\mathrm{ON}}$)	(0.8msec, 1.5msec)
White space Pareto scale (ξ)	0.3095
WSN & CC2420 Properties	
Bandwidth (B_{WSN})	5MHz Zigbee Channel
Tx-Power (P_{WSN})	1dBm
Tx-Rate (R_{WSN})	250kbps
Header Length [Packet overhead] (L_0)	13 bytes
RTS/CTS Length	6 bytes
Minimum SINR (ζ_{SIR})	5dB
Receiver Sensitivity (ψ)	-100 dBm
RSSI Dynamic Range	100dB
Tx/Rx Power Consumption ($P_{\mathrm{ON}}^{\mathrm{WSN}}$)	55mW
Handshake Duration (t_{hs})	768μsec
Channel Sensing Model	
Sampling Frequency (f_s)	5MHz
Sensing Time (t_s)	16μsec
False Alarm Constraint (p_{FA})	10^{-2}

Random Access MAC (RAND), where sensors transmit without any channel sensing before transmission, and a standard 802.15.4-compliant carrier-sense (CSMA) MAC scheme, where WSN nodes perform the standard channel sensing and RTS/CTS handshake. The analytic models for CSMA and RAND are similar to that of COG-MAC and are not presented here due to space limitation. Interested readers can find them in [33], in Sections VII-A, VII-B, respectively. In addition, we evaluate the effect of the channel model on the protocols' performance, by considering log-normal shadowing. Shadowing affects both channel sensing and interference and, therefore, the resulting energy cost. We present the analytic model extensions for COG-MAC and CSMA under log-normal shadowing in Sections VI-C and VII-A of [33], respectively. Finally, to evaluate the importance of the handshake process, we compare COG-MAC performance with and without RTS/CTS mechanism, derived from the model in [28]. For all schemes we consider the normalized energy cost

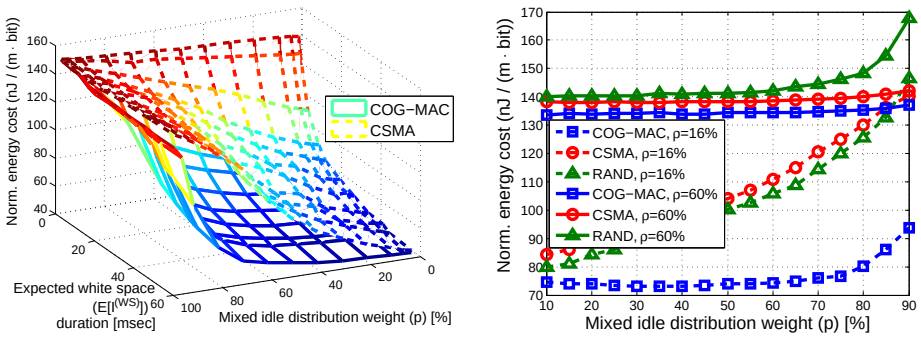
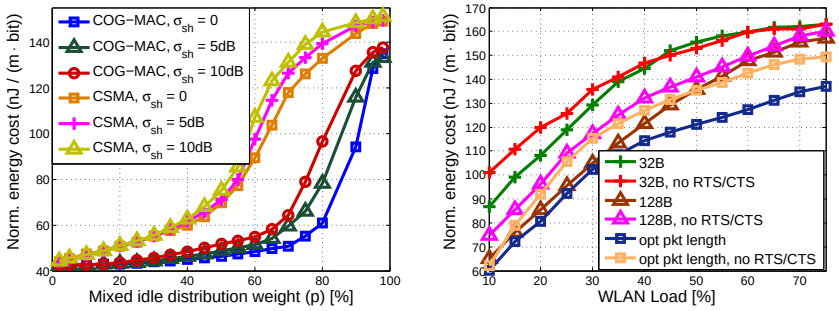


Figure 4: (a) Comparison of Cognitive and CSMA-based MAC over p and σ, and (b) normalized energy cost with respect to p for fixed WLAN Load ($\rho = 16\%$ and 60%)

metric under optimized transmission distance and packet size as defined by (13). The default parameters of our reference evaluation scenario are listed in Table 1.

7.1 Comparison with RAND & CSMA schemes

Fig. 4(a) compares the normalized energy cost of COG-MAC and CSMA with respect to the parameter p, the percentage of short WLAN back-off intervals, and $E\left[I^{(\mathrm{WS})}\right] = \frac{\sigma}{1-\xi}$, the average length of white spaces, controlled by the shape parameter σ of the generalized Pareto distribution. We set the observable load at the receiving sensor at $p^R_{\mathrm{CCA}} = 0.5$. To consider random transmitter location we randomize $\hat{p}^T_{\mathrm{CCA}}$, following a normalized binomial distribution, in $[0, p^R_{\mathrm{CCA}}]$. In general, increasing p or decreasing $E\left[I^{(\mathrm{WS})}\right]$ increases the load and consequently the normalized energy consumption for both protocols. Fig. 4(a), however, shows that COG-MAC significantly outperforms CSMA. In Fig. 4(b) we keep the load constant at $\rho = 16\%$ and 60%, and increase the percentage of the back-off periods p. Even in this case COG-MAC shows better normalized energy consumption compared to both CSMA and RAND. The energy cost of COG-MAC is only marginally affected by the growing percentage of back-off periods. In contrast, RAND and CSMA, due to the fact that they have to optimize transmission parameters for the mixture idle time distribution, cannot provide energy efficient communication for a large range of p. Under high percentage of back-off idle periods RAND exhibits a performance drop due to the absence of the RTS/CTS mechanism, as failed transmission attempts result in a high energy cost at the receiving nodes. On the contrary, handshake-enabled protocols are more energy efficient due to a lower rate of packet transmission failures.



(a) $\overline{I}^{\mathrm{WS}} = 60$ms (b) $p = 0.8$

Figure 5: (a) COG-MAC – CSMA comparison with respect to the normalized energy cost under log-normal shadowing and (b) fixed packet lengths.

7.1.1 Impact of channel shadowing

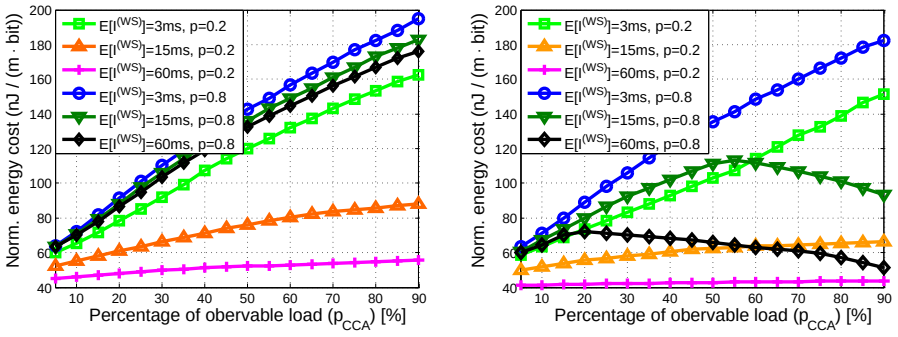
Fig. 5(a) compares COG-MAC to CSMA under log-normal shadowing channel model and for various shadowing standard deviation (σ_{sh}) values. Shadowing on the wireless channel degrades the WSN communication energy efficiency as it adds uncertainty in both the WLAN spatio-temporal model estimation and the interference calculation, and the degradation is significant for COG-MAC at high WLAN load. Still, the large performance gap between the two solutions remains.

7.1.2 Impact of the RTS/CTS handshake mechanism

Fig. 5(b) compares the efficiency of COG-MAC with and without RTS/CTS exchange under increasing WLAN channel load – decreasing average white-spaces duration – and for fixed and optimized WSN packet lengths. We observe that RTS/CTS is always beneficial under optimized packet lengths. As discussed in Section 7.1 the absence of the handshake mechanism degrades the energy efficiency under high channel load. For fixed packet sizes the effect of RTS/CTS handshake depends on the WLAN load. Under high load values, that is, under short expected white-space durations, the increased performance due to efficient white-space discovery is limited, thus, it does not compensate for the additional overhead of the RTS/CTS mechanism.

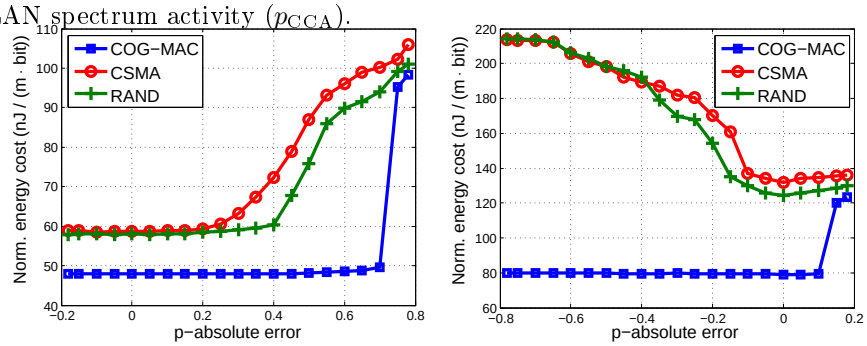
7.1.3 Impact of the receiver observable load

Let us now investigate the effect of the observable load p^R_{CCA} on energy efficiency. Fig. 6 compares the CSMA and COG-MAC normalized energy cost as a function of p^R_{CCA} and for different p and $E\left[I^{(\mathrm{WS})}\right]$ values. For the CSMA scheme (Fig. 6(a)) the energy cost increases monotonically with the observable load, since the interference-free time decreases. We can see similar trends for COG-MAC for low p values in Fig. 6(b). On the contrary, at high p value the COG-MAC energy cost



(a) CSMA (b) COG-MAC

Figure 6: Normalized energy cost with respect to the percentage of the observable WLAN spectrum activity (p_{CCA}).



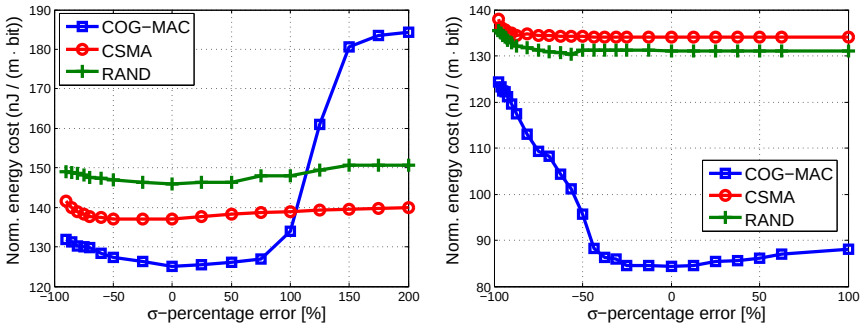
(a) $p = 0.2$ (b) $p = 0.8$

Figure 7: Normalized energy cost with respect to the absolute of error in p-estimation, $[\hat{p} - p]$. $E\left[I^{(\mathrm{WS})}\right] = 36$msec ($\sigma = 0.025$).

decreases at high p^R_{CCA}, because in these scenarios COG-MAC can efficiently filter the short back-off periods. As a result, COG-MAC can provide energy efficient communication despite the limited sensing range, and can decrease the energy cost with up to 66% compared to CSMA.

7.1.4 Sensitivity to the model parameters

COG-MAC may not be able to use optimal packet size and transmission distance due to imperfect WLAN Local View parameter estimation and due to the limited number of options that can be stored in the look-up table in the sensor memory. Here we study the effect of estimation errors on the energy efficiency. We evaluate the energy cost under a given WLAN activity model realization, while COG-MAC variables are optimized in (13) considering erroneous p and σ values. (The results for the other model parameters are similar.) We present comparative results for CSMA



(a) $\sigma = 5 \cdot 10^{-3}$ (7msec) (b) $\sigma = 2 \cdot 10^{-2}$ (30msec)

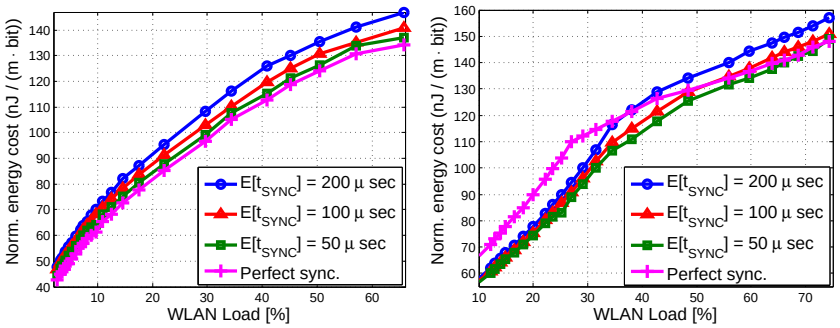
Figure 8: Normalized energy cost with respect to the percentage of error in σ-estimation, $[(\hat{\sigma} - \sigma)/\sigma]$. $p = 80\%$.

and RAND. Fig. 7 shows the effect of the imperfect estimation of p, for low and high p values. We see that COG-MAC is not sensitive to estimation error, unless p is heavily overestimated, since the dual sensing filters out the back-off periods. CSMA and RAND, however, need to take the short back-off periods into account for the optimization, and therefore the imperfect estimation of p deteriorates their performance.

Fig. 8 depicts the sensitivity of the performance on the estimation of σ, the shape parameter of the white space distribution. For the considered scenario CSMA and RAND are not sensitive to estimation errors due to the high p value that makes the estimation of the actual WLAN white spaces less important. COG-MAC, however, transmits, primarily, in the white spaces, and therefore the over- and underestimation of σ leads to increased energy consumption. Still up to 50% error in estimating σ does not have significant effect on the energy cost. Based on [34], this level of accuracy can be achieved by considering 100-1000 idle period samples. This in turn leads to an estimation time in the order of 1-10 seconds, depending on the average lengths of the idle periods. The low sensitivity to estimation errors allows even the use of look-up tables with low granularity. These confirm that the proposed approach with local channel estimation and parameter optimization based cognitive access is a viable solution for sensor networks.

7.2 The effect of loose synchronization

In Fig. 9 we evaluate the energy cost of COG-MAC considering the case of imperfect synchronization of the TR and RR duty-cycles, based on the model in [33], Section VI-B. We consider synchronization gaps uniformly distributed in $(0, t_{\text{sync}}^{\max})$, $E[t_{\text{SYNC}}] = t_{\text{sync}}^{\max}/2$. We show the effect of $E[t_{\text{SYNC}}]$ on the normalized COG-MAC energy cost as a function of the WLAN load, for low and high p values. Since the shifted double sensing procedures require more time, synchronization gaps decrease the probability of successful handshake and reduce the interference-free time for



(a) $p = 20\%$ (b) $p = 80\%$

Figure 9: Normalized communication energy cost with respect to increasing WLAN load (decreasing expected white-space duration), for different average synchronization offsets. $E\left[I^{(\mathrm{WS})}\right] = 60$msec.

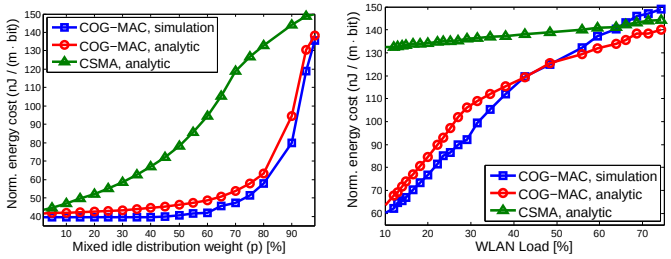
packet transmission, and therefore can increase the energy cost, as demonstrated in Fig. 9(a). Fig. 9(b), however, shows that for high p values and low network load synchronization gaps may slightly improve protocol performance. The time-shift of the TR and RR sensing times increases the chance of detecting a WLAN transmission after undetected active period and back-off time, and consequently increases the probability that the transmission happens in WLAN white space. All in all, synchronization offsets in the order of 100μsec have only a slight impact on the protocol performance.

8 A Simulation Study of COG-MAC

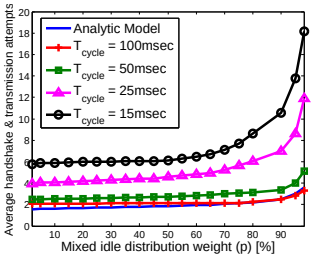
The model-based evaluation in Section 7 is subject to the two assumptions, *i)* the WLAN sources are uniformly distributed around the TR, and *ii)* the consecutive handshake and packet transmission attempts in COG-MAC observe independent WLAN channel status. In addition, to simplify the derived analytical expressions, we made three approximations in the model, in (10), (37) and (45). We present, here, a simulation study of COG-MAC, where the above assumptions and approximations are removed to see their effect on the protocol performance. Moreover, we evaluate COG-MAC in a multihop WSN.

8.1 Implementation and simulation scenario

We simulate the coexisting networks in the NS-Miracle framework [36]. The 802.11b-compliant NS-Miracle module is used for the WLAN nodes. For the WSN COG-MAC module we implemented model estimation, sensing, access control and packet reception model, as described in Sections 3,5. WSN packet losses trigger retransmissions, occurring at consecutive duty-cycles of 50msec length. We consider a single WLAN AP area with a limited set of wireless terminals (WT), operating in



(a) $\bar{I}^{(\mathrm{WS})} = 60$msec (b) $p = 0.8$



(c)

Figure 10: (a), (b) Comparison between numerical and simulation results for COG-MAC under various p, σ parametrization sets and (c) average handshake and transmission attempts with respect to increasing percentage of back-off periods, parameterized by the WSN duty-cycle duration.

the high SNR regime. We inject WLAN traffic by generating a packet stream that creates a sequence of idle and active periods that follow the proposed parameterized Global View model, and assign the packets to the WTs and the AP independently at random. To simulate a practical case we allocate 50% of the injected packets to the AP, assigning the rest, uniformly, at the WTs. We consider saturated buffer at the WSN TRs to minimize simulation time.

8.2 Model validation

We perform controlled experiments for a set of WLAN traffic parameter pairs $(p, E\left[I^{(\mathrm{WS})}\right])$ as follows. For each experiment we first place a single RR sensor uniformly at random at a distance lower than R_{CCA} from the AP and deploy 10 WTs uniformly at random outside the $\mathcal{A}^R_{\mathrm{CCA}}$. Thus the RR observes 50% of the WLAN traffic ($p^R_{\mathrm{CCA}} = 0.5$). We determine the optimal distance r^* and the packet size for possible $\hat{p}^T_{\mathrm{CCA}}$ values by the optimal solution of (13), and place the TR randomly on circle with radius r^* around the RR. Clearly, this topology leads to non-uniform WLAN source distribution around TR. To achieve statistically meaningful averages we randomize the location of the RR and the 10 deployed WTs for 100 simulation runs within each experiment. Each simulation run terminates when

the TR sensor completes the transmission of 500 packets, or, alternatively, when the simulation time exceeds 1000 seconds. The energy cost is calculated based on the number of the handshake and transmission attempts that each packet experiences.

Fig. 10 compares the simulation and analytic results for the normalized energy cost with respect to p (Fig. 10(a)) and with respect to increasing WLAN load ρ, by decreasing the expected white space durations (Fig. 10(b)). Despite a slight overestimation of the communication energy cost by the analytic model, we can conclude that the model sufficiently captures the performance of COG-MAC. The analytically evaluated performance of CSMA is also plotted for the sake of comparison.

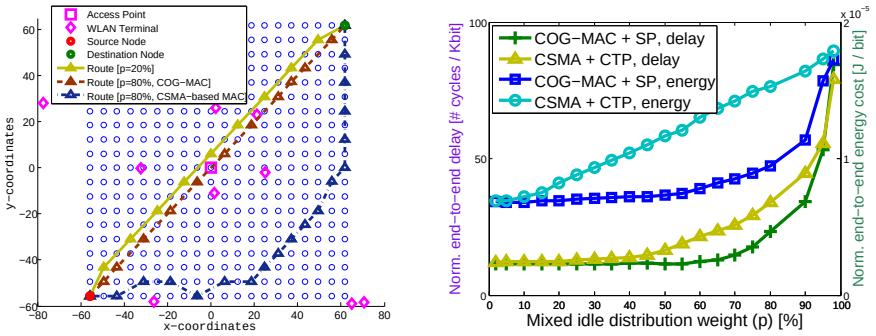
Fig. 10(c) shows the effect of the duration of the WSN duty-cycle on COG-MAC performance, and thus evaluates the modeling assumption of independent WLAN channel status at the consecutive dual sensing events. We consider the average per packet handshake and transmission attempts for duty-cycle periods from 15ms to 100ms. The WLAN load increases with p, while $E[I^{\mathrm{WS}}] = 36$msec. As the duty-cycle duration increases above 50msec, the simulated performance matches closely the analytic one, and for 100ms the performance difference is negligible, that is, the time between successive handshake or transmission attempts is long enough for COG-MAC to experience uncorrelated channel conditions. As the practical WSN duty-cycles are usually much longer than the ones we consider, we conclude that the assumption of independent WLAN status during consecutive COG-MAC cycles is practical.

8.3 Multihop COG-MAC performance analysis

Finally, we study the impact of COG-MAC on the energy efficiency of multihop WSN communication under WLAN interference. COG-MAC with energy optimized shortest path (SP) routing is compared to a benchmark solution with CSMA/CA and the widely accepted Collection Tree Protocol (CTP) [37], that finds the shortest path with the expected number of required transmissions per packet as the link weight. In COG-MAC the optimal packet sizes may differ for the links along a shortest path. To avoid the need for packet fragmentation, the packet size is chosen at the source node as the minimum of all optimal packet lengths along the path. Packet size is selected similarly for the benchmark system.

As shown in Fig. 11, we consider a square WSN grid with 5m inter-node distance, and a source and a destination node in the opposite corners. We place the WLAN AP in the center of the grid, and many of the WTs close to the AP, to generate a heterogeneous spectrum occupancy, with higher load around the center. We compare the performance of the two solutions for a constant $\bar{I}^{(\mathrm{WS})}$ and increasing p value, i.e., increasing WLAN load. Fig. 11(a) shows the transmission paths for two case studies with $p = 0.2$ and $p = 0.8$, respectively. For low p value, that is, low WLAN load the shortest paths are identical and traverse along the line connecting the source and destination node. For high load, however, the CSMA based solution needs to avoid the area around the AP, and redirects the transmission path to the

borders, where the WLAN interference is lower. At the same time COG-MAC can safely transmit along the diagonal. Fig. 11(b) gives the normalized energy and delivery delay per transmitted bit over the source-destination transmission path. The COG-MAC based solution outperforms the benchmark system, particularly when the back-off period percentage increases but the WLAN load is still moderate. We can conclude that COG-MAC leads to significant energy savings and lower delays in multihop WSNs, and, additionally, to optimal routes that are insensitive to WLAN load changes.



(a) (b)

Figure 11: (a) Topology and transmission paths for COG-MAC and CSMA-based cross-layer schemes under different percentage of WLAN back-off periods, p. (b) Normalized end-to-end energy cost per bit and end-to-end delay with respect to p. $\bar{I}^{(\mathrm{WS})} = 36ms$.

9 Conclusion

In this paper we proposed COG-MAC, a cognitive MAC scheme for energy efficient WSN operation under WLAN coexistence. The proposed scheme is based on controlling the interference from the coexisting WLAN by predicting its behavior with a smart channel sensing mechanism that takes into consideration the WLAN channel usage model. Energy cost minimization is achieved by optimizing the WSN single-hop transmission distance and packet length, based on the estimated parameters of the WLAN channel usage model. To solve the optimization problem we derived an analytic model for the successful single-hop WSN packet communication. Through numerical evaluation we showed that COG-MAC significantly outperforms other MAC protocols, especially in case of severe WLAN interference. The evaluation also revealed that both COG-MAC optimization of packet size and transmission distance and smart channel sensing are key mechanisms for increasing energy efficiency. We also presented simulation results to demonstrate the accuracy of the analytic model and to show that COG-MAC achieves significant gains even in multihop environment. Consequently, COG-MAC provides a distributed solution, that

exploits existing functionalities available in current commercial sensor hardware, and archives energy-efficient communications in the presence of coexisting WLAN networks.

References

[1] S. Pollin, I. Tan, B. Hodge, C. Chun, and A. Bahai, "Harmful coexistence between 802.15.4 and 802.11: A measurement-based study," in *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, pp. 1–6, May 2008.

[2] S. Geirhofer, L. Tong, and B. Sadler, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Communications Magazine*, vol. 45, pp. 66 –72, May 2007.

[3] J. Al-Karaki and A. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 11, pp. 6 – 28, Dec. 2004.

[4] A. Bachir, M. Dohler, T. Watteyne, and K. Leung, "MAC essentials for Wireless Sensor Networks," *IEEE Communications Surveys Tutorials*, vol. 12, no. 2, pp. 222–248, 2010.

[5] P. Lin, C. Qiao, and X. Wang, "Medium access control with a dynamic duty cycle for sensor networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 3, March 2004.

[6] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: a short Preamble MAC Protocol for Duty-Cycled Wireless Sensor Networks," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems*, SenSys'06, pp. 307–320, ACM, 2006.

[7] W. Ye, F. Silva, and J. Heidemann, "Ultra-low duty cycle MAC with scheduled channel polling," in *Proceedings of the 4th international conference on Embedded networked sensor systems*, SenSys'06, pp. 321–334, ACM, 2006.

[8] M. Di Francesco, G. Anastasi, M. Conti, S. Das, and V. Neri, "Reliability and Energy-Efficiency in IEEE 802.15.4/ZigBee Sensor Networks: An Adaptive and Cross-Layer Approach," *IEEE Journal on Selected Areas in Communications*, vol. 29, pp. 1508–1524, Sep. 2011.

[9] A. Camillò, M. Nati, C. Petrioli, M. Rossi, and M. Zorzi, "Iris: Integrated data gathering and interest dissemination system for wireless sensor networks," *Ad Hoc Networks*, vol. 11, no. 2, pp. 654–671, 2013.

[10] C. Petrioli, D. Spenza, P. Tommasino, and A. Trifiletti, "A novel wake-up receiver with addressing capability for wireless sensor nodes," in *Distributed Computing in Sensor Systems (DCOSS), 2014 IEEE International Conference on*, pp. 18–25, May 2014.

[11] L. Lo Bello and E. Toscano, "Coexistence Issues of Multiple Co-located IEEE 802.15.4/ZigBee Networks Running on Adjacent Radio Channels in Industrial Environments," *IEEE Transactions on Industrial Informatics*, vol. 5, pp. 157–167, May 2009.

[12] L. Angrisani, M. Bertocco, D. Fortin, and A. Sona, "Experimental study of coexistence issues between IEEE 802.11b and IEEE 802.15.4 Wireless Networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, pp. 1514–1523, Aug. 2008.

[13] S. Y. Shin, H. S. Park, S. Choi, and W. H. Kwon, "Packet error rate analysis of ZigBee under WLAN and Bluetooth Interferences," *IEEE Transactions on Wireless Comm.*, vol. 6, no. 8, pp. 2825–2830, 2007.

[14] M. Hanninen, J. Suhonen, T. Hamalainen, and M. Hannikainen, "Link quality-based channel selection for resource constrained WSNs," *Springer Advances in Grid and Pervasive Computing*, vol. 6646, pp. 254–263, 2011.

[15] K. il Hwang, S.-S. Yeo, and J. H. Park, "Adaptive multi-channel utilization scheme for coexistence of IEEE 802.15.4 LR-WPAN with other interfering systems," in *Proceedings of the 11th IEEE International Conference on High Performance Computing and Communications, 2009. HPCC'09*, pp. 297–304, Jun. 2009.

[16] J. Ansari and P. Mähönen, "Channel selection in spectrum agile and cognitive mac protocols for wireless sensor networks," in *Proceedings of the 8th ACM international workshop on Mobility management and wireless access*, ACM MobiWac'10, pp. 83–90, ACM, 2010.

[17] H. Rahul, N. Kushman, D. Katabi, C. Sodini, and F. Edalat, "Learning to share: Narrowband-friendly wideband networks," *SIGCOMM Comput. Commun. Rev.*, vol. 38, pp. 147–158, Aug. 2008.

[18] K. Chowdhury and I. Akyildiz, "Interferer Classification, Channel Selection and Transmission Adaptation for Wireless Sensor Networks," in *Proceedings of IEEE International Conference on Communications*, pp. 1–5, June 2009.

[19] Y. Wu, G. Zhou, and J. Stankovic, "ACR: Active Collision Recovery in Dense Wireless Sensor Networks," in *Proceedings of IEEE International Conference on Computer Communications*, pp. 1–9, Mar. 2010.

[20] C. Liang, N. Priyantha, J. Liu, and A. Terzis, "Surviving Wi-Fi interference in low power ZigBee networks," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pp. 309–322, 2010.

[21] P. Rathod, O. Dabeer, A. Karandikar, and A. Sahoo, "Characterizing the exit process of a non-saturated ieee 802.11 wireless network," in *MobiHoc*, pp. 249–258, 2009.

[22] J. Misic and V. Misic, "Characterization of idle periods in IEEE 802.11e networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1004–1009, Mar. 2011.

[23] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive medium access: Constraining interference based on experimental models," *IEEE Selected Areas in Communications*, vol. 26, no. 1, 2008.

[24] L. Stabellini, "Quantifying and modeling spectrum opportunities in a real wireless environment," in *Proceedings of Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Apr. 2010.

[25] I. Glaropoulos, A. Vizcaino Luna, V. Fodor, and M. Papadopouli, "Closing the gap between traffic workload and channel occupancy models for 802.11 networks," *Ad Hoc Netw.*, vol. 21, pp. 60–83, Oct. 2014.

[26] J. Huang, G. Xing, G. Zhou, and R. Zhou, "Beyond Co-existence; Exploiting WiFi White Space for ZigBee Performance Assurance," in *Proceedings of IEEE International Conference on Netw. Protocols*, 2010.

[27] S. Geirhofer, J. Z. Sun, L. Tong, and B. M. Sadler, "Cognitive frequency hopping based on interference prediction: theory and experimental results," *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 13, pp. 49–61, Sep. 2009.

[28] I. Glaropoulos, V. Fodor, L. Pescosolido, and C. Petrioli, "Cognitive WSN transmission control for energy efficiency under WLAN coexistence," in *Proceedings of the 6th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, Jun. 2011.

[29] X. Liu, S. Zhang, J. Wang, J. Cao, and B. Xiao, "Anchor supervised distance estimation in anisotropic wireless sensor networks," in *Proceedings of IEEE Wireless Communications and Networking Conference*, pp. 938–943, Mar. 2011.

[30] "2.4GHz IEEE 802.15.4 / ZigBee-ready RF Transceiver," tech. rep., Texas Instruments, 2008.

[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[32] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proceedings of the 1st IEEE International Symposium on new Frontiers in Dynamic Spectrum Access Networks (DySPAN'05)*, pp. 131–136, Nov. 2005.

[33] I. Glaropoulos, "Energy efficient cognitive mac for sensor networks under wlan co-existence - Revised complementary technical report," tech. rep., KTH, Royal Institue of Technology, January 2015.

[34] M. Lagana, I. Glaropoulos, V. Fodor, and C. Petrioli, "Modeling and estimation of partially observed wlan activity for cognitive wsns," in *IEEE Wireless Communications and Networking Conference*, pp. 1526 –1531, Apr. 2012.

[35] I. Glaropoulos and V. Fodor, "Discrete stochastic optimization based parameter estimation for modeling partially observed WLAN spectrum activity," *Infocommunications Journal*, vol. 4, no. 2, pp. 11–17, 2012.

[36] *NS-Miracle: Multi-InteRfAce Cross-Layer Extension library for the Network Simulator*. http://telecom.dei.unipd.it/pages/read/58/.

[37] O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis, "Collection tree protocol," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, SenSys'09, pp. 1–14, ACM, 2009.

Paper C

Discrete stochastic optimization based parameter estimation for modeling partially observed WLAN spectrum activity

Ioannis Glaropoulos and Viktoria Fodor

Published in Infocommunications Journal, 2012.

Discrete Stochastic Optimization Based Parameter Estimation for Modeling Partially Observed WLAN Spectrum Activity

Ioannis Glaropoulos and Viktoria Fodor *

Abstract

Modeling and parameter estimation of spectrum usage in the ISM band would allow the competing networking technologies to adjust their medium access control accordingly, leading to the more efficient use of the shared spectrum. In this paper we address the problem of WLAN spectrum activity model parameter estimation. We propose a solution based on discrete stochastic optimization, that allows accurate spectrum activity modeling and can be implemented even in wireless sensor nodes with limited computational and energy resources.

Index terms— cognitive networks, WLAN spectrum activity, discrete stochastic optimization

1 Introduction

Emerging wireless technologies for local and personal area communication all use the open Industrial, Scientific and Medical (ISM) band. While the variety of introduced solutions increases, the protocol stacks are usually optimized for a given application area, and at the same time assume the exclusive use of the spectrum space. However, most of the time the different technologies coexist, and communication efficiency and performance guarantees can only be achieved, if the networks have cognitive capabilities [1], that is, they are aware of each other and optimize their transmission parameters and communication protocols accordingly.

Key technologies operating in the ISM band are the IEEE 802.11 wireless local area networks (WLANs). As WLAN carrier sensing is designed to detect WLAN signals, it is *blind* towards the low power, narrow band WSN transmissions. Consequently, if the WSN does not adjust itself to the WLAN operation, it will experience

*I Glaropoulos and V. Fodor is with the Access Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. Email: {ioannisg,vfodor}@kth.se.

harmful interference from the WLAN, while the WLAN itself is not affected significantly by the narrow band low power WSN interferers.

Previous work in the area of cognitive WSNs includes proposals for novel carrier sensing and medium access control, and the characterization of the channel usage in WLAN cells. In [2] the interfering technology is identified based on spectral signature. In the case of WLAN interferers, the sensors force the WLAN to back off by sending short, high power jamming signals. The POMDP framework [3] introduces the concept of partial channel knowledge and proposes optimal sensing and channel access strategies considering a Markovian channel occupancy model. A Markovian model, however, may lead to suboptimal WSN operation, and therefore several works deal with a more accurate channel characterization, considering subgeometric [4], hyper-exponential [5] and Pareto [6] idle time distributions.

In [7][8] it is recognized, that the characterization of the idle time can lead to more efficient cognitive access control, if it captures the two basic sources of WLAN inactivity, the short, almost uniformly distributed contention windows and the long, heavy-tailed white space periods, when the WLAN users are inactive. We follow this approach in our previous work, where we propose cognitive medium access control and next hop selection for the WSN [9], given the known WLAN channel idle time distribution. In [10] we define the *Local View* model of WLAN channel activity that extends the solution of [7] and takes into account the limited detection range of the WSN nodes, and propose computationally efficient ways to estimate the model parameters based on time limited continuous sensing at the sensors.

In this paper we provide a deep analysis of the Local View parameter estimation based on discrete stochastic optimization. We follow the approach presented in [11], show that the algorithm converges almost surely to the optimal parameter set, and evaluate how the size of the state space, the size of the sample set and the number of iterations affect the estimation accuracy.

The rest of the paper is organized as follows. Section 2 defines the considered networking scenario along with the WLAN channel activity models and formulates the parameter estimation as an optimization problem. In Section 3 we give an overview of the discrete stochastic optimization algorithm proposed in [11]. In Section 4 we show that the algorithm converges in the case of the considered parameter estimation problem and in 5 we evaluate the performance of the algorithm under practical constraints. We conclude the paper in Section 6.

2 WLAN Idle Time Modeling

We consider an IEEE 802.15.4 compliant WSN operating in the transmission area of an IEEE 802.11 WLAN. The transmission power of the WLAN terminals is orders of magnitude higher than that of the coexisting WSN, and the WLAN terminals are *blind* towards the WSN transmissions. The protocol stack of the energy constrained WSN is enhanced by cognitive functionality to optimize the WSN operation. To

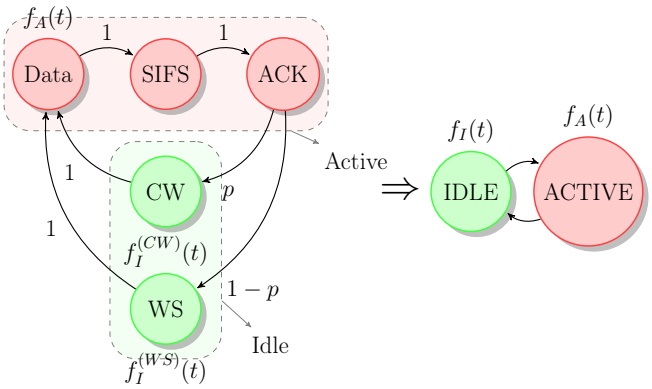


Figure 1: The Global View model with all channel states and the reduced two-state semi-Markovian model.

perform cognitive control, the WSN needs to know the WLAN channel occupancy distribution. For this, the sensors perform continuous sensing and collect samples of busy and idle WLAN period lengths. The sensing is based on the usual Clear Channel Assessment (CCA) process with energy detection, resulting in a limited sensing range.

According to [7][8], the *Global View* of WLAN channel occupancy can be modeled by a semi-Markovian system of Active and Idle periods. Figure 1 depicts all the states of the WLAN channel and their merging into a two-state semi-Markovian chain. The states of Data, SIFS and ACK transmission are grouped together into a single *Active* state, while the states that represent the WLAN Contention Window period (CW) and the WLAN White Space (WS) due to user inactivity are merged into a single *Idle* state. The sojourn times in the Active state can be modeled by the *uniform* distribution $f_A(t)$ within $[\alpha_{\mathrm{ON}}, \beta_{\mathrm{ON}}]$. The idle period distribution, $f_I(t)$, is a *mixture* distribution with a weight p, that is $f_I(t) \triangleq p \cdot f_I^{CW}(t) + (1-p) \cdot f_I^{WS}(t)$. $f_I^{CW}(t)$ is the distribution of the CW periods, and can be modeled with a *uniform* distribution within $[0, \alpha_{\mathrm{BK}}]$. The WS periods, however, exhibit a heavy-tailed behavior, and their distribution $f_I^{WS}(t)$ is well approximated by a zero-location *generalized Pareto* distribution with parameters (ξ, σ).

Thus, the distribution of the sojourn time in the Idle state, $f_I(t)$, is given as:

$$f_I(t) \triangleq \begin{cases} p \cdot \dfrac{1}{\alpha_{\mathrm{BK}}} + (1-p) \cdot \dfrac{1}{\sigma} \left(1 + \xi \dfrac{t}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)} & t \leq \alpha_{\mathrm{BK}} \\ (1-p) \dfrac{1}{\sigma} \left(1 + \xi \dfrac{t}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)} & t > \alpha_{\mathrm{BK}} \end{cases}$$

This Global View, however, is not fully observable at the individual sensors, that can detect WLAN transmissions only within a given detection range. Therefore, in

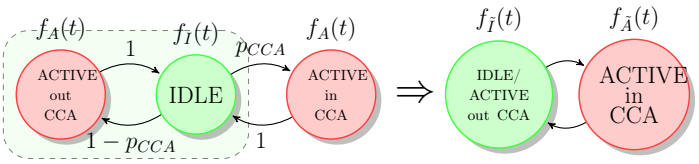


Figure 2: The 3-state semi-Markovian chain and its 2-state equivalent for the Local View channel activity modeling.

[10] we define the *Local View*, that describes the WLAN channel occupancy as seen by an individual sensor. Assuming that consecutive WLAN transmissions are not correlated, we introduce a 3-state semi-Markovian system (Figure 2), distinguishing between detected, and un-detected WLAN activity, that occurs with probabilities p_{CCA} and $(1 - p_{\text{CCA}})$, respectively. To model the *observable* sojourn time distributions $f_{\tilde{A}}(t)$ and $f_{\tilde{I}}(t)$ we define the 2-state *Local View* model by merging the states at which the sensor detects an idle channel. It holds that $f_{\tilde{A}}(t) = f_A(t)$, but $f_{\tilde{I}}(t) \neq f_I(t), \quad \forall p_{\text{CCA}} < 1$.

Our objective is to estimate the parameters of $f_A(t)$ and $f_I(t)$ and the *observable load*, p_{CCA}, from a set of samples of $f_{\tilde{A}}(t)$ and $f_{\tilde{I}}(t)$ obtained through channel sensing.

As the active period distribution, $f_{\tilde{A}}(t)$, is uniform, its parameters α_{ON} and β_{ON} are estimated by the lowest and the largest measured active period according to Maximum Likelihood Estimation (MLE) and the maximum back-off time, α_{BK}, is given by the WLAN specification. The estimation of the rest of the parameters is more difficult. An idle channel period observed by an arbitrary sensor consists of a random number of WLAN "cycles", that is, consecutive idle and un-detected active periods, followed by an additional idle period. The locally observable idle period distribution, $f_{\tilde{I}}(t)$, is, therefore, a function of the idle and active time distributions $f_I(t)$ and $f_A(t)$, and of the observable load, p_{CCA}, and can not be expressed in a closed form, even if $f_A(t)$ and $f_I^{CW}(t)$ are known.

As we show in [10], closed form expression exists in the Laplace domain and, therefore, we propose to estimate the parameters of $f_{\tilde{I}}(t)$ in the Laplace domain. Since according to the semi-Markovian Local View model the number of consecutive WLAN cycles is geometrically distributed, the Laplace Transform (LT) of $f_{\tilde{I}}(t)$ obtains the following form:

$$f_{\tilde{I}}^*(s) = f_I^*(s) \frac{p_{\text{CCA}}}{1 - (1 - p_{\text{CCA}}) f_I^*(s) f_A^*(s)}, \tag{1}$$

where $f_I^*(s), f_A^*(s)$ denote the LT of $f_I(t), f_A(t)$, respectively.

3 An Algorithm for Discrete Stochastic Optimization for Parameter Estimation

In this Section we review the algorithm for stochastic optimization introduced in [11], that we use to estimate the parameters of the Local View model. First we define the necessary notation and then we give the stochastic optimization algorithm, along with the constraint on convergence.

Let us define by $\boldsymbol{\mathcal{K}} \triangleq \{\mathcal{K}_1, \mathcal{K}_2, ..., \mathcal{K}_K\}$ the discrete space of the different alternatives. The number of discrete states, $K = |\boldsymbol{\mathcal{K}}|$, is finite. The optimization problem we aim at solving is of the following form:

$$\mathcal{K}^* = \arg \min_{\mathcal{K}_n \in \boldsymbol{\mathcal{K}}} \{c(n) = E[X_{\mathcal{K}_n}]\}. \tag{2}$$

That is, the function $c(n)$ can not be evaluated analytically and needs to be estimated through a sequence of random samples $\{X_{\mathcal{K}_n}\}$. We denote by:

$$\boldsymbol{\mathcal{L}} = \{\mathcal{L}_1, ..., \mathcal{L}_L\} \subset \boldsymbol{\mathcal{K}} \tag{3}$$

the set of *global* minimizers of the function c, that is:

$$\begin{array}{l} \forall \mathcal{L}_i \in \boldsymbol{\mathcal{L}}, \mathcal{K}_n \in \boldsymbol{\mathcal{K}} \setminus \boldsymbol{\mathcal{L}},\ c(\mathcal{L}_i) < c(\mathcal{K}_n) \text{ and} \\ \forall i, j = 1, 2, ..., L,\ c(\mathcal{L}_i) = c(\mathcal{L}_j). \end{array} \tag{4}$$

In the following we give the original stochastic optimization algorithm as it is proposed in [11] (Algorithm 1). The search process starts from an arbitrary state, $\mathcal{K}_i$. In each iteration step, m, it selects a new state $\mathcal{K}_j$ uniformly at random and obtains the observation of a random variable $Z_{l_m}^{\mathcal{K}_i \to \mathcal{K}_j}$ to compare the two states. The value of this random variable can depend on the two states and on l_m, which is a function of the iteration step m. The algorithm moves to the new state if $Z_{l_m}^{\mathcal{K}_i \to \mathcal{K}_j} > 0$.

Let denote $\mathcal{K}_m$ the state after iteration m and $Q_m(\mathcal{K}_n)$ the "popularity" of state $\mathcal{K}_n \in \boldsymbol{\mathcal{K}}$, i.e. the number of times the algorithm has visited state $\mathcal{K}_n$ until iteration m. The output of the algorithm, $\mathcal{K}^*$, is chosen as the most visited state.

It is shown in [11] that the algorithm converges almost surely to a minimizer, i.e. a member of $\boldsymbol{\mathcal{L}}$, after sufficiently large number of iterations, if the following conditions hold:

Condition 1. *For each $\mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}$ and $l \in \mathbb{N}$, there exists a random variable $Z_l^{(\mathcal{K}_i \to \mathcal{K}_j)}$ such that the limit $\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i \to \mathcal{K}_j)} > 0\}$ exists for all $\mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}$ and for all $\mathcal{K}_i \in \boldsymbol{\mathcal{L}}, \mathcal{K}_j \notin \boldsymbol{\mathcal{L}}, \mathcal{K}_n \neq \mathcal{K}_i, \mathcal{K}_j$, and $l \in \mathbb{N}$,*

$$\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_j \to \mathcal{K}_i)} > 0\} > \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i \to \mathcal{K}_j)} > 0\}, \tag{5}$$

$$\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_n \to \mathcal{K}_i)} > 0\} \geq \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_n \to \mathcal{K}_j)} > 0\}, \tag{6}$$

Algorithm 1 A global search for discrete stochastic optimization [11].

Step 0:

Select a starting point $\mathcal{K}_0 \in \boldsymbol{\mathcal{K}}$.
$Q_0(\mathcal{K}_0) \leftarrow 1$ and $Q_0(\mathcal{K}_n) \leftarrow 0, \forall \mathcal{K}_n \in \boldsymbol{\mathcal{K}}, \mathcal{K}_n \neq \mathcal{K}_0$.
$m \leftarrow 0$ and $\mathcal{K}_m^* \leftarrow \mathcal{K}_0$. Go to Step 1.

Step 1:

Generate a uniform random variable $\mathcal{J}_m$ such that for all $\mathcal{K}_n \in \boldsymbol{\mathcal{K}}, \mathcal{K}_n \neq \mathcal{K}_m$, $\mathcal{J}_m \leftarrow \mathcal{K}_n$ with probability $\frac{1}{K-1}$. Go to Step 2.

Step 2:

Generate an observation R_m of $Z_{l_m}^{\mathcal{K}_m \rightarrow \mathcal{J}_m}$.
if $R_m > 0$ **then**
  $\mathcal{K}_{m+1} \leftarrow \mathcal{J}_m$.
else
  $\mathcal{K}_{m+1} \leftarrow \mathcal{K}_m$.
end if Go to Step 3.

Step 3:

$m \leftarrow m+1$, $Q_m(\mathcal{K}_m) \leftarrow Q_{m-1}(\mathcal{K}_m) + 1$ and $Q_m(\mathcal{K}_n) \leftarrow Q_{m-1}(\mathcal{K}_n)$ for all $\mathcal{K}_n \neq \mathcal{K}_m$.
if $Q_m(\mathcal{K}_m) > Q_m(\mathcal{K}_{m-1}^*)$ **then**
  $\mathcal{K}_m^* \leftarrow \mathcal{K}_m$.
else
  $\mathcal{K}_m^* \leftarrow \mathcal{K}_{m-1}^*$
end if Go to Step 1.

$$\lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i\to\mathcal{K}_n)} \le 0\} \ge \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_j\to\mathcal{K}_n)} \le 0\}. \quad (7)$$

Condition 2. *$\{l_m\}$ is a sequence of positive integers such that $l_m \to \infty$ as $m \to \infty$.*

Condition 3. *The Markov matrix $\mathcal{P}$ defined in the following equations is irreducible.*

$$\mathcal{P}(\mathcal{K}_i, \mathcal{K}_j) = \tfrac{1}{K-1} \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i\to\mathcal{K}_j)} > 0\}$$
$$\forall \mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}, \mathcal{K}_i \neq \mathcal{K}_j,$$
$$\mathcal{P}(\mathcal{K}_i, \mathcal{K}_i) = \tfrac{1}{K-1} \sum_{\mathcal{K}_j\in\boldsymbol{\mathcal{K}}\setminus\{\mathcal{K}_i\}} \lim_{l\to\infty} P\{Z_l^{(\mathcal{K}_i\to\mathcal{K}_j)} \le 0\}$$
$$\forall \mathcal{K}_i \in \boldsymbol{\mathcal{K}}.$$

4 Local View Parameter Estimation

4.1 The Estimation Process

We apply Algorithm 1 to estimate the parameters of $f_{\tilde{I}}(t)$. We discretize the model parameters ξ, σ and p within the reasonable intervals, and define the state $\mathcal{K}_n$ as the set of model parameters:

$$\mathcal{K}_n \triangleq (\xi_n, \sigma_n, p_n)\,.$$

The fourth model parameter, p_{CCA}, is uniquely determined through Moment Evaluation (ME) by the sample mean of the observable idle periods ($\bar{\mu}$) and the rest of the parameters, i.e.

$$p_{\mathrm{CCA}} = \frac{\frac{p\alpha_{\mathrm{BK}}}{2} + \frac{(1-p)\sigma}{1-\xi} + \frac{\alpha_{\mathrm{ON}}+\beta_{\mathrm{ON}}}{2}}{\bar{\mu} + \frac{\alpha_{\mathrm{ON}}+\beta_{\mathrm{ON}}}{2}}.$$

We would like to determine the optimal state, $\mathcal{K}^* \in \boldsymbol{\mathcal{K}}$, that is the optimal model parameter set $\mathcal{K}^* \triangleq (\xi^*, \sigma^*, p^*)$, that minimizes the Mean Square Error (MSE) between the Laplace transform of the idle distribution, $f_{\tilde{I}}^*(s)$, and the LT given by the system state, $f_{\tilde{I}}^*(s; \mathcal{K}_n)$, over $\mathcal{S} = \{s_1, s_2, \ldots s_S\}$, the finite discrete subset of the s-domain, that is,

$$(\xi^*, \sigma^*, p^*) = \arg\min_{\mathcal{K}_n\in\boldsymbol{\mathcal{K}}} \frac{1}{S} \sum_{k=0}^{S} (f_{\tilde{I}}^*(s_k) - f_{\tilde{I}}^*(s_k; \mathcal{K}_n))^2. \quad (8)$$

As $f_{\tilde{I}}^*(s)$ is not known, it needs to be evaluated through the idle period samples obtained by channel sensing.

To ensure fast parameter estimation, we propose to run the estimation process, that is, Algorithm 1 parallel to the channel sensing. That is, in each iteration step,

m, n_m new idle period samples are integrated in the empirical LT. The total number of samples integrated up to iteration m is $N_m = \sum_{k=0}^{m} n_m$. We define the *empirical* LT, $f^*_{\tilde{I}e}(s; N_m)$, of the observable idle time distribution directly from a set of N_m measured idle period samples, $(t_1, ..., t_{N_m})$ as:

$$f^*_{\tilde{I}e}(s; N_m) = \frac{1}{N_m} \sum_{i=1}^{N_m} e^{-st_i}. \tag{9}$$

Comparing the general expression in (2) and the Mean Square Error minimization problem in (8) we have:

$$X_{\mathcal{K}_n} = \mathrm{MSE}_n^{(N)} = \frac{1}{S} \sum_{k=0}^{S} \left[f^*_{Ie}(s_k; N) - f^*_I(s_k; \mathcal{K}_n) \right]^2, \forall \mathcal{K}_n, \tag{10}$$

where $\mathrm{MSE}_n^{(N)}$ denotes the MSE calculated with the N samples, and, consequently,

$$\begin{aligned} c(n) &= E\left[\tfrac{1}{S} \textstyle\sum_{k=0}^{S} [(f^*_{Ie}(s; N) - f^*_I(s; \mathcal{K}_n))^2] \right] = \\ &= \tfrac{1}{S} \textstyle\sum_{k=0}^{S} E[(f^*_{Ie}(s; N) - f^*_I(s; \mathcal{K}_n))^2]. \end{aligned} \tag{11}$$

Accordingly, for Algorithm 1 we select $l_m = N_m$ and define:

$$Z_{l_m}^{(\mathcal{K}_i \to \mathcal{K}_j)} \triangleq X_{\mathcal{K}_i} - X_{\mathcal{K}_j} = \mathrm{MSE}_{\mathcal{K}_i}^{(l_m)} - \mathrm{MSE}_{\mathcal{K}_j}^{(l_m)}, \ \forall \mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}. \tag{12}$$

That is, in our algorithm, the process moves to the new state $\mathcal{K}_j$ if the MSE is decreased this way.

4.2 On the Convergence of the Estimation Process

To prove that the proposed parameter estimation algorithm solves the optimization problem in (8), we proceed as follows. The proof that Condition 2 holds is trivial; since $\{l_m\}$ defines the number of samples that are integrated in the empirical LT calculation until step m, it is a sequence of integers that tends to ∞ as $m \to \infty$. With Lemma 1 we prove that $f^*_{\tilde{I}e}(s; N)$ is an unbiased estimator of $f^*_{\tilde{I}}(s)$, and converges to $f^*_{\tilde{I}}(s)$ as $N \to \infty$. Based on Lemma 1, we prove with Corollary 1 that the minimization of $c(n)$ solves the original problem in (8). Lemma 2 proves that the particular selection of the random variable $Z_{l_m}^{(\mathcal{K}_i \to \mathcal{K}_j)}, \forall \mathcal{K}_i, \mathcal{K}_j \in \boldsymbol{\mathcal{K}}$ satisfies Condition 1. Lemma 3 shows that in our problem Algorithm 1 converges to the optimal state, bypassing the requirement for Condition 3 to hold.

Lemma 1. *The empirical Laplace Transform as a function of N i.i.d. samples $\{t_1, ..., t_N\}$ can be approximated as*

$$f^*_{Ie}(s; N) = \frac{1}{N} \sum_{l=1}^{N} e^{-st_l}.$$

and is an unbiased estimator of $f^*(s)$, converging to $f^*(s)$ as $N \to \infty$.

Proof. We, first, generate the empirical distribution function, $F_e(t;N)$ based on the N collected time period samples, $T = \{t_1, ..., t_N\}$,

$$F_e(t;N) = \frac{\#\text{samples in } T \leq t}{N}.$$

It is known that $F_e(t;N)$ converges almost surely to the actual CDF, $F(t)$, as $N \to \infty$, based on the strong law of large numbers. In addition, $F_e(t;N)$ is an unbiased estimator of $F(t)$, i.e. $E[F_e(t;N)] = F(t)$. $F_e(t_l) - F_e(t_{l-1})$ is, then, an unbiased estimator for $P\{t \in (t_{l-1}, t_l)\} = F(t_l) - F(t_{l-1}), \quad l = 1, 2, ..., N$. We define the empirical density function, $f_e(t;N)$ being non-zero only on the set T, as follows:

$$f_e(t;N) = \sum_{l=1}^{N} \left[F_e(t_l;N) - F_e(t_{l-1};N)\right] \delta(t - t_l),$$

where $\delta(t)$ is the Dirac function and by convention $t_0 = 0, F_e(t_0) = 0$. Clearly,

$$\lim_{N\to\infty, t_l - t_{l-1} \to dt_l} F_e(t_l;N) - F_e(t_{l-1};N) = f(t_l)dt_l$$

and so

$$\begin{aligned} &\lim_{N\to\infty} \textstyle\sum_{l=1}^{N} \left[F_e(t_l;N) - F_e(t_{l-1};N)\right] \delta(t - t_l) = \\ &\quad = \textstyle\int_{t_l} f(t_l)\delta(t - t_l)dt_l = f(t), \end{aligned}$$

consequently $f_e(t;N)$ converges to the actual pdf. The empirical Laplace Transform is defined as

$$f_e^*(s;N) \triangleq \int_0^\infty f_e(t;N)e^{-st}dt.$$

Since $F_e(t_l) - F_e(t_{l-1}) = 1/N$, the above becomes

$$\begin{aligned} &f_e^*(s;N) = \textstyle\int_0^\infty \sum_{l=1}^{N} \frac{1}{N}\delta(t - t_l)e^{-st}dt = \\ &\quad = \textstyle\sum_{l=1}^{N} \frac{1}{N} \int_0^\infty \delta(t - t_l)e^{-st}dt = \frac{1}{N}\sum_{l=1}^{N} e^{-st_l}. \end{aligned}$$

The convergence of $f^*(s;N)$ is ensured due to the convergence of $f_e(t;N)$. Finally,

$$E[f_e^*(s;N)] = \sum_{l=1}^{N} \frac{1}{N} E[e^{-st_l}] = E[e^{-st}] = f^*(s), \tag{13}$$

so $f^*(s;N)$ is an unbiased estimator of the LT transform. □

Corollary 1. *The minimization of $c(n)$ in* (11) *solves the original problem in* (8).

Proof. We have:

$$
\begin{aligned}
c(n) &= \tfrac{1}{S}\sum_{k=0}^{S} E\left[\left(f^*_{\tilde{I}e}(s_k;N) - f^*_{\tilde{I}}(s_k;\mathcal{K}_n)\right)^2\right] = \\
&= \tfrac{1}{S}\sum_{k=0}^{S} E\left[\left(f^*_{\tilde{I}e}(s_k;N)\right)^2 - 2f^*_{\tilde{I}e}(s_k;N)f^*_{\tilde{I}}(s_k;\mathcal{K}_n)\right] + \\
&\quad + \left(f^*_{\tilde{I}}(s_k;\mathcal{K}_n)\right)^2 = \\
&= \tfrac{1}{S}\sum_{k=0}^{S} E\left[\left(f^*_{\tilde{I}e}(s_k;N)\right)^2\right] - 2f^*_{\tilde{I}}(s_k)f^*_{\tilde{I}}(s_k;\mathcal{K}_n) + \\
&\quad + \left(f^*_{\tilde{I}}(s_k;\mathcal{K}_n)\right)^2.
\end{aligned}
$$

For $N \to \infty$ it holds from Lemma 1 that $\mathrm{Var}\left[f^*_{\tilde{I}e}(s_k;N)\right] = 0$, and consequently, $E\left[\left(f^*_{\tilde{I}e}(s_k;N)\right)^2\right] = E\left[f^*_{\tilde{I}e}(s_k;N)\right]^2$, so $c(n)$ converges to $\frac{1}{S}\sum_{k=0}^{S}(f^*_{\tilde{I}}(s_k) - f^*_{\tilde{I}}(s_k;\mathcal{K}_n))^2$. □

Let us now prove, that Condition 1 is satisfied.

Lemma 2. *Let us select a random variable* $Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}$ *as follows:*

$$Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)} = MSE_{\mathcal{K}_i}^{(l_m)} - MSE_{\mathcal{K}_j}^{(l_m)}.$$

The variable $Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}$ *satisfies Condition 1.*

Proof. We start with showing that (5) is satisfied. Let $\mathcal{K}_i \in \mathcal{L}$, $\mathcal{K}_j \notin \mathcal{L}$, so that

$$\frac{1}{S}\sum_{k=0}^{S}(f^*_I(s_k) - f^*_I(s_k;\mathcal{K}_i))^2 < \frac{1}{S}\sum_{k=0}^{S}(f^*_I(s_k) - f^*_I(s_k;\mathcal{K}_j))^2 \tag{14}$$

We show, first, by direct computation that the mean of $Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}$, defined in (12)

is negative.

$$
\begin{aligned}
&E[Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}] = \\
&= E[\mathrm{MSE}_{\mathcal{K}_i}^{(l_m)} - \mathrm{MSE}_{\mathcal{K}_j}^{(l_m)}] = \\
&= E[\tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_i)\right)^2 - \\
&\quad -\tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_j)\right)^2] \\
&= \tfrac{1}{S}E[\textstyle\sum_{k=0}^{S}\left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_i)\right)^2 - \\
&\quad -\left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_j)\right)^2] \\
&= \tfrac{1}{S}E[\textstyle\sum_{k=0}^{S}\left(2f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_i) - f_I^*(s_k;\mathcal{K}_j)\right)\cdot \\
&\quad \cdot\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right)] \\
&= \tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right)\cdot \\
&\quad \cdot E\left[2f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_i) - f_I^*(s_k;\mathcal{K}_j)\right] \\
&\overset{(13)}{=} \tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right)\cdot \\
&\quad \cdot\left(2f_I^*(s_k) - f_I^*(s_k;\mathcal{K}_i) - f_I^*(s_k;\mathcal{K}_j)\right) \\
&= \tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_I^*(s_k) - f_I^*(s_k;\mathcal{K}_i)\right)^2 - \left(f_I^*(s_k) - f_I^*(s_k;\mathcal{K}_j)\right)^2 \\
&\overset{(14)}{<} 0.
\end{aligned}
$$

We now show that $\lim_{l_m\to\infty} Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}$ is symmetric around its mean. We write:

$$
\begin{aligned}
&Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)} = \\
&= \tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_i)\right)^2 - \\
&\quad -\tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_j)\right)^2 \\
&= \tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(2f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_i) - f_I^*(s_k;\mathcal{K}_j)\right)\cdot \\
&\quad \cdot\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right) \\
&= \tfrac{1}{S}\textstyle\sum_{k=0}^{S}2f_{Ie}^*(s_k;l_m)\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right) - \\
&\quad -\tfrac{1}{S}\textstyle\sum_{k=0}^{S}\left(f_I^*(s_k;\mathcal{K}_j) + f_I^*(s_k;\mathcal{K}_i)\right)\cdot \\
&\quad \cdot\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right).
\end{aligned}
$$

The second term is deterministic and, thus, excluded from the calculations. We concentrate on the first term.

$$
W_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)} = \frac{1}{S}\sum_{k=0}^{S}2f_{Ie}^*(s_k;l_m)\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right) \tag{15}
$$

As shown above the empirical LT is generated as $f_{Ie}^*(s_k;l_m) = \frac{\sum_{l=1}^{l_m} e^{-s_k t_l}}{l_m}$. We

rewrite (15) as

$$\begin{aligned}
&W_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)} = \\
&= \tfrac{1}{S}\sum_{k=0}^{S} 2f_{Ie}^*(s_k;l_m)\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right) \\
&= \tfrac{1}{S}\tfrac{1}{l_m}\sum_{k=0}^{S}\sum_{l=1}^{l_m} 2e^{-s_k t_l}\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right) \\
&= \tfrac{1}{S}\sum_{l=1}^{l_m}\left[\tfrac{1}{l_m}\sum_{k=0}^{S} 2e^{-s_k t_l}\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right)\right].
\end{aligned}$$

Since the variables t_l are i.i.d. so are the variables

$$g_l^{ji} = \sum_{k=0}^{S}\frac{2}{S}e^{-s_k t_l}\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_i)\right). \tag{16}$$

As a result, the variable $\frac{1}{l_m}\sum_{l=1}^{l_m} g_l^{ji}$ approaches a Gaussian distribution, due to the Central Limit Theorem, and becomes, thus, symmetric around its mean value. Consequently, the limit $\lim_{l_m\to\infty} W_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}$ is symmetric around its mean, and so does $\lim_{l_m\to\infty} Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}$. Since, additionally, $E\left[Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)}\right] < 0$, it follows that

$$P\{Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)} < 0\} > P\{Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)} > 0\}. \tag{17}$$

From the last equation, along with $P\{Z_{l_m}^{(\mathcal{K}_i\to\mathcal{K}_j)} < 0\} = P\{Z_{l_m}^{(\mathcal{K}_j\to\mathcal{K}_i)} > 0\}$, follows Eq. (5).

We proceed with showing that (6) is satisfied. For that, we need to show, first that the variance of $Z_{l_m}^{(\mathcal{K}_n\to\mathcal{K}_j)}$ is finite. By direct computation we obtain:

$$\begin{aligned}
&\mathrm{Var}\left[Z_{l_m}^{(\mathcal{K}_n\to\mathcal{K}_j)}\right] \\
&= \mathrm{Var}\left[\mathrm{MSE}_{\mathcal{K}_n}^{(l_m)} - \mathrm{MSE}_{\mathcal{K}_j}^{(l_m)}\right] \\
&= \mathrm{Var}[\tfrac{1}{S}\sum_{k=0}^{S}\left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_n)\right)^2 \\
&\quad - \left(f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_j)\right)^2] \\
&= \mathrm{Var}[\tfrac{1}{S}\sum_{k=0}^{S}\left(2f_{Ie}^*(s_k;l_m) - f_I^*(s_k;\mathcal{K}_n) - f_I^*(s_k;\mathcal{K}_j)\right)\cdot\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_n)\right)]
\end{aligned}$$

We neglect the deterministic term in $Z_{l_m}^{(\mathcal{K}_n\to\mathcal{K}_j)}$, resulting in the expression

$$\begin{aligned}
&\mathrm{Var}\left[\tfrac{1}{S}\sum_{k=0}^{S} 2f_{Ie}^*(s_k;l_m)\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_n)\right)\right] \\
&= \mathrm{Var}\left[\tfrac{2}{l_m S}\sum_{k=0}^{S}\sum_{l=1}^{l_m} e^{-s_k t_l}\left(f_I^*(s_k;\mathcal{K}_j) - f_I^*(s_k;\mathcal{K}_n)\right)\right] \\
&= \left(\tfrac{1}{l_m}\right)^2 \mathrm{Var}\left[\sum_{l=1}^{l_m} g_l^{nj}\right],
\end{aligned}$$

where g_l^{nj} is given as in (16). For the time distribution that we consider in [10], $\mathrm{Var}[t_l]$ is finite, as a result $\mathrm{Var}[e^{-s_k t_l}]$ is finite. Consequently, the variance of g_l^{nj} is finite as as summation over non-independent variables indexed by s_k:

$$g_l^{nj}(s_k) = \frac{2}{S} e^{-s_k t_l} \left(f_I^*(s_k; \mathcal{K}_j) - f_I^*(s_k; \mathcal{K}_n) \right)$$

where

$$\begin{aligned} &\mathrm{Var}\left[g_l^{nj}(s_k) \right] = \\ &= \tfrac{4}{S^2} \left(f_I^*(s_k; \mathcal{K}_j) - f_I^*(s_k; \mathcal{K}_n) \right)^2 \mathrm{Var}[e^{-s_k t_l}] < \infty, \end{aligned}$$

$$\mathrm{Var}\left[g_l^{nj} \right] = \sum_{s_k=0}^{S} \sum_{s_o=0}^{S} \mathrm{Cov}\left[g_l^{nj}(s_k), g_l^{nj}(s_o) \right] < \infty.$$

Since $\mathrm{Var}[g_l^{nj}] < \infty$, and the g_l^{nj} are i.i.d., it is clear that

$$\lim_{l_m \to \infty} \frac{1}{l_m^2} \mathrm{Var}\left[\sum_{l=1}^{l_m} g_l^{nj} \right] = 0.$$

Consequently,

$$\lim_{l_m \to \infty} \mathrm{Var}\left[Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_j)} \right] = 0. \tag{18}$$

For $i, n \in \boldsymbol{\mathcal{L}}$ it, then, holds

$$\lim_{l_m \to \infty} P\{ Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_i)} > 0 \} \geq \lim_{l_m \to \infty} P\{ Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_j)} > 0 \} = 0,$$

since

$$E[Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_j)}] < 0.$$

For $n \notin \boldsymbol{\mathcal{L}}$

$$\lim_{l_m \to \infty} P\{ Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_i)} > 0 \} = 1 \geq \lim_{l_m \to \infty} P\{ Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_j)} > 0 \}, \tag{19}$$

since

$$E[Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_i)}] > 0.$$

This proves statement (6). Finally, the proof of (7) follows from the proof of (6) and from

$$\begin{aligned} &P\{ Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_j)} \leq 0 \} = \\ &= P\{ \mathrm{MSE}_{\mathcal{K}_n}^{l_m} - \mathrm{MSE}_{\mathcal{K}_j}^{l_m} \} P\{ Z_{l_m}^{(\mathcal{K}_j \to \mathcal{K}_n)} > 0 \}, \\ &\forall j, n \in \boldsymbol{\mathcal{K}}. \end{aligned} \tag{20}$$

□

Lemma 3. *Algorithm 1 converges almost surely to a minimizer state.*

Proof. Consider, first, the case when Condition 3 holds. Since Conditions 1,2 hold as well, the requirements for convergence, according to Theorem 3.1 in [11] are satisfied and Algorithm 1 leads to mean square error minimization. Consider, now, the case when Condition 3 does not hold. Assume $\mathcal{K}_n \notin \boldsymbol{\mathcal{L}}$. If $\mathcal{K}_n$ is transient, then with probability one the sequence $\{\mathcal{K}_m\}$ of visited states will not converge to $\mathcal{K}_n$ as $m \to \infty$. Assume now that $\mathcal{K}_n$ is positive recurrent. By (19) in Lemma 2 we have that $\lim_{l_m \to \infty} P\{Z_{l_m}^{(\mathcal{K}_n \to \mathcal{K}_i)} > 0\} > 0, \forall \mathcal{K}_i \in \boldsymbol{\mathcal{L}}$. Consequently, $\{\mathcal{K}_m\}$ and all states of $\boldsymbol{\mathcal{L}}$ belong to the same communicating class, denoted by $\boldsymbol{\mathcal{K}^I}$, as well as all the other positive recurrent $\mathcal{K}_i \notin \boldsymbol{\mathcal{L}}$ states. The system is thus reduced to a set of states $\boldsymbol{\mathcal{K}^I}$. Clearly, Condition 1 holds for all states in $\boldsymbol{\mathcal{K}^I}$, and a result, the requirements for Theorem 3.1 in [11] are fulfilled. □

5 Performance evaluation

The performance of the discrete stochastic optimization based parameter estimation depends on the granularity of the state space, $\boldsymbol{\mathcal{K}}$, and on the number an the location of the s-domain points, on which the empirical and the analytic LTs are compared, for the MSE calculation. These parameters affect the accuracy of the parameter estimation, even if the optimal parameter vector is determined by exhaustive search.

In addition, we consider a limited idle period sample size, and terminate the algorithm when all idle period samples are integrated. This on one hand minimizes the time spent for parameter estimation, but on the other hand, does not ensure that the algorithm finds the optimal parameter vector. To evaluate the achievable estimation performance, we perform parameter estimation with exhaustive search and with early termination, considering a large set of model compliant traffic traces. We select 10^4 $(\xi, \sigma, p, p_{\mathrm{CCA}})$ parameter vectors, generate a sequence of idle and active periods for each vector, and run the estimation algorithms. The parameters are randomized according to Table 1, to cover a wide range of traffic patterns. For the evaluations presented here we fix $S = 10^3, s_k \in (10^0, 10^5), 1 \leq k \leq S$, and integrate one new idle period sample in each iteration step.

As stated in Section 2 it is assumed that the $f_A(t)$ parameters can be estimated correctly and α_{BK} is known. We measure the estimation accuracy by calculating the mean absolute error (MAE) of the p and p_{CCA} and the mean percentage error (MPE) of the ξ and σ estimation.

As the number of idle period samples affects the time needed for continuous sensing and in our case even gives the number of iterations of the optimization algorithm, it is one of the main design parameters to be considered. Therefore we evaluate the parameter estimation performance for 10^3 and 10^4 idle period samples and iteration steps. In addition, to evaluate the effect of the size of the state space of the discrete optimization we change the granularity of the state parameters $\{\xi_i, \sigma_i, p_i\}$ from 10^{-1} to 10^{-5}.

Table 1: Model Parameters

Parameter	Distribution	Min	Max	Mean	StdDev
ξ	Truncated Gaussian	0.1	0.4	0.3095	0.1
σ	Truncated Gaussian	1e-4	0.1	0.02	0.2
p	Uniform	0.1	1.0		
p_{CCA}	Uniform	0.1	1.0		
α_{ON}	Uniform	0.0008	0.001		
β_{ON}	Uniform	α_{ON}	0.0015		
α_{BK}	Deterministic			0.0007	

Figure 3 compares the estimation accuracy of $\{\xi, \sigma, p, p_{\mathrm{CCA}}\}$ under exhaustive search and with stochastic optimization with early termination. Considering the number of integrated samples, we can see that the increased number of samples improves the estimation accuracy under exhaustive search. At the same time, an increased state space does not necessarily lead to better estimation accuracy. The estimation accuracy may increase with increased state space for a while, in this interval the minimizer is found, and the increased granularity means lower MSE. However, as the state space is further increased, the minimizer can not anymore be discovered in the limited number of iterations, and therefore the estimation accuracy drops. Therefore, the state space size has to be selected carefully, taking the expected number of samples into account.

The results show that the performance of the proposed algorithm is comparable to the one of the exhaustive search. A number of samples in the range of 10^4 and parameter granularity of $10^{-3} - 10^{-4}$ gives an estimation accuracy that is sufficient for the cognitive control as it was shown in [9], while it allows acceptable sensing times, and a state space size that is implementable on sensor devices with limited memory.

6 Discussion

In the heterogeneous networking environment of the the open ISM band the prediction of the availability of the wireless resources is a key enabler for the design of energy efficient wireless networks. In this paper we considered the issue of WLAN and WSN coexistence. In this case WSN transmissions suffer from WLAN interference, because the WLAN carrier sensing does not detect the low power, narrow band WSN transmissions. The sensor network can avoid this interference, if it can characterize the channel occupancy, and tune its transmission parameters accordingly.

We described a semi-Markovian model of the WLAN channel occupancy, as observed by the individual sensor nodes and proposed a discrete stochastic opti-

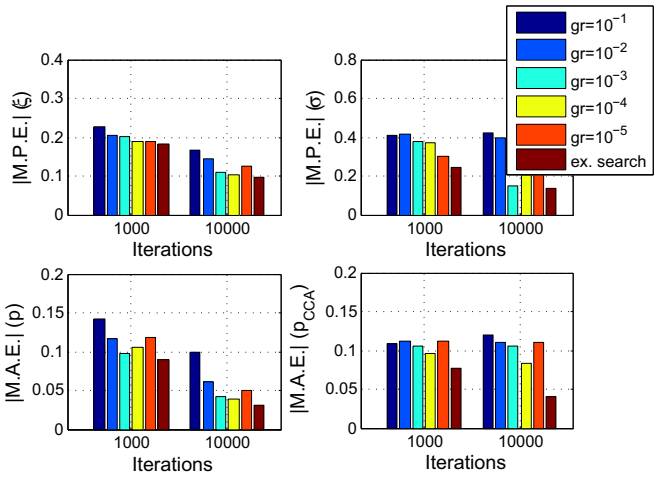


Figure 3: The accuracy of the LT-based estimation with respect to the number of iterations, and the granularity of the state space. Exhaustive search results are shown for comparison.

mization based algorithm to estimate the parameters of the idle time distribution in the Laplace domain. We showed that the proposed solution can achieve the required estimation accuracy by sequentially integrating the measured idle period samples and by simultaneously searching for the optimal parameter vector. We can conclude that the required idle time sample size allows limited sensing times and the parameter granularity can be low enough for the algorithm to be implemented in resource limited sensor nodes. Therefore the proposed algorithm can support the development of cognitive medium access control and routing in WSNs.

References

[1] S. Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201 – 220, February 2005.

[2] K.R. Chowdhury and I.F. Akyildiz. Interferer classification, channel selection and transmission adaptation for wireless sensor networks. In *Proceedings of IEEE ICC'09*, pages 1 –5, June 2009.

[3] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP

framework. *IEEE Journal on Selected Areas in Communications*, 25(3):589 –600, April 2007.

[4] J. Misic and V.B. Misic. Characterization of idle periods in IEEE 802.11e networks. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1004 –1009, March 2011.

[5] L. Stabellini. Quantifying and modeling spectrum opportunities in a real wireless environment. In *Proceedings of Wireless Communications and Networking Conference (WCNC)*, pages 1 –6, April 2010.

[6] J. Huang, G. Xing, G Zhou, and R. Zhou. Beyond co-existence: Exploiting WiFi white space for Zigbee performance assurance. In *Proceedings of 18th IEEE International Conference on Network Protocols*, ICNP 2010, pages 305–314, October 2010.

[7] S. Geirhofer, L. Tong, and B.M. Sadler. Cognitive medium access: Constraining interference based on experimental models. *IEEE Journal on Selected Areas in Communications*, 26(1):95–105, January 2008.

[8] S. Geirhofer, J. Z. Sun, L. Tong, and B. M. Sadler. Cognitive frequency hopping based on interference prediction: theory and experimental results. *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, 13:49–61, September 2009.

[9] I. Glaropoulos, V. Fodor, L. Pescosolido, and C. Petrioli. Cognitive WSN transmission control for energy efficiency under WLAN coexistence. In *2011 Sixth International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, pages 261 –265, June 2011.

[10] M. Lagana, I. Glaropoulos, V. Fodor, and C. Petrioli. Modeling and estimation of partially observed WLAN activity for cognitive WSNs. In *IEEE Wireless Communications and Networking Conference (WCNC), 2012*, pages 1526 – 1531, April 2012.

[11] S. Andradottir. A global search method for discrete stochastic optimization. *SIAM Journal on Optimization*, 6(2):513–530, 1996.

Paper D

Closing the gap between traffic workload and channel occupancy models for 802.11 networks

Ioannis Glaropoulos, Alexandre Vizcaino Luna, Viktoria Fodor, and Maria Papadopouli

Published in Elsevier Journal of Ad Hoc Networks, vol. 21, no. 0, pp. 60-83, 2014.

Closing the gap between traffic workload and channel occupancy models for 802.11 networks

Ioannis Glaropoulos, Alexandre Vizcàino Luna, Viktoria Fodor, and Maria Papadopouli*

Abstract

The modeling of wireless network traffic is necessary to evaluate the possible gains of spectrum sharing and to support the design of new cognitive protocols that can use spectrum efficiently in network environments where diverse technologies coexist. In this paper we focus on IEEE 802.11 wireless local area networks and close the gap between two popular levels of modeling, macroscopic traffic workload modeling and microscopic channel occupancy modeling. We consider traffic streams generated by established traffic workload models and characterize the networking scenarios where a simple, semi-Markovian channel occupancy model accurately predicts the wireless channel usage. Our results demonstrate that the proposed channel occupancy model can capture the channel idle time distribution in most of the scenarios, while the Markovian assumption can not be validated in all cases.

Index terms— 802.11 networks; traffic models; channel occupancy models; statistical validation.

1 Introduction

Spectrum sharing among diverse network technologies has been introduced as a promising solution to increase the efficiency of spectrum utilization in wireless environments, and thus ease the problem of spectrum scarcity. One of the key components of efficient spectrum sharing is cognitive medium access control, building on the knowledge of the channel usage patterns of the coexisting networks [1]. Therefore, traffic workload and channel occupancy models, either considered to be known [2] or derived on-line [3], are necessary for protocol design and channel access optimization. The issue of network coexistence in the open ISM band is particularly relevant due to the proliferation of diverse low-power wireless technologies, all sharing the ISM spectrum with the high-power Wireless Local Area Networks

*I. Glaropoulos, A. Vizcàino Luna, and V. Fodor are with the KTH, Royal Institute of Technology, Sweden. M. Papadopouli is with the University of Crete, Greece.

(WLANs). Accurate WLAN modeling enables these low-power technologies to alleviate harmfull WLAN interference [4] and to ensure an effective use of the shared open spectrum [5][6].

WLAN modeling can be classified in two main categories, based on the considered time scale, *traffic workload* modeling and *channel occupancy* modeling. *Traffic workload* studies involve stochastic analysis and modeling of high-layer traffic statistics, such as user arrival and departure process [7][8][9] and client-generated flow statistics [10][11][12], the characterization of user traffic [8][13] or the user mobility [14][15]. In these studies WLAN measurement data is collected via active probing or passive network monitoring, followed by the statistical processing of the collected data, when analytic probability distributions are fitted to the empirical traces. Traffic workload models are often specific to a given networking scenario, for example, [16] considers a campus-wide WLAN and provides detailed multi-level, campus-wide WLAN traffic modeling where both session and flow statistics are collected and fitted to analytic distributions. Although sufficiently realistic, these approaches capture the behavior of WLANs only at a macroscopic level.

Contrary to traffic workload modeling, *channel occupancy* studies aim at modeling directly the short term temporal behavior of the channel status in WLAN networks. They characterize the periods when the channel is either *active* due to a WLAN packet transmission, or *idle*. Clearly, the distribution of the active times is determined by the packet sizes used by the applications, while the distribution of the idle periods depends on both the process of packet generation and the medium access control. We can distinguish between *analytic* and *measurement-based* studies, depending on whether the spectrum occupancy model is developed based on analytic modeling of user behavior and network protocols, or it is extracted from channel occupancy measurements. The seminal work in [17] gives an analytic model for the impact of the IEEE 802.11 MAC protocol on channel occupancy, and derives the network throughput of a single Access Point (AP) WLAN assuming saturated user traffic, i.e. users always have packets to transmit. The case of a non-saturated single WLAN AP is studied in [18], modeling the packet arrivals at the users as a Bernoulli process. In [19] WLAN output buffers are modeled as M/G/1 queues, resulting in sub-geometric idle period distribution. The generality of these analytic channel occupancy models, is, however, limited, since they are based on specific, simple traffic workload models.

As far as measurement-based approaches are concerned, in [20] a hyper exponential distribution is fitted to the empirical idle period distribution derived by traffic traces from an area with heterogeneous wireless devices. In [21] a Markovian channel occupancy model is developed based on channel measurements extracted from controlled laboratory environments in the 2.4GHz ISM band. In [22][23] the heavy-tailed behavior of the idle channel periods is demonstrated and a mixture distribution is proposed to capture the two basic sources of channel inactivity, the short, almost uniformly distributed contention windows and the long, heavy-tailed white space periods, when the WLAN users are inactive. The simplicity of the re-

sulting semi-Markovian model makes it attractive for analytic performance studies and cognitive protocol design [5][23][24]. This model considers an idle period length distribution with a high number of degrees of freedom and, potentially, good fitting quality, therefore we select it as the candidate channel occupancy model. In [22][23] it has been validated for a limited set of scenarios, under perfect channel conditions and considering constant UDP payload traffic with exponential packet inter-arrival times. In this paper we evaluate it for a wide range of traffic patterns and network scenarios and define the key factors that affect its accuracy. As traffic traces can not provide the diversity we are looking for, we build our evaluation on synthetic traces based on validated traffic workload models. To investigate the generality of the model, we select three scenarios with significantly different traffic workload characteristics, namely, university campus, conference-hall, and industrial-plant WLANs.

Specifically, we validate the proposed heavy-tail idle period distribution and the Markovian *assumption*, that is, the assumption that consecutive idle periods have independent durations. We focus on the idle periods, as the active period distribution is not significantly affected by the medium access control, instead, it is directly determined by the packet sizes in the application mix. The model validation is completed with an evaluation of its accuracy considering a restricted dataset of real WLAN traces.

The main contributions of the paper are summarized as follows.

1. *Traffic workload and channel occupancy modeling.* We define detailed traffic workload models for the three networking scenarios, and parameterize the related semi-Markovian channel occupancy model using extensive simulations of an IEEE 802.11 AP.

2. *Evaluation of the idle period distribution.* We evaluate the fitting quality of the channel occupancy model, specifically considering the distribution of the idle times. Our results indicate that the mixture distribution proposed in [22][23] is valid in a wide range of networking scnearios.

3. *Evaluation of the Markovian assumption.* We evaluate the validity of the Markovian assumption considering the correlation of the consecutive idle time period lengths. We conclude that the idle period lengths may be correlated at low or very high load and when the traffic is highly heterogeneous. Therefore the Markovian assumption has to be applied with care.

4. *Model validation with real WLAN traces.* We evaluate the fitting accuracy of the semi-Markovian model as well as the validity of the Markovian assumption, considering a set of real 802.11 channel occupancy traces captured in diverse WLAN environments. The results are similar to the ones with the synthetic traces, however, they show as well that real occupancy traces can reflect unexpected traffic characteristics.

The remainder of the paper is structured as follows. In Section 2 we review the considered traffic workload and channel occupancy models. In Section 3 we

introduce the networking scenarios under study, along with a detailed description of the multi-layer traffic models. The simulation setup, as well as the employed statistical validation tools are presented in Section 4. Section 5 presents the results of the channel occupancy model validation using synthetic WLAN channel occupancy traces, while Section 6 includes the validation over the real WLAN traceset. Section 7 concludes the paper.

2 Traffic Workload and Channel Occupancy Models

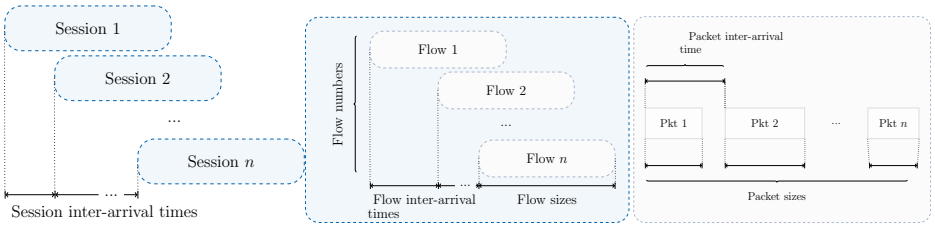


Figure 1: The structure of the multi-layer traffic workload model, comprising of (a) session-, (b) flow- and (c) in-flow processes.

In this section we define the structure of the multi-layer WLAN traffic workload model and the analytic model for WLAN channel occupancy that is considered in the paper.

2.1 Multi-layer WLAN Traffic Workload Model

Fig. 1 depicts the structure of the multi-layer traffic workload model. As suggested by [16], the *sessions* (Fig. 1(a)) on the top of the model hierarchy represent the users of the WLAN. Sessions *arrive* at the network following a stochastic process that is, in general, time-variant. A session is characterized by the *number* of traffic *flows* it generates (Fig. 1(b)), and by the *inter-arrival times* between these flows. The length of the sessions is not part of the model. A session is considered to be terminated, when its last flow ends.

A flow is defined as the unidirectional end-to-end packet sequence from a specific transport-layer source-destination connection or a media stream. Each flow is characterized by its *size* in bytes (Fig. 1(b)), and the *in-flow characteristics* given by the sizes, and the inter-arrival times of the *packets* (Fig. 1(c)). The in-flow characteristics depend on the particular network application that generates the flow [13].

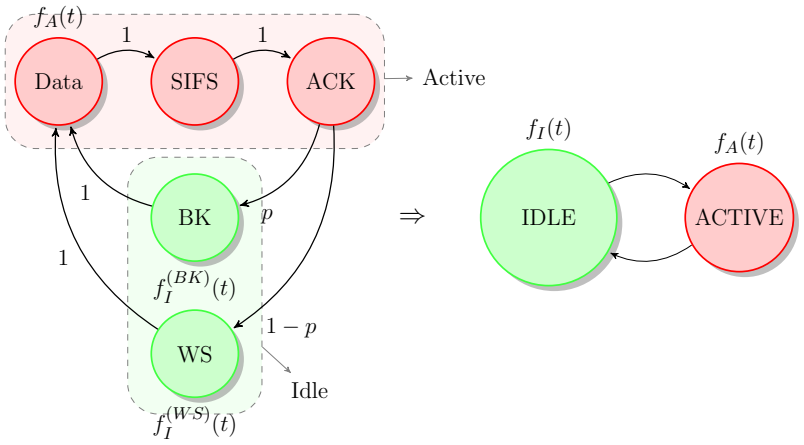


Figure 2: The semi-Markovian channel occupancy model and its two-state representation.

Note that the effect of the medium access control is not considered in the traffic workload model.

2.2 Semi-Markovian Channel Occupancy Models

Fig. 2 depicts the IEEE 802.11 WLAN channel occupancy model, originally proposed in [22][23]. The *states* of the semi-Markovian system correspond to the different phases of the WLAN transmission cycle. The channel is *Active* when there is a packet, either Data or ACK, under transmission. Neglecting the short Inter-frame Space (*SIFS*) period between a Data and an ACK packet, due to its considerably short duration, *Data*, *SIFS* and *ACK* states can be merged together into a single *Active* state. In the absence of packet transmission the WLAN channel is *Idle*, and we distinguish between the short *back-off* idle periods (*BK*) introduced by the IEEE 802.11 contention resolution mechanism, and the significantly longer idle periods due to user inactivity, denoted as WLAN *white spaces* (*WS*). Merging the states *BK* and *WS* into a single *Idle* state, the model is reduced into a two-state, semi-Markovian model with holding times $f_A(t), f_I(t)$, respectively. As proposed in [22] active periods are sufficiently modeled as uniform, while $f_I(t)$ needs to be described by a *mixture* distribution, which aims at capturing both of the sources of channel inactivity, that is, the back-off periods with holding times $f_I^{(\mathrm{BK})}(t)$ and the white space periods with $f_I^{(\mathrm{WS})}(t)$. Consequently, $f_I(t)$ obtains the form:

$$f_I(t) = p f_I^{(\mathrm{BK})}(t) + (1-p) f_I^{(\mathrm{WS})}(t), \quad t \geq 0, \tag{1}$$

with the *mixture* parameter $p \in [0,1]$ defining the probability that an idle period is a short back-off.

Measurement results in [22] suggest that the back-off period durations can be modeled with uniform distribution:

$$f_I^{(\mathrm{BK})}(t) = 1/\alpha_{\mathrm{BK}}, \quad \alpha_{\mathrm{BK}} = 0.07\mathrm{msec}, \tag{2}$$

independently from the network load. Note that α_{BK} is not the maximum possible back-off period in IEEE 802.11 WLANs, however, higher values appeared rarely in the measurement set [22]. The white space periods are suggested to be modeled by a zero-location, generalized Pareto distribution:

$$f_{\mathrm{gP}}(t) = \frac{1}{\sigma}\left(1 + \xi\frac{t}{\sigma}\right)^{-\frac{1}{\xi}-1}, \tag{3}$$

with σ and ξ, being the *scale* and *shape* parameters, that depend on the actual network traffic. To capture the effect of access-point *beaconing*, in this paper we consider a truncated version of (3). As white spaces are limited by the beacon *period*, T_B:

$$f_I^{(\mathrm{WS})}(t) = \frac{1}{\sigma F_{\mathrm{gP}}(T_B)}\left(1 + \xi\frac{t}{\sigma}\right)^{-\frac{1}{\xi}-1}, \; t \in (0, T_B), \tag{4}$$

where $F_{\mathrm{gP}}(T_B) \triangleq Pr[T_B \leq t] = \left[1 - \left(1 + \frac{\xi}{\sigma}t\right)\right]^{-1/\xi}$. The average white space duration is:

$$\overline{I}^{(\mathrm{WS})} = \int_0^{T_B} f_I^{(\mathrm{WS})}(t)dt \approx \frac{\sigma}{1-\xi}, \text{ if } F_{\mathrm{gP}}(T_B) \approx 1. \tag{5}$$

Note, that this channel occupancy model considers good channel conditions without the eventual loss of Data or ACK packets. Under packet losses, the Active and Idle period length distributions are slightly different. Specifically, first, the Active period may not include the ACK packets, and therefore, the Active time distribution is affected by the packet loss probability. Note, however, that the estimation is still straightforward, since the distribution depends only on the Data packet sizes in the original traffic mix and on the packet loss probability. Still, the length of the ACK packets is very short compared to the size of the majority of the Data packets. Second, in the case of Data packet loss ACKs are not transmitted, and the idle period starts with a short idle time determined by the Net Allocation Vector (NAV) timer, typically set as the SIFS+ACK duration, in the order of tens of microseconds, followed by the minimum back-off period. Consequently the maximum of the back-off period (α_{BK}) is not affected. The loss affects p, the ratio of back-off periods, but this is a parameter to be estimated. Third, if the ACK is transmitted, but not received correctly, only p is affected due to the packet retransmission. Consequently, the proposed model fits well also for modeling the idle-time distribution in scenarios with transmission errors.

2.3 Validation of the Channel Occupancy Model

We assess the ability of the aforementioned semi-Markovian channel occupancy model to capture the statistical behavior of a practical WLAN channel usage, focusing on the idle time distribution, $f_I(t)$. This assessment involves the following steps:

1. We generate synthetic traffic streams based on the traffic workload model in Section 2.1, simulate the stream transmissions on a IEEE 802.11-compliant WLAN, and extract the resulting idle time periods sequence.
2. Then, using the collected idle period samples, we determine T_B as the maximum idle period sample, and estimate the parameters of $f_I(t)$, that is ξ and σ, the *shape* and *scale* parameters of the generalized Pareto distribution, and the *mixture* p [25].
3. We preform a detailed statistical analysis over the simulated and analytic idle period traces to evaluate the accuracy of the channel occupancy model.

3 Networking Scenarios

In this paper we focus on three common WLAN networking scenarios, where the traffic workload is expected to be different:

1. campus WLAN,
2. conference-hall WLAN,
3. industrial-plant WLAN.

3.1 Campus WLAN

To model the traffic demand in the campus WLAN, we follow the session arrival and flow models of [16]. Table 1 summarizes the proposed distributions and their parameters. As [16] shows the session arrival intensity varies during the day. However, it does not change significantly within the time frame at which traffic modeling is useful for protocol design and evaluation. Therefore the session arrivals are modeled by a stationary Poisson process with parameter value drawn from a truncated geometric distribution to model arrival intensities reported in [16]. Flow inter-arrival times are Log-normal, while the number of flows per session and the flow sizes are modeled by heavy-tailed BiPareto distributions.

As [16] does not discuss in-flow characteristics, we categorize the flows according to [13] as DNS, Web, FTP, P2P, VoIP, and Video, which gives a highly heterogeneous traffic mix, and give the probability that a flow belongs to a given category. We model the DNS flow as the transmission of a single, uniform-sized packet, since

Traffic object	Distribution	Parameters
Session		
Session arrival [hour]	Stationary Poisson process (λ)	1 (min), 928 (max), 11(median) trunc. Geometric
Flow		
Flow inter-arrival times [sec]	Log-normal	μ=$-1.6355, \sigma$=2.6286
Flow numbers per session	BiPareto	α=0.07, β=1.75, c=295.38, k=1
Flow sizes [bytes]	BiPareto	α=0.00, β=1.02, c=15.56, k=111
Application	Web/FTP/P2P VoIP+Video/Other	P_{Web}=0.25, P_{FTP}=0.12, P_{P2P}=0.33, $P_{\text{V+V}}$=0.21, P_{oth}=0.09

Table 1: Campus WLAN. Traffic workload model parameters.

Application	*Packet Size (bytes)*	*Inter-arrival Times (sec)*
DNS	Deterministic (512)	Single Packet
Web	Uniform $(512, 1536)$	Exponential ($\lambda = 10^{-1}$)
FTP	Uniform $(892, 1152)$	Deterministic (10^{-3})
P2P	Exponential ($\lambda = 512$)	Weibull $(0.53, 0.13532)$
VoIP	Uniform $(128, 384)$	MMDP (λ, μ, T) $\lambda = 0.1, \mu = 0.1, T = 0.03$
Video	Uniform $(384, 768)$	MMUP $(\lambda, \mu, \alpha, \beta)$ $\lambda = 0.1, \mu = 0.1, \alpha = 0.03, \beta = 0.05$

Table 2: Packet size and inter-arrival time distributions for the various network applications.

DNS requests and replies typically have short payload. We characterize the Web flows with uniform packet sizes and exponential packet inter-arrival times, accounting for persistent HTTP connections. FTP flows consist of almost deterministic-size packets, transmitted back-to-back. To characterize P2P flows, we follow the suggestions in [26] assuming BitTorrent traffic. Finally, we model the VoIP and Video stream flows considering that they represent Skype traffic and follow the suggestions in [27] on Skype video and voice traffic flow characteristics. The parameters are summarized in Table 2.

3.2 Conference-hall WLAN

While several conference-hall traffic workload models have been proposed, they do not follow the structure proposed in [16]. Therefore, we consider the conference-hall WLAN model proposed in [7], and convert it to the structure proposed in [16]. The conference-hall model suggests that the session arrival process follows the conference time schedule, and proposes statistical distributions for the characterization

Traffic object	Distribution	Parameters
Session		
Session arrival [sec^{-1}]	ON/OFF MMPP $[\lambda_{\mathrm{ON}}, \lambda_{\mathrm{OFF}}]$	λ_{ON}=38^{-1}, $\lambda_{\mathrm{OFF}} = 0$
Session duration [min]	Pareto	ξ=0.78, $\sigma = 30.76$
Session Workload		
Data rate [kbps]	Uniform *light* ($p_L = 0.25$), *medium* ($p_M = 0.65$), *heavy* ($p_H = 0.1$)	μ_l=15, $\max_{\mathrm{l}}$=60 $\mu_{\mathrm{m}}\in(15-80)$, $\max_{\mathrm{m}}\in(60-175)$, $\max_{\mathrm{h}}$=175, μ_{h}=80
Application	Web (HTTP,HTTPS,SSH) / DNS (ICP,ICMP,DNS)	$P^{\mathrm{fl}}_{\mathrm{Web}}$=0.62, $P^{\mathrm{fl}}_{\mathrm{DNS}}$=0.31 $P^{\mathrm{fl}}_{\mathrm{oth}}$=0.07 $P^{\mathrm{vol}}_{\mathrm{Web}}$=0.68, $P^{\mathrm{vol}}_{\mathrm{DNS}}$=0.11 $P^{\mathrm{vol}}_{\mathrm{oth}}$=0.21

Table 3: Conference-hall WLAN. Traffic workload model parameters.

of the *session arrival process*, the *session durations*, and the *data rate of a session*. It also categorizes the flows according to applications, and gives the probability distribution of the *application mix*. Table 3 summarizes the modeling of the traffic workload. Session arrivals are modeled by an ON/OFF Markov-Modulated Poisson Process (MMPP), where the ON state represents the beginning of a conference event. The session durations are Pareto distributed, while the session data rates are randomly selected within three different rate intervals, representing *light*, *medium* and *heavy* traffic, with probabilities $p_L, p_M,$ and p_H, respectively. Once the rate interval is chosen, the data rate is uniformly selected inside the given interval. The application mix is characterized by the probability that a flow belongs to a given application category and additionally by the portion of the traffic an application category generates.

While this model includes the notion of a traffic flow through the definition of the application mix, it does not give models for the number of flows per session, for the flow size and for the flow arrival process. We, therefore, convert the model to the multi-layer traffic workload model by characterizing the missing model elements as in [16] and by parameterizing their statistical distributions in order to fit with the given session, transmission rate and application mix characterization. For each session we, first, draw the session duration and the data rate according to the model parameters proposed in [7], and calculate the traffic volume (in number of bits) of the session. Given the traffic volume and the predefined application mix, we first determine the number of DNS flows, based on our DNS model in Table 2. Then, from the number of DNS flows we estimate the number of Web and other flows and the respective flow sizes, to fit the application mix parameters. Finally we parameterize the flow inter-arrival time distribution to match the known session duration.

Traffic object	Distribution	Parameters
Session		
Session Number	Fixed	$N \in (5, 20)$
Flow		
Flow inter-arrival time [sec]		
- Monitoring	Uniform (α, β)	$\alpha = 5, \beta = 90$
- Alarm occurrence & actuation	gPareto(ξ, σ)	$\xi \in (0.25, 0.5)$, $\sigma \in (5, 10)$,
Flow size [bytes]	Uniform (α, β)	$\alpha = 512, \beta = 1536$
Packet Statistics		
Packet size [bytes]	Uniform (α, β)	$\alpha = 512, \beta = 1024$
Packet inter-arrival time [msec]	Uniform (α, β)	$\alpha = 1, \beta = 10$

Table 4: Industrial-plant WLAN. Traffic workload model parameters.

3.3 Industrial-plant WLAN

To construct the traffic workload model for the industrial-plant scenario, we consider an 802.11 network with a single access point, serving as a backbone for a sensor network deployed for monitoring and actuation purposes. The WLAN terminals are fixed and forward sensor data towards a data-management center, that is, towards the WLAN AP, and control data to sensors and actuators. To define the traffic workload, we follow the use case of [28]. WLAN terminals cover similar areas in the industrial plant. During normal conditions they receive periodic messages from the sensors. These messages are first merged into larger packets at the WLAN terminals and then forwarded to the WLAN AP. Bursty messages arrive from or transmitted to the AP during alarms.

Consequently, we construct the multi-layer traffic workload model as follows. A fixed number of sessions is used for modeling the stationary WLAN terminals. We define two kinds of flows for each session. Flows representing the periodic monitoring traffic to the WLAN AP exhibit uniform flow inter-arrival times with small support, to reflect the random delay in the sensor packet aggregation process. The bursty traffic of alarms is represented by flow inter-arrival times according to heavy-tailed Generalized Pareto distribution. For both of the flow types we consider uniformly distributed flow size, representing the slightly different amount of information forwarded to or received from the data-management center. Similarly, we consider uniform distribution with small support for the packet sizes and packet inter-arrival times within the flows. Table 4 summarizes the traffic workload parameters for the industrial-plant WLAN.

4 Simulations and Validation Tools

In this Section we describe the simulator framework as well as the analytic tools that we employ for the validation of the semi-Markovian channel occupancy model.

4.1 Simulation Framework

The considered networking scenarios are simulated on the *NSmiracle* [29] platform. In all cases an IEEE 802.11b-compliant network is built, employing existing NSmiracle Channel, PHY, MAC and higher layer modules.

We consider a single access point area with a radius of 100m, inside which 802.11 terminals are uniformly distributed. The channel propagation model depends on the considered networking scenario. For the campus WLAN scenario we assume a path-loss-based channel model with a moderate exponent value, $\theta = 2.5$. The model is enhanced by log-normal shadowing with standard deviation $\sigma_{\text{shd}} = 10\text{dB}$. For the in-door scenarios of conference-hall and industrial-plant the channel attenuation is modeled considering a site-general Indoor ITU model [30] with power decay $N = 38$, and zero floor-penetration attenuation. Unless otherwise noted, we use a simplified PHY layer, where the user terminal transmits with a fixed data-rate of 11Mbps for the industrial-plant scenario, while for the campus and conference-hall WLAN cases the rate is inversely proportional to the distance from the AP, in the 1Mbps-11Mbps interval, to reflect adaptive rate control. In all cases the MAC follows the IEEE 802.11 standard with an IPv4 network layer on the top. Flows are transmitted with TCP or UDP according to their traffic type. The Session module of the simulator implements the multi-layer traffic workload model introduced in Section 3. The directions of the flows are selected randomly, for all cases, apart from the monitoring traffic flows in the industrial-plant scenario, which are always directed from the terminals to the AP.

Furthermore, we implemented a simple protocol stack for a *sensing device*. This device is responsible for continuously measuring the spectrum activity, for collecting a sequence of M samples of idle channel durations, and for building the *empirical* distribution function, $F_{Ie}(t; M)$. It also estimates the parameters of the semi-Markovian channel occupancy model (p, ξ, σ) based on the collected idle period sequences, applying the MLE estimation algorithm in [25]. The outcome of the estimation is the *analytic* distribution function $F_I(t; \xi, \sigma, p)$ of the idle time duration.

4.2 Analytic Validation Tools

Our goal is to validate the semi-Markovian channel occupancy model described in section 2.2, using the statistics of the idle channel durations collected from the simulations with the multi-layer traffic workload model.

We assess the *goodness-of-fit* by evaluating the *D-value* of the *Kolmogorof-Smirnoff* test, defined as the maximum absolute difference between the analytic and empirical idle period distribution functions:

$$D = \sup_{\tau \in \mathcal{T}} |F_I(\tau; \xi, \sigma, p) - F_{Ie}(\tau; M)|, \tag{6}$$

where $\mathcal{T}$ is the set of collected idle periods. A low D-value indicates the good fitting performance of the analytic model. We underline that the D-value is a worst-case metric, as it considers the supremum of the point-wise difference between the two functions, instead of the average.

The goodness-of-fit is, additionally, evaluated by performing *two-sample Kolmogorov-Smirnoff* (K-S) tests [31], where the collected sequences of idle periods, $\mathcal{T}$, are tested against *synthesized* idle period sequences, $\widehat{\mathcal{T}}$, that are generated randomly from the estimated analytic distribution, $F_I(t; \xi, \sigma, p)$. The K-S test assesses the validity of the *null hypothesis*, that is, both idle period series can originate from the same distribution. The test is conducted considering n randomly selected samples of sequences $\mathcal{T}, \widehat{\mathcal{T}}$, denoted as $\mathcal{T}_n, \widehat{\mathcal{T}}_n$, respectively. The evaluation is done by deriving the following two-sample K-S test statistic:

$$K_n = \sqrt{\frac{n}{2}} \sup_{\tau \in \mathcal{T}_n, \hat{\tau} \in \widehat{\mathcal{T}}_n} |F_{Ie}(\tau; n) - F_{Is}(\hat{\tau}; n, \xi, \sigma, p)|, \tag{7}$$

where $F_{Is}(\hat{t}; n, \xi, \sigma, p)$ denotes the empirical distribution of the synthesized random sequence. We assess the null hypothesis by calculating the *p-Value* (p_{KS}) of this test, that is the probability of obtaining a test statistic, K_n, at least as extreme as the one we observe. We reject the null hypothesis at a significance level of $\alpha \in (0, 1)$, if $K_n > K_\alpha$, where K_α is the *critical* value [31] defined as $K_\alpha = k : \Pr\{K_n > k\} < \alpha$.

For a deeper understanding of the results of the goodness-of-fit tests we show typical examples of empirical and fitted analytic distributions as well as *quantile-quantile* (Q-Q) plots.

In addition to the goodness-of-fit study, we evaluate the hypothesis that the lengths of the consecutive idle time periods are uncorrelated, an assumption that is required for the semi-Markovian channel occupancy model. We perform a test of independence, by comparing the *lag-k autocorrelation* values of the obtained sample sequence against a sample sequence that approximates well a *white* noise series with low auto-correlation. We generate the white noise reference by sampling the original idle period sequence with large time gap separations. We repeat the autocorrelation test using different sub-sequences of the original sample series. For each test we record the sign of the difference of the lag-k autocorrelation value of the empirical sample sequence and that of the white noise reference. If the samples of the empirical idle-period sequence are correlated, a large portion of these sign outcomes will be positive. If the samples of the tested time series are indeed independent variables, the autocorrelation values of the empirical sample sequence and the white noise

Distribution	Parameters		
Packet sizes [bytes]			
Uniform (α, β)	(64,192),	(256,768),	(768,1280)
Exponential (λ)	128	512	1024
Deterministic	128	512	1024
Inter-arrival time [sec]			
Uniform (α, β)	$(0.005, 0.015)$	$(0.05, 0.15)$	$(0.5, 1.5)$
Exponential $(1/\lambda)$	10^{-2}	10^{-1}	1
Deterministic (μ)	10^{-2}	10^{-1}	1

Table 5: In-flow stochastic models.

reference bare similar statistical behavior and thus, the sign of their difference gives a Bernoulli (1/2) trial. Therefore, to decide on the validity of the null hypothesis we compute the difference between the positive and negative sign outcomes for all tests. We define the p-value of our test as the probability that a Bernoulli(1/2) sequence gives the same or larger difference as the obtained one. That is, the p-value reflects the probability that the observed correlation metric can occur in a sequence of independent samples. We reject the null hypothesis of independence at a significance level α.

5 Numerical Evaluation

5.1 The Impact of In-flow Characteristics on Fitting Accuracy

Multi-layer workload models in general and also [7][16] characterize the higher-level traffic workload in WLANs, while the modeling of in-flow traffic is not considered. Therefore, in the first part of this Section we evaluate the influence of the in-flow traffic characteristics on the fitting performance of the proposed mixed uniform-Pareto idle period distribution. This evaluation will help us conclude whether the selection of a particular application mix with application-specific packet size and packet inter-arrival time distribution can have a significant effect on the fitting performance.

We consider a single high-level workload *configuration*, that is, we generate sessions (session arrival times) and flows (number of flows per session, flow arrival times and flow sizes) considering the campus WLAN model (see Table 1). We use the same workload configuration for all the experiments. For each experiment we select a packet size and inter-arrival time distribution pair according to Table 5, and perform 50 simulation runs (for all but the deterministic packet size, deterministic inter-arrival time case) collecting $M = 10^4$ idle period samples within each run. Table 6 gives the average D-value for each experiment. The D-values are very low for all cases with both random packet sizes and random inter-arrival times, and do not seem to depend on the actual parameter value of the distributions. The mixed

Pareto distribution, however, does not fit well the empirical data under short, deterministic inter-arrival times, especially when combined with deterministic packet size. We evaluate the reason of the large D-value with the help of Fig. 3, showing the empirical and analytic idle-period distribution functions for two in-flow configurations.

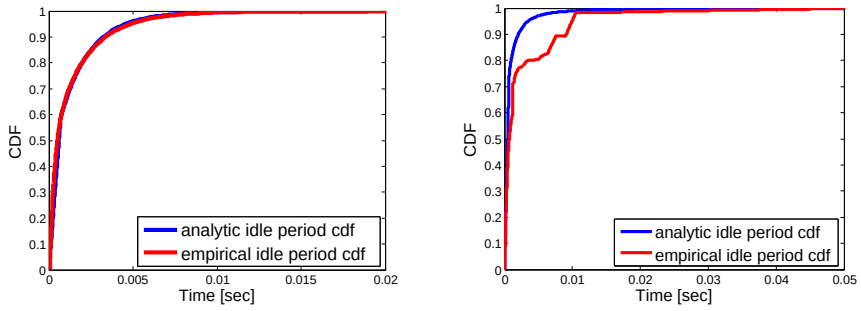


Figure 3: Examples of fitting performance for different in-flow characteristics: (a): D-Value $= 0.005$, Exponential($\lambda^{-1} = 512b$) packets-size, Uniform($\mu = 10^{-2}, \sigma^2 = 10^{-5}$s) packet inter-arrival time. (b): D-Value $= 0.055$, Deterministic($1024b$) packet-size, Deterministic(10^{-2}s) packet inter-arrival time.

Fig. 3(a) shows results with Exponential ($\lambda^{-1} = 512B$) packet size and Uniform ($\mu = 10^{-2}, \sigma^2 = 10^{-5}$s) packet inter-arrival time, exhibiting a D-value $= 0.005$, while Fig. 3(b) with Deterministic ($1024B$) packet size and Deterministic (10^{-2}s) packet inter-arrival time, with D-value $= 0.055$. Under deterministic packet size and packet inter-arrival time the empirical function, $F_{Ie}(t; M)$ is dominated by either 802.11 back-off periods, or idle periods with a duration around 10^{-2}s, generated between successive packets inside flows. Apparently, the random mixing of the deterministic packet streams does not lead to generalized Pareto distributed idle times in this case.

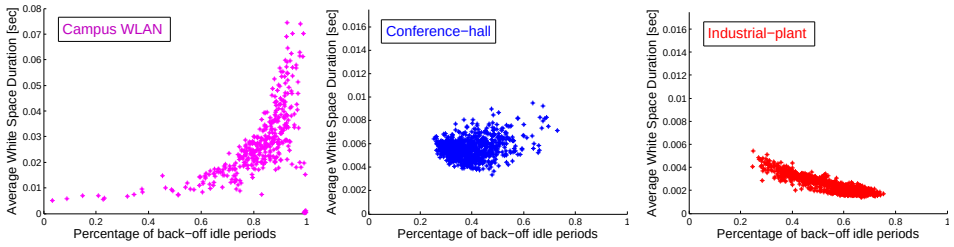
Based on the results in Table 6 we conclude that the in-flow traffic characteristics may have an impact on the fitting accuracy, and therefore, the validation of the channel occupancy model under real WLAN scenarios must take into account realistic, application-dependent in-flow models.

5.2 The Impact of the Networking Scenario on the Fitting Accuracy

Let us now evaluate the accuracy of the idle time distribution model for the three scenarios introduced in Section 3, that is, the campus, the conference-hall and the industrial-pant WLAN, with the multi-layer traffic workload model parameters sum-

Packet size [bytes]		Packet inter-arrival times [sec] Uniform 10^{-2}	Uniform 10^{-1}	Uniform 1	Exponential 10^{-2}	Exponential 10^{-1}	Exponential 1	Deterministic 10^{-2}	Deterministic 10^{-1}	Deterministic 1
U	**128**	.0122	.0151	.0171	.0205	.0165	.0169	**.0945**	.0220	.0344
	512	.0093	.0115	.0177	.0180	.0136	.0176	**.0943**	.0508	.0291
	1024	.0248	.0144	.0181	.0101	.0144	.0171	**.0899**	.0196	.0356
E	**128**	.0169	.0218	.0166	.0170	.0229	.0163	**.0706**	.0287	.0303
	512	.0243	.0217	.0168	.0294	.0221	.0167	**.0574**	.0418	.0280
	1024	.0251	.0409	.0163	.0454	.0416	.0157	**.0912**	.0476	.0314
D	**128**	.0140	.0097	.0183	.0114	.0094	.0179	**.0676**	**.0665**	.0461
	512	.0110	.0105	.0185	.0085	.0092	.0181	**.1486**	**.0817**	.0455
	1024	.0199	.0113	.0178	.0079	.0111	.0178	**.1733**	**.0362**	.0448

Table 6: Expected D-value for the various in-flow configurations.

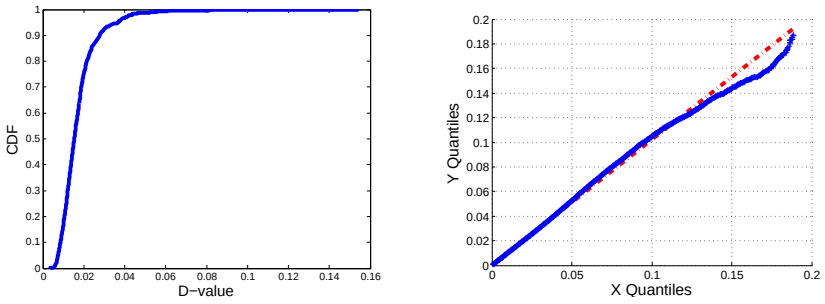


(a) (b) (c)

Figure 4: The resulting parameter space for the idle spectrum period distribution for the three networking scenarios.

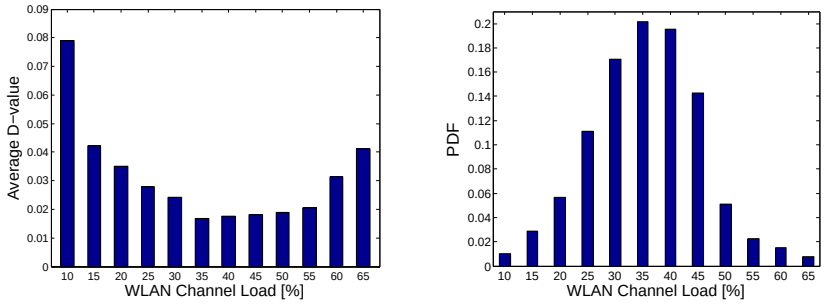
marized in Tables 1, 3, and 4, and the in-flow parameters in Table 2. We simulate the different scenarios with randomized traffic workload parameters and estimate the parameters of the idle time distribution based on an input sequence of $M = 10^4$ idle period samples, allowing for a 10^3-samples warm-up period. Depending on the length of the active and idle time periods this corresponds to different simulation time durations in the order of $10^2 - 10^3$ seconds. We perform 10^3 simulation runs for each of the scenarios.

Fig. 4 shows the parameter space of the idle time distribution of the semi-Markovian model obtained from the simulations, considering p, the percentage of back-off periods, and the average estimated white space duration $\overline{I}^{(\mathrm{WS})}$ defined in (5). The figures show that the three scenarios present significantly different parameter spaces. Comparing the three scenarios we can identify the significant factors that affect the accuracy of the channel occupancy model and determine the model limitations.



(a) (b)

Figure 5: Campus WLAN. (a) Empirical distribution function of the D-value. (b) Q-Q plot for the empirical and synthesized idle period series over all simulation runs.



(a) (b)

Figure 6: Campus WLAN. (a) Average D-value with respect to WLAN channel load, (b) empirical density function of the channel load.

5.2.1 Campus WLAN

Fig. 5(a) shows the empirical cumulative distribution function of the D-value over 10^3 simulation runs. The distribution exhibits a low mean, $\overline{D} = 0.0199$, and moderate variance with 95% of the cases being lower than 0.04, revealing an excellent fitting quality. Fig. 5(b) depicts the quantile-quantile plot for the collected and the synthesized idle series, averaged over all simulation runs. We observe that the quantile-quantile curve follows closely the $x = y$ axis, which indicates that the series of the collected and synthesized idle period samples do come from the same distribution. This is verified as well by the low failing rate of the conducted Kolmogorov-Smirnoff test, shown in Table 7. The average p-value of the test is $p_{\mathrm{KS}} = 0.5714$, with a coefficient of variation of $C_{p_{\mathrm{KS}}} = 0.0923$, while its failing rate is 7.01%, for the standard significance level, $\alpha = 5\%$.

	D-value	p_{KS}	$C_{p_{\mathrm{KS}}}$	$P\{p_{\mathrm{KS}} \leq 5\%\}$
Campus	0.0199	0.5714	0.0923	0.0701
Conference-hall	0.0088	0.5335	0.0214	0.0467
Industrial-plant	0.0334	0.4029	0.2103	0.0934

Table 7: Summary of goodness-of-fit evaluation for the considered networking scenarios.

Even though the average fitting quality is satisfactory, we would like to identify the scenarios when the mixed uniform-Pareto distribution fails to adequately model the idle periods. Therefore, we first classify the simulation results according to the experienced WLAN *channel load*, defined as the percentage of the time the channel is in active state. Fig. 6(a) demonstrates that the channel load affects the D-value. The fitting accuracy is low in light-loaded cases, and the fitting is rather weak even at high load values. This indicates that the idle period distribution model is more suitable for capturing the channel occupancy statistics for moderate load cases. The empirical density function of WLAN channel load is shown in Fig. 6(b). We observe that the cases of extreme channel load, which result in poor model fitting quality, are relatively rare, and therefore we can conclude that the proposed analytic model is, in general, sufficient for channel occupancy modeling in campus WLANs.

Fig. 7 shows the fitting quality of the semi-Markovian model, when the WLAN terminals have higher transmission rate capabilities, considering the same traffic as for Fig. 6. We consider fixed, distance-dependent transmission rate in the 1-54Mbps range in Figs 7(a),(b), and per packet dynamic rate adaptation in the same range in Figs 7(c),(d). In both cases the fitting quality follows a similar D-value – WLAN load trend, as in Fig. 6, indicating that the transmission rates of the terminals do not significantly affect the characterization of the idle channel period durations. The fitting quality improves slightly under per packet rate adaptation due to the additional randomization of active and, consequently, of the idle period lengths. Due to the high transmission rates we observe a few simulation runs where the WLAN channel load is relatively low, (1-10%); in these cases the fitting quality is again relatively weak.

In Fig. 8 we depict the relation between the D-value and the number of *active* sessions, that is, all the sessions that arrive and transmit during the simulation run. Fig. 8(a) suggests that the number of sessions affects the fitting quality of the model, with low number of sessions leading to low fitting quality. As low number of sessions usually means low load, we evaluate whether the load or the number of sessions has dominant effect. In Fig. 8(b) we plot the D-value with respect to the number of sessions, restricting the study for the cases of low or heavy channel load. As simulation runs with very low and very heavy load are rare, we select load regions $\leq 25\%$ and $\geq 45\%$ to include a reasonable number of runs. We observe that under similar channel load, the number of the active WLAN sessions has little impact on the fitting quality, unless both the network load and the number of sessions are very

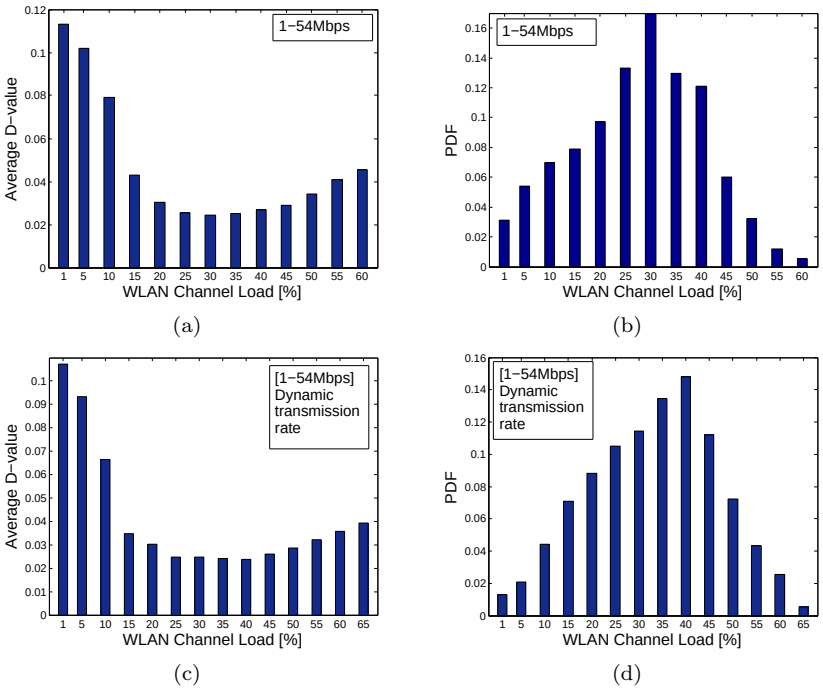
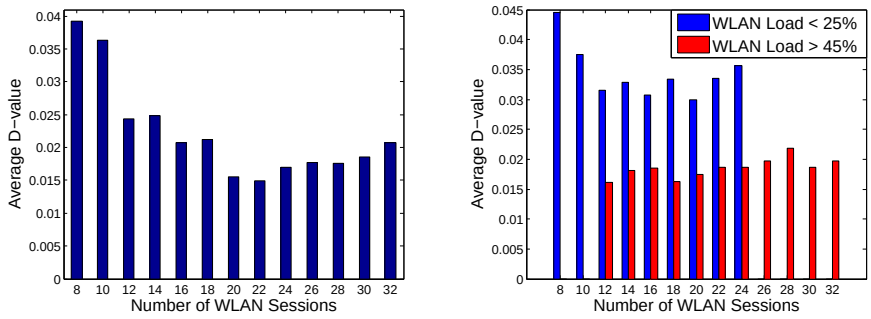


Figure 7: Campus WLAN. (a), (c) Average D-value with respect to WLAN channel load, (b),(d) empirical density function of the channel load for the case of 802.11 terminals with (a)(b) higher transmission rates and (c)(d) the case of 802.11 terminals with dynamic rate adaptation.

low. We evaluate the reason of the high D-value in this region by comparing the empirical and analytic idle period CDFs for a simulation run in Fig. 9(a). As we see, the empirical CDF first increases rapidly, but then shows heavy-tail characteristics. Due to the low number of active sessions, each with low load, flows rarely overlap in time, and the idle period distribution is determined by the in-flow characteristics, that can not be captured by the generalized Pareto distribution. Fig. 9(a) shows, additionally, that in low-load cases the percentage of entirely idle beacon intervals may be significant, therfore, modeling the white-spaces with the truncated density in (4) is essential for achieving high accuracy.

Fig. 10 shows the relation between the D-value and the level of traffic dispersion among the active sessions. We define the *normalized traffic dispersion*, η, as:

$$\eta = \sum_{i=1}^{N} |l_i - \frac{1}{N}|,$$

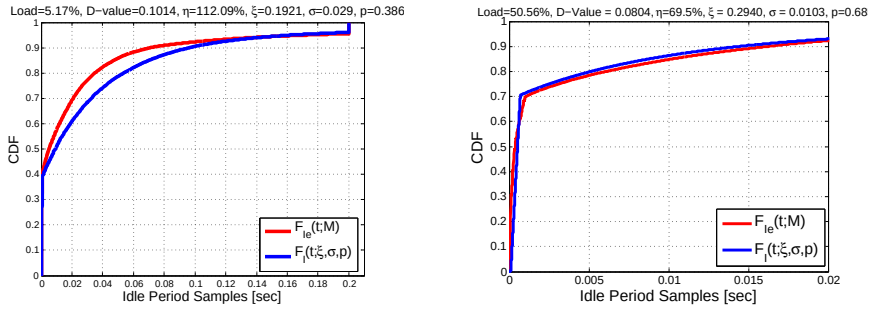


(a) (b)

Figure 8: Campus WLAN. (a) Average D-value with respect to the number of sessions considering (a) all cases and (b) cases with load below 25%, and above 45%.

where N is the number of active sessions in the measurement window, and l_i indicates the ratio of total active duration due to session i, $\sum_{i=1}^{N} l_i = 1$. Thus, η is zero under completely balanced traffic, low when the network traffic is roughly evenly distributed among the sessions, while high η values indicate unbalanced traffic load. Fig. 10(a) suggests that unless it is very high, the traffic dispersion does not have an impact on the fitting quality. However, as it is shown in 10(b), dispersion has different effects in low and in high load regions. For low or moderate channel load the fitting accuracy degrades with increasing traffic dispersion. As at high dispersion the large part of the load is generated by a subset of the sessions, the reason of the low fitting quality is the same as for the low load – low session number case, that is, the idle time distribution is determined directly by the in-flow characteristics. Under high channel load the fitting quality is generally good, apart from the case of very balanced load. We investigate the reason of the high D-value in this case in Fig. 9(b). According to the figure, the empirical and analytic CDFs do not fit at the back-off period interval. High channel load together with balanced traffic means that many sessions access the wireless channel concurrently. This leads to high level of contention and thus to the exponential increase of the user back-off window. The assumption of uniformly distributed back-off period length can not hold in this case, which degrades the performance of the fitting.

Finally, we investigate the impact of transmission errors on the fitting quality. Fig. 10(c) depicts the relation between the D-value and the MAC *error rate*, that is the percentage of packet transmissions, for which an error has occurred, including a CCA failure, that is, terminated transmission attempt after maximum back-off, collision due to hidden terminals, or error in decoding data or ACK packet due to bad channel conditions. Clearly, the fitting quality degrades at high error rate. To determine the reasons behind the fitting performance degradation, we plot in



(a) (b)

Figure 9: Campus WLAN. Comparison between the empirical and the analytic CDF of the idle channel period durations for a single simulation run, considering (a) low WLAN load (5.17%), and (b) high WLAN load (50.56%) with low dispersion ($\eta = 69.5\%$).

Fig. 10(d) the relation between the MAC error rate and the fitting quality for different WLAN load regions. Under low or moderate channel load, the increased MAC error rate, mainly, due to bad channel conditions, does not have an effect on the fitting quality. This is because the packet retransmission scheme of 802.11 does not affect the distribution of the back-off periods and it does not significantly shrink the channel white spaces. As the channel occupancy becomes high, the MAC error rate increases mainly because of CCA failures and collisions due to hidden-terminals, and the increasing D-value reflects that the fitting quality is low due to the non-uniform distribution of the back-off periods, as a consequence of the high-load itself.

We can conclude, that in the scenario of a campus WLAN, the mixture distribution, proposed to characterize the idle period lengths, is accurate for the typical cases with moderate channel load. We have observed that the fitting quality is worse at low load and at low number of sessions or at high level of traffic dispersion among the sessions, because in this case very few packet streams are aggregated and the the in-session flow characteristics determines the idle period distribution. The fitting quality can be low even at very high, well balanced load, and under high MAC error rate, as in these cases the back-off period model is not accurate enough.

5.2.2 Conference-hall WLAN

As Fig. 4(b) shows, the conference-hall WLAN exhibits a set of differences compared to the campus WLAN scenario, with significantly lower variation of the average idle period length and of the percentage of back-offs across the simulation runs. From Fig. 11(a) we can see, that D-values, in general, are lower than in the campus WLAN case. As given in Table 7, their average is $\overline{D} = 0.0088$ and this is reflected

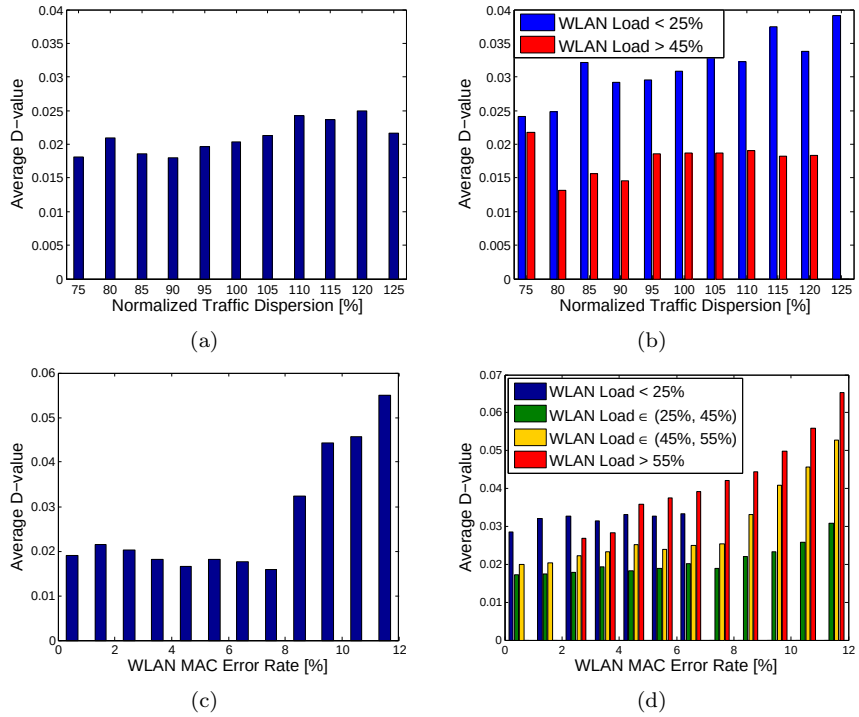
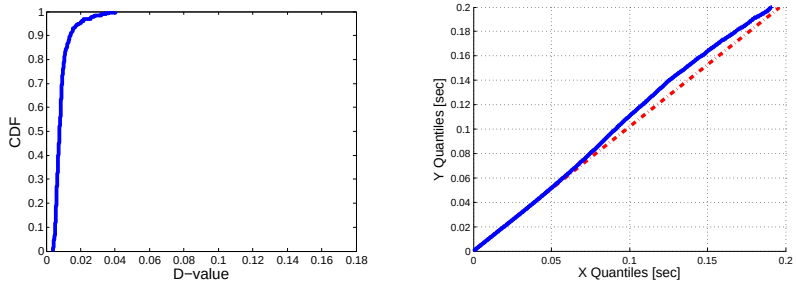


Figure 10: Campus WLAN. Average D-value with respect to the normalized traffic dispersion among the WLAN users considering (a) all cases and (b) cases with load below 25%, and above 45%. and with respect to the MAC error rate, considering (c) all cases and (d) cases with different WLAN load regions.

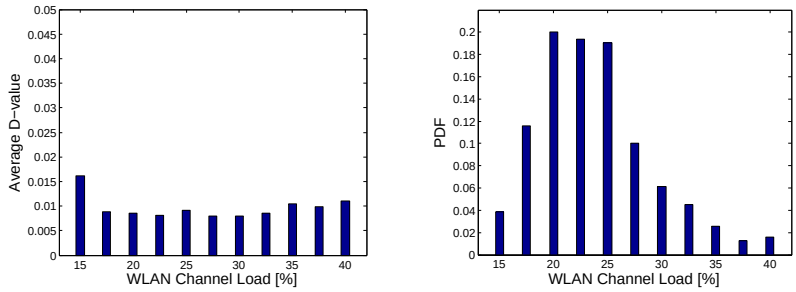
in the respective Kolmogorov-Smirnoff test where the average p-value is 0.5335, but with a low coefficient of variation of 0.0214, and, therefore, a low null hypothesis rejection rate equal to 4.67%. Fig. 11(b) with the Q-Q plot of the collected and synthesized idle period series shows the same good accuracy, with a good fit up to very high time values.

As for the campus scenario, let us evaluate how the load, the number of sessions, the traffic dispersion and the MAC error rate affect the fitting accuracy. Fig. 12(a) depicts the relation between the network load and the resulting D-value. Similarly to the campus WLAN scenario, the fitting accuracy is weaker at very low network load, but after that the D-value becomes low and independent of the load. Moreover, as shown in Fig. 12(b), the WLAN load ranges, mostly, between 15% and 35%, with the majority of the simulation runs showing an average load of 20% – 25%, a range that is significantly shorter than that of the campus WLAN case. The fitting quality is very good under the typical load levels, showing that the analytic idle time model



(a) (b)

Figure 11: Conference-hall. (a) Empirical distribution function of the D-value. (b): Q-Q plot for the empirical and synthesized idle period series over all simulation runs.

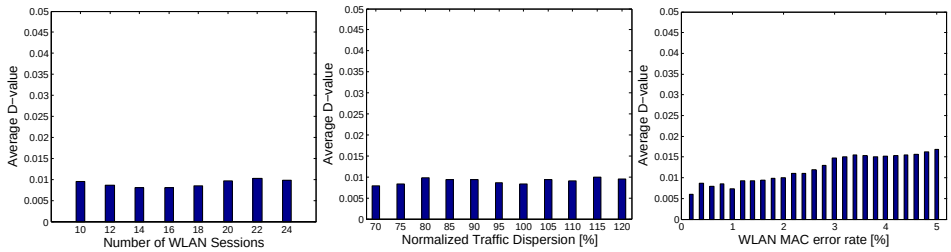


(a) (b)

Figure 12: Conference-hall. (a) Average D-value with respect to WLAN channel load, and (b) Empirical density function of the channel load.

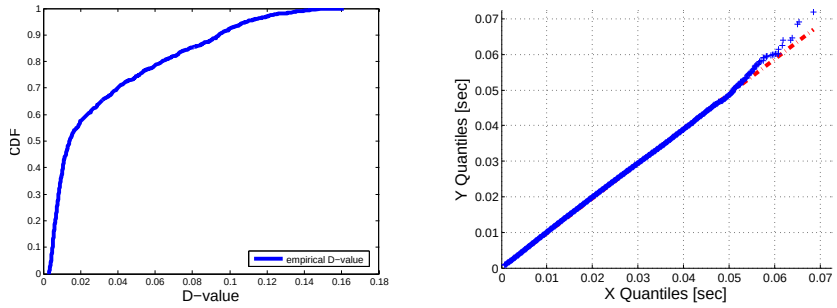
is adequate in the conference-hall scenario. As Fig. 13 shows, the fitting accuracy in the conference-hall case is not affected by the number of sessions or by the load dispersion. Higher MAC error rates lead to a slightly decreased fitting quality. However, contrary to the campus WLAN scenario, both the D-value and the error rates are relatively low. Error rates are low because packet errors and hidden terminals are not as frequent due to better channel conditions – as described in 4.1 – while CCA failures are not present as the load does not reach very high values. Consequently, the low error rates do not significantly affect the fitting accuracy, as discussed earlier in Fig. 10(d).

Comparing the conference-hall results to the campus WLAN ones, we can see that the D-values are lower under similar channel load values, and are less affected by the level of traffic dispersion. This good behavior follows from the in-flow characteristics in the conference-hall scenario. As we can see in Table 3, the traffic is rather



(a) (b) (c)

Figure 13: Conference-hall. Average D-value with respect to (a) the number of sessions, (b) the normalized traffic dispersion among the WLAN users, and (c) the MAC error rate.



(a) (b)

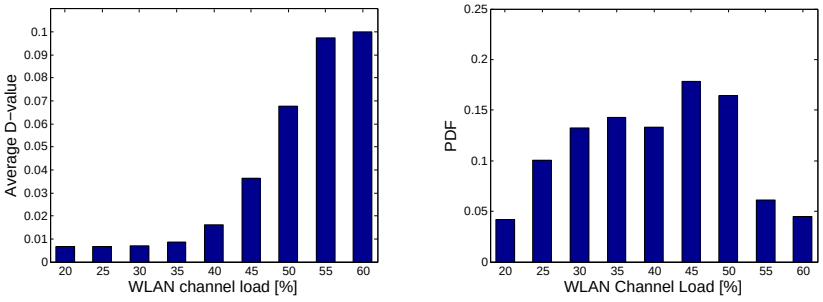
Figure 14: Industrial-plant WLAN. (a) Empirical distribution function of the D-value. (b): Q-Q plot for the empirical and synthesized idle period series over all simulation runs.

homogeneous, dominated by Web flows, each of them with relatively long, exponentially distributed packet inter-arrival times. The mixed uniform-Pareto distribution seems to fit very well the idle-time distribution resulting from the aggregation of a high number of such flows.

We can conclude that the proposed channel occupancy model is accurate for the conference-hall WLAN scenario.

5.2.3 Industrial-plant WLAN

As illustrated in Fig. 4(c) the channel occupancy in the case of the industrial-plant WLAN resembles the conference-hall case, with short idle period durations and low variations in the percentage of back-off idle periods. As shown in Table 7, the average D-value is 0.0321, higher, compared to the campus and conference-

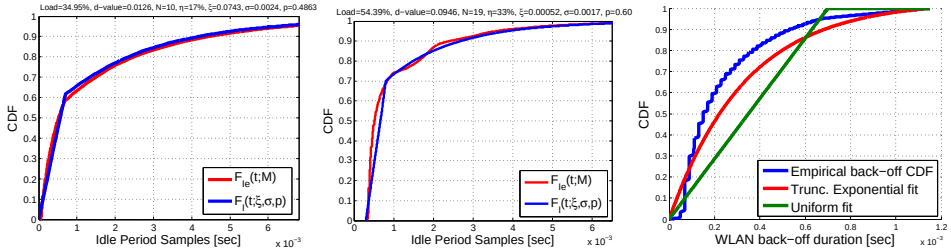


(a) (b)

Figure 15: Industrial-plant WLAN. (a) Average D-value with respect to WLAN channel load, and (b) Empirical density function of the channel load.

hall cases. The average p-value of the K-S goodness-of-fit test is, consequently, the lowest among all cases (0.4029), which, together with a relatively high coefficient of variation (0.2103) results in the highest null hypothesis rejection rate, 9.34%. However, the CDF of the D-value, shown in Fig. 14(a) and the Q-Q plot in Fig. 14(b) still show a good fitting accuracy.

Repeating the evaluation process, we investigate whether the channel load, the number of sessions, the traffic dispersion, or the MAC error rate affect the fitting quality. As shown in Fig. 15(a), contrary to the conference-hall case, the fitting accuracy decreases significantly in cases of high channel load (Fig. 15(a)), moreover, these cases are rather frequent, as shown in Fig. 15(b). We compare the idle period distribution for moderate and high load cases in Fig. 16. Moderate channel load leads to good fitting quality (Fig. 16(a)), even though, due to the periodic monitoring traffic, the idle distribution, $f_{Ie}(t; M)$ is upper bounded, that is, clearly, non heavy-tailed. Fig. 16(b) evaluates the reason for the high D-value under high load. The empirical CDF diverges from the analytic one at very low idle period values, showing that, due to the high contention level, the back-off periods are again non-uniform, as for the Campus case on Fig 9(b). We show the empirical distribution of the idle back-off period sequence in Fig. 16(c). Under high contention, back-off periods extend above the maximum value, $\alpha_{\text{bk}} = 0.7$msec, considered in our model estimation process, which results in fitting error above α_{bk}. However, the larger gap is at the lower time values, where the back-off period distribution is clearly non-uniform. To keep the estimation process feasible, we propose to fit the back-off period distribution with an exponential-like function, which can resemble the linear combination of increasing back-off period length, but requires the estimation of a single parameter. Fig. 16(c) compares the fitting performance of a left-right truncated exponential distribution with the standard uniform density, indicating, clearly, that the former is more capable of capturing the real behavior of the 802.11 back-off periods. We have to notice, however, that the estimator algorithm must



(a) (b) (c)

Figure 16: Industrial-plant WLAN. Comparison between the empirical and the analytic CDF of the idle channel period durations for a single simulation run, considering (a) moderate WLAN load (34%) and (b) high WLAN load (54.39%). (c) compares the exponential-based fitting of the back-off WLAN periods, as opposed to the employed uniform distribution.

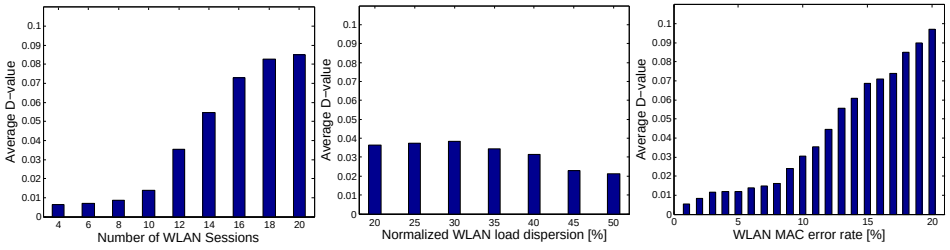
be performed after the estimation of the mixture parameter p, which decreases the achievable accuracy. Still, the empirical CDF diverges from the analytic one even for low white-space period values. We believe that consecutive contention periods under the rather bursty traffic of the industrial plant delay the transmission of consecutive packets for several time-slots, causing the short white-spaces to disappear.

As shown in Fig. 17(a), the fitting accuracy degrades as well with the number of WLAN sessions. However, as in this case the number of sessions and the network load are strongly correlated, the reason for the high D-value is the same as at the high-load case above. The normalized traffic dispersion is low ($\eta \in (20 - 50\%)$), as the monitoring application generates similar traffic on all WLAN sessions (Fig. 17(b)), and has little effect on the model accuracy. According to Fig. 17(c), the MAC error rates can be high, compared to the campus and conference-hall cases, due to the higher transmission rate (11Mbps), that increases the bit error rate probability at larger terminal-AP distances. The fitting accuracy degrades with increasing error rate, as we have seen in the previous scenarios.

To conclude, in the industrial-plant scenario the channel occupancy model is accurate in many cases, but the fitting quality depends heavily on the channel load and the MAC error rate. The mixed uniform-Pareto idle period distribution can be applied with high confidence for networks with up to moderate load and less than 10% MAC error probability.

5.3 Evaluation of the Markovian Assumption

Finally, we investigate whether the durations of the successive WLAN channel idle periods are independently distributed random variables, which is a fundamental assumption for the semi-Markovian occupancy model. For the considered three



(a) (b) (c)

Figure 17: Industrial-plant WLAN. Average D-value with respect to (a) the number of sessions, (b) the normalized traffic dispersion among the WLAN users, and (c) the MAC error rate.

scenarios we evaluate the effect of the channel load, the number of sessions and the traffic dispersion, by applying the test for independence, described in Section 4.2. We consider lag-1 autocorrelation, unless otherwise noted.

For each scenario we conduct 500 simulation runs with randomized traffic workload parameters, thus resulting in different channel load, number of sessions and traffic dispersion values. For each simulation run we collect a sequence of $4 \cdot 10^4$ idle period samples. The long sequence is divided to 100 intervals of 400-samples length. We construct the white noise reference sequence by selecting one sample from each interval. In each interval we select randomly one of the first 100 samples, so as to guarantee both randomness and a minimum time separation of 300 samples. Then we perform the lag-1 test of independence within each of the 400-sample intervals, considering the sub-sequence of the first 100 idle period samples. We finally calculate the p-value of the test over the 100 intervals. Recall, the p-value reflects the probability that the observed correlation metric can occur in a sequence of independent samples. We reject the null hypothesis of independence at a significance level of $\alpha = 5\%$. We repeat the process 100 times by randomizing the starting point of the sub-sequences within the corresponding 400-sample intervals, and calculate the average p-value and the probability that the null hypothesis of independence is rejected, that is, $P(\text{p-value} < 5\%)$.

Fig. 18 shows the average p-value of the independence test for the campus WLAN scenario, along with the percentage of tests with p-value below the 5% significance level. Fig.18(a) shows the p-value with respect to the channel load. Under low or moderate load the test shows a high failure rate, suggesting that the successive idle period durations can not be considered as uncorrelated. The correlation diminishes at high load values. We observe similar trends in Fig. 18(c), where the failure rate is high when the number of sessions is low. Under light load and few sessions only a few flows are intermixed in the AP area. As flows can have very different characteristics in the campus WLAN case, the idle period distribution can change significantly in the event of a flow arrival or departure; this results in correlated successive idle

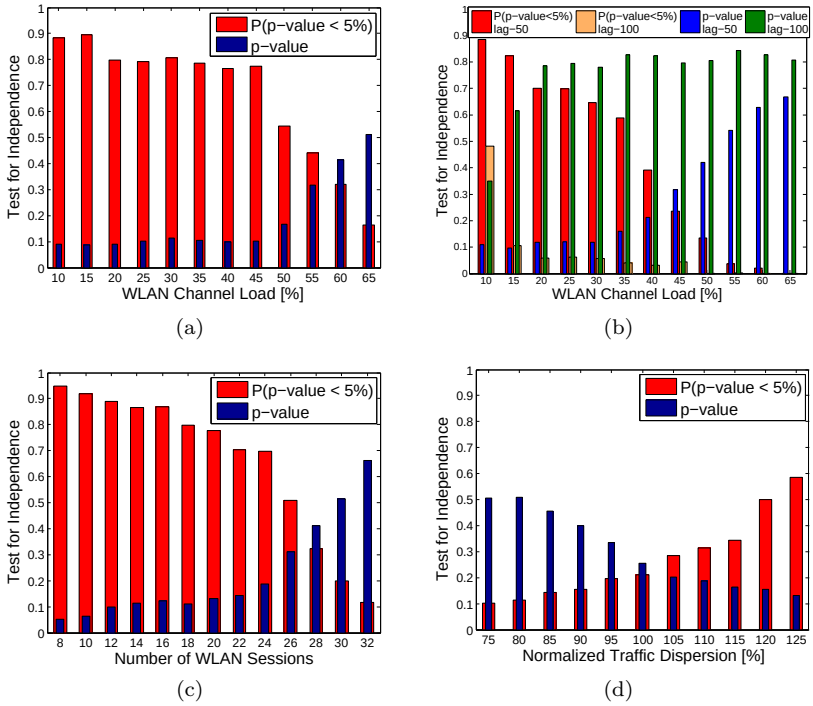
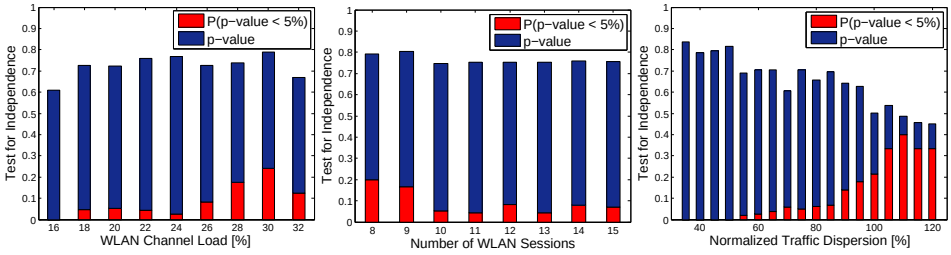


Figure 18: Campus WLAN. Test for lag-1 independence for the idle period process. Average p-value and the percentage of p-values below the 5% significance level, with respect to (a) the WLAN channel load, (c) the number of WLAN sessions, and (d) the normalized traffic dispersion among the sessions. (b) Lag-50 and lag-100 independence tests.

period lengths in parts of the measurement period. At high load, or, as more sessions are active, more flows are transmitted in parallel, and the correlation decreases. Fig. 18(d) shows the results of the independence test as a function of the traffic dispersion. Low level of dispersion typically means several sessions transmitting at an arbitrary point of time, and we experience good independence properties which deteriorate, as the traffic becomes less balanced and a few sessions dominate the traffic.

Since the lag-1 correlation is high in many of the considered cases, it is worth to evaluate how the correlation changes at increased temporal separation. Fig. 18(b) presents the results of the independence test when the autocorrelation is calculated for lag-50 and lag-100. We observe that the correlation remains significant even between temporal separation of 50 samples (lag-50), and it only degrades drastically for lag-100, when the set of active flows is usually changed. This suggests



(a) (b) (c)

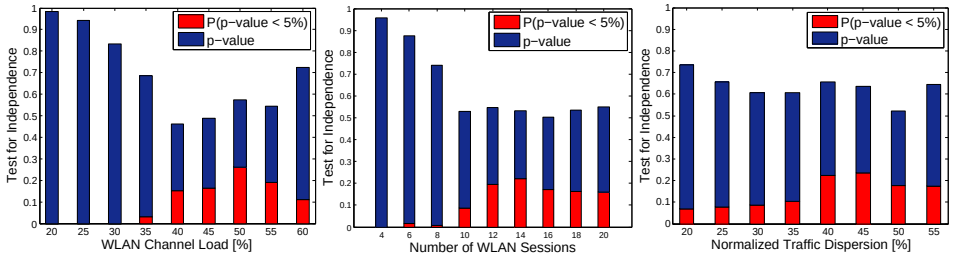
Figure 19: Conference-hall. Test for independence for the idle period process. Average p-value and the percentage of p-values below the 5% significance level, with respect to (a) the WLAN channel load, (b) the number of WLAN sessions, and (c) the normalized traffic dispersion among the sessions.

that idle periods need to have a high temporal distance to be safely considered as independent.

Fig. 19 depicts the p-value of the independence test for the conference-hall scenario. As shown in Fig. 19(a),(b), we observe relatively high p-values, that are only marginally affected by the channel load or the number of sessions. The effect of dispersion, as shown in Fig. 19(c), is similar to the one in the campus WLAN scenario (Fig. 18(d)), that is, higher traffic dispersion decreases the p-value which, however, remains at a rather high level. According to our simulations, the probability of experiencing high dispersion is low, ($P(\eta > 110) < 0.05$), and consequently there are very few test results in the high dispersion range. This allows both the p-value and the hypothesis rejection probability to be around 0.5 in this range.

In general, the hypothesis on the independence of the consecutive idle periods may hold in the conference-hall scenario. Contrary to the campus WLAN case, the majority of the traffic is from Web flows with identical packet inter-arrival processes. Additionally, flow arrival intensities are higher, leading to a high number of concurrent flows intermixed in the AP area; the above reasons result in an aggregated idle period sequence with low autocorrelation.

Fig. 20 presents the results of the industrial-plant scenario. Here the resulting p-values are high, particularly at low and moderate load cases (which also implies low number of sessions in this scenario), as a result of the constant number of sessions and flows with very similar characteristics (Fig. 20(a),(b)). The p-value drops at higher load, when increased contention introduces consecutive back-off idle periods whose statistical properties differ significantly from those of the packet inter-arrival process. The p-value increases again at very high load, when most of the idle-periods are back-offs. Finally, Fig. 20(c) indicates that the idle sequence autocorrelation is almost indifferent to the level of traffic dispersion, which is expected, since the



(a) (b) (c)

Figure 20: Industrial-plant WLAN. Test for independence for the idle period process. Average p-value and the percentage of p-values below the 5% significance level, with respect to (a) the WLAN channel load, (b) the number of WLAN sessions, and (c) the normalized traffic dispersion among the sessions.

	D-value	p_{KS}	$C_{p_{\mathrm{KS}}}$	$P\{p_{\mathrm{KS}} \leq 5\%\}$
CAFETERIA	0.0322	0.4932	0.0903	0.0783
LIBRARY	0.0356	0.5332	0.1032	0.0888
PSU_CS	0.0414	0.3903	0.1895	0.1439
PIONEER	0.0399	0.4598	0.1421	0.0912
POWELLS	0.0674	0.0694	0.0637	0.5532
UG	0.0388	0.3802	0.1453	0.0941

Table 8: Summary of goodness-of-fit evaluation for the considered real trace-based evaluations

traffic dispersion range is very short.

To summarise, the Markovian assumption holds with high probability in the conference-hall scenario and in the industrial-plant WLAN, unless the channel load, and thus the probability of contention are high. In the campus-WLAN scenario, however, the Markovian assumption is only justified under high channel load, when a high number of traffic flows are intermixed. The idle sequence autocorrelation decreases drastically only for a temporal separation in the order of 100 samples. Nevertheless, based on (1) and the experimental results in Fig. 4(a), the corresponding time distance lies below 1 second. This means that the Markovian assumption can be applied for relevant protocol design, e.g. in the case of the coexistence of WLANs and wireless sensor networks with long duty-cycle medium access schemes.

6 Model Validation over Real WLAN Traces

This Section investigates the ability of the proposed 802.11 channel occupancy model to capture the statistical behavior of a set of real WLAN channel usage data. We experiment with a set of real, high time-resolution 802.11 traces, captured with

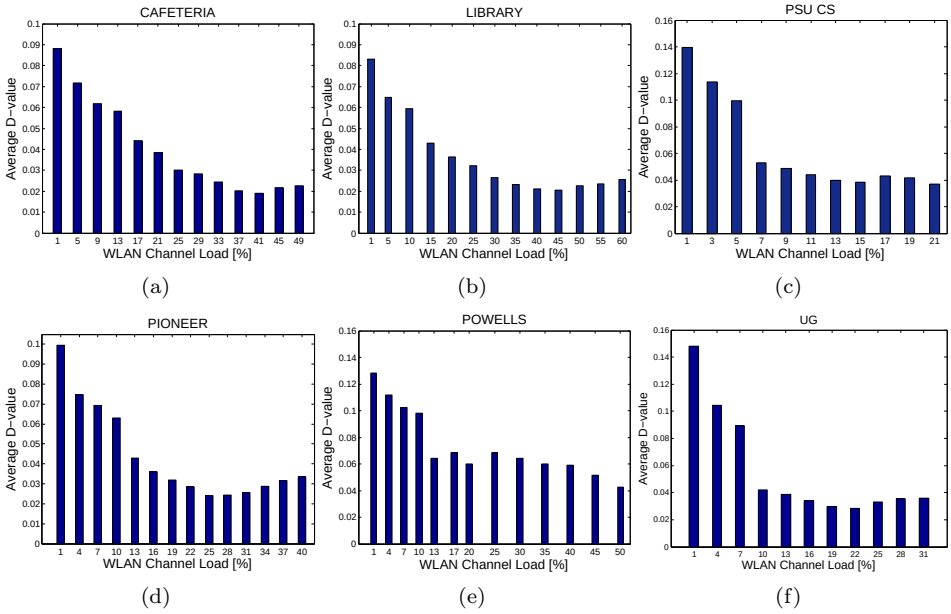
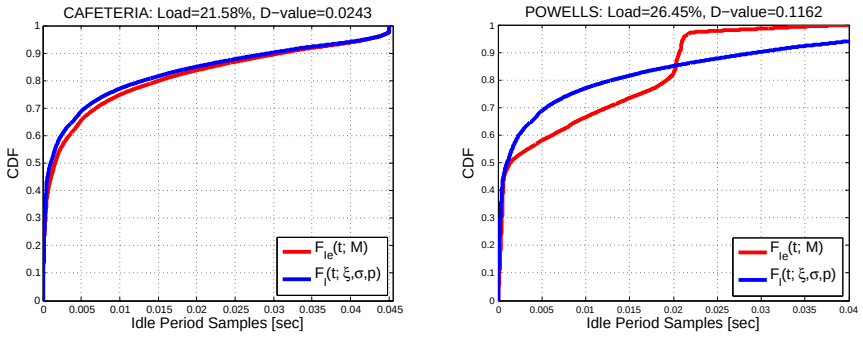


Figure 21: Aggregate results for fitting accuracy for the considered real WLAN traces

a commercial sniffer device in diverse WLAN environments [32]. The considered data set includes traces from campus wireless networks, namely a campus coffee spot (`CAFETERIA`), a university library building (`LIBRARY`) and a university office building (`PSU_CS`), allowing direct comparison with the results with the synthetic Campus WLAN traffic. The dataset includes, additionally, traces from a public hot-spot (`PIONEER`) and two coffee places (`POWELLS`, `URBAN GRIND (UG)`).

The considered traces provide a high (nano-second) resolution timing information on packet arrivals and frame-in-the-air durations as well as MAC control information including beaconing, and 802.11 control packet transmissions. Thus their level of detail is appropriate for generating channel occupancy statistics, and, consequently, for our objective of WLAN channel usage characterization. For all traces there exists an unfiltered version containing all frames correctly deciphered by the radio capturing device as well as a filtered-by-BSSID ("*pcap*") version limiting the captured trace to the traffic associated with the considered WLAN hot-spot. Note, however, that the higher-layer traffic characteristics are not known.

As the traces provide idle period sequences that are in the order of 10^5 samples, we partition them into shorter sub-sequences of $4 \cdot 10^4$ samples each. We perform the model parameterization and fitting validation for all sub-sequences and extract aggregate statistical results, evaluating both the accuracy of the idle-period fitting,



(a) CAFETERIA (b) POWELLS

Figure 22: Examples of empirical distribution fitting for the LIBRARY and the POWELLS traces.

	p-value	$P(p\text{-value} < 5\%)$
CAFETERIA	0.6592	0.0842
LIBRARY	0.8931	0.0702
PSU_CS	0.3931	0.1977
PIONEER	0.5312	0.0951
POWELLS	0.2921	0.3982
UG	0.1932	0.2830

Table 9: Summary of the test for independence for the idle period process in the considered real trace-based evaluations

as well as the validity of the Markovian assumption.

Fig. 21 depicts the fitting accuracy of the trace set with respect to the average channel load. In the majority of the cases the trends are similar to the ones with the synthetic traces.

Table 8 summarizes the fitting accuracy results for the considered traceset. With the exception of the `POWELLS` case the proposed modeling of the idle period duration can effectively capture the statistical behavior of the idle period traces. This is verified by both the low D-value (Fig. 21) of the cdf-fitting process, as well as the low fail rate ($P(p\text{-value} < 5\%)$) of the related K-S test. The resulting $P(p\text{-value} < 5\%)$ for the `LIBRARY` and `CAFETERIA` complies with the fail-rate evaluated for the Campus WLAN scenario (Tab. 7), while for the `PSU_CS` case it is higher, due to the significantly lower WLAN load.

Fig. 22 illustrates a comparative fitting example for sub-sequences taken from the `LIBRARY` and `POWELLS` traces. As shown in Fig. 22(b), the `POWELLS` trace reflects a rather unexpected, significant nearly-periodic WLAN activity, resulting in a high number of WLAN idle periods around 20msec, which does not allow for an accurate fitting with the generalized Pareto-based white-space distribution.

Finally Table 9 summarizes the results of the evaluation of the Markovian assumption. For most of the traces the resulting fail rates of the test for independence and the related p-value are comparable to the ones of the synthetic networking scenarios. As for the fitting accuracy test, the `POWELLS` case with its periodic traffic exhibits a high fail. The fail rate is rather high even for the `UG` rate, showing again that the Markovian assumption has to be handled with care.

7 Conclusions

The modeling of the radio channel occupancy becomes a key issue for the design of future wireless networks sharing common spectrum bands. In this paper we considered the special case of IEEE 802.11 WLANs channel occupancy, since these networks are wide-spread and emerging networks have to adapt their access strategies to WLAN presence.

We addressed the question whether semi-Markovian channel occupancy models already proposed in the literature, but validated only for very limited use cases, can be applied in realistic networking environments. As controlled experiments with large parameter sets are hard to conduct in real testbeds, we performed our study via simulations. For the simulations we selected three networking scenarios, with significantly different traffic workload, the university campus, the conference-hall, and the industrial-plant. We performed detailed study to evaluate whether the proposed WLAN channel occupancy model with heavy-tail idle time distribution and the Markovian assumption on the independence of the consecutive idle times are valid.

Considering the proposed idle time distribution, we can conclude that the accuracy is affected by the traffic mix and the network load. The white space distribution is satisfactory in most of the cases, however, it is not very accurate when the traffic is very heterogeneous and the load is low, or under heavy, nearly periodic traffic. The assumption on uniformly distributed back-off periods necessarily fails under high contention level, for example in cases where the high MAC error rate and consequent retransmissions moves the network to the high load regime. The Markovian assumption holds for many of the considered scenarios, but fails again when the traffic is very heterogeneous and the load is low.

Clearly, both the traffic mix and the networking technology change with time, therefore, it is necessary to discuss the generality of our results. New services introduce flows with new in-flow characteristics, and the weight of the different flow types changes with time. Based on our results we can predict with confidence, that the proposed channel occupancy model will hold in future networking scenarios as well, apart from the cases when the load is low, with some dominant flows. In those cases the in-flow characteristics of the dominant flows determines the idle-time distribution. New networking technologies are expected to increase the transmission rate in general and to use efficient physical layer techniques that will increase spec-

trum efficiency. As our results show, these changes do not affect the accuracy of the proposed channel occupancy model. Power saving options with duty cycling may introduce periodicity in the channel access of some terminals with low traffic. This will not affect the channel occupancy characteristics if the aggregate load of the AP is not very low.

References

[1] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, 2007.

[2] X. Li, Q. Zhao, X. Guan, and L. Tong, "Optimal Cognitive Access of Markovian Channels under Tight Collision Constraints," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 746–756, 2011.

[3] K. W. Choi and E. Hossain, "Opportunistic Access to Spectrum Holes Between Packet Bursts: A Learning-Based Approach," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2497–2509, 2011.

[4] S. Pollin, I. Tan, B. Hodge, C. Chun, and A. Bahai, "Harmful Coexistence Between 802.15.4 and 802.11: A Measurement-based Study," in *Proceedings of ICST International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM '08)*, pp. 1–6, May 2008.

[5] S. Geirhofer, J. Z. Sun, L. Tong, and B. M. Sadler, "Cognitive frequency hopping based on interference prediction: theory and experimental results," *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 13, pp. 49–61, September 2009.

[6] J. Huang, G. Xing, G. Zhou, and R. Zhou, "Beyond Co-existence; Exploiting WiFi White Space for ZigBee Performance Assurance," in *Proceedings of IEEE International Conference on Network Protocols*, 2010.

[7] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless LAN," *SIGMETRICS Perform. Eval. Rev.*, vol. 30, pp. 195–205, June 2002.

[8] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Usage patterns in an urban WiFi network," *IEEE/ACM Trans. Netw.*, vol. 18, pp. 1359–1372, Oct. 2010.

[9] A. Ghosh, R. Jana, V. Ramaswami, J. Rowland, and N. Shankaranarayanan, "Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots,"

in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 2921 –2929, april 2011.

[10] G. He, J. C. Hou, W.-P. Chen, and T. Hamada, "Characterizing individual user behaviors in WLANs," in *Proceedings of ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems (MSWiM '07)*, pp. 132–137, 2007.

[11] X. G. Meng, S. H. Y. Wong, Y. Yuan, and S. Lu, "Characterizing flows in large wireless data networks," in *Proceedings of ACM International Conference on Mobile Computing and Networking (MobiCom '04)*, pp. 174–186, 2004.

[12] H. Feng, Y. Shu, and O. W. Yang, "Research on characterization of wireless LANs traffic," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 8, pp. 3179 – 3187, 2011.

[13] F. Wamser, R. Pries, D. Staehle, K. Heck, and P. Tran-Gia, "Traffic characterization of a residential wireless internet access," *Telecommunications Systems, Springer*, vol. 48, no. 1-2, pp. 5–17, 2010.

[14] J.-K. Lee and J. C. Hou, "Modeling steady-state and transient behaviors of user mobility: formulation, analysis, and application," in *Proceedings of ACM International Symposium on Mobile ad hoc Networking and Computing (MobiHoc '06)*, pp. 85–96, 2006.

[15] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proceedings of ACM International conference on Mobile systems, applications and services (MobiSys '03)*, pp. 303–316, 2003.

[16] F. Hernández-Campos, M. Karaliopoulos, M. Papadopouli, and H. Shen, "Spatio-temporal modeling of traffic workload in a campus WLAN," in *Proceedings of EAI International Workshop on Wireless Internet (WICON '06)*, 2006.

[17] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535 –547, march 2000.

[18] P. Rathod, O. Dabeer, A. Karandikar, and A. Sahoo, "Characterizing the exit process of a non-saturated IEEE 802.11 wireless network," in *Proceedings of ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '09)*, pp. 249–258, 2009.

[19] J. Misic and V. Misic, "Characterization of idle periods in IEEE 802.11e networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '11)*, pp. 1004 –1009, March 2011.

[20] L. Stabellini, "Quantifying and Modeling Spectrum Opportunities in a Real Wireless Environment," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '10)*, pp. 1 –6, April 2010.

[21] C. Ghosh, S. Pagadarai, D. Agrawal, and A. Wyglinski, "A framework for statistical wireless spectrum occupancy modeling," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 38 –44, january 2010.

[22] S. Geirhofer, L. Tong, and B. Sadler, "Cognitive radios for dynamic spectrum access - Dynamic Spectrum Access in the Time Domain: Modeling and Exploiting White Space," *IEEE Communications Magazine*, vol. 45, pp. 66–72, May 2007.

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

[24] I. Glaropoulos, V. Fodor, L. Pescosolido, and C. Petrioli, "Cognitive WSN transmission control for energy efficiency under WLAN coexistence," in *Proceedings of ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM' 11)*, pp. 261 –265, june 2011.

[25] M. Lagana, I. Glaropoulos, V. Fodor, and C. Petrioli, "Modeling and Estimation of Partially Observed WLAN Activity for Cognitive WSNs," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC' 12)*, April 2012.

[26] C. Ciflikli, A. Gezer, A. T. Ozsahin, and O. Ozkasap, "Bittorrent packet traffic features over IPv6 and IPv4," *Simulation Modelling Practice and Theory*, vol. 18, no. 9, pp. 1214 – 1224, 2010.

[27] D. Bonfiglio, M. Mellia, M. Meo, and D. Rossi, "Detailed Analysis of Skype Traffic," *IEEE Transactions on Multimedia*, vol. 11, pp. 117 –127, jan. 2009.

[28] "EU FP7 HYDROBIONETS Deliverable D2.2 - Scenario descriptions and system requirements (v2)," 2013. http://www.hydrobionets.eu/index.php/deliverables.

[29] *NS-Miracle: Multi-InteRfAce Cross-Layer Extension library for the Network Simulator*. http://telecom.dei.unipd.it/pages/read/58/.

[30] T. Chrysikos, G. Georgopoulos, and S. Kotsopoulos, "Site-specific validation of ITU indoor path loss model at 2.4 GHz," in *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks Workshops, (WoWMoM '09)*, pp. 1–6, 2009.

[31] J.-Y. L. Boudec, *Performance evaluation of computer and communication systems*. EPFL Press, 2010.

[32] C. Phillips and S. Singh, "Analysis of wlan traffic in the wild," in *Proceedings of the 6th International IFIP-TC6 Conference on Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet*, NETWORKING'07, (Berlin, Heidelberg), pp. 1173–1178, Springer-Verlag, 2007.

Paper **E**

Enhanced power saving mode for low-latency communication in multi-hop 802.11 networks

Vladimir Vukadinovic, Ioannis Glaropoulos, and Stefan Mangold

Published in Elsevier Journal of Ad Hoc Networks, vol. 23, no. 0, pp. 18-33, 2014.

Enhanced Power Saving Mode for Low-Latency Communication in Multi-Hop 802.11 Networks

Vladimir Vukadinovic, Ioannis Glaropoulos, and Stefan Mangold

Abstract

The Future Internet of Things (IoT) will connect billions of battery-powered radio-enabled devices. Some of them may need to communicate with each other and with Internet gateways (border routers) over multi-hop links. While most IoT scenarios assume that for this purpose devices use energy-efficient IEEE 802.15.4 radios, there are use cases where IEEE 802.11 is preferred despite its potentially higher energy consumption. We extend the IEEE 802.11 Power Saving Mode (PSM), which allows WLAN devices to enter a low-power doze state to save energy, with a traffic announcement scheme that facilitates multi-hop communication. The scheme propagates traffic announcements along multi-hop paths to ensure that all intermediate nodes remain awake to receive and forward the pending data frames with minimum latency. Our simulation results show that the proposed Multi-Hop PSM (MH-PSM) improves both end-to-end delay and doze time compared to the standard PSM; therefore, it may optimize WLAN to meet the networking requirements of IoT devices. MH-PSM is practical and software-implementable since it does not require changes to the parts of the IEEE 802.11 medium access control that are typically implemented on-chip. We implemented MH-PSM as a part of a WLAN driver for Contiki OS, which is an operating system for resource-constrained IoT devices, and we demonstrated its efficiency experimentally.

Index terms— Internet of Things; IEEE 802.11; power saving; ad hoc networks; multi-hop networks

1 Introduction

Nowadays almost every desktop computer, laptop, tablet, and smartphone is connected to the Internet. The emergence of the Internet of Things (IoT) will provide global IP connectivity to a broader variety of devices, such as entertainment electronics, wearable sport gadgets, home appliances, and industrial sensors. Some of these devices are portable, battery-powered, and need to connect wirelessly to surrounding devices and Internet gateways. The wireless communication may significantly contribute to their overall battery consumption, especially in the case of

constrained embedded devices. Therefore, minimizing the energy consumption of wireless interfaces and networking protocols is one of the prerequisites for the IoT.

Figure 1: Application scenario in which smart radio-enabled toys communicate with decorative lightning (© Disney).

Different wireless standards have been proposed for IoT. Zigbee, which is based on the IEEE 802.15.4 standard [1], is often referred to as a wireless technology of choice for home and building automation, smart metering, and IoT in general because of its simplicity and energy-efficiency. Z-Wave [2] is another technology that targets similar applications and environments with emphasis on home automation. Both Zigbee and Z-Wave provide meshing capabilities, which are required by many IoT applications. Although it does not support meshing, Bluetooth Low Energy (BLE) is also a candidate technology for IoT. The advantages of BLE are the low energy consumption and the common presence in smartphone radio stacks. These technologies, however, do not cover the entire spectrum of IoT devices and applications. Wi-Fi, which is based on IEEE 802.11 standard [3], dominates the consumer electronics segment: Any IoT device that needs to connect to smartphones, tablets,

TVs, set-top boxes, game consoles, and toys would benefit from Wi-Fi connectivity. Also, some sensors that operate at high sampling rates, such as those used in seismic monitoring and imaging, may generate large amounts of data that cannot be transmitted using ZigBee due to its limited throughput, but can easily be transmitted by Wi-Fi. The economy of scale and the possibility of reuse of the existing Wi-Fi infrastructure offer key cost savings and facilitate faster deployment with Wi-Fi than with competing technologies. Furtheremore, Wi-Fi has the advantage of native compatibility with IP, which is the key enabler for IoT: IP eliminates the need for expensive gateway solutions to connect IoT devices to the Internet. The feasibility of connecting battery powered sensors to the IoT using commercially available Wi-Fi chips has been demonstrated in [4]. In [5], the authors share their experiences of using off-shelf Wi-Fi modules to connect *things* to the Web of Things.

One of the key challenges for the use of Wi-Fi in IoT objects is its energy consumption, which is relatively high compared to ZigBee. An always-on Wi-Fi interface may quickly drain the battery of a device. Long battery recharge/replacement cycles are preferred for cost and convenience reasons. For example, a survey has shown that 51% of electronic toy consumers are concerned about the battery replacement costs [6]. There have been some notable improvements in hardware and many low-power Wi-Fi chips with energy-efficient radio transceivers have appeared on the market. The 802.11 MAC protocol, however, is inherently energy-hungry. One of the major sources of unnecessary energy consumption in 802.11 MAC is idle listening, which consumes energy even when there is no traffic in the network — the radio must perform idle listening continuously in order to detect arriving packets. The energy consumption of idle listening in 802.11 is comparable to that of packet transmission and reception [7]. To alleviate the problem, the 802.11 standard [3] specifies a Power-Saving Mode (PSM) that allows an idle 802.11 station to transition to a low-power doze state by switching off its radio transceiver. The role of 802.11 PSM is similar to that of Radio Duty Cycling (RDC) in 802.15.4. There are some notable differences: RDC typically operates below MAC, directly on top of the 802.15.4 PHY layer. It may include information from the MAC layer, in which case MAC and RDC are cross-optimized as in [8], but it can also be isolated from MAC. With RDC, a radio can be switched on and then rapidly switched off after a few milliseconds if no activity is detected on the channel. The 802.11 PSM is part of the MAC layer management entity. The intervals in which PSM alternates between doze and awake states are typically measured in tens and hundreds of milliseconds: All 802.11 stations wake up synchronously at the beginning of a beacon interval, listen for traffic announcements from other stations that have data packets destined to them, and announce their own data packets (if any) destined to other stations. If a station does not receive any traffic announcements and it does not have any buffered packets that need to be transmitted in the current beacon interval, it returns to the doze state.

The IEEE 802.11 standard specifies the details of PSM for the infrastructure/BSS mode (Basic Service Set with an access point) and the ad hoc/IBSS mode (Inde-

pendent Basic Service Set without an access point). Since it has been originally designed for single-hop communication in the infrastructure mode (from the access point to a station and vice versa), the PSM performs poorly in the ad hoc mode, especially in multi-hop networks [9, 10, 11]. When a data frame is forwarded over multiple hops, the PSM may significantly increase the delivery delay because only the next-hop station is notified about the pending frame via traffic announcements, while the stations on subsequent hops may remain in the doze state. Therefore, in each beacon interval the frame is forwarded over a single hop and has to be buffered before being forwarded further. Depending on the number of hops, the end-to-end delay may be long enough to affect time-sensitive applications. Another problem of PSM is that a station is occasionally forced to stay awake even though it has no frames to transmit or receive. The reason to stay awake is to respond to probe requests of devices that are actively scanning the medium when attempting to discover and join networks. For example, if there are two stations in an 802.11 ad hoc network, at least one of them would have to remain awake at any time, which limits the sleep time to at most 50%. Hence, the 802.11 PSM is not suitable for low-energy low-latency multi-hop communication, which is a common requirement for the IoT.

In this paper, we address the problem of increased frame delays due to PSM in multi-hop ad hoc networks. We propose a mechanism that wakes up downstream stations so that data frames can be forwarded over multiple hops in a single beacon interval. This is achieved by instructing each station along the path to forward the traffic announcement to its downstream neighbor. The proposed mechanism significantly reduces the end-to-end delay, especially for bursty traffic where intermediate stations may move to the doze state between two consecutive traffic bursts. We also question the 802.11 standard requirement for a station to stay awake to respond to probe requests. We describe a mechanism that enables actively scanning stations to discover an ad hoc network in which no station is required to stay awake for the entire duration of the beacon interval. The proposed mechanisms enhances the standard PSM to what we call Multi-Hop PSM (MH-PSM). MH-PSM does not prevents stations to inter-operate with those that employ standard PSM since it does not alter the state machine, the frame formats, and other important protocol mechanisms. MH-PSM is also software-implementable: It does not require modifications to the parts of the 802.11 MAC protocol that are usually implemented on-chip, such as the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) medium access protocol. We implemented MH-PSM as a part of our WLAN driver for the Contiki operating system [12]. The paper provides a practical demonstration that, with few simple modifications, WLAN ad hoc mode may become a compelling technology for some IoT applications. A concise version of this paper was published in [13].

The rest of the paper is organized as follows: Section 2 provides an overview of the standard 802.11 PSM. In Section 3, we describe MH-PSM and discuss deployment and standard compatibility issues. The performance of MH-PSM is evaluated

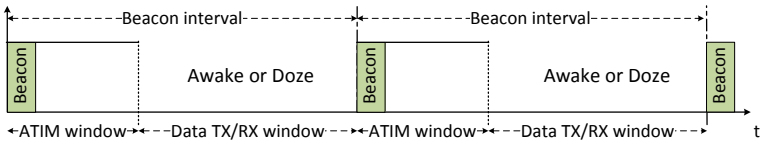


Figure 2: From [13]. The 802.11 PSM divides time into beacon intervals.

in Section 4 using simulations, and in Section 5, we describe our testbed implementation of MH-PSM and experimental results. Section 6 gives a brief overview of related work. Finally, Section 7 concludes the paper.

2 Power-saving mode for 802.11 ad hoc networks

In the standard 802.11 PSM for ad hoc/IBSS networks, time is divided into periods called beacon intervals. Each station wakes up at the beginning of a beacon interval and starts a back-off procedure in an attempt to transmit a beacon. If a station receives a beacon from another station before its back-off timer expires, it cancels the pending beacon transmission. The Timing Synchronization Function (TSF) uses the time-stamped beacons to synchronize clocks among stations, to ensure that all stations wake up at the same time. Following the beacon exchange, each station stays awake for an ATIM window interval as shown in Fig. 2. During the ATIM window, stations announce pending data frames to their neighbors using unicast Announcement Traffic Indication Messages (ATIMs). ATIMs are sent using the 802.11 Distributed Coordination Function (DCF) that operates with a CSMA/CA channel access procedure. A station that receives an ATIM should respond with an ACK. Successful exchange of ATIM-ACK packets between two stations implies that they can now exchange any pending data frames and thus both should stay awake until the next beacon interval. Stations that neither send nor receive any ATIM frame during the ATIM window will move to the doze state for the rest of the beacon interval. After the end of an ATIM window, all stations that remain awake will follow the normal DCF procedure to transmit and receive data frames.

The described PSM protocol has many drawbacks: It uses DCF, which may waste scarce battery resources and bandwidth due to frame collisions and increase the frame delay due to back-offs. A station that has pending data frames must estimate if the receiving station is using PSM. ATIMs should be sent only to stations that are using PSM. Stations that are not in PSM will not respond with an ACK, which will trigger undesirable re-transmissions. The standard however does not specify how to estimate if a station is using PSM or not. When a station successfully transmits or receives an ATIM frame during an ATIM window, it must stay awake during the entire rest of the beacon interval. At low loads, this approach results in a higher energy consumption than necessary. Another shortcoming is that all

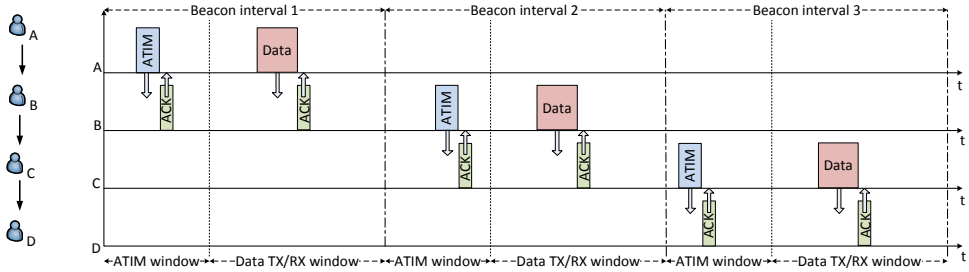


Figure 3: From [13]. Multi-hop forwarding in standard 802.11 PSM may cause a delay of several beacon intervals.

stations in an IBSS must use the same fixed ATIM window size, which is set at the time when the IBSS is created, as well as identical beacon intervals. Since the ATIM window size critically affects the throughput and energy consumption, the fixed ATIM window does not perform well in all situations, as shown in [14]. Some of these drawbacks have been addressed in previous work, which are discussed in Section 6. This paper, however, addresses the problem of end-to-end delay on multi-hop paths, as described in the following.

Consider a typical multi-hop scenario where station A needs to send a single-frame message to station D using intermediate stations B and C as relays (Fig. 3). In the first beacon interval, station A announces the data frame to station B using an ATIM. Station B acknowledges the ATIM an remains awake so that it can receive the data in the period that follows the ATIM window. Assume that station C has not received any traffic announcements and, therefore, it enters the doze state. Since station B is not able to forward the frame to station C in the current beacon interval, it has to wait for the start of the next beacon interval to send an ATIM to station C. Following a successful ATIM-ACK exchange, the frame is forwarded to C. Station D will receive the frame in the third beacon interval. The resulting increased end-to-end delay may considerably affect applications with strict latency constraints, which is undesirable. Therefore, enabling PSM in multi-hop ad hoc networks must be combined with effective mechanisms for mitigating its effect on the resulting packet delays.

3 Enhanced 802.11 PSM for Multi-Hop Communication

In the scenario described above, the data frame sent by A must be buffered at B before it is relayed to C in the following beacon interval. This could be avoided if there was a way for B to, upon receiving the ATIM from A, send an early ATIM

to C and D to inform them about the pending data frame at A. This is what our low-latency multi-hop PSM (MH-PSM) aims to achieve.

Before introducing MH-PSM, we describe the format of ATIM frames. An ATIM frame contains a MAC header, whose structure is common to all management frames as shown in Fig. 4. The frame body of an ATIM is empty. The header includes three address fields: Address 1 contains the MAC address of the ATIM receiver. Address 2 contains the MAC address of the ATIM sender. Address 3 may contain different information depending on the type of the management frame and network mode (BSS, IBSS, or mesh). The Address 3 field of ATIM frames contains the BSSID (BSS identifier) of the IBSS. In case of *group-addressed* (i.e., broadcast) ATIMs, the BSSID is used to verify that the frame originated from a station in the IBSS of which the receiving station is a member. In case of individually addressed (i.e., unicast) ATIMs, the BSSID is not used at the receiver [3].

3.1 Proposed Extension: Multi-Hop PSM (MH-PSM)

We propose that, in order to inform all stations along the path to D about the pending data frame, the station A writes the MAC address of D into the Address 3 field of the ATIM frame sent to B. Methods to resolve the MAC address of D from its IP address are discussed later in this section. Upon receiving the ATIM, B inspects the Address 3 field to derive the final destination of the data frame announced by that ATIM. It retrieves the MAC address of D from the Address 3 field, resolves it to the IP address of D, and consults the routing table to find out that C is the next hop on the path to D. Then B creates an ATIM frame for C with the MAC address of D inside the Address 3 field. When C receives the ATIM from B, it uses the same procedure to create an ATIM for D. In this way, a chain of ATIM transmissions is created along the multi-hop path to wake up all relays and the destination of the data frame. Following the end of the ATIM window, the data frame can be forwarded end-to-end in the current beacon interval since all stations on the path are in the awake state. The procedure is illustrated in Fig. 5. The ATIM chain may not reach the end destination: It may terminate at the end of the ATIM window or upon reaching a station that cannot resolve the MAC address of the destination. In that case, the data frame will be forwarded as far as the furthest station that has

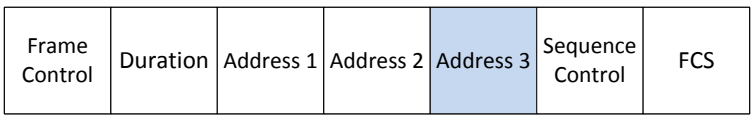


Figure 4: From [13]. Structure of the ATIM frame. The Address 3 field can be used for the MAC address of the end destination.

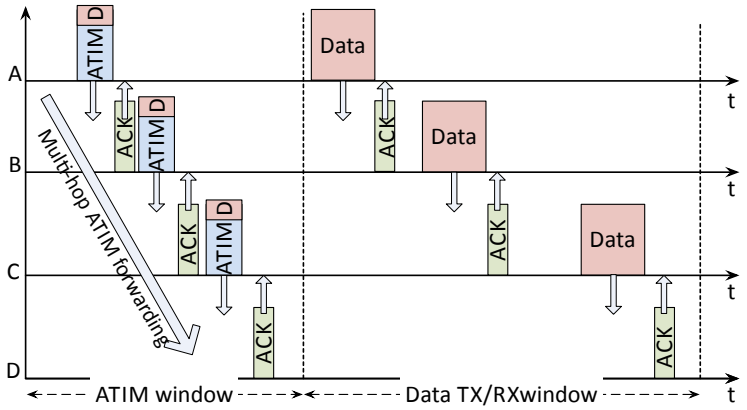


Figure 5: From [13]. The proposed multi-hop forwarding mechanism allows data frames to be forwarded end-to-end in a single beacon interval.

received the ATIM. Nevertheless, MH-PSM may significantly decrease the end-to-end delay in lightly-loaded multi-hop networks because, unlike with the standard PSM, data frames are forwarded over multiple hops in a single beacon interval.

3.2 Address 3 Resolution

The sending station A needs to write the MAC address of the destination D into the Address 3 field of the ATIM sent to B. Therefore, A needs to resolve the MAC address of D from its IP address. Since the paper targets Internet of Things (IoT) and smart toy communication scenarios, we assume that IPv6 is used. The IPv6 protocol suite includes the Neighbor Discovery (ND) protocol [15], which provides address resolution, next-hop determination, and duplicate address detection. Address resolution enables stations to determine MAC addresses of their neighbors given only their IP addresses. The neighbor solicitation messages, which are used for address resolution, are sent via multicast. The ND protocol is not designed with multi-hop ad hoc networks in mind. A node in such network is able to broadcast to other nodes within its radio range, but the communication is non-transitive. Therefore, a wireless ad hoc network is a Non-Broadcast Multi-Access (NBMA) structure with generally no network-wide multicast capabilities. The network solicitation messages are not forwarded in an IBSS. Hence, station A is only able to resolve MAC addresses of its immediate neighbors, but not of D, which is multiple hops away. There are several proposals to extend the capabilities of the ND protocol to multi-hop ad hoc networks [16] and 6LoWPAN (IPv6 over Low power Wireless Personal Area Networks [17]) in particular [18]. These proposals include mechanisms

for multi-hop Duplicate Address Detection (DAD), which allows a station to check the uniqueness of an IP address in an n-hop neighborhood. The multi-hop DAD can also be used for multi-hop address resolution: Station A may initiate multi-hop DAD for the IP address of D. Upon receiving a DAD request, D will respond with a DAD confirmation message that contains its MAC address. It this way, A can resolve the MAC address of D based on its IP address. Each station maintains a cache of resolved addresses, which limits the need for network-wide multi-hop address resolution.

3.3 Backward-Compatibility

Backward-compatibility with the standard PSM for IBSS is ensured since MH-PSM does not violate neither the frame formats nor the protocol operations. Stations that implement standard PSM will not check the Address 3 field of received ATIMs and, therefore, the chain of ATIMs will terminate at such stations. This diminishes some of the delay improvements, but otherwise does not prevent or impair communication.

To better understand how standard PSM and MH-PSM may coexist, consider a scenario where station A sends data to station E using B, C, and D as intermediate relays. Assume that all the stations except station B use MH-PSM. When B receives an ATIM from A (with E's MAC address in the Address 3 field), it will not immediately create an ATIM for C: Instead, it will wait for the data packet to arrive and then, at the beginning of the next beacon interval, it will create an ATIM for C. This ATIM will not contain E's address in its Address 3 field because B runs standard PSM. Once it receives the ATIM, the MH-PSM-enabled station C (and all subsequent downstream stations) will fall back to the standard PSM.

3.4 Support for Network-Wide Broadcasts

As pointed out earlier in this section, in case of broadcast ATIMs, the Address 3 field must contain the BSSID, which is used by the receiver to verify that the frame originated from a station in the IBSS of which the receiver is a member. Therefore, it cannot be used to store the MAC address of the final destination. Broadcast ATIMs are mostly used to announce *link-local* broadcasts (e.g., Node Solicitation messages of the Node Discovery protocol). Hence, they do not need to be forwarded over multiple hops since the announced broadcast is aimed at stations that are one hop away from the sender. However, when *network-wide* broadcasts (e.g., Route Request messages of the AODV routing protocol) are announced, the broadcast ATIMs should be forwarded over multiple hops to ensure that all stations in the network remain awake following the end of the ATIM window. To support such broadcasts in MH-PSM, there should be a field in the ATIM header (other than Address 3 field) that a sender could use to declare if the broadcast ATIM is announcing a link-local or a network-wide broadcast. A possible solution would

be to amend the 802.11 standard to include a new management frame subtype Multihop ATIM (analogous to Multihop Action frames [3]), which would contain an additional field in the header for this purpose.

3.5 Sleep On Beacon Transmission (SoBT)

We propose an additional mechanism to increase the doze time of idle stations. The 802.11 standard mandates that a station that wins the back-off at the beginning of a beacon interval and subsequently transmits a beacon, should remain awake until the end of the beacon interval. This is considered necessary to ensure that the IBSS to which that station belongs can be discovered by the devices that employ active scanning. Most portable battery-powered devices, such as smartphones and tablets, use active instead of passive scanning to conserve energy. Unlike with passive scanning, when the scanning device spends substantial time listening for incoming beacons, with active scanning the device may wake up for a short period of time, transmit Probe Requests on different channels, and return to the doze state. At least one station in the IBSS must reply with a Probe Response containing the SSID of the IBSS, which can then be added to the list of known SSIDs. If all stations in an IBSS move to the doze state, there is no one to reply to the Probe Requests and, therefore, the IBSS might be invisible to the devices that employ active scanning. The probability that a station transmits a beacon increases as the number of neighbors decreases. Hence, in a small and/or sparse IBSS network, a station might win beacon back-offs in many consecutive beacon intervals and is forced to stay awake even though there is no traffic in the network. In the following, we discuss how this requirement of the 802.11 standard can be relaxed in practice.

A device that employs active scanning should repeat the following procedure for each channel to be scanned (see [3] for a full description):

(a) Wait until the ProbeDelay time has expired or a PHYRxStart.indication has been received.

(b) Perform the basic DCF access procedure and send Probe Request to the broadcast destination address.

(c) Clear and start a ProbeTimer.

(d) If PHY-CCA.indication (busy) has not been detected before the ProbeTimer reaches MinChannelTime, then scan the next channel, else when ProbeTimer reaches MaxChannelTime, process all received Probe Responses.

Hence, ProbeDelay is the delay prior to transmitting a Probe Request on a new channel, MinChannelTime and MaxChannelTime are, respectively, the minimum and the maximum amount of time spent on that channel after the Probe Request

transmission. Assuming that PHYRxStart.indication has not been received in a), the device spends at least the time interval

$$T_{min} = ProbeDelay + t_{TX}(ProbeRequest) + MinChannelTime$$

on each channel. If the device receives a beacon during the interval T_{min}, the SSID advertised in the beacon will be added to the list of discovered SSIDs. Since Wi-Fi channels in the 2.4 GHz band are overlapping (which is not the case in the 5 GHz band), the minimum time that the scanning device spends listening on a channel may be even longer: While being tuned to channel i, a Wi-Fi device is able to receive transmissions on channels $[i - k,\ i + k]$, where typically $k = 2$ or $k = 3$ in the 2.4 GHz band. Therefore, assuming that neighboring channels are scanned consecutively, the minimum total time during which the scanning device is able to receive transmissions on channel i is $T^{tot}_{min} = T_{min} \times (2k+1)$ for $k+1 \leq i \leq 13-k$. For channels that are at the edges of the 2.4 GHz band (e.g., $i = 1$ and $i = 13$), $T^{tot}_{min} = T_{min} \times (k+1)$. In case of non-overlapping channels in the 5 GHz band, $T^{tot}_{min} = T_{min}$. Instead of staying awake for the entire beacon interval to respond to Probe Requests, an idle station in PSM mode could wake up periodically and transmit what we call *intra-beacons*. As long as the intra-beacon interval T_{IB} is shorter than T^{tot}_{min} the scanning device will be able to receive an intra-beacon and discover the IBSS. Hence, unlike regular beacons that are sent at the beginning of each beacon interval (if the station wins the back-off), intra-beacons are transmitted during the beacon interval. The station generates intra-beacons only in beacon intervals in which it i) wins the beacon back-off and subsequently transmits a regular beacon and ii) does not receive or send any ATIM frames during the ATIM window. If the beacon interval is shorter than T^{tot}_{min} there is no need to transmit intra-beacons because regular beacons are frequent enough to be received by the scanning device. We call this intra-beaconing mechanism SoBT (Sleep On Beacon Transmission). SoBT increases the doze time at the expense of additional (intra-)beacon transmissions. We define the SoBT overhead as the average number of intra-beacons transmitted per beacon interval. It is calculated as $\lceil T_{BI}/T_{IB} - 1 \rceil \times P_{SoBT}$, where T_{BI} is the beacon interval, T_{IB} is the intra-beacon interval, and P_{SoBT} is the probability that intra-beacons are transmitted in a beacon interval (i.e., i) and ii) are fullfilled).

We investigated the feasibility of the proposed SoBT mechanism for cases where scanning devices run Android and Apple iOS, which today cover more than 90% of the current smartphone market, according to [19]. An Android phone may either perform soft-scanning or hard-scanning depending on whether its Wi-Fi driver implements active scanning or not. If active scanning is not supported by the driver, the phone performs so-called soft-scanning, which is implemented in /net/mac80211/scan.c of the Android kernel. Otherwise, the hard-scanning is performed. With soft-scanning, an Android phone always waits for ProbeDelay before it sends a Probe Request on a new channel regardless if PHYRxStart.indication has been received or not. This is a departure from the step a) of the standard scanning procedure.

The ProbeDelay and MinChannelTime are 30 ms each. Therefore, an Android phone spends $T_{min} = 60$ ms on each channel, assuming that the time to transmit a Probe Request $t_{TX}(ProbeRequest)$ is negligible. The described soft-scanning procedure is however rarely used because Wi-Fi drivers of most Android phones implement hard-scanning. We examined the hard-scanning procedure of a Samsung Galaxy S3 phone, which is equipped with a Broadcom BCM4334 Wi-Fi chip. The Broadcom's *bcmdhd* driver for Android does not wait for ProbeDelay before sending a Probe Request on a new channel (step a) of the standard scanning procedure). When the phone is not connected to any Wi-Fi network, the driver scans for new networks by sending two consecutive Probe Requests and waiting $MinChannelTime = 40$ ms for Probe Responses on each channel. Therefore, the Samsung Galaxy S3 phone (and any other Android phone whose hard-scanning is performed by the *bcmdhd* driver) spends $T_{min} = 40$ ms on each channel, assuming that the time to transmit two consecutive Probe Request is negligible. Apple's iPhone 5 is equipped with the same Broadcom BCM4334 Wi-Fi chip as Samsung Galaxy S3, but the details of the active scanning procedure are not readily available since the driver source code is not available. We performed a set of measurements, which indicate that, when iPhone 5 is not connected to any Wi-Fi network, it scans for new networks by sending one Probe Request and waiting $MinChannelTime = 20$ ms for Probe Responses on each channel. Therefore, $T_{min} = 20$ ms, assuming that $t_{TX}(ProbeRequest)$ is negligible.

Our tests with both Samsung Galaxy S3 and iPhone 5 have shown that, during active scanning on channel i of the 2.4 GHz band, the phones are able to receive beacons on channels $[i - 2,\ i + 2]$. Therefore, in an IBSS that uses channel $i = 1$, it is sufficient to transmit intra-beacons every $T_{IB} \leq 3T_{min}$. Hence $T_{IB} \leq 120$ ms and $T_{IB} \leq 60$ ms for Samsung Galaxy S3 and iPhone 5, respectively. In an IBSS that uses channel $3 \leq i \leq 11$, it is sufficient to transmit intra-beacons every $T_{IB} \leq 5T_{min}$. Hence $T_{IB} \leq 200$ ms and $T_{IB} \leq 100$ ms for Samsung Galaxy S3 and iPhone 5, respectively. We implemented the SoBT scheme in MH-PSM as described in Section 5. To test the scheme, we created an IBSS containing a single dozing station whose inter-beaconing period T_{IB} was set according to the determined values. Our tests confirmed that a single scanning round is sufficient to receive an intra-beacon and, therefore, discover the IBSS without forcing the station to stay awake and continuously listen for Probe Requests [1].

4 Simulation Results

We extensively evaluated MH-PSM and compared its performance to standard PSM using simulation. The performance is measured in terms of end-to-end delay, doze

[1] In case of iPhone 5, the reception of an intra-beacon caused the SSID of the IBSS to be added to the list of known SSIDs. In case of Samsung Galaxy S3, we were able to confirm the reception of an intra-beacon, but the SSID was not added to the list because Android's Wi-Fi Manager (as of version 4.2.2) does not support IBSS mode.

Table 1: Default simulation parameters.

Parameter	Value
Channel model	unit disk
IEEE 802.11 PHY mode	6 Mb/s (802.11b)
MAC buffer size	100 frames
Beacon interval T_{BI}	200 ms
Intra-Beacon interval T_{IB}	100 ms (with SoBT)
ATIM window	20 ms
Data frame payload size	500 Bytes
Traffic model	Poisson(λ), λ=2.5, 5, 10 frames/s

time ratio, ATIM overhead, and packet delivery ratio, as defined below. The simulation setup and the results are described in the following.

End-to-End Delay is the average time required to forward a data frame from a source to its destination over multiple hops. It is averaged over all successfully delivered data frames.

Doze Time Ratio is the percentage of beacon intervals in which a station enters the doze state and it is closely related with the energy consumption. It is averaged over all stations that participate in traffic forwarding.

ATIM Overhead is the average number of ATIM frames sent per one successfully delivered data frame. The relative ATIM overhead of MH-PSM is the ratio of ATIM overheads obtained with MH-PSM and standard PSM.

Packet Delivery Ratio is the percentage of data frames that are successfully delivered to the end destination. A station may drop a data frame if it exceeds the maximum number of retransmissions.

4.1 Simulation Setup

We implemented and evaluated MH-PSM in Jemula802 [20], which is a Java-based 802.11 protocol simulator. We consider a network of static regularly spaced 802.11 stations that are 50 m apart from each other. We assume a simple unit disk radio propagation model. We varied the radio range from 50 m to 150 m to influence the number of hops between source-destination pairs. The beacon interval and ATIM window size are 200 ms and 20 ms, respectively, unless stated otherwise. When SoBT is used, intra-beacons are transmitted every 100 ms. The data traffic is Poisson (exponentially distributed frame interarrival times) with fixed frame sizes of 500 B. The number of traffic flows and mean frame interarrival time are varied to control the load in the network. We ensured that the duration of each simulation run is sufficient to make the variations in time-moving averages insignificant. The main simulation parameters are summarized in Table 1.

4.2 Performance Results

Consider first the simple single-flow scenario shown in Fig. 6, where the station on the far left is sending data frames to the station on the far right over multiple hops. The radio transmission range is set to 50 m, 100 m, and 150 m in different simulation runs to create paths with two, three, and six hops, respectively. On average, the sender is generating one frame every 200 ms ($\lambda = 5$ frames/s).

The results for the average end-to-end frame delay are shown in Fig. 7 (top). As expected, the delay increases with the number of hops. For the standard PSM it takes almost N beacon intervals to forward a frame over N hops. It may happen that a frame is forwarded over multiple hops in a single beacon interval: if its next-hop neighbor is awake, a station may immediately forward the frame to it, without waiting for the next beacon interval to send a traffic announcement. In a lightly loaded network, however, it is likely that the next-hop station is in the doze state, and therefore, the data frame has to be buffered. The results show that the delay is significantly shorter for MH-PSM. Although it slightly increases with the number of hops (due to processing in intermediate stations and increasing probability of collisions/retransmissions caused by hidden stations) the average delay is well below 200 ms, which is the duration of the beacon interval. As the number of hops increases from two to six, the percentage of frames that are forwarded end-to-end within a single beacon interval marginally deceases from 100% to 99%, whereas for standard PSM drops from 92% down to zero.

The average doze time ratio is shown in Fig. 7 (middle). Even without SoBT, MH-PSM significantly increases the energy efficiency by allowing the stations to move to the doze state more often than standard PSM. The reason for this is that MH-PSM prevents excessive buffering of frames in intermediate stations (the number of frames in the station's queue is lower), which effectively decreases the traffic load and the probability of collisions/retransmissions. This shows that MH-PSM provides both shorter delay and lower energy consumption, which is a major improvement over standard PSM whose parametric adjustments/optimizations may only trade shorter delay for higher energy consumption and vice versa. When combined with SoBT, the doze time of MH-PSM surges to 60%. Hence, allowing stations

Figure 6: Simulated network topology with a single flow. The transmission range is set to 50 m, 100 m, and 150 m in different simulation runs to produce paths with two, three, and six hops, respectively.

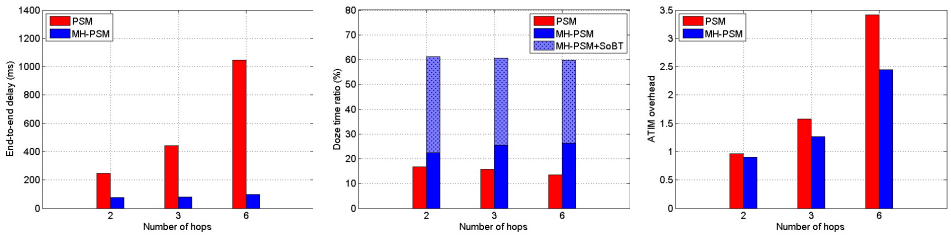


Figure 7: End-to-end delay, doze time ratio, and ATIM overhead for different number of hops.

to sleep after beacon transmissions is very effective in reducing idle listening.[2] Standard PSM forces a station to stay awake after beacon transmission, which can be a major source of energy waste, especially in sparse and in networks with many hidden terminals. Consider our two-hop scenario (A→B→C): If all the stations could hear each others transmissions, the probability that an arbitrary station transmits a beacon would be 1/3. However, since A and C are hidden from each other, they may both transmit beacons if their backoff timers expire before the backoff timer of B. Moreover, if beacon transmissions from A and C overlap in time, they will collide at B and, therefore, B will also transmit a beacon. Hence, without SoBT, there is a high probability that a station has to remain awake because it transmitted a beacon. The SoBT overhead (defined in Section 3.5) is $\lceil T_{BI}/T_{IB} - 1 \rceil \times P_{SoBT}$, where $\lceil T_{BI}/T_{IB} - 1 \rceil = 1$ and the percentage of beacon intervals in which SoBT is invoked P_{SoBT} is given by the difference between doze time ratios with and without SoBT. Hence, for 2, 3, and 6 hops the SoBT overheads are, respectively, 0.39, 0.34, and 0.33 intra-beacons per beacon interval.

In Fig. 7 (bottom), we show the ATIM overhead for both PSM schemes. While the overhead for MH-PSM is comparable to that of standard PSM for the path with two hops, it is almost 30% lower in the six-hop case. To understand the reasons for this, consider a five-hop path from station A to station E via B, C, and D, as shown in Fig. 8. Assume that one frame is buffered at station A and one at station C. In the best-case scenario, it will take four beacon intervals and six ATIMs to deliver both frames to the destination under standard PSM. With MH-PSM however, it will only one beacon interval and four ATIMs to achieve the same because it creates a wave of ATIMs that flushes all buffered frames to the destination, as shown in Fig. 9. There are however scenarios where the ATIM overhead of MH-PSM is higher than that of the standard PSM even for paths with many hops. In standard PSM, a station sends a single ATIM to its neighbor to announce all data frames that it intends to forward to this neighbor, regardless of their end destinations. In MH-PSM, the

[2]Note however, that during SoBT intervals, stations may need to wake up and transmit intra-beacons, which diminishes some of the energy saving gain.

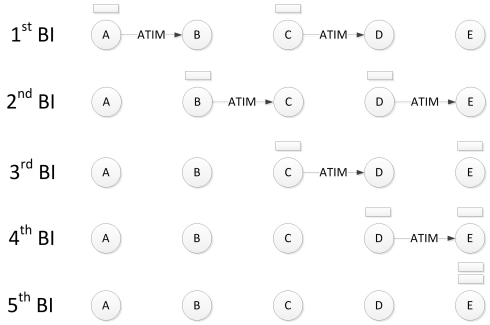


Figure 8: From [13]. Standard PSM requires four Beacon Intervals (BIs) and six ATIMs to deliver the frames buffered at A and C.

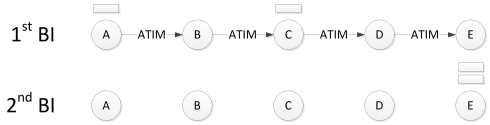


Figure 9: From [13]. MH-PSM requires only one Beacon Interval (BI) and five ATIMs to deliver the frames buffered at A and C.

station may send multiple ATIMs with different Address 3 fields to the neighbor if the pending data frames have different end destinations. For example, consider two flows whose eight-hop paths contain a common subset or relays, as shown in Fig. 10. In MH-PSM, the common relays may need to forward two ATIMs with different Address 3 fields to their next-hop neighbors in the same ATIM window. This is not the case in standard PSM, where only one ATIM is sent. The results in Fig. 11 show that the ATIM overhead of MH-PSM is 20% higher. MH-PSM outperforms standard PSM in all other respects: The end-to-end delay is close to tenfold shorter, the doze time ratio is slightly higher, and the packet delivery ratio is significantly improved. Therefore, the relative ATIM overhead of MH-PSM had no bearing to the key performance metrics.

We next investigate the impact of beacon interval on the performance of PSM and MH-PSM. The results presented so far assume a beacon interval of 200 ms. We changed the interval to 100 ms and 400 ms and repeated the simulations for the basic scenario shown in Fig. 6 with the transmission range of 50 m (i.e. six hops). The average frame interarrival time is 200 ms regardless of the beacon interval. The results are summarized in Table 2. As expected, the frame delay for standard PSM increases linearly with the beacon interval because the time that frames stay buffered in the intermediate nodes is proportional to the beacon interval. The delay for MH-PSM also increases, but remains much shorter than for standard PSM. The increase is due to the fact that MH-PSM does not guarantee that all frames will

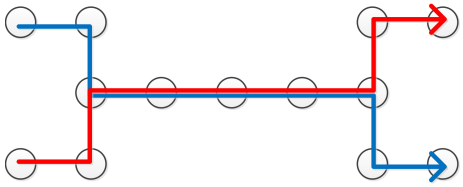


Figure 10: An example of two flows whose paths partially overlap.

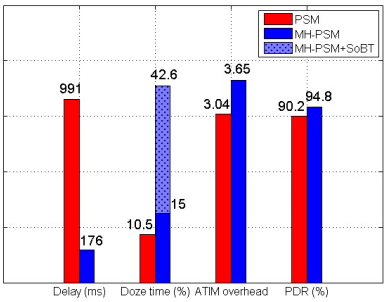


Figure 11: Performance of standard PSM and MH-PSM for the scenario with two flows whose paths partially overlap (Fig. 10).

be delivered end-to-end in a single beacon interval: Some frames may be buffered along the path as in the case of standard PSM. The doze time decreases for both schemes because the number of idle (with no data traffic) beacon intervals decreases as they become longer. Another observation is that the packet delivery ratio of standard PSM decreases significantly for the longer beacon interval (from close to 100% for 100 ms to 93% for 400 ms), while for MH-PSM it remains close to 100%: For standard PSM, the number of buffered frames along the path increases with the duration of the beacon interval, which effectively increases the traffic load in the network and the probability of collisions. With MH-PSM, most frames are delivered end-to-end without buffering in intermediate nodes.

In the last two simulation scenarios, we vary the average frame Interarrival Time (IAT) and the number of flows in the network to evaluate the impact of traffic load on the performance of MH-PSM. The beacon interval is 200 ms regardless of the traffic load.

The results presented so far assume an average frame IAT of 200 ms. We changed the average IAT to 10, 50, 100, and 400 ms and repeated the simulations for the basic scenario shown in Fig. 6 with the transmission range of 50 m (i.e. six hops). The results are summarized in Table 3. The frame delay decreases for higher load

Table 2: Performance of standard PSM and MH-PSM for different beacon intervals. Frame interarrival times is 200 ms. Frames are forwarded over six hops.

BI (ms)	Delay (ms)		Doze time (%)		ATIM overhead	
	PSM	MH-PSM	PSM	MH-PSM	PSM	MH-PSM
100	532	51	22	31+41	4.55	3.02
200	1047	99	13	26+33	3.42	2.45
400	2044	179	8	23+30	2.35	1.45

up to a certain point (IAT $\geq$ 50 ms). This is because the doze time decreases with the traffic load, meaning that frames are more likely to be forwarded without buffering. In case of MH-PSM, nearly 100% of frames are transmitted end-to-end within a single beacon interval for IAT $\geq$ 50 ms. The differences in frame delays for different IATs are due to the initial hold-up at the sender: For the average IAT of 400 ms (twice the beacon interval), it is more likely that the sender is in the doze state when a frame is passed to the 802.11 MAC; therefore, the frame has to wait for the next beacon interval to be transmitted. For average IAT of 100 ms (two frames per beacon interval), many frames are transmitted immediately since the sender is likely to be awake due to earlier packer arrivals. For IAT = 10 ms (20 frames per beacon interval), congestion sets in, frames start colliding, and we observe an increase in the delay due to retransmissions. The ATIM overhead decreases for shorter IATs (higher loads) because it is unnecessary to send ATIMs for some of the frames: Only one ATIM is sent per hop per beacon interval regardless of the number of frames that need to be transmitted in that beacon interval.

We next consider the scenarios with multiple (i.e., 2, 4, and 8) intersecting flows in a grid topology shown in Fig.12. The transmission range is 50 m; therefore, frames are forwarded over six hops. The results in Table 4 show that the performance deteriorates with the number of flows. Transmissions of intersecting nodes are especially prone to collisions because they are surrounded by four active/forwarding stations that do not hear each other's transmissions (hidden stations). The impact of collisions on the performances of standard PSM and MH-PSM is somewhat different:

Table 3: Performance of standard PSM and MH-PSM for different frame interarrival times. Beacon interval is 200 ms. Frames are forwarded over six hops.

IAT (ms)	Delay (ms)		Doze time (%)		ATIM overhead	
	PSM	MH-PSM	PSM	MH-PSM	PSM	MH-PSM
10	999	117	7	11+17	0.32	0.25
50	997	78	9	19+27	1.44	0.81
100	1010	82	9	23+30	2.36	1.49
200	1047	99	13	26+33	3.42	2.45
400	1094	102	22	31+40	4.65	3.55

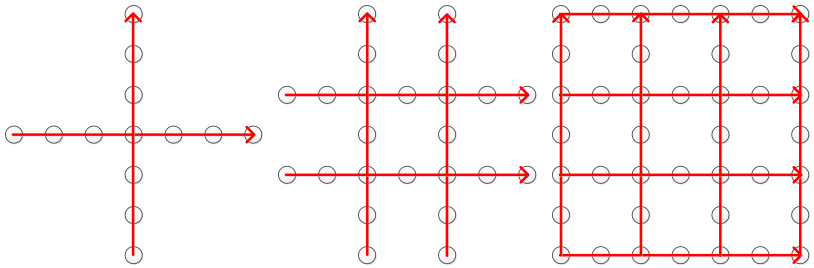


Figure 12: Simulated network topology with 2, 4, and 8 symmetric flows.

While the frame delay for standard PSM remains unaffected by the number of flows, the delay for MH-PSM increases considerably (yet still remains relatively low). The reason is that collisions in intersecting nodes may disrupt the cut-through forwarding of data frames in MH-PSM. In the single-flow scenario, 100% of frames are forwarded end-to-end in a single beacon interval. For the eight-flow scenario, this percentage drops to 63%. The additional hold-up in intersecting nodes does not affect the frame delay in standard PSM so significantly because most frames are anyway forwarded only one hop per beacon interval.

In addition to the regular grid topology, we considered (quasi-)random topologies where 40 nodes are uniformly placed in an 300×300 m area, as shown in Fig. 13. The simulation setup is similar as in the previous scenario: we consider eight intersecting six-hop flows. To control the number of hops, we repeatedly generated random topologies and randomly selected eight source-destination pairs until paths between all eight pairs were six hops long. We collected results from 10 simulation runs. The mean value and the coefficient of variation CoV (standard deviation divided by the mean) over different runs for the delay, doze time ratio, and PDR are shown in Table 5. The results show longer delay and lower doze time ratio compared to the regular grid topology (last row of Table 4) for both standard PSM and MH-PSM. This is because regular grid scenario provides better spatial separation between the

Table 4: Performance of standard PSM and MH-PSM for different numbers of flows. The transmission range is 50 m – frames are forwarded over six hops. Beacon interval (BI) and frame interarrival time (IAT) are 200 ms.

Num. flows	Delay (ms)		Doze time (%)		PDR (%)	
	PSM	MH-PSM	PSM	MH-PSM	PSM	MH-PSM
1	1047	99	14	26+33	99	100
2	1048	171	14	18+28	83	93
4	1055	253	14	15+28	81	90
8	1063	285	12	14+20	77	86

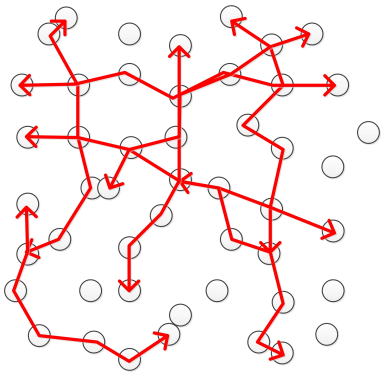


Figure 13: One of the simulated random network topologies with eight six-hop flows.

flows and, therefore, lower number of collision. MH-PSM, however, still performs significantly better than the standard PSM.

Table 5: Performance of standard PSM and MH-PSM for a random topology. The transmission range is 50 m – frames are forwarded over six hops. Beacon interval (BI) and frame interarrival time (IAT) are 200 ms.

	Delay (ms)		Doze time (%)		PDR (%)	
	Mean	CoV	Mean	CoV	Mean	CoV
PSM	1250	0.062	9	0.074	77	0.039
MH-PSM	526	0.072	11+18	0.173+0.048	85	0.050

Finally, we compare the performances of MH-PSM and Link-Indexed Statistical traffic Predictor (LISP), another modification of the standard 802.11 PSM, on the topology shown in Fig. 12 (eight symmetric flows). LISP [21] reduces the end-to-end delay on multihop paths by enabling nodes to establish the correlation between overheard acknowledgments to ATIM frames (ATIM_ACKs) and incoming traffic. The correlation is established when a node overhears an ATIM ACK packet addressed for some other node in a beacon interval (BI), and receives an ATIM in the subsequent BI. Next time when it overhears an ATIM ACK packet again, it takes that as an indication of incoming traffic and herein stays awake through the BI. The node then transmits a pseudo-ATIM_ACK to notify the upstream node of its willingness to stay awake in the current BI and to notify downstream nodes to stay awake as well. On the other hand, if such a conjecture is incorrect and the node receives no packets in the BI, then it erases the recent history and learns from scratch. The basic mechanism may not operate correctly in the presence of multiple connections, as connections may share nodes or links on their routes, and a node

Table 6: Performance comparison of LISP and MH-PSM. The simulated topology is shown in Fig. 12 (eight symmetric flows). The transmission range is 50 m – frames are forwarded over six hops. Beacon interval (BI) and frame interarrival time (IAT) are 200 ms.

Num. flows	Delay (ms)		Doze time (%)		PDR (%)	
	LISP	MH-PSM	LISP	MH-PSM	LISP	MH-PSM
8	299	285	7+19	14+20	88	86

may receive/forward packets from different connections in an interleaving fashion. As a result, the correlation between traffic indicators and arrivals of packets for one connection has to be differentiated from those for other connections. A method by which LISP accomplishes this task is described in [21].

The results shown in Table 6 show that there is no significant difference in the end-to-end delays of MH-PSM and LISP. This is because the flows are stationary (long-lived), so the nodes that are participating in frame relaying soon learn to correlate ATIM_ACKs with incoming traffic. The doze time ratio is higher with MH-PSM because correlation establishment/learning in LISP is not perfect: it may happen that nodes stay awake for no reason. The ATIM overhead of LISP is lower than that of MH-PSM because the relaying nodes stop sending ATIMs once they establish the correlation: only the pseudo-ATIM_ACKs are transmitted.

5 Experimental Results

MH-PSM is software-implementable: The parsing and generation of ATIM frames are not time-critical operations that have to be implemented on-chip. This enables the integration of MH-PSM into an 802.11 device driver without modifications of the lower-level MAC operations. Our experimental MH-PSM implementation is described in the following.

5.1 MH-PSM Implementation

We implemented the proposed MH-PSM as a part of our WLAN driver module for Contiki [12], an open source operating system for the Internet of Things. The used hardware platform consists of an Arduino Due board (Cortex-M3 MCU, 96 KB of SRAM) connected via USB interface to an 802.11n transceiver based on Atheros AR9001U-2NX chipset [22], as shown in Fig. 14. The AR9001U-2NX chipset contains an AR9170 MAC/baseband and an AR9104 (dual-band 2×2) radio chip. Atheros has released the firmware of AR9170 as open source, which enables us to write the Contiki WLAN driver. The open source firmware provides a direct access to the lower-MAC program that runs on the AR9170 chip, which greatly simplifies driver debugging. The used Contiki driver is partially based on the

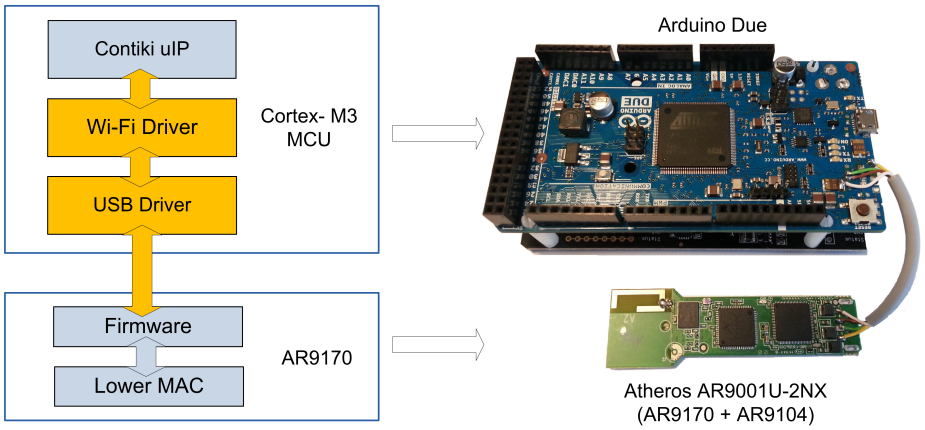


Figure 14: HW/SW platform for MH-PSM evaluation consists of an Arduino Due board and an Atheros 802.11n transceiver. We ported Contiki OS to the Arduino Due and implemented WLAN and USB drivers for the Atheros transceiver.

otus driver [23], a depreciated Linux driver for Atheros devices (replaced by the carl9170 driver as of kernel version 2.6). It is fully integrated with the Contiki's uIP protocol stack for TCP/IP and supports standard PSM and MH-PSM in ad hoc mode. Our goal was not only to validate MH-PSM, but to build a flexible open-source platform for experimentation with 802.11 MAC Layer Management Entity (MLME) algorithms (power saving, beaconing, time synchronization, scanning, association, authentication) for future IoT-ready WLAN-enabled embedded devices.[3] While there is abundance of low-power WLAN modules for embedded devices (e.g., Roving Networks RN-131C, Gainspan GS2100M, Texas Instruments CC3000, Broadcom BCM4390, etc.) and WLAN enabled development boards for IoT applications (WiSmart EC32Lxx, RTX41xx, Spark Core, Flyport WLAN, etc.) they are not suitable for experimentation with 802.11 MAC layer management algorithms: Typically, their proprietary network stack implementations are provided as binary firmwares and only high-level communication and configuration APIs are disclosed. Moreover, their stack implementations often do not support IBSS mode and/or IPv6.

5.2 Experimental Setup

Our experimental setup consists of up to seven WLAN nodes lined up 10 m apart from each other[4] in a quiet alley at the back of an office building compound, as

[3] Lower-MAC functions with strict timing constraints, such as DCF, are implemented on the AR9170 chip.

[4] We could not place the nodes further apart due to space constraints.

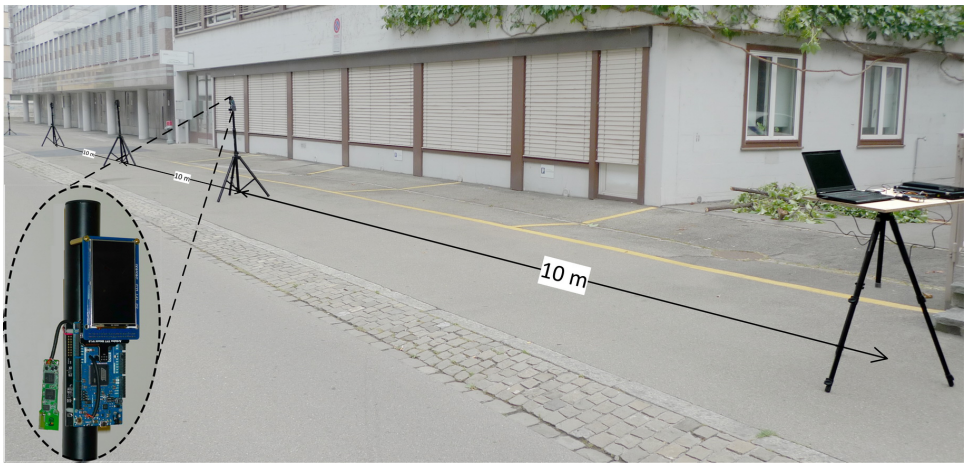


Figure 15: Experimental setup: Seven WLAN nodes lined up 10 m apart from each other. The experiments were run overnight to minimize the interference from surrounding access points. An LCD screen is connected to each node to monitor the status.

shown in Fig. 15. The experiments were run overnight, when the interference from the neighboring access points was negligible: WLAN spectrum monitors picked up only control and management traffic from other networks at those hours. The transmit power of the nodes was reduced to the minimum of 0 dBm, which resulted in the transmission range of roughly 30 m. Therefore, nodes that were up to three hops away could still observe and decode each others transmissions. Hence, there were fewer hidden terminals in the experimental setup than in the simulation setup. The goal was not to exactly replicate the simulation results, but to compare MH-PSM and standard PSM in the described experimental setup.

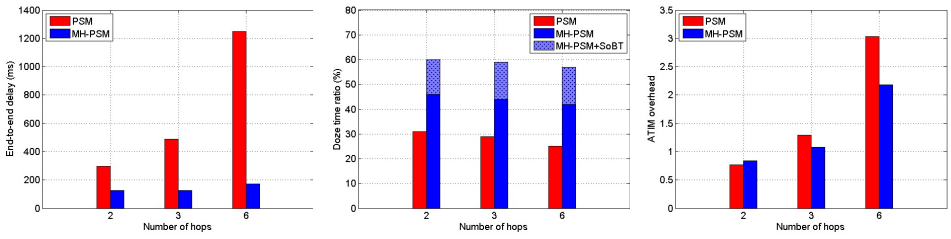


Figure 16: End-to-end delay, doze time ratio, and ATIM overhead for different number of hops.

5.3 Performance Results

We first consider scenarios with two, three, and six hops that are created by placing, respectively, three, four, and seven nodes in a line. The beacon interval and ATIM window size were set to 200 ms and 20 ms. We used the same Poisson traffic generator with fixed packet sizes as in the simulations. The sender (rightmost node in Fig. 15) generated on average one frame every 200 ms ($\lambda = 5$ frames per second). For each scenario we performed multiple (typically three to five) runs with 300 packets each. The results were then averaged over those runs.

The average end-to-end delay, doze time ratio, and ATIM overhead are shown in Fig. 16. The results follow the pattern seen in the simulation (Fig. 7). For the standard PSM it takes almost N beacon intervals to forward a frame over N hops, while for the MH-PSM the delay is much shorter. The absolute delays are somewhat higher then in the simulations because packet generation and processing/parsing in intermediate nodes takes some time. Our hardware/software platform can be further optimized for this task. The doze time ratio is still significantly higher with MH-PSM than with standard PSM. The absolute values are higher than in the simulations (Fig. 7) because nodes have more neighbors (e.g., node A is able to hear not only B, but also C). This reduces the number of beacon intervals in which a node wins the beacon transmission and, therefore, the number of beacon intervals in which the node must stay awake. Hence, the doze time ratio is higher in the experiments. For the same reason, the gain of SoBT is lower, but it is still significant. The SoBT overheads for 2, 3, and 6 hops are, respectively, 0.14, 0.15, and 0.15 intra-beacons per beacon interval. The ATIM overhead of MH-PSM is comparable to that of standard PSM in the two-hop and almost 30% lower in the six-hop case, as predicted by the simulations. The absolute values are lower than in simulation because fewer ATIMs had to be retransmitted due to collisions (there were fewer hidden terminals in the experiments than in the simulations).

We also measured the performance of PSM and MH-PSM for different beacon intervals. The results presented so far assume the beacon interval of 200 ms. We changed the interval to 100 ms and 400 ms and repeated the measurements for the six-hop scenario. The average frame interarrival time is 200 ms regardless of the beacon interval. The results are summarized in Table 7. All our observations based on the simulation results (Table 7 apply to measurement results too: While the frame delay increases linearly with the beacon interval for the standard PSM, it increases moderately and remains comparably short for the MH-PSM. Even without SoBT, MH-PSM outperforms standard PSM in terms of the doze time ratio. As expected, the doze time decreases with the length of beacon intervals because fewer intervals are idle. The improvement in terms of ATIM overhead is also significant.

Finally, we varied the average frame Interarrival Time (IAT) $1/\lambda$ to evaluate the impact of traffic load in the six-hop scenario. The beacon interval is 200 ms regardless of the IAT. The results are summarized in Table 8. With MH-PSM, close to 80% of frames are transmitted end-to-end within a single beacon interval regardless

Table 7: Performance of standard PSM and MH-PSM for different beacon intervals. Frame interarrival times is 200 ms. Frames are forwarded over six hops.

BI (ms)	Delay (ms)		Doze time (%)		ATIM overhead	
	PSM	MH-PSM	PSM	MH-PSM	PSM	MH-PSM
100	586	125	40	52+17	4.11	2.78
200	1249	173	25	42+15	3.03	2.18
400	2234	281	19	41+15	2.12	1.23

of the traffic load (0% with standard PSM). The slight increase in the delay for longer IATs is due to the initial hold-up at the sender, as discussed in Section 4.2. As expected, the doze time increases when the traffic load decreases (i.e., for longer IATs). There is a mismatch between experimental and simulation results (in terms of absolute values, due to different network setups), but trends, relative performance, and conclusions are the same: MH-PSM significantly outperforms standard PSM in all scenarios considered in this paper.

6 Related Work

The power saving mode (PSM) of 802.11 has been thoroughly investigated in the past. However, vast majority of proposed improvements targets infrastructure (BSS) networks, where data is transmitted over single hop between an AP and a station [24, 25, 26, 27, 28, 29, 30]. The IEEE 802.11ah proposal [31] defines a low power medium access method that optimizes standard 802.11 PSM for battery-powered devices used in smart metering and machine-to-machine communication. However, the optimization also focuses on infrastructure networks.

A number of solutions have also been proposed to optimize the PSM in IBSS (ad-hoc) networks. Some of them focus on minimizing the duration of idle listening by introducing mechanisms for early transition to the doze state [9, 10, 11, 32, 33]. For example, in [9, 10], ATIM announcements are modified to include the number of pending frames to allow the receiving station to move to the doze state immediately after it receives the last announced frame without waiting for the end of the

Table 8: Performance of standard PSM and MH-PSM for different frame interarrival times. Beacon interval is 200 ms. Frames are forwarded over six hops.

IAT (ms)	Delay (ms)		Doze time (%)		ATIM overhead	
	PSM	MH-PSM	PSM	MH-PSM	PSM	MH-PSM
100	1226	166	17	37+14	1.89	1.32
200	1249	173	25	42+15	3.03	2.18
400	1285	185	40	52+17	4.11	3.03

beacon interval. In [10], the early transition to the doze state is combined with the dynamic adjustment of the ATIM window duration based on the traffic conditions in the IBSS in order to transit to sleep earlier in case of low network traffic. Considering a low-traffic scenario, [32] proposes a scheme where the absence of traffic is declared by transmitting a delayed beacon, so that stations can skip idle listening during the ATIM window. Similarly, [33] proposes a scheme where transmitting stations announce their intention of sending ATIM frames in a short time period at the beginning of the beacon interval. Stations that do not send or receive any announcements do not have to stay awake for the entire ATIM window. The problem of long end-to-end latency on multi-hop paths is not addressed or evaluated in these works. Entering the doze state as soon as all frames announced by the ATIM are received may increase the latency for frames that arrive to the sender after the ATIM transmission because they cannot be delivered in the current beacon interval. Hence, the multi-hop end-to-end latency may be even worse than that of the standard 802.11 PSM. In [34, 35], the authors propose a topology-aware power-saving algorithm based on the overhearing of the ATIM frames transmitted by the neighbors. By extracting the source addresses from the received ATIM acknowledgments, a station can defer from transmitting ATIMs to stations known to remain awake after the expiration of the ATIM window. This scheme can efficiently decrease the required ATIM window size in a fully-connected IEEE 802.11 mesh network, but it is less effective in multi-hop IBSS network topologies. In [36], a modified 802.11 PSM is proposed, where beaconing nodes act as a centralized schedulers that sends out explicit transmission ordering information to the neighbors. The beacon transmission is moved to the end of the ATIM window. The transmission schedule is computed upon receiving the ATIM announcements information in the ATIM window. The schedule is then announced via beacon transmission at the end of the ATIM window. The objective is to avoid channel contention and maximize the dozing time. In [37], the authors propose an improvement to standard 802.11 PSM, which addresses the problems of clock synchronization and neighbor discovery. The objective is to ensure robust performance in the presence of clock drift, and to provide up-to-date neighbor information to upper layer protocols (e.g., routing). The problem of long end-to-end latency on multi-hop paths is not addressed or evaluated. The schemes described in [36, 37] do not address the latency problem on multi-hop links. In [38], an optimization of the standard 802.11 PSM called SIMPA is proposed. SIMPA decouples power state transitions from beacon intervals: Depending on the traffic, a station may transit to the doze state inside a beacon interval or extent its active state beyond one beacon interval. Every station maintains packet arrival history to decide if the active state should be prolonged. In the case of a high traffic load, stations will stay awake, which expedites the data delivery on multi-hop paths. Therefore, even though SIMPA is not design to minimize latency, it may provide shorter latency than standard 802.11 if the traffic intensity is high enough. MH-PSM does not maintain any traffic history and

provides reduced end-to-end latency regardless of the traffic load, even for sporadic data frames.

The problem of end-to-end latency latency in multi-hop 802.11 networks has been explicitly addressed in several works. In [39], the authors propose an on-demand power management algorithm, which is not specifically targeting 802.11 MAC, but it is applicable and demonstrated on an example of an 802.11 ad hoc network. The scheme integrates routing information from on-demand ad hoc routing protocols and power management capabilities from the MAC layer. It judiciously enables and disables the power saving mode of specific nodes depending on the traffic in the network. The scheme can be used to reduce the latency on multi hop paths by disabling the PSM of the stations along the path. The scheme may provide better performance than MH-PSM, but it is less practical because, unlike MH-PSM, it requires the knowledge of traffic patterns and cross-layer coordination. In [11], the latency is reduced by organizing the IBSS network hierarchically, so that always-awake master stations create a backbone that relays the multi-hop traffic between PSM-enabled slave stations. MH-PSM does not rely on any sort of always-awake mesh routers, we consider an ad hoc network where every station is in power save mode. A low-latency routing algorithm that forwards packets via stations that are known to be awake in the current beacon interval is presented in [40]. Our solution operates at the MAC layer, it is independent of routing, and does not require any a priori knowledge of the power state of the stations. A history-based prediction mechanism by which a station infers if it needs to stay awake to forward incoming packets is described in [41]. In case of sporadic packet bursts and short-lived flows, wrong prediction may cause the station to stay awake for no reason. In [42], the authors propose a fast flooding algorithm that propagates ATIMs to allow broadcast packets to travel multiple hops in a single beacon interval. Conceptually, this is similar to what we propose with MH-PSM. However, unlike MH-PSM, the algorithm in [42] is applicable to broadcast transmissions only.

7 Conclusions

The Internet of Things will connect not only Zigbee-enabled devices, such as industrial sensors, but also consumer electronics that typically uses Wi-Fi for network connectivity. The power saving mechanisms of the IEEE 802.11 MAC have to be optimized to enable low-cost battery-powered devices to connect to each other ad hoc, without infrastructure support. In this paper, we proposed MH-PSM, an extension of the standard IEEE 802.11 PSM that enables low-latency ad hoc communication over multiple hops. MH-PSM also increases the doze time ratio of the devices compared to the standard PSM to further extend their battery lifetime. MH-PSM is software implementable since it does not require changes to the lower MAC. It is also backward-compatible with the standard PSM, which guarantees interoperability with legacy devices. We implementing the scheme on an embedded

open source platform and demonstrated its effectiveness using both simulations and experiments. We are planning to investigate the interaction of MH-PSM with upper layer protocols (i.e., RPL/loadNG routing in particular) to further optimize the IoT communication protocol stack.

8 Acknowledgement

This work was partially funded by the European Union Seventh Framework Programme (FP7-ICT/2007-2013) under grant agreement number 288879 (Calipso; see http://www.ict-calipso.eu/).

References

[1] "IEEE standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks. Specific requirements. Part 15.4: Wireless Medium Access Control and Physical Layer Specifications for Low-Rate Wireless Personal Area Networks," 2006.

[2] "Z-Wave Alliance."
http://www.z-wavealliance.org.

[3] "IEEE standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks. Specific requirements. Part 11: Wireless LAN Medium Access Control and Physical Layer specifications," 2012.

[4] S. Tozlu, M. Senel, W. Mao, and A. Keshavarzian, "Wi-fi enabled sensors for internet of things: A practical approach," *Communications Magazine, IEEE*, vol. 50, pp. 134–143, June 2012.

[5] B. Ostermaier, M. Kovatsch, and S. Santini, "Connecting things to the web using programmable low-power wifi modules," in *Proceedings of the Second International Workshop on Web of Things*, WoT '11, pp. 2:1–2:6, 2011.

[6] "Consumer perception of electronic toys," 2007.

[7] X. Zhang and K. G. Shin, "E-mili: Energy-minimizing idle listening in wireless networks," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 9, pp. 1441–1454, 2012.

[8] A. Dunkels, "The ContikiMAC radio duty cycling protocol," Tech. Rep. T2011:13, Swedish Institute of Computer Science, Dec. 2011.

[9] D.-Y. Kim and C.-H. Choi, "Adaptive power management for IEEE 802.11-based ad hoc networks," in *Proc. 5th World Wireless Congress*, (San Francisco, USA), May 2004.

[10] E.-S. Jung and N. Vaidya, "An energy efficient MAC protocol for wireless LANs," in *Proc. IEEE Infocom*, (New York, USA), 2002.

[11] S. Yongsheng and T. A. Gulliver, "An energy-efficient MAC protocol for ad hoc networks," *Wireless Sensor Network*, vol. 1, no. 5, pp. 407–416, 2009.

[12] "Contiki OS." http://www.contiki-os.org/. [Apr-2013].

[13] I. Glaropoulus, S. Mangold, and V. Vukadinovic, "Enhanced IEEE 802.11 Power Saving for Multi-Hop Toy-to-Toy Communication," in *IEEE Internet of Things (iThings)*, (Beijing, P.R.China), 2013.

[14] H. Woesner, J.-P. Ebert, M. Schlager, and A. Wolisz, "Power-saving mechanisms in emerging standards for wireless LANs: the MAC level perspective," *IEEE Personal Comm.*, vol. 5, no. 3, pp. 40–48, 1998.

[15] T. Nartem, E. Nordmark, W. Simpson, and H. Soliman, "Neighbor discovery for IP version 6 (IPv6)," RFC 4861, Sept. 2007.

[16] M. Grajzer, "Nd++ – an extended IPv6 Neighbor Discovery protocol for enhanced duplicate address detection to support stateless address auto-configuration in IPv6 mobile ad-hoc networks," Internet-Draft, March 2011.

[17] "Ipv6 over low power wireless personal area networks, ietf working group on." http://datatracker.ietf.org/wg/6lowpan/charter/. [Apr-2013].

[18] Z. Shelby, S. Chaktabarti, and E. Nordmark, "Neighbor discovery optimization for Low Power and Lossy Networks (6LoWPAN)," Internet-Draft, Aug. 2012.

[19] "IDC's Worldwide Quarterly Mobile Phone Tracker." http://www.idc.com/tracker/showproductinfo.jsp?prod_id=37. [June-2013].

[20] S. Mangold, "Jemula802." https://github.com/schmist/Jemula802. [Apr-2013].

[21] C. Hu and J. Hou, "Lisp: a link-indexed statistical traffic prediction approach to improving ieee 802.11 psm," in *Proc. Int. Conf. Distributed Computing Systems*, 2004.

[22] "Atheros AR9001U." http://wikidevi.com/files/Atheros/specsheets/AR9001U.pdf. [Apr-2013].

[23] "Otus." http://linuxwireless.org/en/users/Drivers/otus. [Apr-2013].

[24] D. Qiao and K. Shin, "Smart power-saving mode for ieee 802.11 wireless lans," in *Proc. IEEE Infocom*, March 2005.

[25] C.-H. Gan and Y.-B. Lin, "An effective power conservation scheme for ieee 802.11 wireless networks," *Vehicular Technology, IEEE Transactions on*, vol. 58, pp. 1920–1929, May 2009.

[26] Y. He and R. Yuan, "A novel scheduled power saving mechanism for 802.11 wireless lans," *Mobile Computing, IEEE Transactions on*, vol. 8, pp. 1368–1383, Oct 2009.

[27] Y. Xie, X. Luo, and R. Chang, "Centralized psm: An ap-centric power saving mode for 802.11 infrastructure networks," in *Sarnoff Symposium, 2009. SARNOFF '09. IEEE*, pp. 1–5, March 2009.

[28] P. Agrawal, A. Kumar, J. Kuri, M. Panda, V. Navda, and R. Ramjee, "Opsm - opportunistic power save mode for infrastructure ieee 802.11 wlan," in *Communications Workshops (ICC), 2010 IEEE International Conference on*, pp. 1–6, May 2010.

[29] X. Chen, S. Jin, and D. Qiao, "M-psm: Mobility-aware power save mode for ieee 802.11 wlans," in *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, pp. 77–86, June 2011.

[30] X. Perez-Costa and D. Camps-Mur, "A protocol enhancement for ieee 802.11 distributed power saving mechanisms no data acknowledgement," in *Mobile and Wireless Communications Summit, 2007. 16th IST*, pp. 1–7, July 2007.

[31] S. Aust, R. Prasad, and I. G. M. M. Niemegeers, "IEEE 802.11ah: Advantages in standards and further challenges for sub 1 GHz Wi-Fi," in *Communications (ICC), 2012 IEEE International Conference on*, pp. 6885–6889, 2012.

[32] J.-M. Choi, Y.-B. Ko, and J.-H. Kim, "Enhanced power saving scheme for ieee 802.11 dcf based wireless networks," in *Personal Wireless Communications* (M. Conti, S. Giordano, E. Gregori, and S. Olariu, eds.), vol. 2775 of *Lecture Notes in Computer Science*, pp. 835–840, Springer Berlin Heidelberg, 2003.

[33] N. Rajangopalan and C. Mala, "Modified power save model for better energy efficiency and reduced packet latency," *American Journal of Engineering and Applied Sciences*, vol. 5, no. 3, pp. 237–242, 2012.

[34] W. Akkari, A. Belghith, and A. Ben Mnaouer, "Enhancing power saving mechanisms for ad hoc networks using neighborhood information," in *Proc. Int. Wireless Comm. and Mobile Comp. Conf. (IWCMC)*, (Crete, Greece), pp. 794–800, Aug. 2008.

[35] A. Belghith and W. Akkari, "Neighborhood aware power saving mechanisms for ad hoc networks," in *Proc. IEEE Conf. Local Computer Networks (LCN)*, (Montreal, Canada), IEEE, Oct. 2008.

[36] M. H. Ye, C. T. Lau, and A. B. Premkumar, "A modified power saving mode in ieee 802.11 distributed coordinator function," *Comput. Commun.*, vol. 28, pp. 1214–1224, June 2005.

[37] Y.-C. Tseng, C.-S. Hsu, and T.-Y. Hsieh, "Power-saving protocols for ieee 802.11-based multi-hop ad hoc networks," in *Proc. IEEE Infocom*, vol. 1, pp. 200–209, 2002.

[38] C. Hu, R. Zheng, J. C. Hou, and L. Sha, "A microscopic study of power management in ieee 802.11 wireless networks," *Int. Journal Wireless Mobile Computing*, vol. 1, pp. 165–178, Feb. 2006.

[39] R. Zheng and R. Kravets, "On-demand power management for ad hoc networks," in *IEEE INFOCOM*, vol. 1, pp. 481–491, March 2003.

[40] R.-H. Hwang, C.-Y. Wang, C.-J. Wu, and G.-N. Chen, "A novel efficient power-saving MAC protocol for multi-hop MANETs," *Int. Journal of Comm. Systems*, vol. 26, pp. 34–55, Jan. 2013.

[41] N. Rajagopalan and C. Mala, "Modified power save model for better energy efficiency and reduced packet latency," *Am. J. Eng. Applied Sci.*, vol. 5, no. 3, pp. 237–242, 2012.

[42] M.-H. Seo, H.-J. Yoon, and J.-S. Ma, "Fast flooding in Power Save Mode of IEEE 802.11 DCF based mobile ad hoc networks," in *NETWORKING 2004*, vol. 3042 of *Lecture Notes in Computer Science*, pp. 1464–1469, Springer Berlin Heidelberg, 2004.

Paper F

The Stability of Multiple Objective RPL Tree Formation

Ioannis Glaropoulos and Viktoria Fodor

submitted to IFIP Annual Mediterranean Ad Hoc Networking Workshop, 2015.

The Stability of Multiple Objective RPL Tree Formation

Ioannis Glaropoulos, and Viktoria Fodor
Access Linnaeus Center, KTH, Royal Institute of Technology,
Stockholm, Sweden

Abstract

We address the problem of RPL tree formation in self-organized, multihop, wireless sensor networks, where resource-constrained nodes may independently select their routing paths that maximize their performance. We study the result of the tree formation applying a non-cooperative game-theoretic model, and show that multiple objectives may lead to unstable Nash graphs with unwanted traffic cycling. To ensure stability we propose an extension of the node's strategy space, denoted as selective routing, that efficiently eliminates non-acyclic formations from the set of Nash equilibria, while the resulting routing decisions comply with standard RPL.

1 Introduction

Wireless networks formed by low power wireless devices are expected to be key enablers for the emerging Internet of Things. Apart from the limitation on the energy resources, this area of networking is also characterized by a high heterogeneity of connected devices, of the provided networked services and of the physical environment. Therefore, to be successful, the designed networking technologies have to support this heterogeneity. Considering suitable routing protocols, IETF proposes RPL, a Routing Protocol for Low-power and Lossy Networks [1], that supports heterogeneity through the high flexibility in defining routing objectives. In general, these routing objectives need to reflect the requirements of the networking scenario itself, including the long uninterrupted operation and low operational cost. This translates to low, and balanced energy consumption, such that the individual network nodes, as well as the multihop transmission paths to the RPL root have long lifetime. Moreover, routing objectives have to support the application requirements, in terms of end-to-end quality of service, such as delay or packet loss.

In this paper we consider the scenario, where network nodes are selfish, that is, they aim at maximizing their individual performance by independently selecting their routing paths. Our objective is to characterize the topologies that emerge

in Nash Equilibria (NE), for a large class of generic utility functions that reflect multiple performance objectives. We show that the NE topologies are trees, if only end-to-end performance objectives, like delay or packet loss are considered. However, cycles may emerge, if the utilities depend even on the forwarded traffic, like in the case of path lifetime. Therefore, we propose routing over an extended strategy space, allowing some level of routing differentiation according to the traffic source, and prove that in this case the Nash equilibria are stable tree topologies without cycles.

The remainder of the paper is organized as follows. Section 2 introduces the networking scenario and the related analytic models for traffic routing. Section 3 introduces the network formation game, followed by a topology stability analysis under non-congestive and congestive node utility functions. Section 4 presents our main contribution, studying the topology stability under an extended routing strategy space for the network nodes. We present the related work on tree formation games in Section 5 and conclude the paper in Section 6.

2 Networking Scenario and Model

2.1 WSN Topology and Network Stack

We consider a finite set of stationary WSN nodes deployed, arbitrarily, in the area of the network, utilizing a single-channel wireless link layer. The nodes aim at transmitting data traffic to a *gateway* node over multi-hop paths, employing IPv6 and RPL [1]. RPL is a well-known routing protocol, particularly designed for multi-hop, low-power and lossy wireless networks. RPL forms directed acyclic graphs (DAGs) – often denoted as trees for simplicity – rooted at a gateway node. The root node initiates the RPL tree by broadcasting the DAG information object (DIO) message, including among others DAG identifiers, node rank and routing metrics. Upon receiving DIO messages, network nodes select their parent or parent set according to some utility function, select a rank value that is larger than that of the parents, update and re-broadcast the DIO message. RPL preserves the acyclic structure, by forbidding upward traffic forwarding to a node with larger rank. If such a transmission is attempted, RPL initiates a local or a global repair process where the topology is re-organized. To avoid the costly repair procedures the tree construction process should lead to stable RPL tree formations, that is, formations that do not contain cycles.

We denote by $\mathcal{V}$ the finite set of WSN nodes, with $|\mathcal{V}| < \infty$. The matrix $\mathcal{E} \in \mathbb{R}\times\mathbb{R}$ denotes the edge matrix of the network topology, where the element E_{v_i,v_j} denotes the cost of the link between nodes v_i and v_j. The link cost is a decreasing function of the link quality, with an infinite cost for non-existing links. The wireless network is, therefore, represented by a *graph* $\mathcal{G}(\mathcal{V},\mathcal{E})$. Let $v_0 \notin \mathcal{V}$ denote the gateway node. In the context of RPL networks, v_0 will be the root of the RPL instance, and all nodes in $\mathcal{V}$ need to establish routing paths towards v_0. The network is assumed to

have symmetric links, i.e. $E_{v_i,v_j} = E_{v_j,v_i}, \forall v_i, v_j \in \mathcal{V} \cup \{v_0\} = \tilde{\mathcal{V}}$. We denote by N_{v_i} the set of neighbors of node v_i, i.e. $N_{v_i}(\mathcal{E}) = \{u_k \in \tilde{\mathcal{V}} : E_{v_i,u_k} < \infty\}$. We define the *path* between two nodes in $\tilde{\mathcal{V}}$ as follows:

Definition 1. *A path $P_{v_i,v_j}(\mathcal{E})$ of **length** K between nodes $v_i, v_j \in \tilde{\mathcal{V}}$, $i \neq j$ is defined as a sequence of nodes $(u_k), k = 1, \dots K$ such that $u_1 = v_i, u_K = v_j, u_k \in P_{v_i,v_j}(\mathcal{E})$, $\forall k$ and it holds: $E_{u_k,u_{k+1}} < \infty$, $\forall k = 1, \dots, K-1$.*

We assume the existence of a path, P_{v_i,v_j} between any nodes $v_i, v_j \in \tilde{\mathcal{V}}$. In other words, the network topology is *connected*. In general there may exist multiple paths between a pair of nodes, and a path may include cycles, i.e. nodes visited more than one times. Let $\mathcal{P}_{v_i,v_j}(\mathcal{E})$ denote the set of all possible paths between nodes v_i, v_j. We simply denote the path as $P_{v_i}(\mathcal{E})$ when $v_j = v_0$. For node v_i, a *RPL* path, $PL_{v_i} \in \mathcal{P}_{v_i}(\mathcal{E})$ is a path that is used to forward traffic to the root node, v_0. For each node u_k in the path, the next node, u_{k+1} is an RPL *parent* node, while u_k is denoted as *child*. Obviously, $u_{k+1} \in N_{u_k}$. We denote by $\Pi_{v_i}(PL_{v_i})$ the parent of node v_i for that particular RPL path PL_{v_i}. We denote by $\mathcal{PL}_{v_i}(\mathcal{E})$ – or, simply, $\mathcal{PL}_{v_i}$ – the set of all paths of node v_i. Then, the set of parents for v_i is given as

$$\Pi_{v_i} = \bigcup_{PL_{v_i} \in \mathcal{PL}_{v_i}} \Pi_{v_i}(PL_{v_i}). \tag{1}$$

Considering RPL path PL_{v_i}, we define the set of *extended parents* of node v_i as the union of nodes that belong to PL_{v_i}, and denote the set by $\tilde{\Pi}_{v_i}(PL_{v_i})$.

Definition 2. *[Extended parents on a RPL path] The set of indirect parents of node v_i considering path PL_{v_i}:*

$$\tilde{\Pi}_{v_i}(PL_{v_i}) = \{u_k : u_k \in PL_{v_i}\}. \tag{2}$$

The total set of extended parents for node v_i is derived, considering all RPL paths:

$$\tilde{\Pi}_{v_i} = \bigcup_{PL_{v_i} \in \mathcal{PL}_{v_i}} \tilde{\Pi}_{v_i}(PL_{v_i}). \tag{3}$$

A cycle in the RPL network is formed if at least two network nodes include each other in the set of extended parents:

Definition 3. *[Cycle formation] A cycle is formed in a RPL network if there exist two nodes, v_i, v_j for which it holds: $v_j \in \tilde{\Pi}_{v_i}, v_i \in \tilde{\Pi}_{v_j}$.*

2.2 Traffic routing

Let $\mathcal{L} \triangleq \{L_{v_1}, \dots, L_{v_V}\}$ denote the vector of traffic injected by the nodes in the network. Each node v_i forwards incoming traffic from its *children set* and the locally generated workload, L_{v_i}, exclusively via its *parents set*.

We model the result of the parent selection process at the wireless nodes by defining a node's *routing vector* $\mathcal{R}_{v_i} = \{R_{v_i}(u_k)\},\ u_k \in \tilde{\mathcal{V}} \setminus \{v_i\}$, where the element $R_{v_i}(u_k) \in [0,1]$ denotes the portion of incoming or locally generated traffic that is forwarded towards the root via node u_k. Clearly, $R_{v_i}(u_k) = 0,\ \forall u_k \notin \Pi_{v_i}$. $\mathcal{R}_{v_i}$ is, additionally, stochastic, that is $\sum_{u_k \in \tilde{\mathcal{V}} \setminus \{v_i\}} R_{v_i}(u_k) = 1$. $\boldsymbol{\mathcal{R}} \triangleq [\mathcal{R}_{v_i}]$ denotes the routing matrix of the RPL network. We define $L^T_{v_i}$, the *traffic workload* of node v_i recursively as the sum of all incoming and locally injected traffic:

$$L^T_{v_i} = L_{v_i} + \sum_{v_i \in \Pi_{u_k}} R_{u_k}(v_i) L^T_{u_k},\ \forall u_k \in \mathcal{V} \tag{4}$$

While the edge matrix, $\mathcal{E}$, is an input for out network, the routing matrix is the outcome of nodes' decisions on their routing vectors.

2.3 Node objectives

A node v_i decides on its routing vector, $\mathcal{R}_{v_i}$, based on a set of objectives related to its individual performance. We characterise the node objectives as follows. First, the nodes aim at maximizing the quality of service (QoS), when forwarding their own traffic to the gateway. QoS is evaluated over end-to-end performance metrics, such as packet delay, delivery rate, throughput or reliability. Second, under limited energy resources, nodes aim at maximizing their lifetime, i.e at extending for as long as possible their ability to deliver their traffic to the gateway node. Therefore, nodes aim at minimizing their own energy consumption as well as that of their extended parents set. Due to the first objective, a node v_i has an incentive of joining the RPL tree at a level as close as possible to the root node. Such decision, however, may result in v_i having a large sub-tree with high volumes of traffic to be forwarded, clearly contradicting the second objective. In addition, to balance the energy consumption of the nodes in the extended parent set, nodes have an incentive to select less congested paths.

As the outcome of routing decisions according to these multiple objectives is non-trivial, we model the node interactions as non-cooperative strategic game, and evaluate the resulting Nash topologies.

3 RPL Tree Formation Game

3.1 RPL tree formation assumptions

We make the following assumptions that affect the possible actions of the players, that is, the WSN nodes. First, the nodes are able to select a set of preferred parent nodes among their direct neighbors. Most RPL implementations restrict nodes to select a single preferred parent, however, such restriction is not enforced by the standard, and multiple parents are beneficial for balancing the network load.

Supported by the standard, we assume, here, that nodes have no control over their children set, and nodes are forced to advertise the RPL network to their neighbors.

3.2 Non-cooperative routing tree formation

In the *tree formation game* the WSN nodes aim at maximizing their network performance by optimizing their parent selection policy, considering the possible parent selection decisions of the rest of the nodes in the network. The *action* of a node v_i is a routing vector instance, $\mathcal{R}_{v_i}$, and an *action profile* of the game is a vector $\{\mathcal{R}_{v_1}, \ldots, \mathcal{R}_{v_V}\}$. The preference relations of node v_i over all profiles are determined by the node's *utility*, U_{v_i}, that quantitatively evaluates the node's performance as function of the action profile and the *input* parameters, such as the network topology, $\mathcal{G}(\mathcal{V}, \mathcal{E})$, and the traffic workload vector, $\mathcal{L}$. We assume that each node has a prior knowledge of the input parameters and the utility functions [2]. As a result of the above assumption, we model the parent selection game among the nodes in $\mathcal{V}$ as a non-cooperative, strategic game with perfect information [3]. The formulation of the tree formation game is given below.

Definition 4 (Tree formation game)**.** *The considered RPL tree formation game is a finite-player game with the following properties:*

1. *V is the **number** of players*
2. *$Q_{v_i} \triangleq \mathcal{R}_{v_i}, i = 1, 2, \ldots V$ are the actions (or **strategies**) of the players, i.e. the routing vectors. (Q_{v_i}, Q_{-v_i}) denotes a **strategy profile** for the tree formation game.*
3. *$U_{v_i}(Q_{v_i}, Q_{-v_i}), \forall v_i$ describes the player **utility**, that is, in general, a function of the routing vectors of all nodes. RPL nodes prefer an action profile that minimizes their utility function, given the actions of the other nodes.*

The outcome of the game in Def. 4 is the set of routing vectors. Even though, according to the RPL protocol the nodes need to play their actions – join the tree and advertise their routes and the RPL instance, – sequentially, the selected game model is valid, due to the assumption of full prior knowledge of all input parameters. In practice, the nodes will acquire the input parameter values, play the game, and then join the RPL network, as soon as they receive DIO messages from their intended parents. We apply the notion of Nash Equilibrium (NE) as a traditional game theoretic tool of analyzing the behavior of the network nodes and of predicting the outcome of the game.

As nodes can select multiple parents, that is, can follow a mixed strategy, we have the following fundamental property.

Property 1. *The game defined in Def. 4 has at least one mixed strategy NE due to Nash's Existence Theorem.*

3.3 Congestive and Non-congestive utility functions

Consider the generic utility function $U_{v_i}\left(Q_{v_i}, Q_{-v_i}\right),\ \forall v_i \in \mathcal{V}$ for the tree formation game in Def. 4. U_{v_i} may jointly reflect multiple (and possibly counter-acting) performance objectives of the network nodes, e.g. end-to-end delay, throughput, reliability, energy consumption, or lifetime. Therefore, U_{v_i} is, in general, a combination, $\mathcal{H}$, of a set of W sub-utility functions $f_{v_i}^1, \ldots f_{v_i}^W$, each one reflecting a single objective,

$$\begin{aligned} &U_{v_i}\left(Q_{v_i}, Q_{-v_i}\right) \triangleq \\ &\quad \triangleq \mathcal{H}\left(f_{v_i}^1\left(Q_{v_i}, Q_{-v_i}\right), \ldots, f_{v_i}^W\left(\left(Q_{v_i}, Q_{-v_i}\right)\right)\right). \end{aligned} \tag{5}$$

(5) allows for arbitrary sub-utility combinations, with the natural requirement that the aggregate utility function, U_{v_i}, is *strictly monotonic* with each of the sub-utilities, that is:

$$\begin{aligned} &\mathcal{H}\left(Q_{v_i}, Q_{-v_i}\right) < \mathcal{H}\left(Q_{v_i}^{'}, Q_{-v_i}^{'}\right), \quad \text{if} \\ &\quad f^w\left(Q_{v_i}, Q_{-v_i}\right) \leq f_{v_i}^w\left(Q_{v_i}^{'}, Q_{-v_i}^{'}\right), \forall w = 1, \ldots, W, \\ &\text{and} \quad \exists w, \text{s.t. } f_{v_i}^w\left(Q_{v_i}, Q_{-v_i}\right) < f_{v_i}^w\left(Q_{v_i}^{'}, Q_{-v_i}^{'}\right). \end{aligned} \tag{6}$$

In other words, a node's utility function decreases, if at least one sub-utility function has a strictly decreased value, while the remaining sub-utilities do not increase their values. We consider, as a convention, all sub-utilities of the root node to be zero. Under mixed strategy profile the node's sub-utility can be expressed as a combination of the sub-utilities evaluated over each RPL path $PL_{v_i} \in \mathcal{PL}_{v_i} \triangleq \mathcal{PL}_{v_i}\left(Q_{v_i}, Q_{-v_i}\right)$,

$$f_{v_i}^w\left(Q_{v_i}, Q_{-v_i}\right) = \mathcal{F}\left(f_{v_i}^w\left(PL_{v_i,1}\right), \ldots, f_{v_i}^w\left(PL_{v_i,|\mathcal{PL}_{v_i}|}\right)\right). \tag{7}$$

Additionally, we have,

$$U_{v_i}(PL_{v_i}) = \mathcal{H}(f_{v_i}^1(PL_{v_i}), \ldots, f_{v_i}^W(PL_{v_i})). \tag{8}$$

3.3.1 Non-congestive functions

Utility functions that reflect traditional end-to-end network performance metrics – such as packet delivery delay, throughput or loss rate – depend on the length of the formed paths towards the root node, while they do not incorporate any notion of network resource (such as energy) sharing. We denote, here, such functions as *non-congestive*.

Definition 5 (Non-congestive function)**.** *A sub-utility function $f_{v_i}^w\left(Q_{v_i}, Q_{-v_i}\right)$ is a non-congestive function, when it satisfies $\forall\left(Q_{v_i}, Q_{-v_i}\right)$ and $\forall v_i \in \mathcal{V}$ the following two conditions:*

1. $f_{v_i}^w\left(Q_{v_i}, Q_{-v_i}\right) = f_{v_i}^w\left(Q_{v_i}^{'}, Q_{-v_i}^{'}\right),\ \forall Q_{-v_i}^{'} \neq Q_{-v_i}, \quad$ *if*

$$Q_{u_k} = Q'_{u_k},\ \forall u_k \in \{v_i\} \cup \tilde{\Pi}_{v_i}, \quad \textit{and}$$
$$2. \quad f^w_{v_i}(PL_{v_i}) > f^w_{v_i}(PL_{u_k}),\ \textit{if} \quad PL_{v_i} \supset PL_{u_k} \qquad . \tag{9}$$

In other words, a node retains its utility under a different strategy profile, when it does not alter its own strategy, while no node in his extended parents set modifies its strategy profile. In addition, non-congestive functions are strictly *monotonic*, that is, intermediate nodes on a path to root exhibit strictly lower utility values compared to the starting node of the path. We restrict the analysis to linear combinations of the sub-utilities over each formed path. Thus (7) becomes:

$$\begin{aligned} &f^w_{v_i}\left(Q_{v_i}, Q_{-v_i}\right) = \\ &\quad = \textstyle\sum_{PL_{v_i} \in \mathcal{PL}_{v_i}\left(Q_{v_i}, Q_{-v_i}\right)} q\left(PL_{v_i}\right) f^w_{v_i}\left(PL_{v_i}\right), \end{aligned} \tag{10}$$

where $q\left(PL_{v_i}\right)$ denotes the percentage of traffic forwarded via path PL_{v_i}, and $\sum_{PL_{v_i} \in \mathcal{PL}_{v_i}\left(Q_{v_i}, Q_{-v_i}\right)} q\left(PL_{v_i}\right) = 1$. Non-congestive utility functions include, but are not limited to, *additive* functions, where it, additionally, holds:

$$f^w_{v_i}(PL_{v_i}) = f^w_{u_k}(PL_{u_k}) + f^w_{v_i}(E_{v_i,u_k}), \quad u_k = \Pi_{v_i}(PL_{v_i}). \tag{11}$$

In the following we show a characteristic property of the tree formation game when the node utility function comprises solely of non-congestive sub-utilities.

Property 2. *Consider the game of Def. 4 where the node utilities satisfy* (6)*, while the sub-utilities,* f^w *are non-congestive (Def. 5). Any NE of this game corresponds to an acyclic graph.*

Proof. Assume that there exist a NE which is not an acyclic graph. Then, there exist at least two nodes, u_1, u_2 which participate in a cycle formation. This is equivalent to u_1, u_2 having each other in their extended parents set: $u_1 \in \tilde{\Pi}_{u_2}, u_2 \in \tilde{\Pi}_{u_1}$. Denote by PL_{u_1}, PL_{u_2} the alternative RPL paths of u_1, u_2 to root, respectively. For the alternative paths it holds, $u_1 \notin PL_{u_2}, u_2 \notin PL_{u_1}$. Assume, without loss of generality, that

$$U_{u_1}\left(PL_{u_1}\right) \leq U_{u_2}\left(PL_{u_2}\right), \tag{12}$$

where $U_{u_k}(PL_{u_k}) = \mathcal{H}\left(f^1(PL_{u_k}), \ldots, f^W(PL_{u_k})\right)$. For the utility of u_1 the following inequality holds, due to (9), and as a result of traffic circulation between u_1 and u_2:

$$U_{u_1} > q(PL_{u_1}) U_{u_1}(PL_{u_1}) + (1 - q(PL_{u_1})) U_{u_1}\left(PL_{u_2}\right),$$

Due to (12), it holds

$$U_{u_1} > U_{u_1}(PL_{u_1}),$$

so u_1 benefits by deviating all traffic via its own path, instead of the cycle, violating the assumption of a NE. □

We show a second property of the RPL tree formation game that holds for additive, non-congestive utility functions. We start by giving the following definition for the minimum-distance tree.

Definition 6. *A Minimum-Distance RPL Tree (mDRT) is an RPL DAG where parents are selected so that each node has a minimum utility, U, towards the root, where U is a function $\mathcal{H}$ of non-congestive sub-utilities. We denote the minimum distance path-to-root of node v_i as $PL_{v_i}^{(min)}$.*

The following property holds for non-congestive functions that, additionally, satisfy (11).

Property 3. *Consider the game of Def. 4 where the node utilities satisfy* (6)*, while the node sub-utilities are non-congestive (Def. 5) and additive* (11)*. Any NE of this game corresponds to an acyclic tree formation, which results in a minimum distance RPL tree.*

Proof. Consider a NE of the tree formation game. Due to Property 2 the NE is an acyclic graph. Assume that the NE does not correspond to a minimum distance RPL tree. Then there exists a path $PL_{u_k} \in \mathcal{PL}_{u_k}(Q_{u_k}, Q_{-u_k})$, for some node u_k, that is not a minimum distance path, i.e. $PL_{u_k} \neq PL_{u_k}^{(\min)}$, while all paths of the parents of u_k are minimum distance paths, that is:

$$PL_{u_j} = PL_{u_j}^{(\min)},\ \forall PL_{u_j} \in \mathcal{PL}_{u_j},$$
$$\forall u_j \in \bigcup_{PL_{u_k} \in \mathcal{PL}_{u_k}} \Pi_{u_k}(PL_{u_k}).$$

Since we have assumed a NE, PL_{u_k} has been selected as a best response:

$$\begin{aligned} &PL_{u_k} = \\ &= \{PL_{u_k} : U_{u_k}(PL_{u_k}) \leq U_{u_k}(PL_{u_k}^{'}), \forall PL_{u_k}^{'} \in \mathcal{PL}_{u_k}\} \\ &= \{PL_{u_k} : \mathcal{H}(f_{u_k}^1(PL_{u_k}), \ldots, f_{u_k}^W(PL_{u_k}) \leq \\ &\quad \leq \mathcal{H}(f_{u_k}^1(PL_{u_k}^{'}), \ldots, f_{u_k}^W(PL_{u_k}^{'})), \forall PL_{u_k}^{'} \in \mathcal{PL}_{u_k}\}. \end{aligned}$$

For any sub-utility f^w, however, it holds:

$$f_{u_k}^w(PL_{u_k}^{'}) = f_{\Pi_{u_k}}^w(PL_{\Pi_{u_k}}^{'}) + f_{\Pi_{u_k}(L_{u_k}, \Pi_{u_k})}^w < f_{\Pi_{u_k}}^w(PL_{\Pi_{u_k}}^{'})$$

where any $PL_{\Pi_{u_k}}^{'}$ is a minimum distance path. Consequently,

$$f_{u_k}^w(PL_{u_k}^{'}) \leq f_{u_k}^w(PL_{\Pi_{u_k}}^{''})$$

for any $PL_{\Pi_{u_k}}^{''}$ that is not a minimum-distance tree. Due to the above inequalities,

and the monotonic behavior of $\mathcal{H}$, we can write:

$$\begin{aligned}
&PL_{u_k} = \\
&= \{PL_{u_k} : U_{u_k}(PL_{u_k}) \leq U_{u_k}(PL^{'}_{u_k}), \forall PL^{'}_{u_k} \in \mathcal{PL}_{u_k}\} \\
&= \{PL_{u_k} : \mathcal{H}(f^1_{u_k}(PL_{u_k}), \ldots, f^W_{u_k}(PL_{u_k}) \leq \\
&\quad \leq \mathcal{H}(f^1_{u_k}(PL^{''}_{u_k}), \ldots, f^W_{u_k}(PL^{''}_{u_k})), \forall PL^{''}_{u_k} \in \mathcal{PL}_{u_k}\}.
\end{aligned}$$

Consequently, PL_{u_k} is a minimum distance tree. □

3.3.2 Congestive node sub-utilities

Utility functions that reflect a notion of resource sharing (congestion) in the wireless network are for example a node's communication energy cost, or path lifetime. We denote such functions as *congestive* utilities. As the utility values depend on the amount of traffic generated or forwarded by the node or the path, the utility may change if any node sharing a part of a common path towards the RPL root changes its strategy.

Definition 7 (Congestive function)**.** *A sub-utility function $f^w_{v_i}(Q_{v_i}, Q_{-v_i})$ is a congestive function, when it satisfies $\forall v_i$ and $\forall (Q_{v_i}, Q_{-v_i})$ the following condition:*

$$f^w_{v_i}(Q_{v_i}, Q_{-v_i}) = f^w_{v_i}\left(Q_{v_i}, Q'_{-v_i}\right),\ Q_{-v_i} \neq Q'_{-v_i}, \quad \textit{if}$$

$$\begin{aligned}
&Q_{u_k} = Q'_{u_k},\ \forall u_k : u_k \in \{\{v_i\} \cup \tilde{\Pi}_{v_i}\}\ \textit{and} \\
&\forall u_k : \tilde{\Pi}_{u_k} \cap \tilde{\Pi}_{v_i} \neq \varnothing \wedge \\
&\qquad \exists u_j : u_j \in \tilde{\Pi}_{u_k} \cap \tilde{\Pi}_{v_i} \wedge u_j \notin \textstyle\bigcap_{\mathcal{PL}_{u_k}} PL_{u_k}
\end{aligned} \tag{13}$$

In other words, a node v_i retains its utility if *i)* it itself and its extended parents do not change their strategies and *ii)* any other node that has common extend parent with v_i, but this extended parent is not part of all its paths, does not change its strategy. If all nodes considered in (13) do not change their actions, then due to load preservation, formally defined in in (4), neither v_i, nor any of its extended parents will experience a different traffic workload value, thus they will retain their utilities. In the context of this work we focus, exclusively, on congestive sub-utilities, whose values are increasing functions of the traffic workload:

$$f^w_{v_i}(L^T_{v_i}) > f^w_{v_i}(L^{T\,\prime}_{v_i}) \text{if } L^T_{v_i} > L^{T\,\prime}_{v_i}. \tag{14}$$

Unfortunately, under congestive sub-utilities, the tree formation game may result in Nash equilibria that include traffic cycles. As an example, consider the two nodes, v_1 and v_2 (Fig. 1), with paths to the root, PL_{11} and PL_{22}, respectively. Assume that

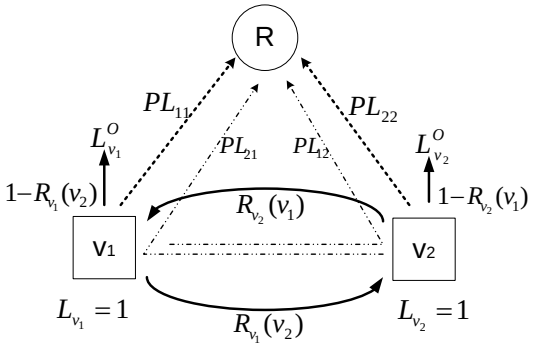


Figure 1: Explanatory figure illustrating an example of a non-acyclic NE.

the nodes' utility function is a linear combination of a non-congestive sub-utility, $f^{NC}_{v_i}, i = 1, 2$, and a congestive one, $f^C_{v_i}$, for which it holds:

$$f^C_{v_i} = \max_{\forall PL_j \in \mathcal{PL}_{v_i}} f^C_{v_i}(PL_j), \tag{15}$$

where $f_{v_i}(PL_j) = \max_{u_k \in PL_j} f^C_{u_k}(L^T_{u_k})$. Consider, finally, $U_{v_i} = \alpha f^{NC}_{v_i} + \beta f^C_{v_i}$, $\alpha, \beta > 0$. Consider a possible NE strategy profile, under which $R_{v_1}(v_2) \neq 0, R_{v_2}(v_1) \neq 0$, that is, the nodes form a cycle under Def. 3. Consider a symmetric scenario with $f^C_{v_1}(PL_{11}) = f^C_{v_1}(PL_{12}) = f^C_{v_2}(PL_{21}) = f^C_{v_2}(PL_{22})$ and $f^{NC}_{v_1}(PL_{11}) = f^{NC}_{v_2}(PL_{22})$. Assuming unit traffic loads at v_1, v_2, the outgoing traffic workload at v_1 will be:

$$L^O_{v_1} = \frac{(1 + R_{v_2}(v_1))}{R_{v_2}(v_1) + (1 - R_{v_2}(v_1))/(1 - R_{v_1}(v_2))} \tag{16}$$

Any strategy $R'_{v_1}(v_2) \neq R_{v_1}(v_2)$ will change the traffic $L^O_{v_1}, L^O_{v_2}$, exiting the nodes, v_1, v_2. As the sum of the exiting traffic needs to remain constant, one of these will increase, increasing the related congestive sub utility due to (14), and consequently, the congestive utility of v_1 due to (15).

If $R'_{v_1}(v_2) > R_{v_1}(v_2)$, the non-congestive sub-utility of v_1 increases, as v_1 forwards more traffic via the cycle. Consequently, U_{v_1} increases. Instead, if $R'_{v_1}(v_2) < R_{v_1}(v_2)$, the non-cogestive sub-utility of v_1 decreases, at maximum by $(R_{v_1}(v_2) - R'_{v_1}(v_2)) \cdot (f^{NC}(PL_c))$, where $f^{NC}(PL_c)$ is the non-cogestive sub-utility cost due to circulation of traffic. Due to (14), $f^C_{v_1}(PL_{11})$ will increase with a rate proportional to $L^{O\,\prime}_{v_1} - L^O_{v_1}$, which increases with a rate proportional to $(R_{v_1}(v_2) - R'_{v_1}(v_2))$:

$$L^{O\,\prime}_{v_1} - L^{O}_{v_1} = \frac{(1+R_{v_2}(v_1))}{R_{v_2}(v_1)+(1-R_{v_2}(v_1))/(1-R'_{v_1}(v_2))} - \\ - \frac{(1+R_{v_2}(v_1))}{R_{v_2}(v_1)+(1-R_{v_2}(v_1))/(1-R_{v_1}(v_2))} \\ > \frac{(R_{v_1}(v_2)-R'_{v_1}(v_2))(1+R_{v_2}(v_1))}{(1-R'_{v_1}(v_2))R_{v_2}(v_1)+(1-R_{v_2}(v_1))} \\ > (R_{v_1}(v_2) - R'_{v_1}(v_2))(1 + R_{v_2}(v_1)),$$

If β/α is sufficiently large, the increase in $f^{C}_{v_1}$ can not be compensated by the decrease in $f^{NC}_{v_1}$. As a result, v_1 retains its strategy. Applying the same argumentation for v_2 concludes that the considered routing strategy profile with traffic cycling, can be a NE.

4 Selective RPL-based routing

4.1 Selective routing

We propose a modification of the RPL parent selection and routing policies, with the aim of avoiding cycle formation in NE. We denote the proposed scheme as *selective routing*. Under selective routing the nodes are allowed to route the incoming traffic through different RPL paths, that are selected based on the origin of the incoming traffic. In terms of modeling, selective routing expands the action *space* of the nodes in the tree formation game of Def. 4. Instead of routing vectors, a node's action is an entire routing array that determines the relaying of the incoming traffic.

Definition 8. *[Selective routing] The action of node v_i is determined by the selective routing array $\boldsymbol{\mathcal{R}}_{v_i} \in (0,1)^{V-1} \times (0,1)^{V-1}$, where the array element $\mathcal{R}_{v_i}(v_j, v_k) \in (0,1)$ declares the maximum of the portion of traffic workload generated at remote node v_j that v_i forwards to its parent v_k.*

For each v_i and v_j it holds that $\sum_{v_k \in \tilde{\mathcal{V}} \setminus \{v_i\}} R_{v_i}(v_j, v_k) = 1, \quad \forall v_j \in \mathcal{V}$. The above definition does not specify whether a routing selection is stochastic or deterministic. In the context of this work, we assume that selective routing is applied on each arriving (or generated) packet, once the packet first arrives at a node. Any time this packet needs to be forwarded, this is done based on the initial routing decision.

Consider an arbitrary node v_j. Denote by $P_{j,i} \in [0,1]$ the ratio of the traffic workload generated by v_j (denoted as L_{v_j}) that arrives at node v_i. Then, the amount of v_j traffic v_i actually relays to v_k is calculated as: $\min\{P_{j,i}, \mathcal{R}_{v_i}(v_j, v_k)\} \cdot L_{v_j}$, that is, it will be lower than the $\mathcal{R}_{v_i}(v_j, v_k) L_{v_j}$, if $P_{j,i} < \mathcal{R}_{v_i}(v_j, v_k)$.

We consider a specific implementation of selective routing, where node v_i categorizes the handled traffic into two classes. The first class is denoted as the *local* traffic, and includes all traffic generated at v_i, as well as any incoming traffic that does not reach the root through any other node that is involved in cycle shared with

v_i. On the contrary, the *remote* traffic includes all traffic generated by nodes in the cycle, as well as all traffic reaching the root via any other node in the formed cycle. Note, that nodes not being part in a cycle have only local traffic.

4.2 Analysis

The following property gives the main result of this work regarding the efficiency of applying selective routing for cycle avoidance under the tree formation game.

Property 4. *When nodes are allowed to implement selective routing, as specified in Def. 8, a Nash Equilibrium tree formation under the game defined in Def. 4 is an acyclic graph.*

Proof. We prove this statement by contradiction, eliminating the possibility of a NE considering all possible cases of cycle formation. Consider a simple topology with nodes u_1, u_2 (Fig. 2(a)-2(e)). Based on Def. 4, the actions of u_1, u_2 are the tuples $(R_{1,0}, R_{1,2}), (R_{2,0}, R_{2,1})$, where, $R_{1,2} = R_{u_1}(u_2), R_{1,0} = 1 - R_{1,2}$, and $R_{2,1} = R_{u_2}(u_1), R_{2,1} = 1 - R_{2,0}$ and $R_{u_1}(.), R_{u_2}(.)$ are the parent selection vector elements defined in Section 2.2.

Consider a NE tree formation, under which a set of two nodes u_1, u_2 form a cycle, that is, $R_{1,2}, R_{2,1} \neq 0$. Due to the assumption of a connected network, at least one of u_1, u_2 has a valid RPL path to the root. We distinguish among the different cases regarding the paths of nodes u_1, u_2 to the root.

Case 1: Node u_1 with path to the root. Consider the case where only node u_1 has a path to the root node (Fig. 2(a)). In this case, u_1 benefits from setting $R_{1,1} = 0$, as any non-congestive metric strictly decreases, due to (9), while, additionally, all nodes in $\tilde{\Pi}_{u_1}$ retain the values of any congestive sub-utility, as the traffic workload does not change, and u_1 strictly decreases its own congestive utility by avoiding traffic circulation.

Case 2: Nodes u_1, u_2 with a common path to root. Consider, now, the case where u_1, u_2 share the same path to the root node, that is, $\tilde{\Pi}_{u_1} = \tilde{\Pi}_{u_2}$. (Fig. 2(b)). Assume, without loss of generality, that any non-congestive utility, evaluated over the path PL_{u_1} – that includes the E_{u_1,u_3} link – has a lower value, than if it is evaluated over PL_{u_2}. Then, u_1 prefers setting $R_{1,0} = 1$, as it decreases any non-congestive utility function. Similarly to Case 1, any node in the common path retains its congestive sub-utility values when $R_{1,0} = 1$, as (13) is satisfied, while u_1 achieves a lower metric for any congestive sub-utility due to cycle avoidance that decreases the incoming workload.

Case 3: nodes u_1, u_2 with the same bottleneck node to root. Consider the scenario in Fig. 2(c), where u_1, u_2 have different paths to the root, with a common *bottleneck* node u_b that forwards all their traffic to the root, and for which it holds: $L^T_{u_b} \geq L^T_{u_l}, \forall u_l \in PL_{u_1} \cup PL_{u_2}$. We assume, without loss of generality that node u_1 has a better or similar path to root with respect to any non-congestive utility function, compared to u_2. (We will consider this assumption in all the remaining cases of

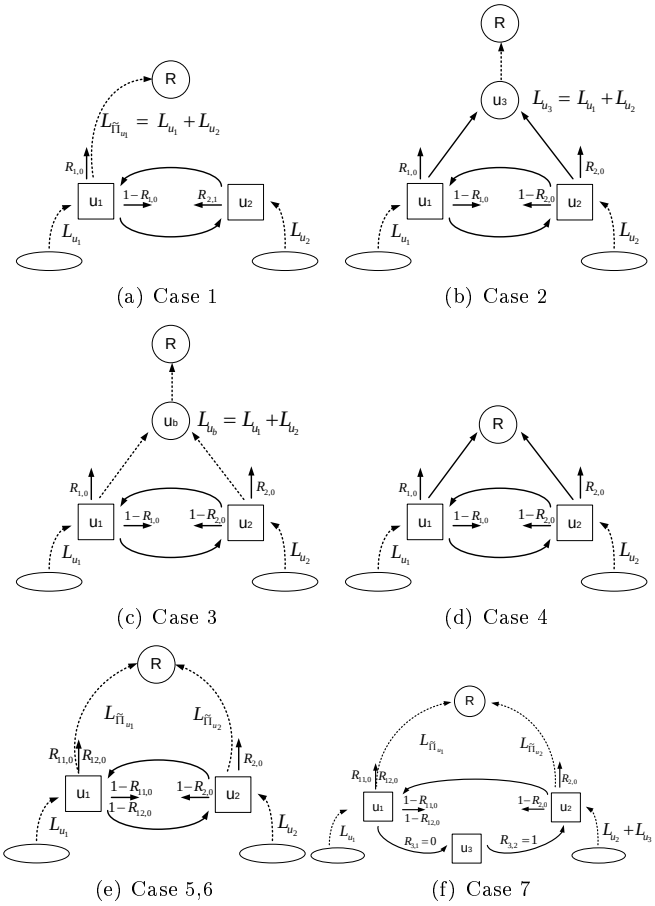


Figure 2: Explanatory diagrams for the different scenarios consider in the proof of Property 4.

this proof.) Then u_1 benefits from setting $R_{1,0} = 1$. First, it strictly decreases its non-congestive utilities. Second, it decreases even its total load, $L^T_{u_1}$, while it keeps the congestive sub-utilities over the paths PL_{u_1} and PL_{u_2} the same, since the total load $L^T_{u_b}$ remains unaffected.
Case 4: nodes u_1, u_2 with direct connection to root. This scenario (Fig. 2(d)) is similar to Case 2, as any congestive sub-utility is determined by the total traffic workload at the nodes u_1, u_2. Therefore, node u_1 benefits from setting $R_{1,0} = 1$, decreasing both its non-congesting utilities, as well as its handled traffic workload by avoiding traffic circulation between u_1, u_2.
Case 5: nodes u_1, u_2 with disjoint paths to root. This is the case of the example in Section 3.3.2, that is, cycles may exist if selective routing is not applied. In this case (Fig. 2(e)), in contrast to Case 4, the congestive sub-utilities depend on the traffic forwarded via each of the disjoint paths to the root. Denote by $\hat{L}_{u_1}, \hat{L}_{u_2}$ the *local* traffic workload at nodes u_1, u_2, as defined in Section 4.1:

$$\hat{L}_{u_1} = L_{u_1} + \sum_{u_1 \in \Pi_{u_j}, u_j \neq u_2} R_{u_j}(u_1) L^T_{u_j}$$

$$\hat{L}_{u_2} = L_{u_2} + \sum_{u_2 \in \Pi_{u_j}, u_j \neq u_1} R_{u_j}(u_2) L^T_{u_j}$$

Then, in the absence of selective routing, the amount of traffic exiting the cycle from each path to the root is calculated as:

$$L_{\tilde{\Pi}_{u_1}} = \frac{R_{1,0}\hat{L}_{u_1} + R_{1,0}\hat{L}_{u_2}(1 - R_{2,0})}{1 - (1 - R_{1,0})(1 - R_{2,0})} \tag{17}$$

$$L_{\tilde{\Pi}_{u_2}} = \frac{R_{2,0}\hat{L}_{u_2} + R_{2,0}\hat{L}_{u_1}(1 - R_{1,0})}{1 - (1 - R_{1,0})(1 - R_{2,0})} \tag{18}$$

Consider that u_1 implements selective routing for its *local* and *remote* traffic. $\mathcal{R}_{11,0}$ denotes the portion of *local* traffic that can be sent directly over the u_1 path to root, while $\mathcal{R}_{12,0}$ represents the ratio of *remote* traffic that can be relayed over the same path. Formally, the action profile of u_1 is the vector of elements, $\mathcal{R}_{u_1}(u_j, u_k), j, k \neq 1$, where:

$$\mathcal{R}_{u_1}(u_j, u_k) = \begin{cases} \mathcal{R}_{11,0}, u_k = \Pi_{u_1}(PL_{u_1}), \forall u_j \text{ outside the cycle}, \\ \mathcal{R}_{12,0}, u_k = \Pi_{u_1}(PL_{u_1}), \forall u_j \text{ inside the cycle}, \\ 1 - \mathcal{R}_{11,0}, u_k = u_2, \ \forall u_j \text{ outside the cycle}, \\ 1 - \mathcal{R}_{12,0}, u_k = u_2, \ \forall u_j \text{ inside the cycle}, \\ 0, \text{ otherwise} \end{cases}$$

Then, (17), (18) become:

$$L^{(S)}_{\tilde{\Pi}_{u_1}} = \mathcal{R}_{11,0}\hat{L}_{u_1} + \min\{1 - R_{2,0}, \mathcal{R}_{12,0}\}\hat{L}_{u_2}, \tag{19}$$

$$L^{(S)}_{\tilde{\Pi}_{u_2}} = (1 - R_{11,0})\hat{L}_{u_1} + (1 - \min\{1 - R_{2,0}, \mathcal{R}_{12,0}\})\,\hat{L}_{u_2}, \tag{20}$$

where $(.)^{(S)}$ indicates the load after selective routing is implemented. In (17), we have:

$$\begin{aligned} \tfrac{R_{1,0}\hat{L}_{u_2}(1-R_{2,0})}{1-(1-R_{1,0})(1-R_{2,0})} = \tfrac{R_{1,0}\hat{L}_{u_2}(1-R_{2,0})}{R_{1,0}+R_{2,0}-R_{1,0}R_{2,0}} = \\ = \tfrac{\hat{L}_{u_2}(1-R_{2,0})}{1+R_{2,0}/R_{1,0}-R_{2,0}} < \hat{L}_{u_2}(1 - R_{2,0}). \end{aligned} \tag{21}$$

Consequently, selecting $\mathcal{R}_{11,0} = \frac{R_{1,0}}{1+(1-R_{1,0})(1-R_{2,0})}$, and, $\mathcal{R}_{12,0} = \frac{(1-R_{2,0})}{1+R_{2,0}/R_{1,0}-R_{2,0}}$, we get from (19),

$$L^{(S)}_{\tilde{\Pi}_{u_1}} = L_{\tilde{\Pi}_{u_1}}. \tag{22}$$

However, as a result of (21), and since $\mathcal{R}_{11,0} \in (0,1)$ and $\min\{1 - R_{2,0}, \mathcal{R}_{12,0}\} \in (0,1)$, there is always a pair of values $\mathcal{R}_{11,0} > \frac{1}{1+R_{2,0}/R_{1,0}-R_{2,0}}$, and $\mathcal{R}_{12,0} < \frac{(1-R_{2,0})}{1+R_{2,0}/R_{1,0}-R_{2,0}}$, such that (22) holds. Such action profile will decrease any non-congestive sub-utility for node u_1, as it deviates a higher portion of local traffic through a better path to the root node. Moreover, any congestive sub-utility function, evaluated over the u_1 path to root, will retain its value because of (22). Finally, the traffic workload arriving at node u_1 from the cycle (node u_2) will be:

$$L^{C}_{u_1} = \frac{\hat{L}_{u_1}(1 - \mathcal{R}_{11,0})}{1 - (1 - R_{2,0})} + \frac{\hat{L}_{u_2}(1 - \min\{1 - R_{2,0}, \mathcal{R}_{12,0}\})}{1 - (1 - R_{2,0})}. \tag{23}$$

Clearly, due to (19), it holds $L^{C}_{u_1} = \text{const.}$, for any pair of $\mathcal{R}_{11,0}, \mathcal{R}_{12,0}$, for which (22) is satisfied. As a result, u_1 benefits from implementing selective routing. As $\mathcal{R}_{11,0}, \mathcal{R}_{12,0} \in [0,1]$, we concentrate on two boundary cases below.

Case 5.1: $\mathcal{R}_{12,0} = 0$.

In this scenario node u_1 implements selective routing, but forwards no remote traffic through its own path to root. In this case u_2 responds better by setting $R_{2,0} = 1$, thus, deviating all its incoming traffic through its own path to root, which decreases any non-congestive usb-utility due to (9), while congestive sub-utilities do not increase, as the workload exiting through path PL_{v_2}, $L_{\tilde{\Pi}_{u_2}}$ is unaffected.

Case 5.2: $\mathcal{R}_{11,0} = 1, \mathcal{R}_{12,0} \neq 0$.

In this case u_1 implements selective routing, forwarding both local and remote traffic via its path to root. If $1 - R_{2,0} \leq \mathcal{R}_{12,0}$, there is no traffic cycling. If $1 - R_{2,0} > \mathcal{R}_{12,0}$, then a non-zero portion, $1 - R_{2,0} - \mathcal{R}_{12,0} > 0$, of local ($u_2$) traffic is subject to cycling between nodes u_1, u_2, until it is forwarded to root via PL_{u_2}. Consequently, u_2 benefits from setting $R_{2,0} = \mathcal{R}_{12,0}$, strictly decreasing its non-congestive sub-utilities, due to (9), while the congestive sub-utilities over the outgoing paths PL_{u_1}, PL_{u_2} are unaffected and the local traffic workload at node u_2 is decreased due to cycle elimination.

Case 6: u_1, u_2 *with disjoint paths to root and* u_2 *implementing selective routing.*

Case 5 shows that the node with a shorter path to root benefits from implementing selective routing, which, eventually leads to cycle elimination. Consider, now, that u_2 implements selective routing with $\mathcal{R}_{21,0}, \mathcal{R}_{22,0} \in (0,1)$. If u_1 does not implement selective routing (and $R_{1,2} \neq 0$), the case is similar to 5.1, 5.2, where u_1 benefits from setting $R_{1,2} = 0$. Assume, now, that u_1 implements selective routing with $\mathcal{R}_{11,0}, \mathcal{R}_{12,0} \neq 0$. If $\mathcal{R}_{12,0} \geq 1 - \mathcal{R}_{22,0}$ and $\mathcal{R}_{21,0} \geq 1 - \mathcal{R}_{11,0}$ there is no cycle. Otherwise, due to the definition of selective routing, there exists a non-zero portion of traffic that is never forwarded to the root node, which results in infinite traffic workload, thus, an infinite value of the congestive sub-utilities due to (14). Therefore, either node benefits from breaking the cycle, so as to obtain a finite value for its congestive sub-utilities.

The above analysis shows that selective routing, as defined in Def 8, eliminates cycle formation between two nodes in a NE. The last part of the proof will show that cycles containing additional nodes can not be formed under a NE. We distinguish between two sub-cases:
Case 7.1: Intermediate nodes in the cycle with no alternative paths to root.
The scenario of intermediate nodes in the cycle (Fig. 2(f)) can be reduced to case 5, where each of u_1, u_2 groups the incoming traffic from the intermediate nodes together with the rest of the remote traffic they handle.
Case 7.2: Intermediate nodes in the cycle with alternative paths to root. This case generalizes the scenarios with only two nodes having alternative paths to root. There always exists a node with a shorter path to the root, with respect to the non-congestive sub-utilities. If no node implements selective routing, 7.2 can, then, be reduced to Case 5, where this particular node will deviate and implement selective routing triggering responses that eliminate the cycle formation. If any node implements selective routing, 7.2 reduces to Case 6. □

4.3 Discussion

Property 4, together with Property 1, ensure that if the parent selection game can converge, the resulting RPL topology is always a stable, acyclic graph, for the wide class of monotonic utility functions. Note, however, that Property 4 does not suggest an algorithmic solution for convergence to the NE strategy profile. Game convergence to Nash trees has been studied, primarily, in the basis of best-response dynamics [4][5]. It lies, however, outside the scope of our analysis.

While the solution to avoid cycles under generic utility functions is the introduction of selective routing in the route selection game, the proposed implementation of the selective routing strategy, distinguishing local and remote traffic, ensures that the resulting Nash equilibrium – without cycles and, therefore, without any remote traffic at the nodes – does not apply selective routing. As a consequence, the construction of the RPL DAG with the DIO and DAO messages and the data routing itself can be performed according to the RPL standard.

5 Related work

Game theory has been extensively applied in self-organized wireless networks as a method of studying the effect of network nodes acting as independent agents, towards a broad range of network objectives that involve connectivity and topology control [6][7][8], parent admission control [9], routing [10][11][12], and energy-balanced [4] and secure [13] network design. In routing, or network formation games, the employed game theoretic models include, predominantly, strategic, non-cooperative games, [14][13], or repeated games [4][15][5], depending on whether the wireless nodes converge on a routing matrix after a sequence of game iterations.

Power control with requirements of connectivity is considered in [6]. It shows that NE exists and can be efficiently computed, and gives an upper bound on the total power cost. [15] defines a utility function that reflects both the reward and the cost of maintaining a network link. Considering a repeated network formation game, the authors show that Nash equilibia exist, however, finding the nodes' best response is in general NP-hard. [5] considers a scenario where link formations are agreed by parent and child nodes, and shows that convergence into a Nash tree network is guaranteed under a finite number of game iterations. A similar scenario is addressed in [16], where nodes establish links on a contract-based bilateral negotiation basis, where cost is a function of the relayed traffic, the negotiated link price and the eventual lack of connectivity, a parameter that is shown to crucially affect the stability of network formation. [4] discusses network formation games under the requirement of maximizing the lifetime of the network. The authors propose an iterative best response-based algorithm that is shown to always converge to a Nash equilibrium, where local energy cost at the nodes is minimized. Similarly, [17] focuses in utilites that capture both the reward of connectivity and the cost of preserving network links and propose an iterative improvement algorithms that guarantees convergence into Nash trees.

Significant attention is devoted to scenarios concerning self-organized networks with conflicting node objectives [14][18], showing that pure strategy NE may not always exist. Mixed strategy Nash equilibria are considered in [20] for bottleneck games, where the utility of the nodes depends on the amount of traffic congestion on their routing paths. [20] shows that the quality of a mixed strategy NE is equal to the one of the social optimum if nodes select routing paths with minimum number of bottleneck links.

6 Conclusion

We addressed the problem of RPL tree formation considering selfish network nodes. We formulated the topology formation problem as a non-cooperative strategic game, where nodes unilaterally select their set of RPL parents, and studied the resulting Nash equilibria with respect to the characteristic properties of the nodes' utility

function. We showed that under utility functions, comprised solely of non-congestive sub-utilities, the resulting Nash topology is an acyclig graph, however, if also congestive sub-utilities are included there may exist Nash topologies that include cycles among the nodes, leading to unstable RPL instances. Therefore, we proposed an extension of the nodes' strategy profile space, denoted as selective routing, and proved that it eliminates cycle formations in Nash equilibrium topologies. Moreover, in the resulting equilibria, selective routing strategies are never preferred by the nodes, which allows to use the standard RPL DAG construction and data forwarding protocols.

References

[1] T. Winter (Ed.), P. Thubert (Ed.), and RPL Author Team. RPL: IPv6 Routing for Low Power and Lossy Networks. RFC 6550, March 2012.

[2] S. Vasudevan, D. Towsley, D. Goeckel, and R. Khalili. Neighbor Discovery in Wireless Networks and the Coupon Collector's Problem. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, pages 181–192, New York, NY, USA, 2009.

[3] M. J. Osborne and A. Rubinstein. *Course in Game Theory*. MIT Press, Mass., 1994.

[4] N. Sadagopan, M. Singh, and B. Krishnamachari. Decentralized Utility-based Sensor Network Design. *Mobile Networks and Applications*, 11(3):341–350, 2006.

[5] W. Saad, Zhu Han, T. Basar, M. Debbah, and A. Hjorungnes. Network Formation Games Among Relay Stations in Next Generation Wireless Networks. *IEEE Transactions on Communications*, 59(9):2528–2542, Sept. 2011.

[6] S. Eidenbenz, V. S. A. Kumar, and S. Zust. Equilibria in topology control games for ad hoc networks. *Mob. Netw. Appl.*, 11(2):143–159, April 2006.

[7] W. Saad, Z. Han, M. Debbah, and A. Hjorungnes. A distributed merge and split algorithm for fair cooperation in wireless networks. In *IEEE International Conference on Communications Workshops, (ICC Workshops)*, pages 311–315, 2008.

[8] W. Saad, Zhu Han, M. Debbah, and A. Hjorungnes. Network Formation Games for Distributed Uplink Tree Construction in IEEE 802.16j Networks. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1–5, Nov 2008.

[9] D. Niyato and E. Hossain. Qos-aware bandwidth allocation and admission control in ieee 802.16 broadband wireless access networks: A non-cooperative game theoretic approach. *Computer Networks*, 51(11):3305 – 3321, 2007.

[10] E. Stai, S. Papavassiliou, and J. S. Baras. A coalitional game based approach for multi-metric optimal routing in wireless networks. In *IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1930–1934, Sept 2013.

[11] R. Kannan and S.S. Iyengar. Game-theoretic models for reliable path-length and energy-constrained routing with data aggregation in wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 22(6):1141–1150, Aug 2004.

[12] W. Saad, Quanyan Zhu, T. Basar, Z. Han, and A. Hjørungnes. Hierarchical Network Formation Games in the Uplink of Multi-hop Wireless Networks. In *Proceedings of the 28th IEEE Conference on Global Telecommunications*, GLOBECOM, pages 2390–2395, 2009.

[13] W. Saad, Xiangyun Zhou, B. Maham, T. Basar, and H.V. Poor. Tree Formation with Physical Layer Security Considerations in Wireless Multi-Hop Networks. *IEEE Transactions on Wireless Communications*, 11(11):3980–3991, Nov 2012.

[14] A. Nahir, A. Orda, and A. Freund. Topology Design of Communication Networks: A Game-Theoretic Perspective. *IEEE/ACM Transactions on Networking*, 22(2):405–414, apr 2014.

[15] J. Derks, J. Kuipers, M. Tennekes, and F. Thuijsman. Local dynamics in network formation. *Proc. Third World Congress of The Game Theory Society*, 2008.

[16] R. Johari, S. Mannor, and J. N. Tsitsiklis. A Contract-Based Model for Directed Network Formation. *Games and Economic Behavior*, 56, 2003.

[17] R.S. Komali, A.B. MacKenzie, and Robert P. Gilles. Effect of Selfish Node Behavior on Efficient Topology Design. *IEEE Transactions on Mobile Computing*, 7(9):1057–1070, Sept 2008.

[18] A. Nahir and A. Orda. The energy-delay tradeoff in wireless networks - System-wide optimization and game-theoretic perspectives. Technical report, Department of Electrical Engineering, Technion, Israel, 2007.

[19] A. Nahir, A. Orda, and A. Freund. Topology Design and Control: A Game-Theoretic Perspective. In *IEEE INFOCOM*, pages 1620–1628, April 2009.

[20] R. Banner and A. Orda. Bottleneck Routing Games in Communication Networks. *IEEE Journal on Selected Areas in Communications*, 26(6):1173–1179, Aug 2007.