Estimation of Primary Traffic Statistics Based on Spectrum Sensing

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy by

Ahmed Abdulkareem Jaafar Al-Tahmeesschi

October 2018

Abstract

Cognitive Radio (CR) systems can benefit from the knowledge of the activity statistics of primary channels, which can use this information to intelligently adapt their spectrum use to the operating environment and work more efficiently and reduce interference on primary users. Particularly relevant statistics are the minimum, mean and variance of the on/off period durations, the channel duty cycle and the governing distribution. The main aim of this thesis is to improve the estimation of the primary user statistics under different environments. At the beginning of operation, the CR does not have any information about the primary traffic statistics. Spectrum sensing is one of the key methods to obtain this knowledge. Unfortunately, the estimation of primary traffic statistics based on spectrum sensing suffers from some flaws, which are investigated in detail in this thesis.

In general, two main working environments for the CRs can be identified based on the primary signal power, namely low and high signal-to-noise ratio (SNR) at the secondary users. For the high SNR scenario, an analytical model to link the sensing period with the observed spectrum occupancy and quantify its impact is proposed. Simulation results show that the proposed model captures with reasonable accuracy the spectrum occupancy observed at the CR. Moreover, the effect of the sample size (number of on/off periods) on the estimated accuracy is studied as well. Closed form expressions to estimate the statistics of the primary channel to a certain desired level of accuracy are derived to link such sample size with the accuracy of the observed primary activity statistics. The accuracy of the obtained analytical results is validated and corroborated with both simulation and experimental results, showing a perfect agreement.

For the low SNR scenario, both local and cooperative estimation are considered based on the number of SUs performing the estimation. For the single estimation scenario, three novel algorithms are proposed to enhance the estimation of primary user activity statistics under imperfect spectrum sensing given the knowledge of minimum transmission time. Simulation results show that the proposed methods enable an accurate estimation for the primary user statistics. For the cooperative estimation scenario, a new reporting mechanism is proposed in order to increase the spectrum and energy efficiency of the cooperative network and improve resilience under Byzantine attacks. The proposed method is compared in terms of efficiency with methods proposed in the literature and the default periodic reporting method. Simulation results show that the proposed scheme not only reduces significantly the signalling overhead, but with a minor modification it can estimate the primary user distribution under Byzantine attacks with high accuracy.

In summary, this thesis contributes a holistic set of mathematical models and novel methods for an accurate estimation of the primary traffic statistics in CR networks based solely on spectrum sensing.

Acknowledgements

Another part of my study life comes to an end. Looking back makes me realise how lucky I am for meeting many people which in one way or another, helped me to reach this point. My deepest gratitude to all those who were with me and gave support during the last four years.

First of all, I want to express my deep thanks and gratitude to my supervisor, Dr. Miguel López-Benítez for the kind support, guidance, and encouragement over the years. His valuable comments have definitely improved the quality of this thesis. I feel deeply indebted for everything. It has been a great pleasure to make this long journey under his supervision.

I would like to thank Dr. Kenta Umebayashi, Dr. Janne. Lehtomäki, Dr. Dhaval Patel, Dr.Valerio Selis and Dr. Xu Zhu for their contributions and collaborations in editing and enriching the quality of the published work and the thesis. Also, I want to thank the members of the Advanced Network Group for providing such a lovely environment for research.

Finally, special thanks to my mother, brother, the soul of my father, family, and friends, for their unconditional support and understanding during my long journey. This journey would have never been possible without them.

Contents

List of Figures

List of Tables

List of Abbreviations

Symbol	Description
CDF	Cumulative Distribution Function
CR	Cognitive Radio
CTMC	Continuous-Time Markov Chain
CTSMC	Continuous-Time Semi-Markov Chain
DC	Duty Cycle
DEM	Direct Estimation Method
Diff	Differential
DSA	Dynamic Spectrum Access
DTMC	Discrete-Time Markov Chain
E	Exponential Distribution
ECMA	European Computer Manufacturers Association
ED	Energy Detection
FC	Fusion Centre
FCC	Federal Communications Commission
G	Gamma Distribution
GE	Generalised Exponential Distribution
GP	Generalised Pareto Distribution
GSM	Global System for Mobile Communications
IEEE	Institute of Electrical and Electronics Engineers
ISS	Imperfect Spectrum Sensing
KS	Kolmogorov-Smirnov
LTE	Long-Term Evolution
LTM	Longer Than the Minimum
MD	Missed Detection
MMoM	Modified Method of Moments
MoM	Method of Moments
MU	Malicious User
P	Pareto Distribution
PDF	Probability Distribution Function
PECAS	Prototype for the Estimation of Channel Activity Statistics
PMF	Probability Mass Function
PSS	Perfect Spectrum Sensing

PU	Primary User
RF	Radio Frequency
SDR	Software Defined Radio
SINR	Signal to Noise plus Interference Ratio
SNR	Signal to Noise Ratio
SSDF	Spectrum Sensing Data Falsification
STM	Shorter Than the Minimum
SU	Secondary User
W	Weibull Distribution
WLAN	Wireless Local Area Network
WRAN	Wireless Regional Area Network
WS	White Space

Chapter 1

Introduction

Over the years, wireless technologies have evolved significantly to cope with increasingly challenging demands and trends ever since the first radio was invented by Nicola Tesla [1], blessed with the wide range of applications for wireless communications from wireless controlling devices to satellite systems and smart-phones. Wireless communications have had a huge impact on human lifestyle and having an internet access at any-time and any place has become essential. This popularity is followed by an exponential growth in the number of wireless connected devices. For instance, the expected number of connected devices per person by 2020 is 6.58 [2]. These devices could include smart-phones, TVs, computers and a wider range.

The growth in the number of users, applications, and the required bandwidth of modern wireless communication systems has resulted in the Radio Frequency (RF) spectrum becoming increasingly crowded and plagued with interference. Hence, governmental agencies took control of RF management among transmitters [3].

Several field measurements of spectrum usage have demonstrated that the allocated spectrum is underutilized [4–8], with variations depending on access time (day/night) and geographical region (urban/rural), which means that the spectrum scarcity problem is a direct result from a fixed spectrum allocation. Therefore, new spectrum management paradigms are essential to efficiently access the radio resources. This situation motivated the introduction of more flexible spectrum allocation policies to overcome the shortcomings of the static allocation. As a result, the Dynamic Spectrum Access (DSA) paradigm based on the Cognitive Radio (CR) technology [9] has gained popularity to overcome the drawbacks and shortcomings of the currently inefficient static allocation schemes.

1.1 Dynamic Spectrum Access

The term Cognitive Radio was first proposed by Joseph Mitola [9]. In essence, a CR is a smart wireless device that is capable of tuning its communication parameters to adapt to the surrounding radio-environment. The DSA principle [10–12], relying on the CR paradigm, has been proposed to improve spectral efficiency [13]. In a CR network,

Secondary Users (SU) access the Primary User (PU) spectrum that is not being utilised (also known as *spectrum holes*) in an opportunistic manner to support spectrum reuse and increase spectrum efficiency [11, 14]. DSA/CR is a promising solution for the spectrum scarcity problem given the huge increase in the demand of wireless connected devices and sensors. It is expected from a CR to sense the primary channel in a periodic manner, to detect the presence of a PU. When a PU signal is detected, the SU has to stop operating at that frequency and has to search for a new empty frequency band. It is essential for a DSA/CR system to operate with minimal, non-harmful interference to the PU network even though PUs may use different modulations, transmission rates and powers, which adds more complexity to the operation of DSA/CR networks. DSA/CR networks are required to enable SUs to:

1. Determine the PU signal availability over the licensed spectrum.
2. Select the most appropriate channel for SU transmission.
3. Coordinate the access to available (idle) primary channels between SUs.

The access of CRs to TV white space is one of the examples of DSA systems [15], where the aim is to have CRs share the geographically unused TV bands in a non-interfering manner. Several standards for CR systems have been designed and deployed in TV white space including:

- The IEEE 802.22 Wireless regional area network (WRAN), which is the first cognitive radio standard [16, 17]. The IEEE 802.22 targets the use of UHF/VHF TV bands (54-862MHz) by SUs on a non-interfering basis, thus spectrum sensing is essential. In [18], different sensing methodologies for compatibility with 802.22 standard were investigated.
- While the IEEE 802.11 standard is well known for Wireless Local Area Network (WLAN) systems, the IEEE 802.11af standard aims to enable WLAN to operate in TV WS in an opportunistic manner [19].
- The IEEE 802.15.4m standard aims to enable ZigBee systems (low power and complexity devices) to operate in television WS [20]. Some example applications include smart grid/utility, advanced sensor networks, and machine-to-machine networks.
- The ECMA 392 standard aims to provide access to personal devices on television WS [21, 22]. Information on the TV channel occupancy can be obtained from databases accumulated by spectrum sensing to ensure interference-free coexistence with TV signal. The ECMA 392 standard fixes the available bandwidth to 6 MHz, 7 MHz, or 8 MHz, unlike IEEE 802.11af which provides flexibility on the SU available bandwidth.

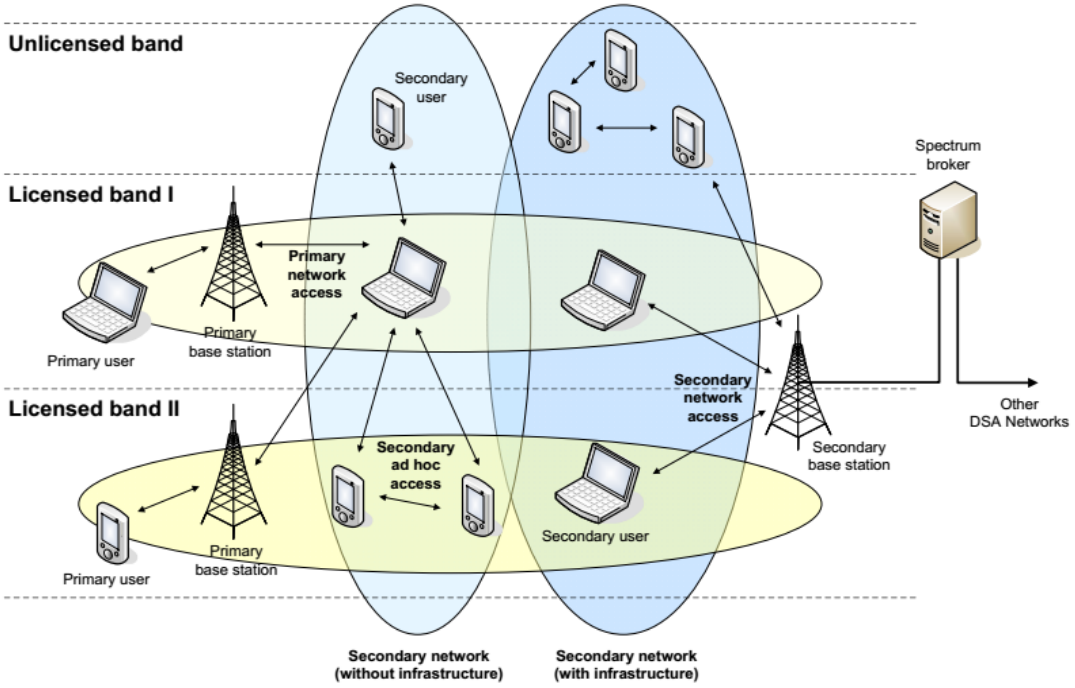


FIGURE 1.1: DSA/CR network architecture [12].

Other examples include CRs access underutilised cellular network channels with higher channel capacity and lower intercell interference [23–26]. Benefiting from the ability to learn information on the primary network and adapt to it (by using previous channel knowledge to select the most appropriate channel with both higher SINR and WS) to create a cognition-inspired 5G cellular network [25]. Another example is cellular networks accessing the WS (unlicensed WiFi spectrum) dynamically in an opportunistic manner without causing interference or unfairness to WiFi networks as preserving fair coexistence is a main goal for DSA systems [27, 28].

1.2 Cognitive Network Architecture

The generic architecture of a DSA/CR network is shown in Fig. 1.1. Two main network components can be identified: the primary network and secondary network [12].

- *Primary network.* A primary network or a licensed network is an existing network infrastructure which has the right to access a specific spectrum band. For instance, cellular networks and TV broadcast are common examples of primary networks. A primary network consists of:
  - *Primary user.* Primary or licensed is a user with a license to operate in a licensed band. The access to the channel is organised by the primary base station with a minimum interference by secondary users. For example, the

primary user is the TV receiver in the licensed TV band or the mobile phone in cellular networks.

- *Primary base-station*. Primary or licensed base station is the access point in the fixed infrastructure of the primary network. Primary base-station may have the capability to support both legacy and new protocols for network access by secondary users.

- *Secondary network*. A secondary or unlicensed network is a network with fixed infrastructure but without a license to operate in a spectrum band, thus accesses the spectrum in an opportunistic manner. A secondary network consists of:

  - *Secondary user*. A secondary or unlicensed user is a network user without a license over the spectrum band. It can only access the spectrum in an opportunistic manner during spectrum holes (PU idle time or vacant areas).
  - *Secondary base-station*. A secondary or unlicensed base station is part of the fixed infrastructure component which provides SUs with spectrum access capabilities without a spectrum license. As for cooperative spectrum sensing, the secondary base-station also serves as a fusion centre (FC) which gathers information from cooperative SUs to provide a global decision on spectrum availability.
  - *Spectrum broker*. A spectrum broker or a scheduling server is a central network node that is in-charge of allocating spectrum resources to SUs (not necessarily from the same network).

1.3 Dynamic Spectrum Access Approaches

Hierarchical spectrum access allows the SU to access the primary spectrum under strict interference restrictions. There are three sub-categories of this scheme as illustrated in Fig. 1.2.

1. *Interweave*[1]. The SU is required to identify spectrum holes (when the PU channel is idle) and utilise this time and geographical location for transmission. As a result, the interference on the PU network would not exist since PU and SU transmissions are orthogonal. Such sharing model is not required to follow any constraints on transmission power. However, SUs are required to periodically monitor the PU channel to determine the activity status (i.e., busy or idle) and vacate the channel when PU is active [29].

2. *Underlay*. The SU accesses the PU channel at any time but with low transmission power to reduce harmful interference on PU [30–32]. Such sharing model can be utilised for short range communication devices. In this model, the tolerable

[1]Sometimes refereed to as overlay in the literature.

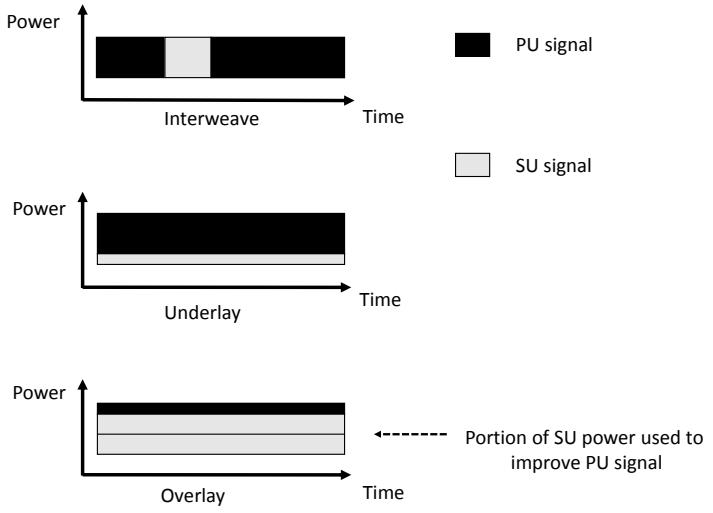


FIGURE 1.2: Dynamic spectrum access techniques.

interference level at PU is defined by the interference temperature which is specified by the Federal Communications Commission (FCC) [33].

3. *Overlay.* The SU also accesses the PU channel at any time but some of the SU power is utilised to assist the transmission of PU and compensate for the interference generated by the remaining power allocated to the SU transmission [34].

This work focuses on the interweave approach.

1.4 Spectrum Sensing

Spectrum sensing is a key enabling technology for CR operation in interweave mode, as it allows SUs to detect the presence/absence of PU traffic, which is essential to reduce the interference [14, 35–39]. Spectrum sensing can also be utilised to capture PU traffic activity for the analysis stage of spectrum characterization. Spectrum sensing can be sub-characterised to two different techniques: non-cooperative and cooperative based sensing.

1.4.1 Non-Cooperative Sensing

The main hypothesis here is the presence of the PU signal. Usually, the received signal at the SU is composed of PU's signal plus noise when the PU is active. Otherwise, the SU receives only noise when the PU is idle. There are three main methods for PU detection.

- *Energy detection.* Energy detection (ED) provides an optimal solution for detection when no information on the signal is available [40]. However, the performance of ED is highly dependent on the received signal energy and noise level [41, 42]. The received signal energy is compared with a predefined threshold and a decision is made stating that PU is active if the signal energy is higher than the threshold. Otherwise, the PU is assumed to be idle. The ED proved to be the most practical one, as most of the time spectrum sensing is performed using low cost devices in practice and this method works regardless of the PU signal format.

- *Matched filter detection.* Matched filter approach provides an optimal solution for additive noise type as it maximises the SNR, however it requires a prior knowledge of the PU signal, for example the modulation type, the pulse shape, and the packet format [12]. Matched filter techniques proved to deliver a good detection performance under low SNR environment [43, 44].

- *Feature detection.* Feature detection targets exploiting the partial knowledge of PU signal. For instance cyclostationary feature detection utilises the periodicity in modulated signals (periodic for modulated signals and aperiodic for noise) [45, 46]. The autocorrelation based detection assumes a wide-sense stationary PU signal and exploits the autocorrelation feature to identify PU signal [47].

1.4.2 Cooperative Sensing

Single based sensing faces reliability problems during the detection of weak signals (due to fading/shadowing) at levels well below the noise floor [48, 49]. Cooperative sensing exploits the spatial diversity of SUs to enhance spectrum sensing [50–52]. In general, cooperative sensing can be classified into distributed cooperative sensing and centralized cooperative sensing. In the distributed cooperative sensing, the SUs exchange their local sensed PU states [53]. While in the centralized cooperative sensing, the SUs report their local sensed PU states to a central server known as the fusion centre (FC) for a global decision based on a specific fusion rule [54].

1.5 Spectrum Availability Modelling

An SU access the licensed user spectrum during idle time in an opportunistic and non-interfering manner. Hence, the performance of an SU is highly dependent on PU activity. Several approaches have been proposed in the literature to model PU activity pattern [11, 55]. In this section, time-dimension models based on Markov chain approach are discussed as it is the most popular modelling method for PU activity pattern.

1.5.1 Two State Markov Chain

At any given time instant, the primary channel may be either busy or idle (binary state). The binary state space for PU channel may be denoted by $S = \{s_0, s_1\}$, where the s_0

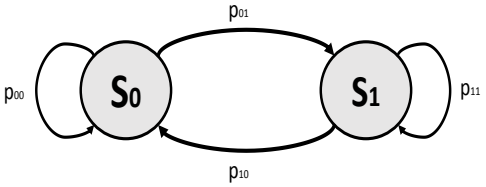


FIGURE 1.3: Discrete-Time Markov chain model.

indicates an idle channel and the s_1 indicates a busy channel. At any given time t, the channel state can be either $S(t) = \{s_0\}$ or $S(t) = \{s_1\}$. The two state Markov chain process can be modelled as either discrete-time or continuous-time based on the time index t. The discussion on both these models are provided below.

1.5.1.1 Discrete-Time Markov Chain

In a Discrete-Time Markov chain (DTMC), the PU channel state changes at discrete time intervals (i.e., $t = t_x = xT_c$, where x is a non-negative integer representing the step number and T_c the time duration between two consecutive state changes [56]. The probability of state transition is defined as p_{ij}. Hence, if the current state is i, the probability of state change j is defined as p_{ij} [57]. The transition matrix of probabilities can be expressed as:

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1-\Psi & \Psi \\ 1-\Psi & \Psi \end{bmatrix}, \tag{1.1}$$

where Ψ is the duty cycle of the PU. p_{ij} represents the probability of system transition between idle S_0 and busy S_1 states.

1.5.1.2 Continuous-Time Markov Chain

In the Continuous-Time Markov Chain (CTMC), the time index takes any continuous value. The channel remains in one state for a random time before shifting the state. The next transition is independent of the state history. The channel holding times (activity time) are modelled as an exponentially distributed random variable. In the literature, CTMC is one of the most commonly deployed models for DSA/CR [58–62].

Field measurements for real wireless communications have shown that the exponential distribution is not the most accurate model for PU holding times. Instead, the Continuous-Time Semi-Markov Chain (CTSMC) model is used to model the occupancy activity [63–65]. In CTSMC, the PU holding times can follow any arbitrary distribution [11].

Other differences between the CTMC/CTSMC and DTMC models is in CTMC/CTSMC it is possible to control the mean, variance and minimum transmission duration of the produced occupancy periods, while this is not possible for DTMC as only the

duty cycle is controllable. In this thesis, both the CTMC and CTSMC are considered to model PU activity time.

1.6 Estimation of Primary Statistics

The estimation of PU traffic statistics (duration of the idle/busy periods of a channel) is inferred from the spectrum sensing decisions. SUs sense the channel with a finite sensing period T_s (shorter than the minimum period duration). At every sensing event, a binary decision is made, H_0 for idle and H_1 for busy. When a change in the observed state occurs, the time interval elapsed since the last is computed to make the estimation of the original real period. The process is repeated for a sufficiently large number of periods until enough data is available. The set of observed periods is then used by the CR to estimate the activity statistics of the PU channel. Some examples of relevant statistics commonly used in the DSA/CR literature are the minimum period duration, the moments (mainly, the mean and the variance), the channel duty cycle and the underlying distribution. The estimation of these statistics will be investigated in this thesis.

Unfortunately, spectrum sensing suffers from practical impairments. The impairments include: i) the estimation of primary statistics based on a finite sensing period, which determines the resolution of estimation for each individual period and imposes a fundamental limit on the accuracy of estimation of individual periods and the corresponding statistics; ii) the finite number of observed periods given that the time required to observe the channel to obtain accurate estimation of the statistics is limited; iii) the presence of sensing errors in the estimation process. The individual estimated period durations will be affected significantly by these practical impairments, which could lead to either under/over estimation of statistics. The impact of all these degrading effects will be investigated in this thesis.

1.7 Motivation and Objectives

The knowledge of the primary activity statistics such as the minimum, mean and variance of the on/off period durations, the channel duty cycle, and the underlying distribution can be employed to access the spectrum more effectively and improve the CR system performance [66–68]. This can be achieved by selecting the most appropriate channel for transmission [69], reducing the switching time delay [70, 71], adapting the parameters of CRs' medium access control (MAC) layer [72], selecting an appropriate threshold in case of using energy detection [73], forecasting the primary occupancy pattern to minimise the interference [74–76], or fight against attacks [77] for the case of cooperative spectrum sensing and thus increase the overall spectrum efficiency.

An example on the benefits of having knowledge of the primary distribution is when a packet needs to be transmitted. The packet size is used to estimate the approximated

time that its transmission would require based on known operation parameters (e.g., channel bandwidth, modulation and coding schemes and so on). Once the estimated transmission time required for the packet is known, the DSA/CR system selects the primary channel with the highest probability to provide a continuous idle/off period that is at least as long as the time required for the packet transmission. Such probability can be obtained from the distribution of idle/off times.

This information could be obtained from spectrum sensing or other alternatives such as databases. However, the sensing-based approach has significant advantages including lower cost and complexity, independence of external systems and better suitability for highly dynamic radio environments [78]. In this thesis, the problem of estimating the PU traffic statistics based on spectrum sensing is investigated.

In the literature, several studies have considered the problem of estimating primary activity statistics based on sensing decisions, mostly focusing on the estimation of the channel duty cycle. In [79], the estimation of the channel occupancy rate (duty cycle) based on different approaches was studied analytically in the presence of sensing errors. A mathematical analysis on the estimation of the mean on/off durations, as well as the duty cycle under DSA, was presented in [80]. In [10] several methods for the classification of such distribution were proposed under the assumption of no sensing errors. To overcome the degrading effect of sensing errors on the estimated primary activity statistics, several algorithms were proposed in [81]. Nevertheless, several questions remain without sufficient answers such as closed form expressions for the estimation of the distribution of busy/idle periods based on spectrum sensing, how big the sample size of busy/idle periods should be to have an accurate estimation of the statistics, and new methods to improve the estimation of the statics under sensing errors.

In this context, this thesis aims at filling the existing gaps by providing answers to these questions. The objective in this line is to provide closed-form expressions that characterise the accuracy of the estimated statistics as function of the practical impairments such as the use of finite sensing period and/or sample size (for the PSS case) and develop new methods to accurately estimate the PU traffic statistics in the presence of sensing errors (for the ISS case).

Moreover, the potential benefits arising from cooperative estimation of the PU statistics is investigated as well, as a means to improve the detection performance taking advantage of the spatial diversity for multiple cooperating nodes [82, 83], it is utilised here to provide an accurate estimation of the primary traffic under imperfect spectrum sensing (ISS). Unfortunately, the improvement in performance achieved by cooperation is hindered by the increase of cooperation overhead. Several studies have aimed at improving energy efficiency in cooperative spectrum sensing by reducing the consumed power at each step of the cooperative sensing operation [84]. For instance, reducing the power consumed during the sensing stage [85, 86], or at the reporting stage [87–89] by selecting the most useful SUs for local states reporting to the fusion centre (FC). Another problem that has not attracted enough attention is the estimation of primary

traffic statistics under spectrum sensing data falsification (SSDF) attacks. CR systems are more susceptible to SSDF attacks (also known as Byzantine attacks) and to the presence of greedy users who send false reports to gain more access to primary channels. Multiple studies have considered the effect of attacks on the sensing process [90–92], however their main aim is the estimation of the probability of primary signal detection instead of the PU traffic statistics. In this thesis, another objective is to investigate the effect of these attacks on the estimation of primary traffic statistics and develop a new reporting mechanism with the aim of reducing the number of required transmissions at each reporting stage while guaranteeing a secure sensing reporting and therefore an accurate estimation of the PU traffic statistics.

1.8 Thesis Contributions

This thesis aims to address the fundamental problem of accurately estimating the primary traffic statistics under different scenarios. The main contributions can be summarised as follows:

1. An analytical model is developed to link the spectrum sensing period with the observed PU traffic statistics and quantify the effects of the spectrum sensing period on the resulting estimation accuracy.

2. A novel approach based on a modified version of the Method of Moments (MoM) is proposed to remove the impact of using a finite sensing period on the estimated PU distribution and improve its accuracy.

3. An analytical model is developed to link the sample size with the observed PU traffic statistics and quantify the effects of the observed sample size on the resulting estimation accuracy.

4. Three novel algorithms are proposed to enhance the estimation of primary activity statics under imperfect spectrum sensing given the knowledge of the minimum transmission duration. Moreover, the impact of different primary distributions on the performance is investigated as well.

5. A detailed study on the cooperative estimation of the PU activity statistics (in particular, the distribution of the channel holding times) under both spectrum sensing errors and SSDF attacks is carried out, which shows improvements in the estimation accuracy. A novel algorithm to reduce both the signalling overhead and estimation errors under SSDF attacks in the cooperative scenario is proposed and compared to other algorithms from the literature showing performance improvements.

1.9 Thesis Outline

The remainder of this thesis is organised as follows.

Chapter 2, investigates the impact of the sensing period on the accuracy of the estimated primary actively statistics. Closed-form expressions for the PDF and CDF of the periods observed at the SU as a function of the original distribution at the PU and the sensing period employed by the SU are derived. Then, closed form expressions for the maximum observed error as a function of the sensing period and distribution parameters are provided as well. New methods to improve the estimation of the distribution are proposed. Finally, the simulation and experimental results of the proposed methods are analysed thoroughly.

The problem of sample size effect on the distribution estimation accuracy is discussed in Chapter 3. First, a detailed mathematical analysis on the sample size required to provide an arbitrarily accurate estimation of the minimum period, the mean and variance of the observed periods, the channel duty cycle and the underlying distribution of the observed periods is provided. The obtained analytical results depend on the real/actual parameters of the PU traffic, which are unknown to the SU; this problem is overcome with an iterative stopping algorithm that enables an accurate estimation of the required sample size in practical implementations. Finally, the simulation and experimental methodology employed to validate the correctness and accuracy of the analytical results is described along with the obtained results.

In Chapter 4, the formal description of the problem of estimating the PU activity statistics under imperfect spectrum sensing is first provided. The novel proposed algorithms to mitigate the impact of sensing errors are explained in detail along with other previous methods proposed in the literature. The performance of the proposed algorithms (obtained by simulation) are analysed and compared with other algorithms along with the discussion on the configuration of spectrum sensing based on the obtained results. Then, different statistical distributions to model the PU activity are introduced and the performance of the proposed algorithms is assessed through simulations under the different considered PU traffic models.

In Chapter 5, the cooperative estimation of primary traffic statistics is considered. First, the structure of the cooperative system is considered along with the cooperative estimation of primary signal durations and cooperative algorithms utilised. Different estimation methods for the cooperative estimation of the primary governing distribution are described. The problem of increasing overhead along with an efficient reporting mechanism are described as well and the problem of SSDF attacks and how to protect against them are also discussed. The performance of the proposed methods is analysed thoroughly by means of simulation and hardware experiments.

Finally, Chapter 6 highlights the main findings and conclusions of this thesis and suggests some ideas for extension in future work.

1.10 List of Publications

Relevant Journal Publications

1. **A. Al-Tahmeesschi**, M. López-Benítez, D. Patel, J. Lehtomäki and K. Umebayashi, "On the Sample Size for the Estimation of Primary Activity Statistics Based on Spectrum Sensing," IEEE Transactions on Cognitive Communications and Networking. (Accepted)

2. **A. Al-Tahmeesschi**, M. López-Benítez, V. Selis, D. Patel, J. Lehtomäki and K. Umebayashi, "Cooperative Estimation of Primary Traffic Under Imperfect Spectrum Sensing and Byzantine Attacks," IEEE Access. (Accepted)

3. M. López-Benítez, **A. Al-Tahmeesschi**, D. Patel, J. Lehtomäki and K. Umebayashi, "Estimation of Primary Channel Activity Statistics in Cognitive Radio Based on Periodic Spectrum Sensing Observations," IEEE Transactions on Wireless Communications. (Under revision)

Relevant Conference Publications

1. **A. Al-Tahmeesschi**, M. López-Benítez, J. Lehtomäki and K. Umebayashi, "Investigating the Estimation of Primary Occupancy Patterns under Imperfect Spectrum Sensing," 2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), San Francisco, CA, 2017, pp. 1-6

2. **A. Al-Tahmeesschi**, M. López-Benítez, K. Umebayashi and J. Lehtomäki, "Analytical study on the estimation of primary activity distribution based on spectrum sensing," 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, 2017, pp. 1-5.

3. **A. Al-Tahmeesschi**, M. López-Benítez, J. Lehtomäki and K. Umebayashi, "Accurate estimation of primary user traffic based on periodic spectrum sensing," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6.

4. **A. Al-Tahmeesschi**, M. López-Benítez, J. Lehtomäki and K. Umebayashi, "Improving primary statistics prediction under imperfect spectrum sensing," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6.

5. M. López-Benítez, **A. Al-Tahmeesschi**, K. Umebayashi and J. Lehtomäki, "PECAS: A Low-Cost Prototype for the Estimation of Channel Activity Statistics in Cognitive Radio," 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, 2017, pp. 1-6.

6. M. López-Benítez, **A. Al-Tahmeesschi** and D. Patel, "Accurate Estimation of the Minimum Primary Channel Activity Time in Cognitive Radio Based on Periodic Spectrum Sensing Observations," European Wireless 2018; 24th European Wireless Conference, Catania, Italy, 2018, pp. 1-6.

Other Publications

1. Z. Wei, X. Zhu, S. Sun, Y. Huang, **A. Al-Tahmeesschi** and Y. Jiang, "Energy-Efficiency of Millimeter-Wave Full-Duplex Relaying Systems: Challenges and Solutions," in IEEE Access, vol. 4, pp. 4848-4860, 2016.

2. Z. Wei, X. Zhu, S. Sun, Y. Jiang, **A. Al-Tahmeesschi** and M. Yue, "Research Issues, Challenges, and Opportunities of Wireless Power Transfer-Aided Full-Duplex Relay Systems," in IEEE Access, vol. 6, pp. 8870-8881, 2018.

Chapter 2

Impact of the Sensing Period under Perfect Spectrum Sensing

2.1 Introduction

DSA/CR users can utilize spectrum sensing decisions to obtain information on PU channel activity. The PU channel is sensed periodically by DSA/CR users to decide the channel state (busy or idle) at every sensing event based on a signal detection algorithm [36]. These spectrum decisions can be used to estimate the durations of the idle and busy periods. Unfortunately, the estimation of PU activity periods and statistics by means of spectrum sensing (periodic channel observations) suffers from practical limitations, such as finite sensing duration and limited observation time. These limitations reduce the accuracy of PU parameters estimated by DSA/CR users as discussed in Section 1.6.

The main interest and focus of this chapter is to first analyse the impact of the spectrum sensing period on the accuracy of the estimated PU activity statistics in particular, the distribution of PU busy/idle periods. A complete characterisation of the PU activity statistics is provided. Despite being an elemental problem of crucial importance for CR systems, this has never been considered or analysed before in the existing literature. Second, to improve the estimation of PU activity statistics (focusing on the distribution of PU busy/idle periods), a novel approach is proposed based on the Method of Moments (MoM) to improve the PU distribution estimation under finite spectrum sensing. The impact of sensing errors (i.e., false alarms and missed detections) is out of the scope of this chapter and hence a high signal-to-noise ratio (SNR) scenario with no sensing errors is considered. The imperfect sensing scenario will be addressed in Chapters 4 and 5.

The contributions of this chapter can be summarised as follow:

1. Analytical expressions are derived for both the Probability Density Function (PDF) and the Cumulative Distribution Function (CDF) observed at the SU taking into account the effect of the spectrum sensing period.

2. Analytical expressions are derived for the maximum error in the distribution taking into account the effect of the spectrum sensing period.

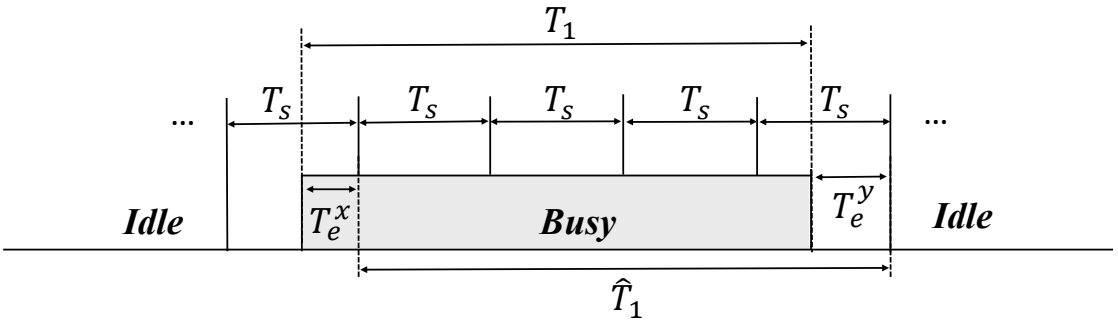


FIGURE 2.1: Considered model. $T_s, T_1, \widehat{T}_1$ represent the sensing period, original busy period duration and estimated busy period duration, respectively. T_e^x and T_e^y are the errors in period estimation.

3. The effect of the spectrum sensing period on the distribution observed at the SU is studied.

4. A novel method to palliate the effects of spectrum sensing on the distribution estimation is proposed. Experimental results are provided to validate the proposed method and simulations.

2.2 System Model

A single SU is considered to detect PU activity. The SU performs spectrum sensing decisions with periodicity T_s time units (t.u.) to detect the presence/absence of PU signal on a specific frequency band. The results of the decisions are introduced as a binary alternating state: busy when the PU signal is present at the SU and idle when the PU signal is absent at the SU. The computed elapsed time (at SU) between two PU state changes is considered as an estimation $\widehat{T}_i$ of the real period duration T_i ($i = 0$ for idle periods and $i = 1$ for busy periods) as illustrated in Fig. 2.1, where the estimation of the duration of a busy period is shown (idle periods can be estimated using the same method). The estimated period durations are integer multiples of T_s (i.e., $\widehat{T}_i = kT_s$, with $k \in N^+$, where k represents the number of sensing events within the estimated period. A similar model was considered in [81].

As discussed in Section 2.1, a high SNR scenario is assumed with no sensing errors so that the only degrading effect considered in this study is the impact of the finite sensing period T_s, which is the aspect of interest in this chapter. The PU activity periods T_i can be sensed accurately in case the channel is sensed exactly at the points of PU state change. In practice the SU is de-synchronised with the PU channel activity and the PU channel is sensed at arbitrary time instants every T_s time units (t.u.). As a result,

the estimated periods $\widehat{T}_i$ depend not only on the original periods T_i but also on the employed sensing period T_s. The first main objective of this chapter is to explore the relation between the original periods T_i and the estimated periods $\widehat{T}_i$ as a function of the sensing period T_s. To this end, closed-form expressions are developed for the PDF/CDF of $\widehat{T}_i$ as a function of the PDF/CDF of T_i and T_s.

2.3 Distribution of the Estimated Periods

2.3.1 Calculation of the Estimated distribution

The estimated periods $\widehat{T}_i$ can be expressed as a function of the original periods T_i as $\widehat{T}_i = T_i + T_e$, where T_e is the error component, which according to the model of Fig. 2.1 is given by $T_e = T_e^y - T_e^x$. As it can be appreciated from Fig. 2.1, both T_e^x and T_e^y can take any value between 0 and T_s. Reasonable and intuitive assumptions that both of them are independent and follow a uniform distribution (i.e., T_e^x and $T_e^y \sim U(0, T_s)$). Both assumptions can be verified from Fig. 2.2, which was obtained by simulating the sensing of a sufficiently high number of exponentially distributed periods T_i using a sensing period $T_s = 5$ t.u., recording the error components T_e^x and T_e^y, and computing their normalized histograms (i.e., PDFs). As it can be observed, the assumptions of uniform distribution and independences for the T_e^x and T_e^y error components are correct. In some specific cases, this assumption might not be completely accurate, depending on the particular value of the involved parameters. The effect, however, was observed to be minimal, with just a small ripple in the shape of the PDF of T_e which is negligible. Moreover, the assumption of independency between T_e^x and T_e^y is necessary to make the problem under study analytically tractable.

The PDF for the triangular distribution of T_e is:

$$f_{T_e}(t) = \begin{cases} 0 & t < -T_s \\ \frac{T_s + t}{T_s^2} & -T_s \leq t \leq 0 \\ \frac{T_s - t}{T_s^2} & 0 \leq t \leq T_s \\ 0 & t > T_s. \end{cases} \tag{2.1}$$

This model can be verified from simulation results as shown in Fig. 2.3.

The PU state holding times (T_0 and T_1) are random variables assumed to be independent and exponentially distributed [71]. The exponential distribution is the most common model used to describe the periods of the on/off states in the literature [80, 93] even though it is proven not to be the most accurate since other distributions provide better fit for real scenarios such as the generalized Pareto, Gamma or even more complicated distributions [64]. The exponential distribution is utilised because of its analytical traceability. The PDF and CDF for the exponential distribution are given as [94]:

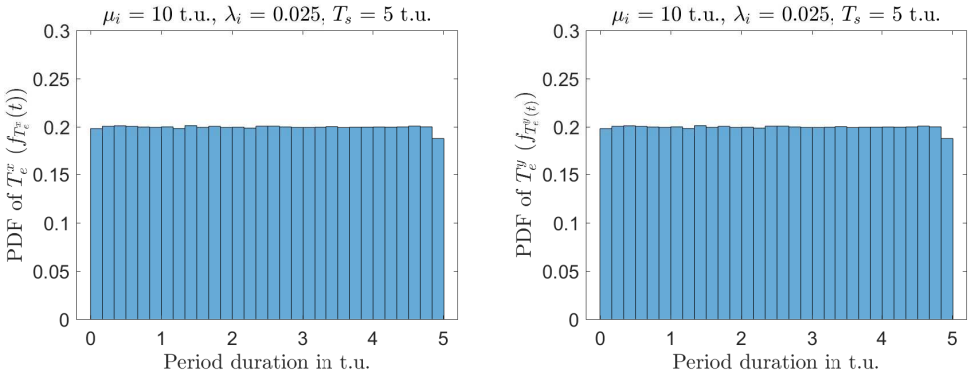


FIGURE 2.2: The PDF of the error components: (a) $f_{T_e^x}(t)$, (b) $f_{T_e^y}(t)$.

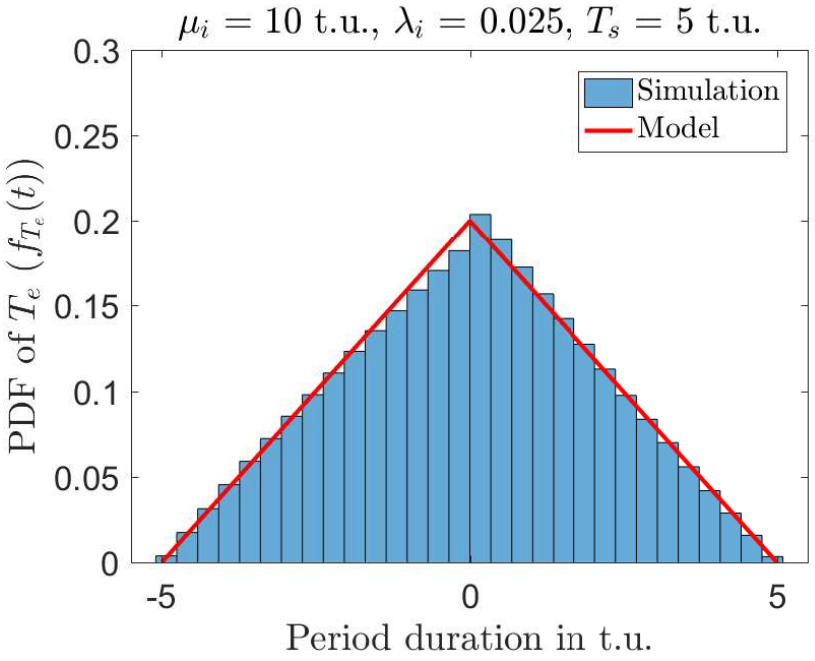


FIGURE 2.3: The PDF of the combined error component $f_{T_e}(t)$.

$$f_{T_i}(t) = \begin{cases} 0 & t < \mu_i \\ \lambda_i e^{-\lambda_i(t-\mu_i)} & t \geq \mu_i, \end{cases} \tag{2.2}$$

$$F_{T_i}(t) = \begin{cases} 0 & t < \mu_i \\ 1 - e^{-\lambda_i(t-\mu_i)} & t \geq \mu_i, \end{cases} \tag{2.3}$$

where λ_i is the distribution scale parameter and μ_i is the distribution location parameter (also the smallest value for the PU activity period).

Since $\widehat{T}_i = T_i + T_e$, the PDF of the estimated periods can be obtained as [95]:

$$f_{\widehat{T}_i}(t) = f_{T_i}(t) * f_{T_e}(t) = \int_{-\infty}^{\infty} f_{T_i}(\tau) \cdot f_{T_e}(t-\tau) d\tau, \tag{2.4}$$

where $f_{T_i}(t)$ and $f_{T_e}(t)$ are given by (2.2) and (2.1) respectively. The operator $*$ refers to the convolution operation. The resulting expression for the PDF $f_{\widehat{T}_i}(t)$ is shown in (2.5) while the CDF $F_{\widehat{T}_i}(t)$ can be obtained through the direct integration of $f_{\widehat{T}_i}(t)$ as shown below:

$$F_{\widehat{T}_i}(t) = \int_{-\infty}^{t} f_{\widehat{T}_i}(\tau) d\tau. \tag{2.6}$$

The final CDF expression can be seen in (2.7).

Note that the distributions in (2.5) and (2.7) have a continuous domain, while the actual distributions of the periods observed at a SU are discrete since the periods estimated from spectrum sensing as shown in Fig. 2.1 are integer multiples of the employed sensing period (i.e., $\widehat{T}_i = kT_s, k = 1, 2, 3 \ldots$). Such discrete distribution can be obtained by evaluating (2.5) and (2.7) at the right points of each interval/bin of the PDF and CDF, respectively, as:

$$g_{\widehat{T}_i}(k) = f_{\widehat{T}_i}(kT_s), \tag{2.8}$$

$$G_{\widehat{T}_i}(k) = F_{\widehat{T}_i}((k + 1/2)T_s). \tag{2.9}$$

The set of obtained expressions provide closed-form relations between the distributions of the original periods T_i resulting from the PU transmission (and its parameters μ_i, λ_i), the distribution of the estimated periods $\widehat{T}_i$ as observed by the SU based on spectrum sensing decisions, and the employed sensing period T_s. These mathematical results are useful to evaluate the impact of the employed sensing period on the accuracy of the distributions estimated by the SU and can find many practical applications such as mathematical analysis, simulation or system design (e.g., determine the maximum value of T_s required for a given level of estimation accuracy).

2.3.2 Error of the Estimated Distribution

To better understand the sensing period effect on the observed distribution, we utilize the well-known Kolmogorov-Smirnov (KS) distance. This is the most commonly used metric to quantify the error between two distributions. The KS distance is defined as the largest absolute error between two continuous CDFs and given as follows [96]:

$$D_{KS} = \sup_t \left| F_{T_i}(t) - F_{\widehat{T}_i}(t) \right|. \tag{2.10}$$

To find the value of t that maximises the distance (D_{KS}), the partial derivative of the absolute difference in the KS distance is taken and equated to zero as follows:

$$\frac{\partial [F_{T_i}(t) - F_{\widehat{T}_i}(t)]}{\partial t} = 0. \tag{2.11}$$

$$f_{\widehat{T}_i}(t) = \begin{cases} 0 & t < \mu_i - T_s \\ \dfrac{T_s + t - \mu_i}{T_s^2} - \dfrac{1}{\lambda_i T_s^2}\left[1 - \dfrac{1}{\lambda_i} f_{T_i}(t + T_s)\right] & \mu_i - T_s \leq t < \mu_i \\ \dfrac{T_s - t + \mu_i}{T_s^2} + \dfrac{1}{\lambda_i T_s^2}\left[1 + \dfrac{1}{\lambda_i} f_{T_i}(t + T_s) - \dfrac{2}{\lambda_i} f_{T_i}(t)\right] & \mu_i \leq t \leq \mu_i + T_s \\ \dfrac{1}{(\lambda_i T_s)^2}\left[f_{T_i}(t + T_s) - 2 f_{T_i}(t) + f_{T_i}(t - T_s)\right] & t > \mu_i + T_s \end{cases} \tag{2.5}$$

$$F_{\widehat{T}_i}(t) = \begin{cases} 0 & t < \mu_i - T_s \\ \dfrac{t^2 - (\mu_i - T_s)^2}{2T_s^2} - \dfrac{\left[1 - \lambda_i(T_s - \mu_i)\right](t + T_s - \mu_i)}{\lambda_i T_s^2} + \dfrac{1}{(\lambda_i T_s)^2} F_{T_i}(t + T_s) & \mu_i - T_s \leq t < \mu_i \\ \dfrac{\mu_i^2 - t^2}{2T_s^2} - \dfrac{2 - \lambda_i T_s}{2\lambda_i T_s} + \dfrac{\left[1 + \lambda_i(T_s + \mu_i)\right](t - \mu_i)}{\lambda_i T_s^2} + \dfrac{1}{(\lambda_i T_s)^2}\left[F_{T_i}(t + T_s) - 2F_{T_i}(t)\right] & \mu_i \leq t \leq \mu_i + T_s \\ 1 + \dfrac{1}{(\lambda_i T_s)^2}\left[F_{T_i}(t + T_s) - 2F_{T_i}(t) + F_{T_i}(t - T_s)\right] & t > \mu_i + T_s \end{cases} \tag{2.7}$$

The largest difference occurs at $t = \mu_i$ which is found through numerical methods. Since $F_{T_i}(\mu_i)$ is zero at $t = \mu_i$, the final expression for the KS distance will be:

$$\begin{aligned} D_{KS} &= F_{\widehat{T}_i}(\mu_i) \\ &= \frac{1}{2} - \frac{1}{\lambda_i T_s} + \frac{1 - e^{-\lambda_i T_s}}{(\lambda_i T_s)^2}. \end{aligned} \tag{2.12}$$

Expression (2.12) provides an easy and accurate tool to mathematically calculate the KS distance between the estimated and original CDFs as a function of the employed sensing period. Moreover, expression (2.12) can be used to calculate the T_s required for a given target estimation error.

2.3.3 Numerical Results

In this section, first the accuracy of the proposed model for both PDF and CDF will be assessed, then the effect of the sensing period on the distribution estimation. For all the considered cases the sensing period is lower than the minimum PU activity time ($T_s < \mu_i$). This is required to ensure that no activity periods are missed in the sensing process (the shortest detectable period is T_s), which would otherwise lead to significant estimation errors. Notice that this consideration implicitly assumes that the minimum PU activity time μ_i is known to the SU so that the value of T_s can be configured not to exceed μ_i. This assumption is realistic since the value of μ_i is available for some well-known standardised radio technologies (e.g., the time-slot duration of GSM or LTE) or can be obtained with other methods such as blind recognition/estimation [97] or from PU beacon signals [98].

The simulation results are obtained by following steps listed below:

1. Generate idle/busy periods' lengths T_i following a generalized exponential distribution.
2. Perform idle/busy sensing decisions H_0/H_1 on the generated sequence in step 1 every T_s time units (t.u.).
3. Calculate the idle/busy lengths $\widehat{T}_i$ using H_0/H_1 sequence from step 2 estimated under PSS.
4. Compute the CDF/PDF of the idle/busy lengths obtained in step 3, and compare with the CDF/PDF of the original lengths in step 1.

On the other hand, analytical results are obtained by applying the system parameters into the derived expressions. For example, 2.12 is used find the error in the CDF (KS distance).

Fig. 2.4 shows the busy periods PDFs $f_{\widehat{T}_i}(t)$ obtained from simulation and analytical expression versus the original distribution $f_{T_i}(t)$ for multiple values of sensing periods (T_s

= 1, 3 and 5 t.u.). The discrete expression $g_{\widehat{T}_i}(k)$ has not been included for clarity but its corresponding values can be easily obtained as the values of the analytical expression $f_{\widehat{T}_i}(t)$ at kT_s. It can be appreciated that the closed form analysis provides an excellent fit with the simulation results for all the considered scenarios, which verifies the validity of the mathematical expression obtained for the PDF. Moreover, Fig. 2.4 shows the effect of sensing period T_s on the discrete estimated PDF $g_{\widehat{T}_i}(k)$. High sensing periods give higher estimation errors and vice versa.

Fig. 2.5 shows the busy periods CDFs obtained from simulation $G_{\widehat{T}_i}(t)$ (discrete) and analytical expression $F_{\widehat{T}_i}(t)$ versus the original distribution $F_{T_i}(t)$ for multiple values of sensing periods (T_s = 1, 3 and 5 t.u.). The closed form analysis provides an excellent fit with the simulation results for all the considered scenarios, which verifies the validity of the mathematical expression obtained for the CDF. The stair shape of the observed CDF $G_{\widehat{T}_i}(t)$ represents the effect of the spectrum sensing operation and the resulting discrete observed periods.

Fig. 2.6 shows the KS distance for the simulated and analytical CDF with respect to the original distribution. The x-axis represents the duration of sensing period in time units and the y-axis represents the KS distance. Since the sensed CDF is a discrete distribution $G_{\widehat{T}_i}(t)$, it is to be transformed to a continuous form for comparison purposes. To this end, the CDF frequency polygons are utilised [99], where the mid points of the discrete CDF are joined together and extended to include the zero frequency cases from left of the normalised histogram and hence obtain the continuous form of the CDF. As it can be appreciated from Fig. 2.6, the analytical expression (2.12) gives an excellent prediction of the estimation error. High T_s values will result in larger errors in the estimation of the PU activity pattern, however the resulting estimation error can be reduced by decreasing T_s.

Fig. 2.7 analyses the impact of different λ_i values ($\lambda_i = 0.15, 0.25, 0.35$ and 0.45) on the KS distance based on (2.12). Fig. 2.7 implies that not only the value of T_s has an impact on the estimated error (KS distance) but also the value of λ_i (distribution scale). The KS distance increases with higher values of λ_i. The analytical result in (2.12) can be used as shown in Fig. 2.7 to determine the maximum value of T_s required for a given level of estimation accuracy of the distribution.

2.4 Methods for Accurate Estimation of the Distribution

This section proposes new methods to overcome the impact of a finite sensing period on the estimated distribution. To this end, it is first necessary to analyse how the sensing period affects the estimation of the minimum period, as well as the mean and variance of the estimated periods.

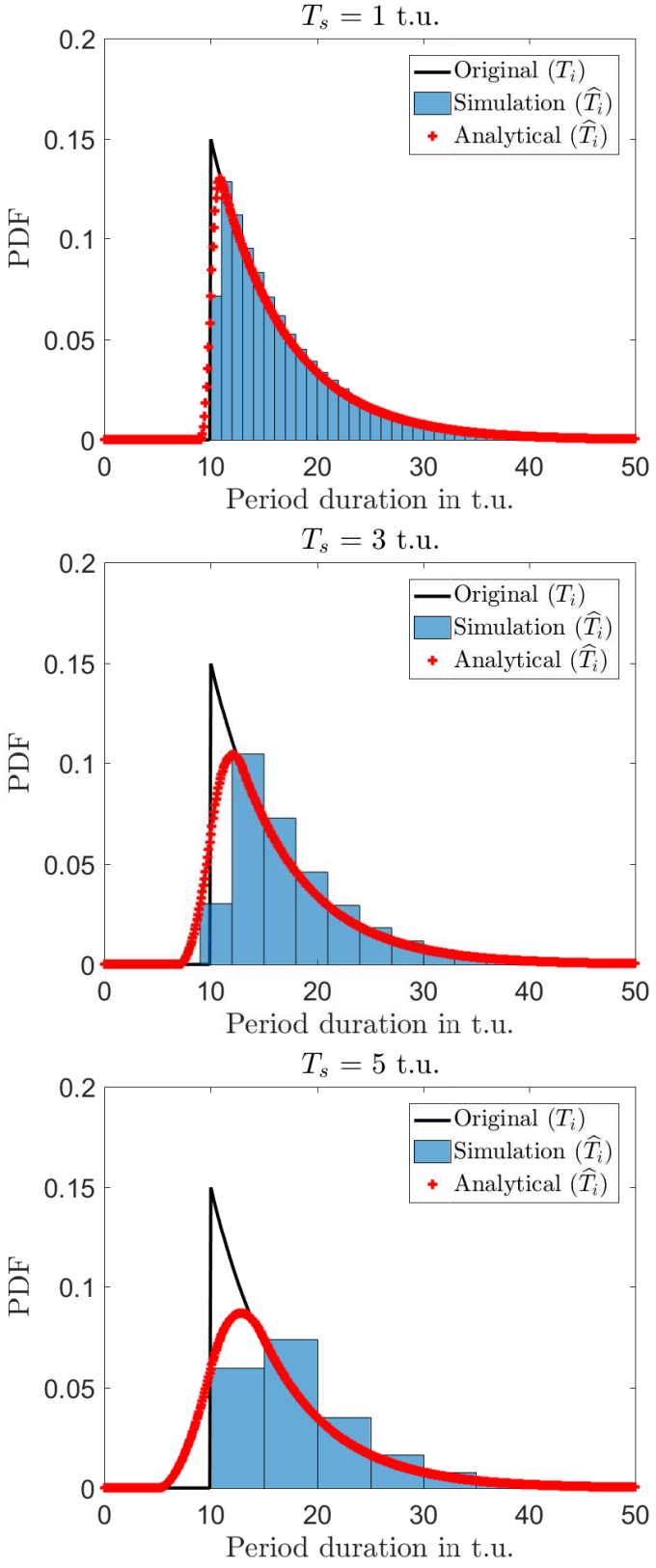


FIGURE 2.4: Validation of the PDF of the estimated periods ($\lambda_1 = 0.15, \mu_1 = 10$ *t.u.*, $\mathbb{E}\{T_1\} = 16.66$ *t.u.* and $\Psi = 0.5$.)

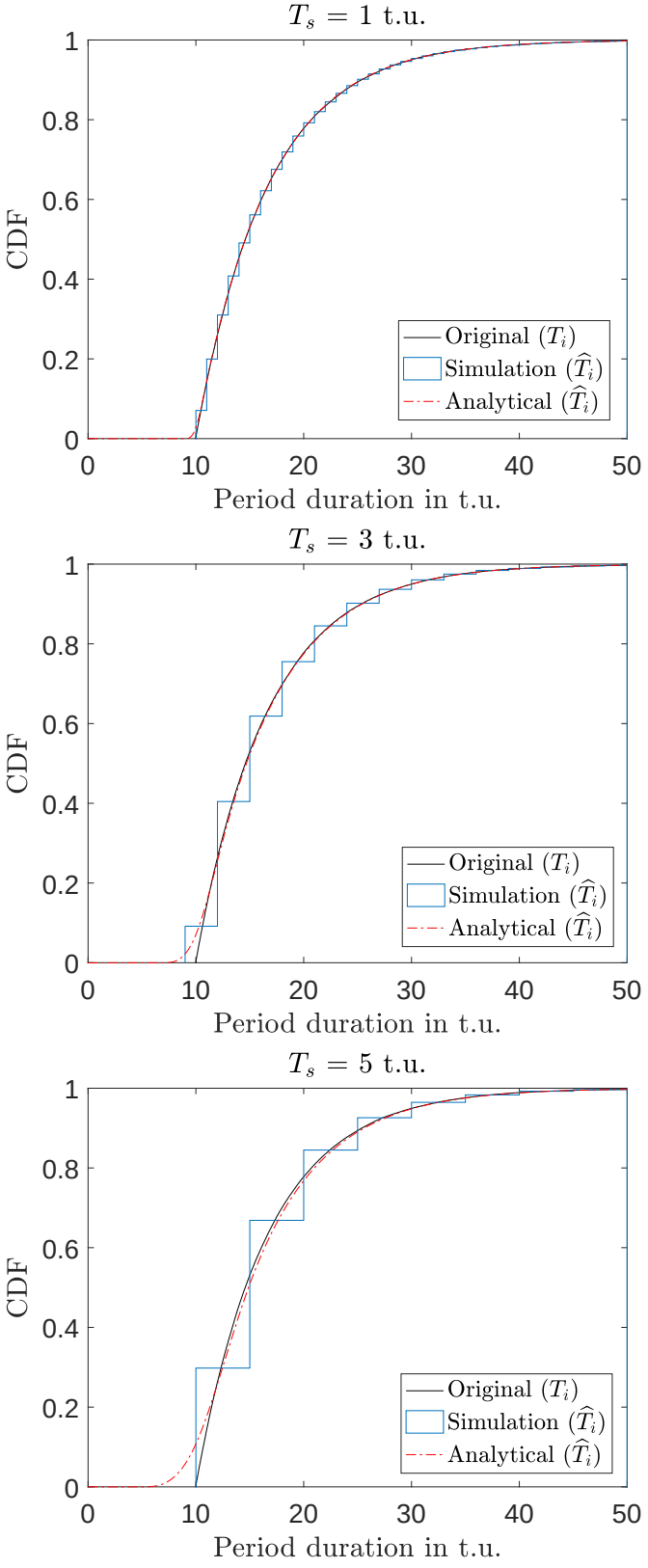


FIGURE 2.5: Validation of the CDF of the estimated periods ($\lambda_1 = 0.15, \mu_1 = 10$ *t.u.*, $\mathbb{E}\{T_1\} = 16.66$ *t.u.* and $\Psi = 0.5$.)

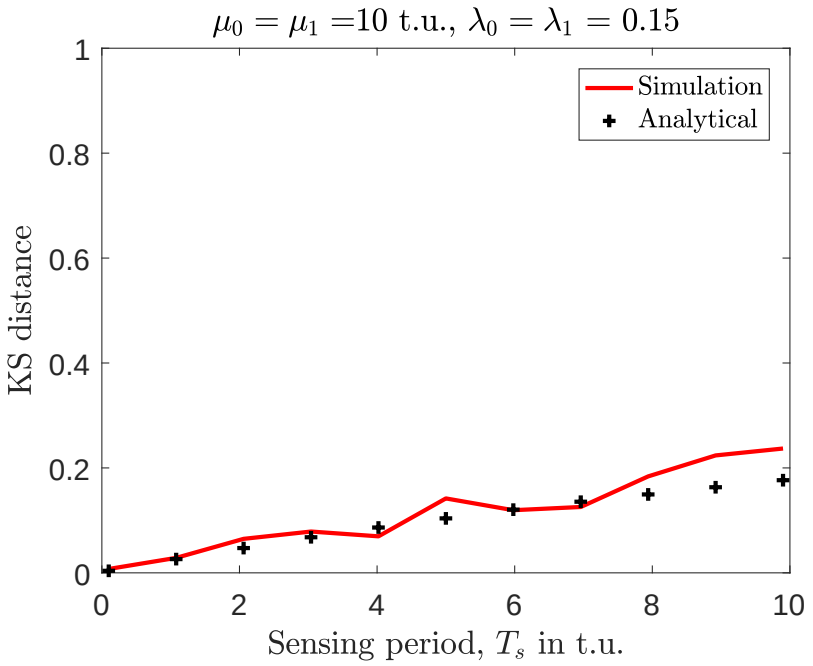


FIGURE 2.6: KS distance for the observed and analytical model CDFs.

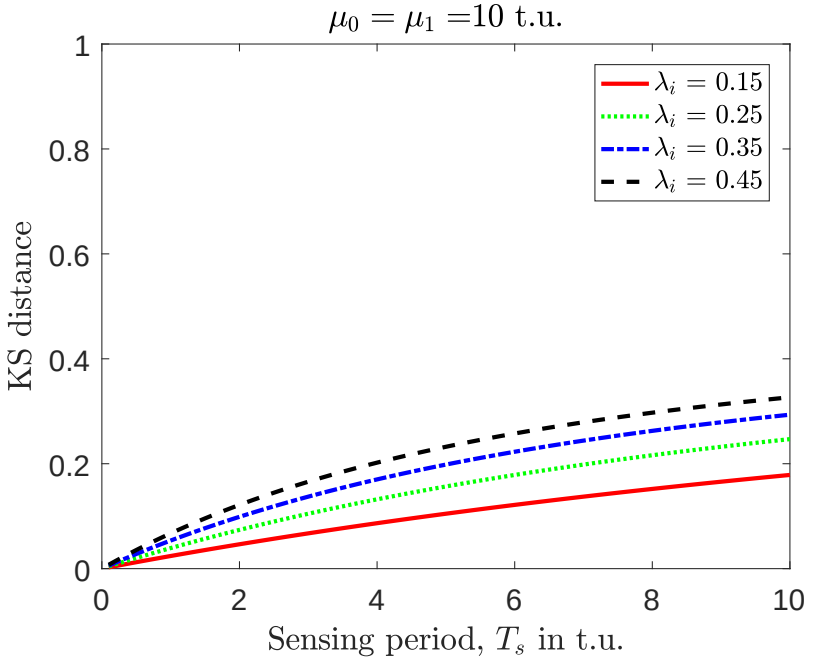


FIGURE 2.7: KS distance for the observed and analytical model CDFs.

2.4.1 Estimation of the Minimum Period Duration

As mentioned earlier, CR systems estimate the PU activity pattern based on the discrete observation periods $\widehat{\tau}_i = \{\widehat{T}_{i,n}\}_{n=1}^{N}$, where N represents the number of observed periods (sample size of $\widehat{\tau}_i$). In order to have accurate estimations of the distribution, the sample size N has to be sufficiently large. The minimum activity duration was simulated in [81], nevertheless a closed form expression for the estimation of minimum PU activity time ($\widehat{\mu}_i$) is provided with a link to the sensing period (T_s). It can be shown that the estimated periods can also be expressed as:

$$\widehat{T}_i = \left(\left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s, \tag{2.13}$$

where $\lfloor \cdot \rfloor$ denotes the floor operator and $\xi \in \{0, 1\}$ is a Bernoulli random variable [57], introduced to reflect the fact that the same original period T_i can lead to two possible estimated periods, either $\widehat{T}_i = kT_s$ or $\widehat{T}_i = (k+1)T_s$, depending on the relative (random) position of the sensing events with respect to the beginning/end of T_i. Hence, $\widehat{\mu}_i$ can be expressed as:

$$\begin{aligned} \widehat{\mu}_i = \min(\widehat{\tau}_i) \approx \min(\widehat{T}_i) &= \min \left[\left(\left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s \right] \\ &= \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s, \end{aligned} \tag{2.14}$$

note that the minimum value in (2.14) corresponds to $\min(T_i) = \mu_i$ and $\min(\xi) = 0$.

2.4.2 Estimation of the Mean and Variance of Period Durations

Given a set of N (discrete) estimated periods $\widehat{\tau}_i = \{\widehat{T}_{i,n}\}_{n=1}^{N}$, the mean $\mathbb{E}(\widehat{T}_i)$ and variance $\mathbb{V}(\widehat{T}_i)$ of the provided durations can be estimated based on the corresponding sample moments:

$$\mathbb{E}(\widehat{T}_i) = \frac{1}{N} \sum_{n=1}^{N} \widehat{T}_{i,n}, \tag{2.15}$$

$$\mathbb{V}(\widehat{T}_i) = \frac{1}{N-1} \sum_{n=1}^{N} (\widehat{T}_{i,n} - \mathbb{E}(\widehat{T}_i))^2. \tag{2.16}$$

The impact of T_s on the estimated moments (first and second) can be determined as follows:

$$\mathbb{E}(\widehat{T}_i) = \mathbb{E}(T_i) + \mathbb{E}(T_e) = \mathbb{E}(T_i), \tag{2.17}$$

$$\mathbb{V}(\widehat{T}_i) = \mathbb{V}(T_i) + \mathbb{V}(T_e) = \mathbb{V}(T_i) + \frac{T_s^2}{6}, \tag{2.18}$$

where T_i and T_e are assumed to be mutually independent, and $\mathbb{E}(\widehat{T}_e)$ and $\mathbb{V}(\widehat{T}_e)$ have been replaced with the mean and variance of the triangular distribution in (2.1). The triangular distribution in this case has a mean value of zero (i.e., $\mathbb{E}(T_e) = 0$), which means that the duration of T_s does not affect the calculation of the mean value. On

the other hand, the calculation of the variance is affected by a factor of $T_s^2/6$. Based on (2.18), the effect of T_s can be minimized by applying to (2.16) the appropriate correction factor:

$$\mathbb{V}(\widetilde{T}_i) = \mathbb{V}(\widehat{T}_i) - \frac{T_s^2}{6}, \tag{2.19}$$

where $\mathbb{V}(\widetilde{T}_i)$ is the observed variance after correction. This approach eliminates the impairments imposed by the sensing operation with duration of T_s in the estimated moments and is able to provide an accurate estimation of the real moments of T_i, based on the estimated period durations $\widehat{\tau}_i$, as long as the sample size N is sufficiently large.

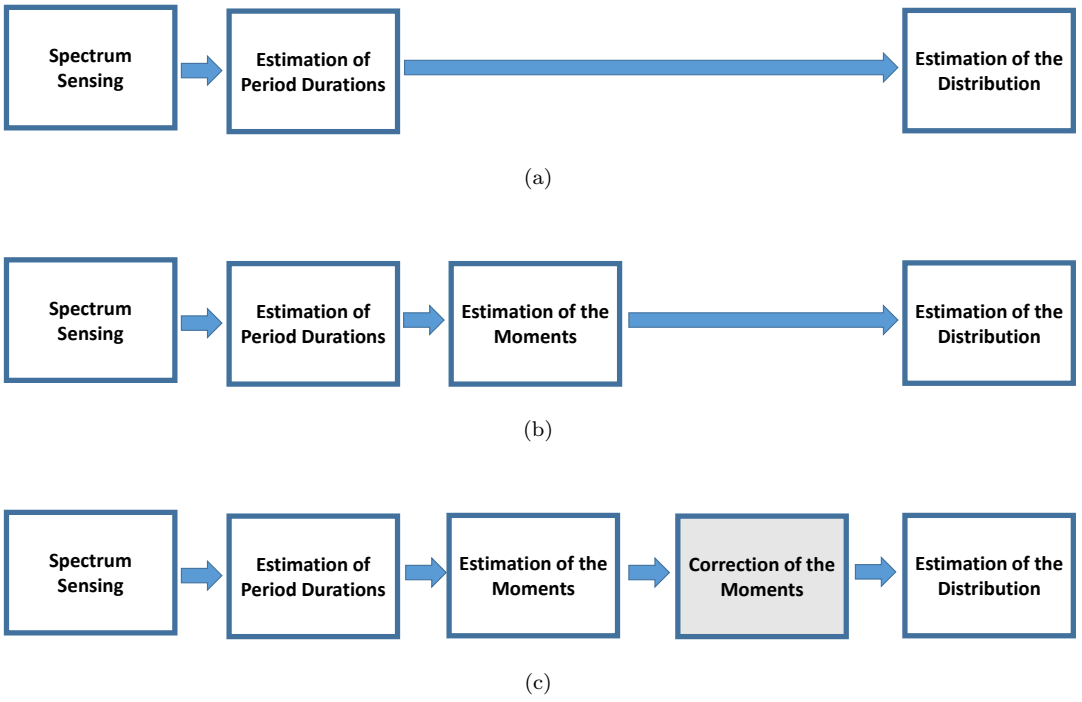


FIGURE 2.8: Distribution estimation methods : (a) Direct estimation, (b) Method of Moments (MoM), (c) Modified Metod of Moments MMoM.

2.4.3 Considered Estimation Methods

Three methods are considered to estimate the distribution.

2.4.3.1 Direct Estimation

The direct estimation method is based on the calculation of the empirical cumulative distribution function (`ecdf` function in MATLAB). The `ecdf` function calculates the Kaplan-Meier estimate of the provided samples [100]. The flowchart of this strategy is illustrated in Fig. 2.8(a). The main drawback of this method is that the estimated distribution is discrete as the values of the estimated periods are integer multiples of the

sensing period (i.e., $\widehat{T_i} = kT_s, k = 1, 2, 3, \ldots$). Moreover, it is not possible to apply any correction factors to this estimation method, which affects its accuracy.

2.4.3.2 Estimation based on Method of Moments (MoM)

To overcome the first drawback of the direct estimation method, a solution based on the Method of Moments (MoM) is utilised. Instead of estimating the distribution of the PU activity periods directly from the observed periods themselves, which produces a discrete distribution, this method computes first the moments (mean and variance of the PU activity periods) and then estimates the parameters of the distribution based on the MoM assuming a certain distribution model. The flowchart of this strategy is illustrated in Fig. 2.8(b). The sample moments are equated to the distribution moments and then by solving the resulting equations the distribution parameters are obtained. As opposed to the previous method, the resulting distribution with this approach is continuous instead of discrete, thus offering the possibility to minimize the impact of sensing period.

Various methods have been proposed to estimate the distribution parameters besides MoM such as, Maximum likelihood Estimation and Least Squares Estimation [101, 102]. In this work, we only consider MoM-based solutions. Even though other methods might provide a better distribution parameters fit, they require the complete history of past observed period durations while with MoM the distribution moments can be estimated from sample moments, which can be computed recursively based on last samples. As a result, the practical implementation of MoM-based solutions would result in significantly lower computation and memory cost for CR devices.

Here, we assume the state holding times of PU (T_0 and T_1) follow a Generalized Pareto (GP) distribution, which was proven to give best accuracy fit with a reasonable complexity in comparison with other more complex distributions [64]. The busy and idle durations are also assumed to be independent of each other [71]. The PDF and CDF for the GP distribution are given, respectively, as [94]:

$$f_{T_i}(t) = \begin{cases} 0 & t < \mu_i \\ \frac{1}{\lambda_i}\left[1 + \frac{\alpha_i(t-\mu_i)}{\lambda_i}\right]^{-(1/\alpha_i+1)} & t \geq \mu_i, \end{cases} \tag{2.20}$$

$$F_{T_i}(t) = \begin{cases} 0 & t < \mu_i \\ 1 - \left[1 + \frac{\alpha_i(t-\mu_i)}{\lambda_i}\right]^{-1/\alpha_i} & t \geq \mu_i, \end{cases} \tag{2.21}$$

where α_i and λ_i are the shape and scale of the GP distribution respectively, and μ_i is the location (also the minimum PU activity duration). Moreover, $T_i \geq \mu_i, \alpha_i \geq 0, \lambda_i \geq 0$. The mean and variance of the GP distribution are expressed as:

$$\mathbb{E}(T_i) = \mu_i + \frac{\lambda_i}{1 - \alpha_i}, \tag{2.22}$$

$$\mathbb{V}(T_i) = \frac{\lambda_i^2}{(1-\alpha_i)^2(1-2\alpha_i)}. \tag{2.23}$$

The expressions needed to estimate the parameters of the GP distribution from the sample moments can be obtained by solving (2.22) and (2.23) for such parameters, which yields:

$$\widehat{\mu}_i = \min(\widehat{T}_i), \tag{2.24}$$

$$\widehat{\alpha}_i = \frac{1}{2}\left(1 - \frac{(\mathbb{E}(\widehat{T}_i) - \widehat{\mu}_i)^2}{\mathbb{V}(\widehat{T}_i)}\right), \tag{2.25}$$

$$\widehat{\lambda}_i = \frac{1}{2}\left(1 + \frac{(\mathbb{E}(\widehat{T}_i) - \widehat{\mu}_i)^2}{\mathbb{V}(\widehat{T}_i)}\right)\left(\mathbb{E}(\widehat{T}_i) - \widehat{\mu}_i\right), \tag{2.26}$$

where $\widehat{\lambda}_i$ and $\widehat{\alpha}_i$ are the estimated values of original λ_i and α_i. Introducing the MoM estimates provided by (2.24), (2.25) and (2.26) into (2.20) and (2.21) provides a continuous estimation of the distribution of PU activity periods.

Notice that the location parameter μ_i can be estimated as shown in (2.24) since it corresponds to the minimum period duration. However, such estimation will be affected by the employed sensing period T_s. In many cases SU may be able to have a perfect knowledge of this parameter, for example in the case of primary systems that use some form of regional beacon signals with real-time information [98] or when the radio technology of the primary system is standardised and known (e.g., the slot duration of GSM).

2.4.3.3 Estimation based on Modified Method of Moments (MMoM)

The MoM solution discussed in the previous section solves the problem of the estimation error introduced by the discrete distribution resulting from the direct estimation method. However, as shown in the analysis of Section 2.4.2, the estimated moments may have an error component resulting from the use of a finite sensing period T_s. This motivates the introduction of a Modified Method of Moments (MMoM) solution. The flowchart of this proposed strategy is illustrated in Fig. 2.8(c). The main difference with respect to the MoM method is the correction of the estimated moments, which is shaded in Fig. 2.8(c)

Based on (2.19), the new distribution parameters can be estimated as follows:

$$\widetilde{\alpha}_i = \frac{1}{2}\left(1 - \frac{(\mathbb{E}(\widehat{T}_i) - \widehat{\mu}_i)^2}{\mathbb{V}(\widetilde{T}_i)}\right), \tag{2.27}$$

$$\widetilde{\lambda}_i = \frac{1}{2}\left(1 + \frac{(\mathbb{E}(\widehat{T}_i) - \widehat{\mu}_i)^2}{\mathbb{V}(\widetilde{T}_i)}\right)\left(\mathbb{E}(\widehat{T}_i) - \widehat{\mu}_i\right). \tag{2.28}$$

Notice that (2.27) and (2.28) are similar to their counterparts in (2.25) and (2.26), respectively, but are based on a corrected version of the moments. In particular, the

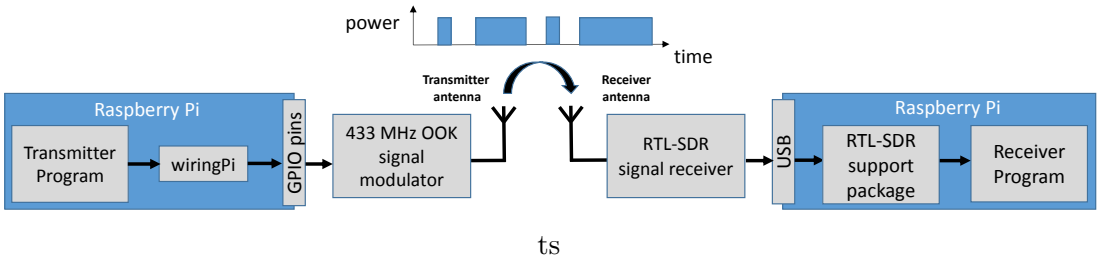


FIGURE 2.9: Block diagram of the PECAS prototype employed for hardware experiments [103].

corrected variance $\mathbb{V}(\widetilde{T}_i)$ is used instead of the sample variance $\mathbb{V}(\widehat{T}_i)$, while the sample mean does not need correction as inferred from (2.17).

2.4.4 Numerical and Experimental results

This study assumes that the sensing period is lower than or equal to the minimum PU activity time ($T_s \leq \mu_i$). This is required to ensure that no PU activity periods are missed in the sensing process (the shortest detectable period is T_s), which would otherwise lead to significant estimation errors. Notice that this consideration implicitly assumes that the minimum activity time of the PU, μ_i, is known to the SU so that the value of T_s can be configured not to exceed μ_i.

The simulation results are obtained by following steps listed below:

1. Generate idle/busy periods' lengths T_i following a generalized Pareto distribution, which has been proven to provide the best fit to empirical spectrum data [64].

2. Perform idle/busy sensing decisions H_0/H_1 on the generated sequence in step 1 every T_s time units (t.u.).

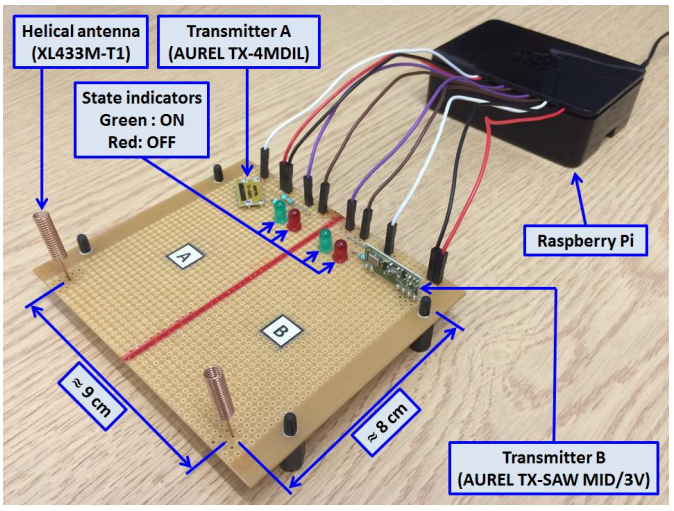
3. Calculate the idle/busy lengths $\widehat{T}_i$ using H_0/H_1 sequence from step 2 estimated under PSS.

4. Process the sequence of period lengths resulting from step 3 to reconstruct the original distribution by using one of the three methods considered in this chapter.

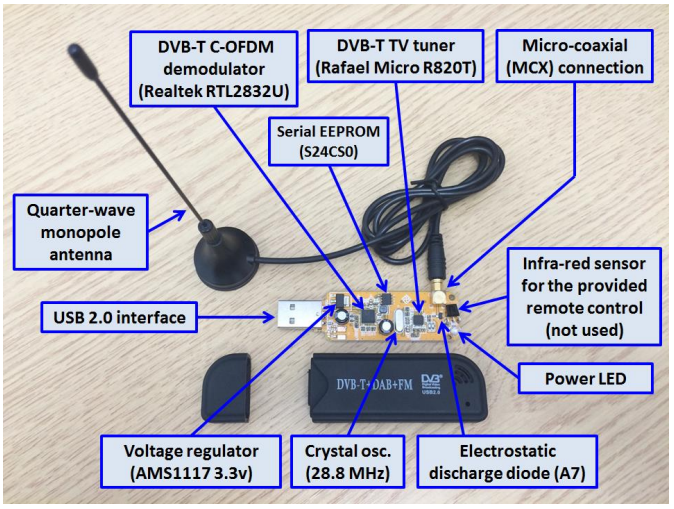
5. Compute the CDF of the idle/busy lengths obtained in steps 3 & 4, and compare with the CDF of the original lengths in step 1.

The analytical results are obtained by applying the system parameters into the derived expressions.

The hardware experiments were conducted with a Prototype for the Estimation of Channel Activity Statistics (PECAS) [103]. This prototype is implemented with common low-cost components with an approximated total cost of £60/$80. The aim to reproduce a realistic scenario with inexpensive DSA/CR devices and introduce typical



(a)



(b)

FIGURE 2.10: PECAS hardware implementation [103]: (a) Transmitter, (b) Receiver.

hardware sources of error and inaccuracies. The prototype is based on free open source code [1].

The hardware experiments are conducted following the same principle as the simulations but by utilising a real transmitter and a real receiver (i.e., PECAS). The block diagram of PECAS is shown in Fig. 2.9. The transmitter (primary user) sends a sequence of GP-distributed idle/busy periods utilising a 433 MHz ON-OFF Keying (OOK) modulator with an output power of 2 dBm (controlled from a C program based on the `wiringPi` library). The receiver (secondary DSA/CR user), placed 1 metre apart, uses a Software-Defined Radio (SDR) with a gain of 20 dB to monitor the transmitter activity (idle/busy) at 433 MHz every T_s seconds. Fig. 2.10 shows the hardware implementation of the transmitter and receiver parts of PECAS.

[1] Available at: `www.lopezbenitez.es/misc/PECAS.zip`

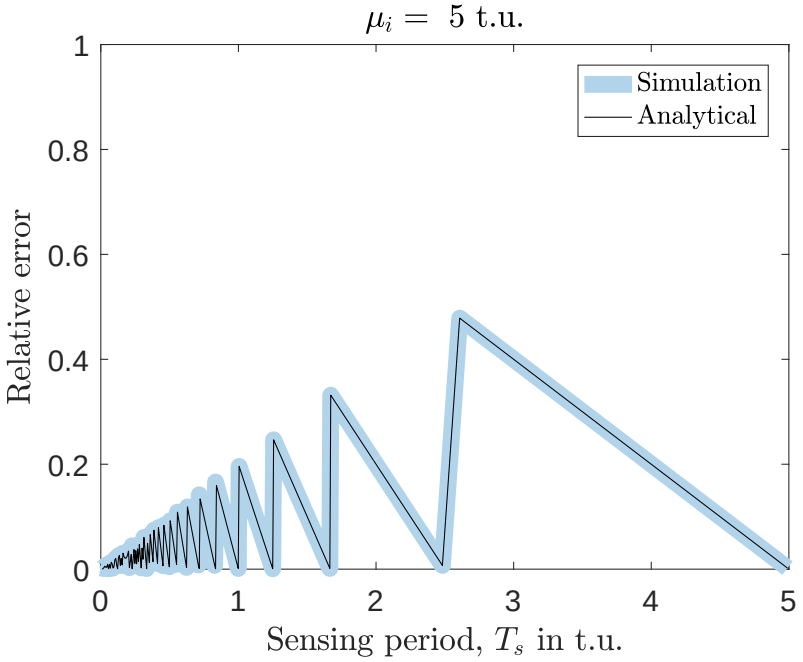


FIGURE 2.11: Relative error of the estimated minimum period $\widehat{\mu}_i$.

At every sensing event, signal samples are captured at a sample rate of 10^6 samples per second, which are processed to decide the instantaneous channel state (idle/busy) using energy detection. The outcomes of the energy detection decisions are used to estimate the durations of the observed idle/busy periods as shown in Fig. 2.1 and compute the primary activity statistics. While transmitter and receiver are controlled by C programs running on the same Raspberry Pi 3 microcomputer. Both programs (For transmission and reception at the same Raspberry Pi 3) run independently without synchronisation (as it would be the case of primary/secondary users in a real scenario). Real-time operation is achieved by a patched version of the Linux kernel and running the programs with real-time priority. More details on PECAS can be found in [103].

First, the effect of desynchronized spectrum sensing on the observed minimum period will be discussed. In order to quantify this effect, the relative error (R_e) metric is utilized and calculated as $R_e = \left(|\widehat{\mu}_i - \mu_i|\right)/\mu_i$. Fig. 2.11 shows the relative error between the minimum sensed period $\widehat{\mu}_i$ versus the original minimum μ_i as a function of the sensing period. As it can be appreciated from Fig. 2.11, the analytical expression provides a perfect agreement with the simulation results. The estimated value of $\widehat{\mu}_i$ depends on T_s and the relative error shows an oscillating pattern with zeros at T_s values that are integer sub-multiples of the original μ_i.

Fig. 2.12 shows the relative error of the calculated variance with (2.16) and (2.19). As appreciated, the variance estimated using (2.16) is accurate only for low values of the sensing period T_s, while the estimation obtained using (2.19) is accurate regardless of the value of T_s, which proves that the correction factor in (2.19) can reduce significantly the estimation error of the variance.

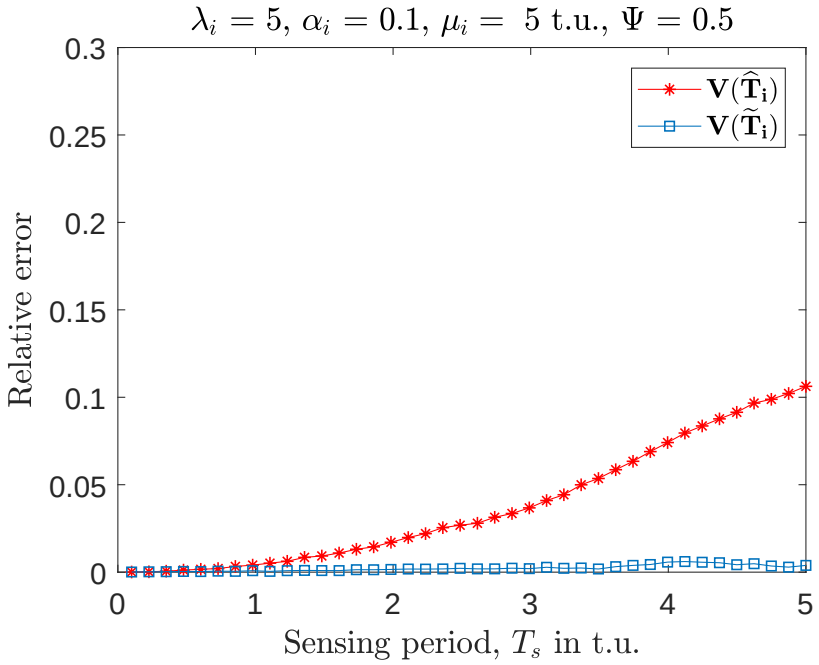


FIGURE 2.12: Relative error of the calculated variance using (2.16) and (2.19).

Fig. 2.13 shows the KS distance as a function of the sensing period for the considered distribution estimation methods. Two cases are shown for the MoM and MMoM methods, one where the value of the location parameter is estimated from spectrum sensing observations as indicated in (2.24) (labelled as 'with $\widehat{\mu}_i$') and another where the true value of the location parameter is assumed to be known by the SU (labelled as 'with μ_i').

First thing to notice in Fig. 2.13 is that the direct estimation method results in a significantly higher estimation error than the MoM and MMoM methods. In fact, the direct estimation method can provide an accurate estimation of the distribution of PU activity periods only if the employed sensing period is very short. The MoM and MMoM methods can provide in general more accurate estimations over the whole range of T_s values. The results obtained for the MoM and MMoM methods indicate that the estimation error is zero when the employed sensing period T_s is an integer sub-multiple of the true minimum PU activity time μ_i. Notice that the same behaviour is observed for the relative error of the estimated minimum $\widehat{\mu}_i$ in Fig. 2.11, which suggests that an accurate estimation (or perfect knowledge) of the value of μ_i can improve significantly the accuracy of the estimated distribution. To corroborate this observation, the accuracy of the MoM and MMoM methods is shown in Fig. 2.13 for the cases where the minimum PU activity time is known by the SU ('with μ_i'), and when it is estimated as indicated in (2.24) and is therefore not perfectly accurate ('with $\widehat{\mu}_i$'). As it can be observed, the distribution can be estimated more accurately when the minimum PU activity time μ_i is known accurately. However, for the MoM method this is not enough to provide a sufficient level of accuracy as seen in Fig. 2.13 since the effect of the finite sensing period T_s has not been removed. Only the MMoM method, which corrects and overcomes the

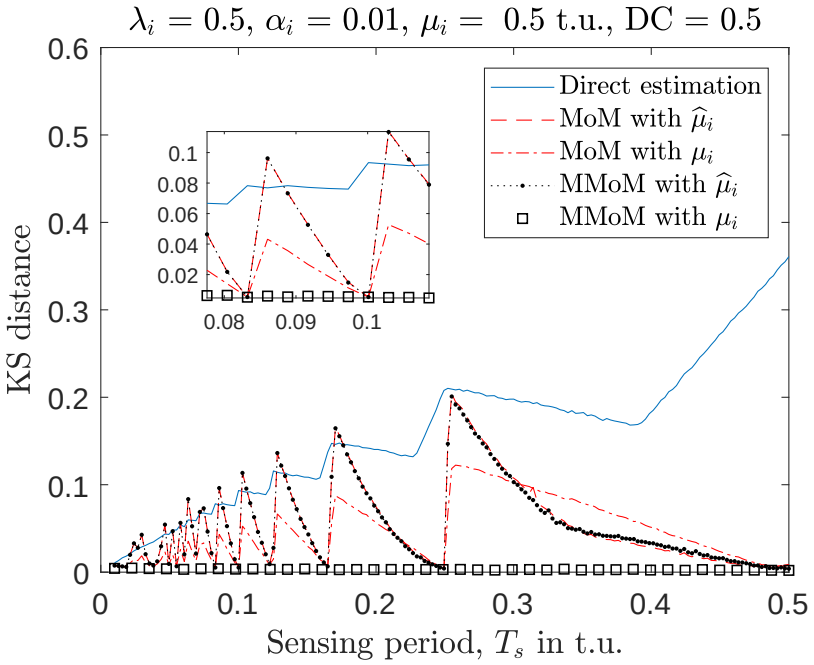


FIGURE 2.13: The KS distance using direct estimation, MoM and MMoM.

impact of the finite sensing period T_s on the estimation of the moments (in particular, the variance) can provide a nearly zero estimation error. These results demonstrate that only the proposed MMoM method can provide a nearly perfect estimation of the distribution of PU activity periods from spectrum sensing observations under realistic operation conditions.

Table 2.1 shows the KS distance as a function of the sensing period for the considered distribution estimation methods from simulations, versus the experimental KS distance from PECAS. The first point to note is that the experimental results match the ones obtained through simulations. The results shown in Table 2.1 prove that the proposed methods are accurate for CR systems. Moreover, these results have a significant importance for practical CR system design since they indicate that the proposed estimation methods allow an accurate estimation of primary traffic statics even with low-cost hardware devices as it the case of the PECAS prototype.

Next, the effect of sample size (i.e., number of idle/busy periods used to estimate the moments and subsequently the distribution) on the estimation accuracy of both MoM and MMoM methods is analysed. Fig. 2.14 shows the maximum KS distance for both methods versus the sample size for $T_s = 0.2$ *t.u.* and an accurate knowledge of the minimum PU activity time (this is the case that provides the best possible accuracy). As it can be observed, the MMoM method requires a lower sample size. For example, for a target KS distance of $D_{KS} = 0.1$, the MoM method requires the observation of 900 periods approximately while the MMoM method only requires around 300 period samples. Moreover, the MMoM estimation error decreases monotonically with the sample size, meaning that it can provide an arbitrarily accurate estimation of the distribution provided that a sufficiently large sample size is available, while the MoM method shows

TABLE 2.1: KS distance estimated using PECAS (experiment) versus simulation.

$\mathbf{T_s}$ **in seconds**	**KS distance**									
	Direct estimation		MoM with $\widehat{\mu}_i$		MoM with μ_i		MMoM with $\widehat{\mu}_i$		MMoM with μ_i	
	Sim	Exp	Sim	Exp	Sim	Exp	Sim	Exp	Sim	Exp
0.05	0.0431	0.0491	0.0095	0.0020	0.0081	0.0020	0.0094	0.0059	0.0038	0.0021
0.10	0.0934	0.0947	0.0052	0.0018	0.0052	0.0018	0.0053	0.0021	0.0034	0.0021
0.15	0.1086	0.1074	0.0439	0.0428	0.0288	0.0299	0.0431	0.0391	0.0036	0.0021
0.20	0.1403	0.1399	0.0777	0.0723	0.0583	0.0572	0.0761	0.0706	0.0029	0.0022
0.25	0.2079	0.2111	0.0051	0.0021	0.0045	0.0021	0.0042	0.0037	0.0030	0.0008
0.30	0.1979	0.1961	0.1076	0.1055	0.1022	0.1018	0.1030	0.1009	0.0029	0.0024
0.35	0.1825	0.1817	0.0446	0.0419	0.0789	0.0785	0.0462	0.0447	0.0024	0.0018
0.40	0.1851	0.1840	0.0282	0.0270	0.0507	0.0500	0.0330	0.0319	0.0024	0.0015
0.45	0.2714	0.2737	0.0127	0.0108	0.0239	0.0222	0.0159	0.0143	0.0023	0.0018
0.50	0.3609	0.3626	0.0090	0.0079	0.0089	0.0079	0.0015	0.0010	0.0020	0.0010

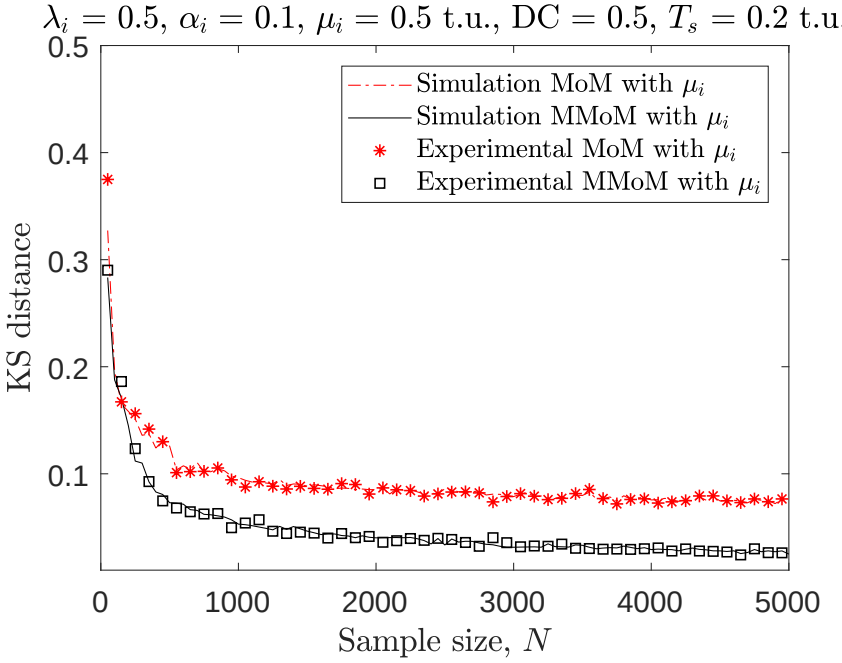


FIGURE 2.14: KS distance for MoM and MMoM versus the sample size, both of the methods assumed with perfect μ_1 knowledge.

a lower bound below which it is impossible to reduce the estimation error no matter how large the sample size is. The impact of the sample size will be analysed in detail in Chapter 3.

2.5 Summary

CR utilizes spectrum sensing to periodically monitor PU channel activity states. A CR benefits from this knowledge to improve the general system/device performance. However spectrum sensing uses a finite sensing period which imposes limitations on the measured durations of busy/idle periods and hence the time resolution to which the resulting distribution for PU activity can be estimated.

This chapter has focused on two aspects. First, the analytical perspective of how this limitation affects the estimation of PU distribution. Closed form expressions are derived to show the relationship between the employed sensing period and the resulting estimated distribution under finite sensing periods, as well as the corresponding estimation error in terms of the KS distance. The analytical results showed a good agreement with simulation results and can be used in the design and analysis of CR systems. Second, methods have been proposed to enable SU to obtain an accurate estimation of the distribution of PU activity periods, which plays a crucial role in improving the performance of CR systems and reducing the interference to primary networks. A modified version of the Method of Moments has been proposed to improve the primary distribution estimation. Simulation results have shown that the proposed method outperforms the conventional approach based on the direct estimation by means of empirical CDF calculation as well as the approach based on the standard Method of Moments. It has

been found out that an accurate estimation or knowledge of the minimum PU activity time is essential to achieve an accurate estimation of the distribution of PU activity periods. Provided that the minimum PU activity time is known or can be estimated to a sufficient degree of accuracy, the proposed MMoM method constitutes an ideal solution to provide an accurate (nearly perfect) estimation of the PU activity statistics based on spectrum sensing observations.

Chapter 3

Impact of the Sample Size under Perfect Spectrum Sensing

3.1 Introduction

As discussed earlier, information on primary traffic statistics can be employed to access the spectrum more effectively and reduce harmful interference on primary network. Even though PU traffic information could be obtained from other alternatives such as databases. However, the sensing-based approach has significant advantages including lower cost and complexity, independence of external systems and better suitability for highly dynamic radio environments [78].

In Chapter 3, closed form expressions to model the impact of spectrum sensing duration on the estimated primary distribution along with algorithm to improve the estimation were presented. In this context, an important practical question is how many busy/idle (on/off) periods need to be observed by a DSA/CR system (i.e., the observation sample size) in order to estimate the primary channel activity statistics (in this chapter: minimum, mean and variance of on/off times, channel duty cycle and the underlying distribution) to a certain desired level of accuracy. The only work that has considered this problem is the study presented in [104], where only the channel duty cycle is considered (estimated based on individual on/off sensing decisions) and the required number of individual sensing events is analysed. On the other hand, this chapter considers a more general approach where primary statistics are estimated based on a set of observed busy/idle time durations (instead of individual sensing decisions), which can be used to estimate a broader range of primary activity parameters (not only the channel duty cycle).

This chapter presents a comprehensive analytical study on the estimation of primary channel statistics based on spectrum sensing decisions and determines the relation between the number of observed on/off periods and the accuracy of the resulting primary activity statistics. To the best of author's knowledge, this problem has not received a rigorous treatment before in the literature. The accuracy of the derived closed-form

expressions is validated and corroborated with simulation results as well as experimental results from a hardware prototype developed to this end.

The main contributions of this chapter are outlined as follows:

1. Closed-form expressions are derived for the maximum error of the minimum, mean and variance of the estimated on/off periods, the channel duty cycle and the distribution as a function of the observation sample size (and other relevant parameters such as the sensing period). These analytical results are useful to determine the number of observed periods/samples required to guarantee a desired maximum error in the estimated statistics. Expressions for the required observation sample sizes are derived as well.

2. Practical validation of the obtained analytical results, not only with simulations but also with experiments, is provided using a hardware prototype specifically designed to replicate inexpensive low-end DSA/CR devices. This scenario is closer to real-life DSA/CR scenarios than using advanced and costly laboratory equipment.

3. A comprehensive study about the effect of the observation sample size on the estimation of primary activity statistics is carried out considering a practical primary activity model based on the Generalised Pareto distribution, which is proven to be a more accurate model compared to the commonly used exponential distribution models [64].

3.2 System Model and Problem Formulation

Following the same system model proposed in Section 2.2 which is shown in Fig. 2.1, the SU estimates a set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ of N observed periods of the same type, which is used to calculate the primary activity statistics of interest. A wide range of statistics is considered in this chapter, including the minimum, mean and variance of the observed periods $\widehat{T}_i$, the PU channel duty cycle, and the underlying distribution (this includes the activity statistics most commonly used in the DSA/CR literature).

The main objective of this chapter is to explore the relation between the number of observed periods N (the observation sample size) and the accuracy of the estimated primary activity statistics mentioned above. To this end, explicit closed-form expressions are derived for the estimation error of each primary activity statistic as a function of the sample size N (and other involved parameters). This is essential to enable DSA/CR systems determine whether the estimated statistics of the primary user activity are sufficiently accurate, or more observations are required instead.

3.3 Estimation of the Minimum Period

The estimation of the minimum primary activity time (both busy/on and idle/off times) has a great deal of importance when it comes to spectrum sensing and DSA/CR in

general [105, 106], as it determines the minimum amount of time that a DSA/CR will have to wait before the primary channel is available (minimum busy time) and the minimum amount of time it will be available for transmission (minimum idle time). In some cases, the true minimum period is known, for example in the case of primary systems that use some form of regional beacon signals with real-time information [78] or when the radio technology of the primary system is standardised (e.g., the slot duration of GSM). Otherwise, it needs to be estimated.

Based on a set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ of N observed periods $\widehat{T}_i$ estimated as shown in Fig. 2.1, the minimum period duration or minimum primary (in)activity time, denoted as $\widehat{\mu}_i$, can be obtained as:

$$\widehat{\mu}_i = \min_n \left(\{\widehat{T}_{i,n}\}_{n=1}^{N} \right). \tag{3.1}$$

Note that the periods $\widehat{T}_i$ estimated as shown in Fig. 2.1 are integer multiples of the sensing period T_s, however the same original period T_i can lead to two possible estimated periods, either $\widehat{T}_i = \lfloor T_i/T_s \rfloor T_s = kT_s$, $k \in \mathbb{N}^+$ (where $\lfloor \cdot \rfloor$ is the floor operator) or $\widehat{T}_i = \lceil T_i/T_s \rceil T_s = (k+1)T_s$, $k \in \mathbb{N}^+$ (where $\lceil \cdot \rceil$ is the ceil operator). The actual estimated period depends on the relative (random) position of the sensing events with respect to the beginning/end of the original period T_i. Based on this observation, the estimated periods can be modelled as:

$$\widehat{T}_i = \left(\left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s, \tag{3.2}$$

where $\xi \in \{0, 1\}$ is a Bernoulli random variable. Introducing (3.2) into (5.3), and noting that $\min(T_i) = \mu_i$ and $\min(\xi) = 0$, it can be seen that the estimated minimum period is given by:

$$\widehat{\mu}_i = \min(\widehat{T}_i) = \min \left[\left(\left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s \right] = \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s. \tag{3.3}$$

The main question that this chapter aims to answer is how many primary periods N need to be observed in order to estimate each primary activity statistic to a certain degree of accuracy. From (3.3) it can be observed that, in the particular case of the estimation of the minimum period duration, the estimation error is mainly given by the employed sensing period T_s and increasing the sample size N will not improve the accuracy of the estimated minimum $\widehat{\mu}_i$. However, if the observation sample size N is not sufficiently large, a longer period $\widehat{T}_i > \widehat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$ might be selected as the minimum observed period, thus potentially leading to a less accurate estimation. Therefore, in the particular case of this section, the relevant question is how many primary periods N need to be observed to ensure that the estimated minimum period is the most accurate possible estimation, in other words, ensure that at least one instance of the period $\widehat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$ is observed in the set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$.

To answer this question, let first determine the probability that an observed period $\widehat{T}_i$ is equal to the best possible estimation of the minimum period given by (3.3), i.e., $\widehat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$. A period with duration $\widehat{T}_i = \widehat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$ will be observed if

$T_i \in [\lfloor \mu_i/T_s \rfloor T_s, \lceil \mu_i/T_s \rceil T_s]$ and $\xi = 0$, therefore the probability to observe one instance of $\widehat{\mu}_i$ can be obtained as:

$$\begin{aligned} P\left(\widehat{T}_i = \widehat{\mu}_i\right) &= P\left(\left\lfloor \tfrac{\mu_i}{T_s} \right\rfloor T_s \leq T_i \leq \left\lceil \tfrac{\mu_i}{T_s} \right\rceil T_s\right) \mathbb{E}\left(P(\xi = 0)\right) \\ &= P\left(\mu_i \leq T_i \leq \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s\right) \mathbb{E}\left(P(\xi = 0)\right) \chi_0 \\ &= \left[F_{T_i}\left(\left\lceil \frac{\mu_i}{T_s} \right\rceil T_s\right) - F_{T_i}\left(\mu_i\right)\right] \mathbb{E}\left(P(\xi = 0)\right) \chi_0 \\ &= F_{T_i}\left(\left\lceil \frac{\mu_i}{T_s} \right\rceil T_s\right) \mathbb{E}\left(P(\xi = 0)\right) \chi_0, \end{aligned} \tag{3.4}$$

where $F_{T_i}(\cdot)$ is the CDF of the original periods T_i and $F_{T_i}(\mu_i) = 0$, $P(\xi = 0)$ is the probability that $\xi = 0$ in the model of (3.2) which can be calculated as:

$$P(\xi = 0) = P\left(\widehat{T}_i = \left\lfloor \frac{T_i}{T_s} \right\rfloor T_s\right) = \frac{\left\lceil \frac{T_i}{T_s} \right\rceil T_s - T_i}{T_s} = \left\lceil \frac{T_i}{T_s} \right\rceil - \frac{T_i}{T_s}, \tag{3.5}$$

the expected value of which is given by:

$$\begin{aligned} \mathbb{E}\left(P(\xi = 0)\right) &= \int_T P(\xi = 0)\, f_{T_i}(T)\, dT \\ &= \sum_{m=0}^{\infty} (m+1)\left[F_{T_i}((m+1)T_s) - F_{T_i}(mT_s)\right] - \tfrac{\mathbb{E}(T_i)}{T_s}, \end{aligned} \tag{3.6}$$

where $f_{T_i}(\cdot)$ is the PDF of the original periods T_i, and χ_0 is a correction factor for $P(\xi = 0)$ given by:

$$\chi_0 = \frac{\left\lceil \frac{\mu_i}{T_s} \right\rceil T_s - \mu_i}{T_s} = \left\lceil \frac{\mu_i}{T_s} \right\rceil - \frac{\mu_i}{T_s}. \tag{3.7}$$

Notice that for any arbitrary period T_i, the width of the interval $[\lfloor T_i/T_s \rfloor T_s, \lceil T_i/T_s \rceil T_s]$ is T_s. However, around the minimum period μ_i it holds that $T_i \in [\mu_i, \lceil \mu_i/T_s \rceil T_s]$ (since $T_i \geq \mu_i \geq \lfloor \mu_i/T_s \rfloor T_s$) and the width of such interval is $\lceil \mu_i/T_s \rceil T_s - \mu_i$ instead of T_s. Therefore, a scaling coefficient χ_0 is required for the probability $P(\xi = 0)$ as shown in (3.7).

The probability to observe at least one instance of $\widehat{\mu}_i$ in the N observed periods can be related to the binomial distribution:

$$P_{\text{obs}}^{\widehat{\mu}_i} = 1 - \left[1 - P\left(\widehat{T}_i = \widehat{\mu}_i\right)\right]^N. \tag{3.8}$$

Finally, by specifying a probability of occurrence of the minimum, $P_{\text{obs}}^{\widehat{\mu}_i}$, the minimum number of periods required to ensure the observation of the minimum period $\widehat{\mu}_i$ is:

$$N_{\widehat{\mu}_i} = \frac{\log\left(1 - P_{\text{obs}}^{\widehat{\mu}_i}\right)}{\log\left(1 - P\left(\widehat{T}_i = \widehat{\mu}_i\right)\right)}. \tag{3.9}$$

Notice that increasing the sample size will not improve the accuracy of the estimated minimum $\widehat{\mu}_i$ itself, but the probability that a period $\widehat{\mu}_i$ is observed (otherwise a longer period $\widehat{T}_i > \widehat{\mu}_i$ might be selected as the minimum period, thus potentially leading to a more inaccurate estimation).

3.4 Estimation of the Mean and Variance

Given a set of N observed periods $\{\widehat{T}_{i,n}\}_{n=1}^{N}$, the mean $\mathbb{E}(\widehat{T}_i)$ and variance $\mathbb{V}(\widehat{T}_i)$ of the provided durations can be estimated directly from the (unbiased) sample moments:

$$\mathbb{E}(\widehat{T}_i) \approx \widehat{m}_i = \frac{1}{N}\sum_{n=1}^{N}\widehat{T}_{i,n}, \tag{3.10}$$

$$\mathbb{V}(\widehat{T}_i) \approx \widehat{v}_i = \frac{1}{N-1}\sum_{n=1}^{N}\left(\widehat{T}_{i,n} - \widehat{m}_i\right)^2. \tag{3.11}$$

Since the estimated periods $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ are integer multiples of the sensing period T_s as discussed in Section 3.3, the sample moments obtained as shown in (3.10) and (3.11) will be affected by an error associated with the employed finite sensing period T_s. For the purposes of the analysis carried out in this section, the estimated periods can be modelled as $\widehat{T}_i = T_i + T_e$, where T_e represents the above mentioned estimation error. Such error can be represented by the sum of two error components T_e^x and T_e^y shown in Fig. 2.1 (i.e., $T_e = T_e^y - T_e^x$). Both error components can take values within the interval $[0, T_s]$ and can be assumed to be uniformly distributed (i.e., $T_e^x, T_e^y \sim U(0, T_s)$); the analysis of simulation results indicated that this assumption is valid. The impact on the estimated moments can thus be modelled as:

$$\begin{aligned}\mathbb{E}(\widehat{T}_i) &= \mathbb{E}(T_i) + \mathbb{E}(T_e)\\ &= \mathbb{E}(T_i) + \mathbb{E}(T_e^y) - \mathbb{E}(T_e^x) = \mathbb{E}(T_i),\end{aligned} \tag{3.12}$$

$$\begin{aligned}\mathbb{V}(\widehat{T}_i) &= \mathbb{V}(T_i) + \mathbb{V}(T_e)\\ &= \mathbb{V}(T_i) + \mathbb{V}(T_e^y) + \mathbb{V}(T_e^x) = \mathbb{V}(T_i) + \frac{T_s^2}{6},\end{aligned} \tag{3.13}$$

where $\mathbb{E}(T_e^y) - \mathbb{E}(T_e^x) = 0$ since T_e^x and T_e^y are identically distributed, $\mathbb{V}(T_e) = \mathbb{V}(T_e^y) + \mathbb{V}(T_e^x)$ assuming that T_e^x and T_e^y are independent and $\mathbb{V}(T_e^x) = \mathbb{V}(T_e^y) = T_s^2/12$ is the variance of the uniform distribution $U(0, T_s)$ of T_e^x and T_e^y. As appreciated in (3.12), the estimated sample mean $\widehat{m}_i$ is not affected by the employed sensing period. On the other hand, as observed in (3.13), the estimated sample variance $\widehat{v}_i$ is affected by an error factor of $T_s^2/6$, which is constant and known. Based on (3.13), the effect of T_s can be removed by applying to (3.11) the appropriate correction factor:

$$\widetilde{v}_i = \widehat{v}_i - \frac{T_s^2}{6}, \tag{3.14}$$

where $\widetilde{v}_i$ is the observed variance after correction (i.e., the corrected sample variance). This approach eliminates the impairments imposed by the use of a finite sensing period T_s in the estimated moments and is able to provide an accurate estimation of the real moments of T_i based on the estimated period durations $\{\widehat{T}_{i,n}\}_{n=1}^{N}$, provided that a sufficiently large number of periods N is captured.

Notice that the estimation error T_e resulting from the use of a finite sensing period T_s cannot be reduced by increasing the observation sample size N, therefore it is necessary to first correct the sample moments (in this case, the sample variance) as discussed above in order to remove such error. Once the sample moments have been corrected, the resulting estimations can then be made arbitrarily close to the true population moments of the original periods T_i by taking a sufficiently large number of period observations N. Thus, the question to answer in this section is how large does the set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ need to be (i.e., what is the required value of N) so that the sample mean $\widehat{m}_i$ and the corrected sample variance $\widetilde{v}_i$ of the set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ are as close as desired to the original population moments, i.e., $\widehat{m}_i \approx \mathbb{E}(T_i)$ and $\widetilde{v}_i \approx \mathbb{V}(T_i)$.

Since the estimators $\widehat{m}_i$ in (3.10) and $\widetilde{v}_i$ in (3.14) are unbiased (i.e., $\mathbb{E}(\widehat{m}_i) = \mathbb{E}(T_i)$ and $\mathbb{E}(\widetilde{v}_i) = \mathbb{V}(T_i)$), the sample size required for a certain estimation error can be determined based on the standard errors of the estimators, which are related to their variances [107, p.229] as:

$$\mathbb{V}(\widehat{m}_i) = \frac{\mathbb{V}(\widehat{T}_i)}{N}, \tag{3.15}$$

$$\mathbb{V}(\widetilde{v}_i) = \mathbb{V}(\widehat{v}_i) = \frac{1}{N}\left(\mathbb{M}_4(\widehat{T}_i) - [\mathbb{V}(\widehat{T}_i)]^2\frac{N-3}{N-1}\right), \tag{3.16}$$

where $\mathbb{M}_4(\widehat{T}_i)$ is the fourth central moment of $\widehat{T}_i$, given by:

$$\begin{aligned}\mathbb{M}_4(\widehat{T}_i) &= \mathbb{E}\big([\widehat{T}_i - \mathbb{E}(\widehat{T}_i)]^4\big)\\ &= \mathbb{E}\big([T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e)]^4\big)\\ &= \mathbb{M}_4(T_i) + 6\mathbb{V}(T_i)\mathbb{V}(T_e) + \mathbb{M}_4(T_e)\\ &= \mathbb{M}_4(T_i) + \mathbb{V}(T_i)T_s^2 + \frac{T_s^4}{15},\end{aligned} \tag{3.17}$$

where $\mathbb{V}(T_e) = \mathbb{V}(T_e^x)+\mathbb{V}(T_e^y) = T_s^2/6$ and $\mathbb{M}_4(T_e) = \mathbb{M}_4(T_e^x)+6\mathbb{V}(T_e^x)\mathbb{V}(T_e^y)+\mathbb{M}_4(T_e^y) = T_s^4/15$, since $\mathbb{V}(T_e^x) = \mathbb{V}(T_e^y) = T_s^2/12$ and $\mathbb{M}_4(T_e^x) = \mathbb{M}_4(T_e^y) = T_s^4/80$.

Given an estimator ω, it is possible to define a confidence interval of κ standard deviations around the expected value of the estimator such that the estimated values are within that interval with a minimum probability ρ (confidence level):

$$P\left(|\omega - \mathbb{E}(\omega)| \leq \kappa\sqrt{\mathbb{V}(\omega)}\right) \geq \rho. \tag{3.18}$$

If the estimator ω is unbiased, its relative error can then be bounded by $\varepsilon_{r,\max}^{\omega} \approx \kappa\sqrt{\mathbb{V}(\omega)}/\mathbb{E}(\omega)$. Based on this, the relative errors of the estimators $\widehat{m}_i$ in (3.10) and $\widetilde{v}_i$

in (3.14), which are unbiased (i.e., $\mathbb{E}(\widehat{m}_i) = \mathbb{E}(T_i)$, $\mathbb{E}(\widetilde{v}_i) = \mathbb{V}(T_i)$) are:

$$\varepsilon_{r,\max}^{\widehat{m}_i} \approx \frac{\kappa}{\mathbb{E}(T_i)} \left[\frac{1}{N} \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) \right]^{\frac{1}{2}}, \tag{3.19}$$

$$\varepsilon_{r,\max}^{\widetilde{v}_i} \approx \frac{\kappa}{\mathbb{V}(T_i)} \left[\frac{1}{N} \left(\mathbb{M}_4(T_i) - \frac{N-3}{N-1}[\mathbb{V}(T_i)]^2 + \right. \right.$$
$$\left. \left. + \frac{2N}{3(N-1)} T_s^2 \mathbb{V}(T_i) + \frac{7N+3}{180(N-1)} T_s^4 \right) \right]^{\frac{1}{2}}, \tag{3.20}$$

which relates the maximum relative error of the estimators $\widehat{m}_i$ and $\widetilde{v}_i$ to the observation sample size N (as well as the sensing period T_s and the moments of the original periods T_i).

The value of κ for a certain confidence level ρ can be derived from concentration inequalities (some examples are shown in Table 3.1). However, this approach usually leads to loose upper bounds on the maximum relative error as it will be shown in Section 3.8. A much tighter result can be found by noting that $\widehat{m}_i$ and $\widetilde{v}_i$ can be assumed to be normally distributed by the central limit theorem. The inequality in (3.18) can be rewritten for a normal distribution as follows:

$$\begin{aligned} P\left(\varepsilon_{\text{abs}}^{\omega} \leq \varepsilon_{\text{abs,max}}^{\omega}\right) &= P\left(|\omega - \mathbb{E}(\omega)| \leq \kappa\sqrt{\mathbb{V}(\omega)}\right) \\ &= \int_{\mathbb{E}(\omega)-\kappa\sqrt{\mathbb{V}(\omega)}}^{\mathbb{E}(\omega)+\kappa\sqrt{\mathbb{V}(\omega)}} \frac{e^{-\frac{1}{2}\left(\frac{\omega-\mathbb{E}(\omega)}{\sqrt{\mathbb{V}(\omega)}}\right)^2}}{\sqrt{2\pi\mathbb{V}(\omega)}} d\omega \\ &= \operatorname{erf}\left(\frac{\kappa}{\sqrt{2}}\right) \geq \rho. \end{aligned} \tag{3.21}$$

Solving (3.21) for κ yields the relation $\kappa \geq \sqrt{2}\operatorname{erf}^{-1}(\rho)$. It is worth noting that the approach employed to determine the relation between κ and ρ has a significant impact on the accuracy of (3.19) and (3.20) as well as the mathematical results derived later on for other PU activity statistics. The relations shown in Table 3.1 are concentration inequalities and therefore provide *bounds* on the true value of the maximum estimation error, while the relation $\kappa \geq \sqrt{2}\operatorname{erf}^{-1}(\rho)$ obtained from (3.21) is an *approximation* to the true value of the maximum estimation error. As such, the latter can be expected to be more accurate. This will be shown and discussed in detail in Section 3.8.

The observation sample size required to guarantee a predefined maximum relative error follows from (3.19) and (3.20):

$$N_{\widehat{m}_i} \approx \left(\frac{\kappa}{\mathbb{E}(T_i)\ \varepsilon_{r,\max}^{\widehat{m}_i}} \right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right), \tag{3.22}$$

$$N_{\widetilde{v}_i} \approx \left(\frac{\kappa}{\mathbb{V}(T_i)\ \varepsilon_{r,\max}^{\widetilde{v}_i}} \right)^2 \left(\mathbb{M}_4(T_i) - [\mathbb{V}(T_i)]^2 + \frac{2T_s^2\mathbb{V}(T_i)}{3} + \frac{7T_s^4}{180} \right), \tag{3.23}$$

TABLE 3.1: Relation between κ and ρ for various concentration inequalities [108].

Inequality	**Relation**
Chebyshev	$\kappa \geq 1/\sqrt{1-\rho}$
Cantelli	$\kappa \geq \sqrt{\rho/(1-\rho)}$
Vysochanskij-Petunin	$\kappa \geq 2/3\sqrt{1-\rho}$
Sobolevev	$\kappa \geq \sqrt{-4\ln((1-\rho)/2)}$
Bernstein	$\kappa \geq \sqrt{-2\ln(1-\rho)}$

where $N \gg 3$ has been assumed in (3.20). This assumption is reasonable as a relatively large observation sample size is usually required for an accurate estimation.

It is worth noting that the required sample size for the estimation of moments (sample mean $\widehat{m}_i$ and corrected sample variance $\widetilde{v}_i$) may in some cases be relatively high, depending on the desired level of estimation accuracy, the statistics of the original periods T_i and the employed sensing period T_s. In DSA/CR devices with limited memory capabilities, this problem can be overcome by computing the sample moments based on recurrence formulae [109] instead of storing the complete history of the N past observed period durations.

3.5 Estimation of the Duty Cycle

The duty cycle is an important statistic commonly used to characterise the level of occupancy of a primary channel or frequency band (defined as the probability that the primary channel is busy, i.e., occupied by a PU signal).

The duty cycle can be estimated based on individual spectrum sensing decisions as the ratio of the number of sensing events with a busy/on (H_1) decision to the total number of sensing events. The work reported in [104] provides an analytical study on the sample size (understood in this case as the number of individual sensing events) required for an arbitrarily accurate estimation of the duty cycle based on this approach. However, taking into account that the duty cycle, denoted as Ψ, can be related to the mean value of idle and busy periods as follows:

$$\Psi = \frac{\mathbb{E}(T_1)}{\mathbb{E}(T_0) + \mathbb{E}(T_1)}, \tag{3.24}$$

an estimation thereof, denoted as $\widehat{\Psi}$, can also be obtained based on the sample mean estimator $\widehat{m}_i$ in (3.10) as shown below:

$$\widehat{\Psi} = \frac{\widehat{m}_1}{\widehat{m}_0 + \widehat{m}_1}, \tag{3.25}$$

where $\widehat{m}_0$ and $\widehat{m}_1$ represent the sample mean of idle/off and busy/on periods, respectively. This section provides an analytical study on the observation sample size N (i.e.,

number of periods in the observed set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$) required for an arbitrarily accurate estimation of the duty cycle based on (3.25).

The standard error (i.e., the standard deviation, or equivalently the variance) of the estimated sample mean can be propagated through (3.24)–(3.25) to obtain the standard error (or variance) of the estimated duty cycle [110, 111]:

$$\mathbb{V}(\widehat{\Psi}) = \left(\frac{\partial\widehat{\Psi}}{\partial\widehat{m}_0}\right)^2 \mathbb{V}(\widehat{m}_0) + \left(\frac{\partial\widehat{\Psi}}{\partial\widehat{m}_1}\right)^2 \mathbb{V}(\widehat{m}_1). \tag{3.26}$$

Based on (3.18), the relative error of the duty cycle estimated from (3.25) is obtained as:

$$\begin{aligned}\varepsilon_{r,\max}^{\widehat{\Psi}} &\approx \frac{\kappa}{\mathbb{E}(\widehat{\Psi})}\sqrt{\mathbb{V}(\widehat{\Psi})} \\ &= \frac{\kappa}{\Psi}\left[\frac{1}{N[\mathbb{E}(T_0)+\mathbb{E}(T_1)]^4}\left\{[\mathbb{E}(T_1)]^2\left(\mathbb{V}(T_0)+\frac{T_s^2}{6}\right)+\right.\right. \\ &\quad \left.\left.+\,[\mathbb{E}(T_0)]^2\left(\mathbb{V}(T_1)+\frac{T_s^2}{6}\right)\right\}\right]^{\frac{1}{2}},\end{aligned} \tag{3.27}$$

where $\mathbb{E}(\widehat{\Psi}) = \Psi$ and $\mathbb{V}(\widehat{\Psi})$ is obtained by introducing (3.13) and (3.15) into (3.26) and solving the derivatives.

The observation sample size required to guarantee a given maximum relative error follows from solving (3.27) for N:

$$N_{\widehat{\Psi}} \approx \left(\frac{\kappa}{\Psi\ \varepsilon_{r,\max}^{\widehat{\Psi}}}\right)^2 \frac{\Psi^2\left[\mathbb{V}(T_0)+\frac{T_s^2}{6}\right]+(1-\Psi)^2\left[\mathbb{V}(T_1)+\frac{T_s^2}{6}\right]}{[\mathbb{E}(T_0)+\mathbb{E}(T_1)]^2}, \tag{3.28}$$

where κ can be obtained from concentration inequalities (Table 3.1) or the normal approximation as described in Section 3.4.

3.6 Estimation of the Distribution

The distribution of the busy/on and idle/off times provides a complete characterisation of the PU activity statistics and its accurate estimation is therefore of great importance in the context of DSA/CR systems. Several methods can be used to estimate the distribution of the PU busy/idle times based on a finite set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ of N observed periods $\widehat{T}_i$. The most commonly used method is the direct calculation of the empirical CDF of the set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$. The main drawback of this method is that the estimated distribution has a discrete domain (even though the original distribution is in general continuous) since the periods $\widehat{T}_i$ estimated as shown in Fig. 2.1 are integer multiples of the sensing period T_s. This leads to an irreducible estimation error that depends on the employed sensing period T_s and cannot be improved by increasing the sample size N [105]. This limitation can be overcome by the method of moments proposed in [106],

where the parameters of the primary distribution are estimated based on the sample moments of the set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$. This alternative method relies on the assumption of a particular model for the PU distribution.

A common assumption frequently employed in the literature is that idle/busy periods follow an exponential distribution [80, 112–114], which simplifies analytical studies. Field measurements, however, have shown that this model is unrealistic [115–117]. A more realistic model over a broad range of frequency bands is the Generalised Pareto (GP) distribution [64]. According to this model, the CDF of the original periods T_i is given by [118, ch. 20]:

$$F_{T_i}(T) = 1 - \left[1 + \frac{\alpha_i(T-\mu_i)}{\lambda_i}\right]^{-1/\alpha_i} \quad T \geq \mu_i, \tag{3.29}$$

where $\mu_i > 0$, $\lambda_i > 0$, $\alpha_i \in \mathbb{R}$ are the location, scale and shape parameters, respectively. Based on (3.29), the method proposed in [106] provides an estimated distribution:

$$F_{\widehat{T}_i}(T) = 1 - \left[1 + \frac{\widehat{\alpha}_i(T-\widehat{\mu}_i)}{\widehat{\lambda}_i}\right]^{-1/\widehat{\alpha}_i} \quad T \geq \widehat{\mu}_i. \tag{3.30}$$

Parameter $\widehat{\mu}_i$ represents the estimated minimum period and its value is assumed to be known to a reasonable level of accuracy ($\widehat{\mu}_i \approx \mu_i$). According to the method of moments, the scale and shape parameters are estimated as [118, ch. 20]: .

$$\widehat{\lambda}_i = \frac{1}{2}\left(1 + \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widetilde{v}_i}\right)(\widehat{m}_i - \widehat{\mu}_i), \tag{3.31a}$$

$$\widehat{\alpha}_i = \frac{1}{2}\left(1 - \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widetilde{v}_i}\right), \tag{3.31b}$$

where the sample mean $\widehat{m}_i$ and the corrected sample variance $\widetilde{v}_i$ are obtained as shown in (3.10) and (3.14), respectively, and $\widehat{\mu}_i \approx \mu_i$. Introducing the MoM estimates provided by (3.31) into (3.30) provides a continuous estimation of the distribution of PU on/off times. Such estimation can be made arbitrarily close to the true distribution in (3.29) by increasing the number of periods N used to calculate $\widehat{m}_i$ and $\widetilde{v}_i$.

To determine the observation sample size N required for an arbitrarily accurate estimation, it is necessary to express the error in the estimated distribution as a function of N. This can be achieved by propagating the standard error of $\widehat{m}_i$ and $\widetilde{v}_i$, $\mathbb{V}(\widehat{m}_i)$ and $\mathbb{V}(\widetilde{v}_i)$ in (3.15) and (3.16) respectively, through (3.31):

$$\mathbb{V}\big(\widehat{\lambda}_i\big) = \left(\frac{\partial\widehat{\lambda}_i}{\partial\widehat{m}_i}\right)^2 \mathbb{V}(\widehat{m}_i) + \left(\frac{\partial\widehat{\lambda}_i}{\partial\widetilde{v}_i}\right)^2 \mathbb{V}(\widetilde{v}_i) + 2\frac{\partial\widehat{\lambda}_i}{\partial\widehat{m}_i}\frac{\partial\widehat{\lambda}_i}{\partial\widetilde{v}_i}\mathbb{C}(\widehat{m}_i, \widetilde{v}_i), \tag{3.32a}$$

$$\mathbb{V}\big(\widehat{\alpha}_i\big) = \left(\frac{\partial\widehat{\alpha}_i}{\partial\widehat{m}_i}\right)^2 \mathbb{V}(\widehat{m}_i) + \left(\frac{\partial\widehat{\alpha}_i}{\partial\widetilde{v}_i}\right)^2 \mathbb{V}(\widetilde{v}_i) + 2\frac{\partial\widehat{\alpha}_i}{\partial\widehat{m}_i}\frac{\partial\widehat{\alpha}_i}{\partial\widetilde{v}_i}\mathbb{C}(\widehat{m}_i, \widetilde{v}_i). \tag{3.32b}$$

Since the sample mean and sample variance are not independent, the covariance between both, denoted as $\mathbb{C}(\widehat{m}_i, \widetilde{v}_i)$, needs to be included in (3.32). Such covariance is given by $\mathbb{C}(\widehat{m}_i, \widetilde{v}_i) = \mathbb{M}_3(T_i)/N$ [119] where $\mathbb{M}_3(T_i)$ is the third central moment of the population distribution.

The standard errors obtained from (3.32) can be further propagated through (3.30) to obtain (assuming $\widehat{\mu}_i \approx \mu_i$):

$$\mathbb{V}\big(F_{\widehat{T}_i}(T)\big) = \left(\frac{\partial F_{\widehat{T}_i}(T)}{\partial \widehat{\lambda}_i}\right)^2 \mathbb{V}\big(\widehat{\lambda}_i\big) + \left(\frac{\partial F_{\widehat{T}_i}(T)}{\partial \widehat{\alpha}_i}\right)^2 \mathbb{V}\big(\widehat{\alpha}_i\big), \tag{3.33}$$

and the absolute error of the estimated distribution can then be written as:

$$\left|F_{T_i}(T) - F_{\widehat{T}_i}(T)\right| = \kappa\sqrt{\mathbb{V}\big(F_{\widehat{T}_i}(T)\big)}. \tag{3.34}$$

A common metric typically employed to quantify the difference between two CDFs is the Kolmogorov-Smirnov (KS) distance, which is defined as [96]:

$$D_{KS} = \sup_T \left|F_{T_i}(T) - F_{\widehat{T}_i}(T)\right|, \tag{3.35}$$

and can be obtained from (3.34) by solving:

$$\frac{\partial}{\partial T}\left|F_{T_i}(T) - F_{\widehat{T}_i}(T)\right| = \frac{\partial}{\partial T}\left[\kappa\sqrt{\mathbb{V}\big(F_{\widehat{T}_i}(T)\big)}\right] = 0. \tag{3.36}$$

Unfortunately (3.36) cannot be solved in closed form. However, it can be shown that the value of (3.33) is mainly dominated (and can be approximated) by its first term. As a result, an approximated result to (3.36) can be obtained by solving:

$$\frac{\partial}{\partial T}\left|F_{T_i}(T) - F_{\widehat{T}_i}(T)\right| \approx \frac{\partial}{\partial T}\left[\kappa\sqrt{\left(\frac{\partial F_{\widehat{T}_i}(T)}{\partial \widehat{\lambda}_i}\right)^2 \mathbb{V}\big(\widehat{\lambda}_i\big)}\right] = 0, \tag{3.37}$$

which is maximised for $T = \widehat{\mu}_i + \widehat{\lambda}_i$.

Combining (3.30)–(3.35) and evaluating the resulting expression in $T = \widehat{\mu}_i + \widehat{\lambda}_i$ yields the final expression for D_{KS}:

$$D_{KS} = \kappa\ (1+\alpha_i)^{-\frac{1}{\alpha_i}-1}\left[\frac{1}{\lambda_i^2}\mathbb{V}\big(\widehat{\lambda}_i\big) + \frac{[(1+\alpha_i)\ln(1+\alpha_i) - \alpha_i]^2}{\alpha_i^4}\mathbb{V}\big(\widehat{\alpha}_i\big)\right]^{\frac{1}{2}}, \tag{3.38}$$

where $\mathbb{V}\big(\widehat{\lambda}_i\big)$ and $\mathbb{V}\big(\widehat{\alpha}_i\big)$ are obtained from (3.32) and given by (3.44). The relevant moments in (3.44) for the GP distribution are:

$$\mathbb{E}(T_i) = \mu_i + \frac{\lambda_i}{1-\alpha_i}, \tag{3.39}$$

$$\mathbb{V}(T_i) = \frac{\lambda_i^2}{(1-\alpha_i)^2(1-2\alpha_i)}, \tag{3.40}$$

$$\mathbb{M}_3(T_i) = \frac{2\lambda_i^3(\alpha_i+1)}{(1-\alpha_i)^3(1-2\alpha_i)(1-3\alpha_i)}, \tag{3.41}$$

$$\mathbb{M}_4(T_i) = \frac{3\lambda_i^4(2\alpha_i^2+\alpha_i+3)}{(1-\alpha_i)^4(1-2\alpha_i)(1-3\alpha_i)(1-4\alpha_i)}. \tag{3.42}$$

The result in (3.38) provides a closed-form relation between the observation sample size, N, and the error of the estimated CDF, $F_{\widehat{T}_i}(T)$, in terms of the KS distance with respect to the true distribution $F_{T_i}(T)$. The observation sample size required to guarantee a given KS distance follows from solving (3.38) for N:

$$N_{F_{\widehat{T}_i}(T)} = \left(\frac{\kappa}{D_{KS}}\right)^2 (1+\alpha_i)^{-\frac{2}{\alpha_i}-2}\left[\frac{1}{\lambda_i^2}\Omega(T_i) + \frac{[(1+\alpha_i)\ln(1+\alpha_i)-\alpha_i]^2}{\alpha_i^4}\Upsilon(T_i)\right], \tag{3.43}$$

where $\Omega(T_i)$ and $\Upsilon(T_i)$ are obtained from (3.44), assuming $N \gg 3$ and therefore $(N-3)/(N-1) \approx 1$, and are given by (3.45).

3.7 Iterative Stopping Algorithm

The analytical results obtained in previous sections provide closed-form expressions for the required observation sample size as a function of the desired estimation error for the minimum period (3.9), mean (3.22), variance (3.23), duty cycle (3.28) and distribution (3.43). Notice that such expressions depend not only on the desired estimation error but also on the real moments and/or distribution parameters of the PU traffic, which are unknown to the SU (and indeed the parameters to be estimated). As a result, such expressions cannot be used in a real implementation directly, without further considerations.

To overcome this problem, an iterative stopping algorithm is here proposed. The proposed algorithm is composed of three steps that are executed every time a new idle/busy period is observed (i.e., every time a new sample becomes available)[1]:

1. Update the calculated sample moments (mean, variance, third/fourth central moments) based on appropriate recurrence equations [109] and apply any required correction factors as appropriate (see the appendix).

[1] As discussed in Section 3.3, the estimated minimum period cannot be made arbitrarily accurate by increasing the sample size. Therefore the algorithm proposed in this section is applicable to the rest of PU activity statistics (i.e., mean/variance of the estimated periods, channel duty cycle and distribution.)

$$\mathbb{V}\big(\widehat{\lambda}_i\big) = \frac{1}{N}\left\{\frac{1}{4}\left(1+\frac{3[\mathbb{E}(T_i)-\mu_i]^2}{\mathbb{V}(T_i)}\right)^2\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right) - \frac{\mathbb{M}_3(T_i)}{2}\frac{[\mathbb{E}(T_i)-\mu_i]^3}{[\mathbb{V}(T_i)]^2}\left(1+\frac{3[\mathbb{E}(T_i)-\mu_i]^2}{\mathbb{V}(T_i)}\right) + \right.$$
$$\left. + \frac{1}{4}\frac{[\mathbb{E}(T_i)-\mu_i]^6}{[\mathbb{V}(T_i)]^4}\left[\mathbb{M}_4(T_i)+\mathbb{V}(T_i)T_s^2+\frac{T_s^4}{15}-\frac{N-3}{N-1}\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right)^2\right]\right\} \tag{3.44a}$$

$$\mathbb{V}\big(\widehat{\alpha}_i\big) = \frac{1}{N}\left\{\left(\frac{\mathbb{E}(T_i)-\mu_i}{\mathbb{V}(T_i)}\right)^2\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right) - \left(\frac{\mathbb{E}(T_i)-\mu_i}{\mathbb{V}(T_i)}\right)^3\mathbb{M}_3(T_i) + \right.$$
$$\left. + \frac{1}{4}\left(\frac{\mathbb{E}(T_i)-\mu_i}{\mathbb{V}(T_i)}\right)^4\left[\mathbb{M}_4(T_i)+\mathbb{V}(T_i)T_s^2+\frac{T_s^4}{15}-\frac{N-3}{N-1}\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right)^2\right]\right\} \tag{3.44b}$$

$$\Omega(T_i) = \frac{1}{4}\left(1+\frac{3[\mathbb{E}(T_i)-\mu_i]^2}{\mathbb{V}(T_i)}\right)^2\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right) - \frac{\mathbb{M}_3(T_i)}{2}\frac{[\mathbb{E}(T_i)-\mu_i]^3}{[\mathbb{V}(T_i)]^2}\left(1+\frac{3[\mathbb{E}(T_i)-\mu_i]^2}{\mathbb{V}(T_i)}\right) +$$
$$+ \frac{1}{4}\frac{[\mathbb{E}(T_i)-\mu_i]^6}{[\mathbb{V}(T_i)]^4}\left[\mathbb{M}_4(T_i)+\mathbb{V}(T_i)T_s^2+\frac{T_s^4}{15}-\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right)^2\right] \tag{3.45a}$$

$$\Upsilon(T_i) = \left(\frac{\mathbb{E}(T_i)-\mu_i}{\mathbb{V}(T_i)}\right)^2\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right) - \left(\frac{\mathbb{E}(T_i)-\mu_i}{\mathbb{V}(T_i)}\right)^3\mathbb{M}_3(T_i) +$$
$$+ \frac{1}{4}\left(\frac{\mathbb{E}(T_i)-\mu_i}{\mathbb{V}(T_i)}\right)^4\left[\mathbb{M}_4(T_i)+\mathbb{V}(T_i)T_s^2+\frac{T_s^4}{15}-\left(\mathbb{V}(T_i)+\frac{T_s^2}{6}\right)^2\right] \tag{3.45b}$$

2. Evaluate the expression for the estimation error – i.e., (3.19) for the mean, (3.20) for the variance, (3.27) for the duty cycle, and (3.38) for the distribution – based on most recent the sample estimates of the moments obtained in previous step (instead of the true moments as shown in such expressions) and the current sample size (number of observed periods).

3. If the value obtained in previous step is lower than the desired estimation error then stop, otherwise continue.

An example of how this algorithm would be implemented in a real system for the estimation of the mean period is shown in Algorithm 1 (the algorithm would need to be adapted to other PU activity statistics according to the different obtained analytical results but the operation principle would be the same). The key idea is that the theoretical error expressions are calculated by replacing the true moments (unknown to the DSA/CR system) with their last (known) sample estimates. In the shown example, (3.19) is evaluated by replacing $\mathbb{E}(T_i)$ and $\mathbb{V}(T_i)$ with their corresponding sample estimates $\widehat{m}_i$ and $\widetilde{v}_i$, respectively. It can be verified that the estimation error calculated in Step 2 decreases as the number of observed periods used to calculate the sample estimates increases. When such value is lower that the target/desired estimation error, the algorithm will indicate that the number of observed periods is sufficient to estimate the PU activity statistic of interest with the desired level of accuracy. The SU receiver can then use the set of observed periods to produce an accurate estimation, after which the value of N can be reset to zero in order to start capturing a new set of samples for the next estimation (notice that when the proposed algorithm is executed in real-time the value of N at which it will stop may not necessarily be identical in every execution, even though it will be similar). Notice that all the required input information would be known in a real implementation (desired estimation accuracy in terms of the target error $\varepsilon_{r,\max}^{\widehat{m}_i}$ and confidence interval ρ along with the employed sensing period T_s). Therefore, this iterative stopping algorithm enables the practical implementation of the theoretical results obtained in this work in a practical context.

3.8 Simulation and Experimental Results

In this section, a comprehensive analysis of the obtained analytical results as well as their validation with simulation and experimental results are presented. Simulations were performed in MATLAB by generating several sequences with a sufficiently large number of interleaved on/busy and off/idle periods from a GP distribution with predefined location (μ_i), scale (λ_i) and shape (α_i) parameters. The generated periods T_i were sensed with a specified sensing period T_s in order to calculate the corresponding sequence of estimated periods $\widehat{T}_i$ that would be observed by a DSA/CR receiver following the principle shown in Fig. 2.1. The set of observed periods, $\{\widehat{T}_{i,n}\}_{n=1}^{N}$, was used to estimate the primary activity statistics as described in previous sections and compared to the original true statistics as a function of the observation sample size

Algorithm 1: Iterative stopping algorithm (for the mean period)

Input : $\varepsilon_{r,\max}^{\widehat{m}_i}$, ρ, T_s
Output: N

continue = `true`, $N = 1$
while *continue* == *true* **do**
  Sense PU channel until a set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ is available
  Calculate $\widehat{m}_i = \frac{1}{N}\sum_{n=1}^{N}\widehat{T}_{i,n}$ from (3.10)
  Calculate $\widetilde{v}_i = \left[\frac{1}{N-1}\sum_{n=1}^{N}\left(\widehat{T}_{i,n} - \widehat{m}_i\right)^2\right] - \frac{T_s^2}{6}$ from (3.14)
  Calculate $\varepsilon_{r,\text{new}}^{\widehat{m}_i} = \frac{\sqrt{2}\,\text{erf}^{-1}(\rho)}{\widehat{m}_i}\left[\frac{1}{N}\left(\widetilde{v}_i + \frac{T_s^2}{6}\right)\right]^{\frac{1}{2}}$ from (3.19)
  if $\varepsilon_{r,\text{new}}^{\widehat{m}_i} \leq \varepsilon_{r,\max}^{\widehat{m}_i}$ **then**
    continue = `false`
  else
    $N = N + 1$
  end
end

N. Hardware experiments were conducted using the same PECAS platform described in Section 2.4.4.The value considered for each parameter is shown in the title of each figure in terms of generic time units (t.u.). In the case of experimental results, where a particular time unit needs to be selected according to the real-time capabilities of the employed hardware platform, the reference unit is the second (i.e., 1 t.u. = 1 second).

Fig. 3.1 shows the required observation sample size for the estimation of the minimum period as a function of the sensing period based on (3.9). As it can be observed, the required sample size increases with the desired probability of observation $P_{\text{obs}}^{\widehat{\mu}_i}$. It is worth noting that local minima are observed for T_s values that are integer submultiples of the true minimum (i.e., for $T_s = \mu_i/k$ with $k \in \mathbb{N}^+$). This can be explained by the fact that for such values of T_s it is possible to provide an exact estimation of the true minimum [81]. However for slightly higher values the required sample size tends to infinity, since in such a case $\chi_0 \approx 0$, $P\left(\widehat{T}_i = \widehat{\mu}_i\right) \approx 0$, and the denominator of (3.9) tends to zero as well. It is worth noting that the analytical result in (3.9) provides a perfect match with both simulation and experimental results.

Fig. 3.2 shows the maximum relative error of the estimated mean, $\varepsilon_{r,\max}^{\widehat{m}_i}$, observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size when the channel duty cycle is $\Psi = 0.5$ (i.e., idle and busy periods have the same average duration). As it can be appreciated, when κ is determined based on concentration inequalities as the ones shown in Table 3.1 the result in (3.19) represents a loose upper bound to the true relative error. On the other hand, if κ is calculated assuming that the sample mean estimates are normally distributed as shown in (3.21), then the result in (3.19) represents a very accurate expression for the relative error of the estimated mean (as corroborated by both simulation and experimental results), which can then be used to precisely determine the observation sample size required for an accurate estimation as indicated in (3.22). The

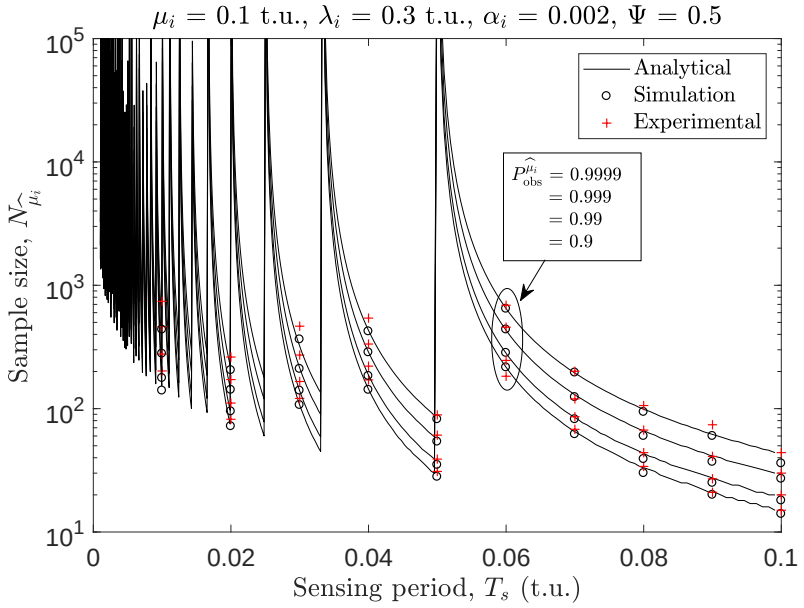


FIGURE 3.1: Required observation sample size for the estimation of the minimum period as a function of the sensing period (duty cycle $\Psi = 0.5$).

counterpart results for the estimated variance are shown in Fig. 3.3. In this case, the analytical result in (3.20) does not follow the simulation and experimental results when the observation sample is low, even if the normal approximation is considered for the calculation of κ. This can be explained by the fact that the assumption of normally distributed values of the sample variance (based on the central limit theorem) considered in (3.21) is valid only for a sufficiently large number of samples. As a result, if the sample size is not sufficiently large (for the particular evaluation conditions considered in the example of Fig. 3.3 this corresponds approximately to $N < 2000$) then the expression in (3.20) differs slightly from the true relative error. However, for a sufficiently large observation sample size (approximately $N > 2000$ in the example of Fig. 3.3) the result in (3.20) provides a very tight approximation for the true relative error, as shown by the perfect agreement with simulation and experimental results in this region of the figure. Since an accurate estimation of primary activity statistics will in general require a large sample size, the result in (3.20) is in practice accurate where it needs to be, and the observation sample size required for an accurate estimation of the variance can therefore be determined precisely based on (3.23).

Fig. 3.4 shows the maximum relative error of the estimated duty cycle observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size when the channel duty cycle is $\Psi = 0.5$. Since the duty cycle is calculated based on the estimated mean value of idle and busy periods as shown in (3.25), the observed results and the accuracy of (3.27) and (3.28) can be explained based on the same arguments as those for Fig. 3.2.

Fig. 3.5 shows the maximum KS distance of the estimated distribution at the 95% percentile ($\rho = 0.95$) as a function of the sample size when the channel duty cycle is $\Psi = 0.5$. The normal approximation for the calculation of κ is depicted when the value

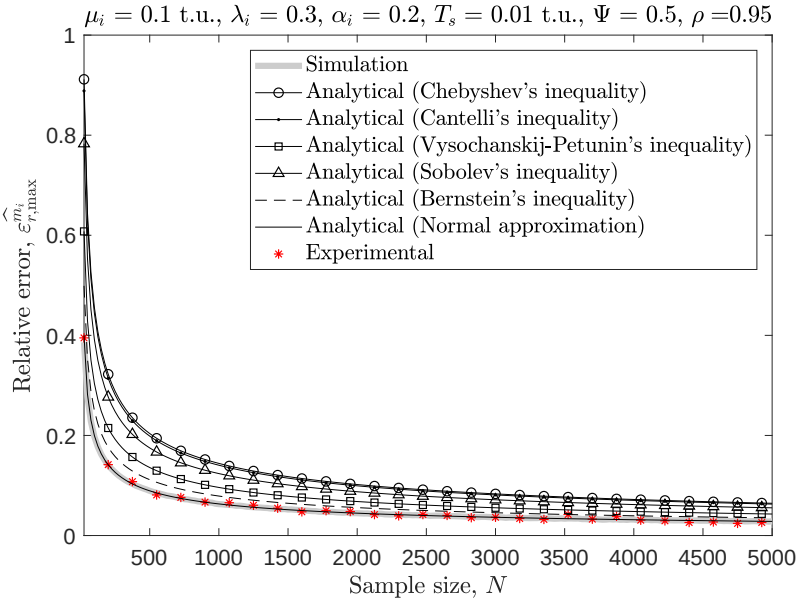


FIGURE 3.2: Maximum relative error of the estimated mean observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

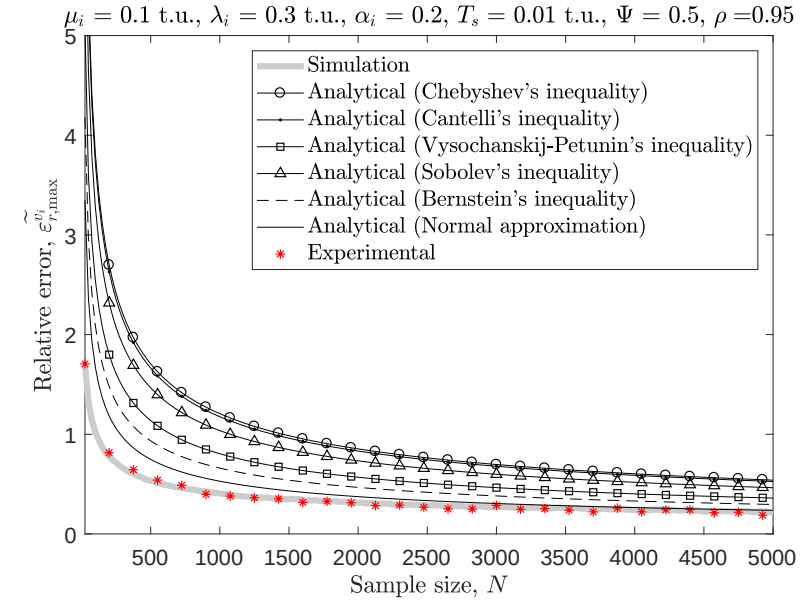


FIGURE 3.3: Maximum relative error of the estimated variance observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

of T that maximises (3.35) is calculated by numerical evaluation of (3.36) (numeric optimum) and analytically based on the approximation considered in (3.37) (analytic optimum). As it can be appreciated, the approximation considered in (3.37) leads to very accurate results. The results observed in Fig. 3.5 show that the analytical result in (3.38) provides a very accurate evaluation of the KS distance for sufficiently large sample sizes (approximately $N > 2000$), similar to the trend observed in Fig. 3.3. Following the same argument, it can be stated that the analytical result in (3.43) provides in practice

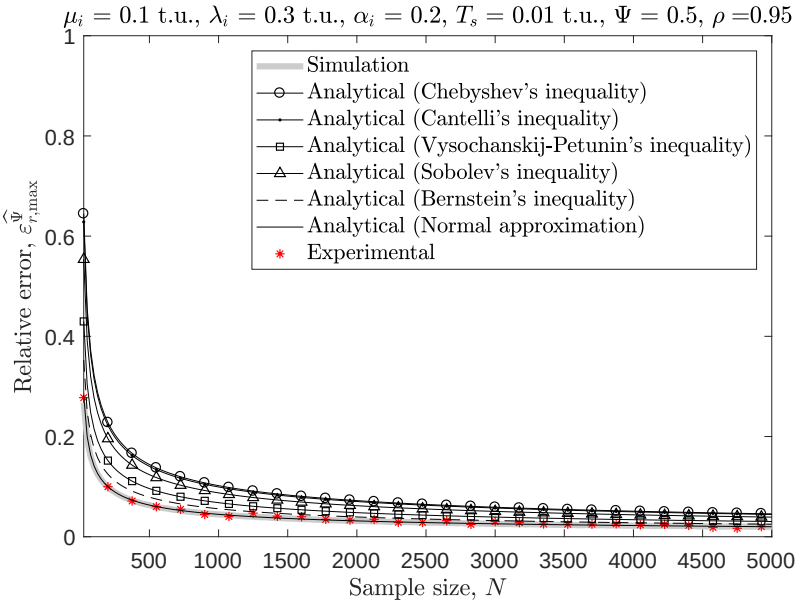


FIGURE 3.4: Maximum relative error of the estimated duty cycle at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

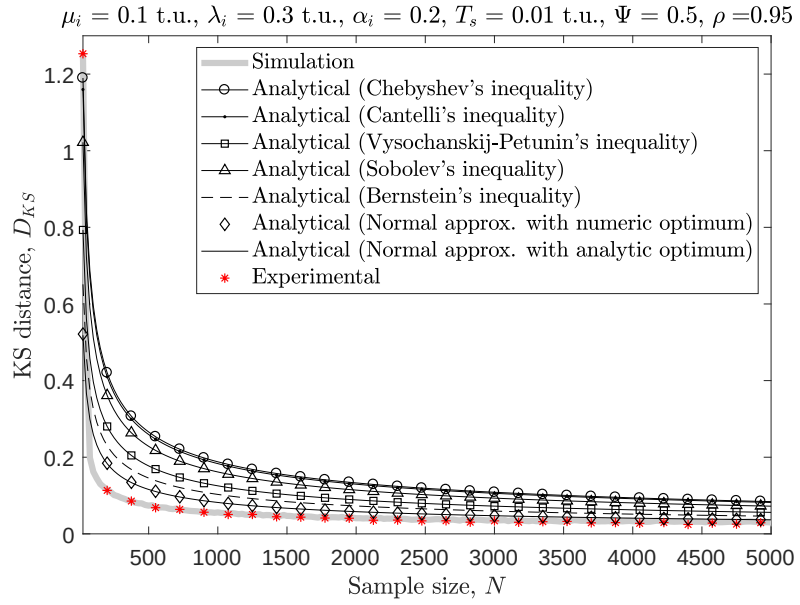


FIGURE 3.5: KS distance of the estimated distribution at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

a close prediction of the observation sample size required for an accurate estimation of the distribution.

Fig. 3.6 compares the analytical predictions of the required sample size as a function of the desired estimation error for the mean (3.22), variance (3.23), duty cycle (3.28) and distribution (3.43) with the estimations provided by the algorithm of Section 3.7 (based on simulations and experiments). As appreciated, the execution of the algorithm

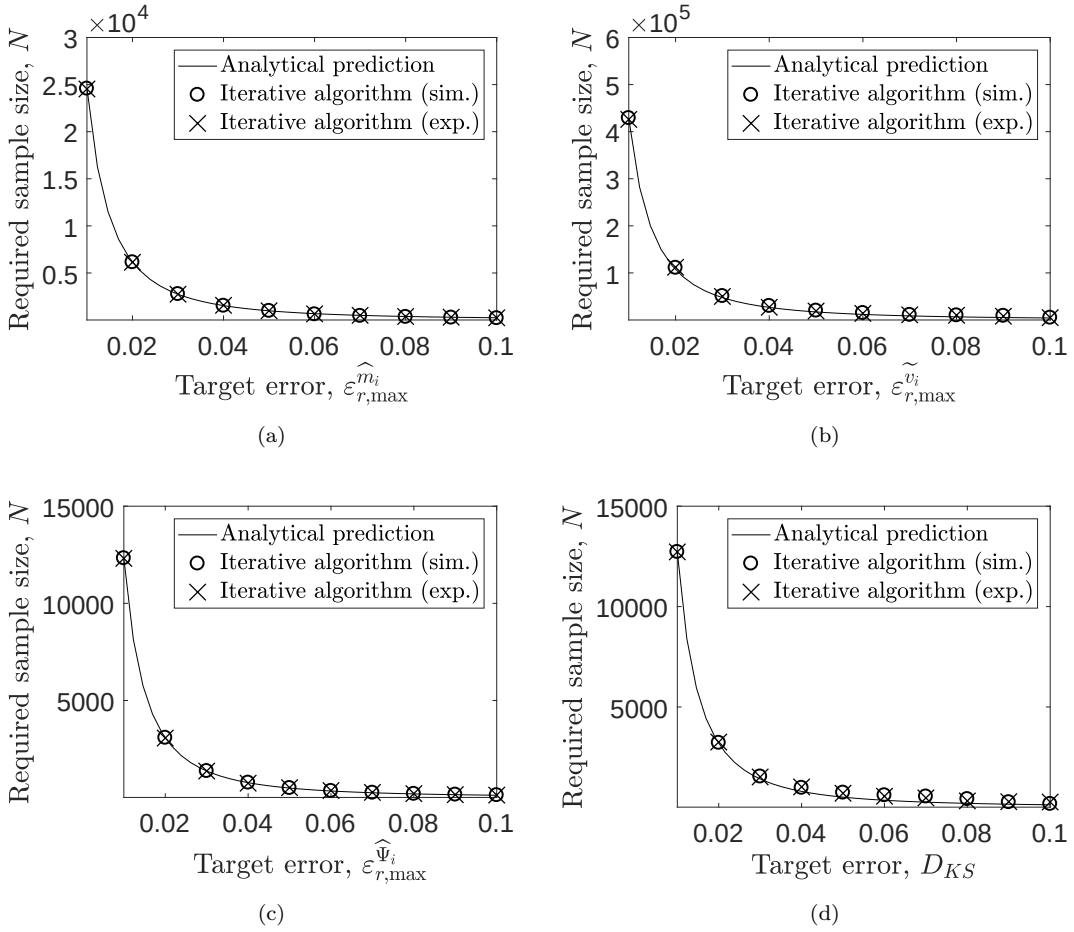


FIGURE 3.6: Required sample size as a function of the desired estimation error for: (a) mean, (b) variance, (c) channel duty cycle, and (d) distribution ($\mu_i = 0.1$ t.u., $\lambda_i = 0.3$ t.u., $\alpha_i = 0.05$, $T_s = 0.01$ t.u., $\Psi = 0.5$, $\rho = 0.95$).

terminates at values of N very close to the analytical predictions. Therefore the proposed algorithm can be used by real SU devices to determine in real-time how many samples (observed periods) are required to accurately estimate the activity statistics of an unknown PU channel.

3.9 Discussion of practical aspects

This section discusses several important aspects related with the practical application of the analytical results obtained in this work in a real context.

An important practical aspect is the degree to which the accuracy of the estimated primary activity statistics can affect the performance of DSA/CR systems. To illustrate this, let us consider as a practical example the problem of channel selection, which is one of the cases where primary activity statistics can be useful. A simple channel selection approach is to select the channel that provides the highest expected opportunistic bit-rate ($\widehat{R}_b$) which can be expressed as a function of the estimated duty cycle ($\widehat{\Psi}$) as $\widehat{R}_b = (1 - \widehat{\Psi})W_b\eta$, where W_b is the primary channel bandwidth and η is the spectrum

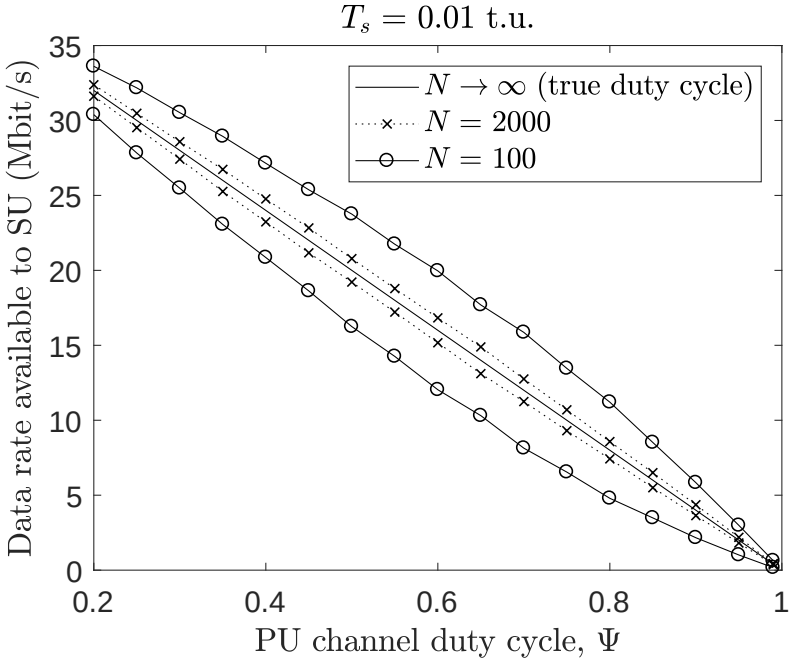


FIGURE 3.7: Estimated available data rate as a function of the estimated duty cycle.

efficiency associated to the modulation and coding schemes used by the DSA/CR system. Fig. 3.7 shows the estimated available bit-rate (with $W_b = 20$ MHz and $\eta = 2$ bit/s/Hz) as a function of the estimated channel duty cycle for three cases: $N = 100$ (insufficient sample size), $N = 2000$ (sufficiently large sample size) and $N \to \infty$ (infinite sample size and therefore perfect duty cycle estimation). The lines shown represent the worst-case upper and lower bounds corresponding to $\widehat{\Psi} = \Psi\,(1 \pm \varepsilon_{r,\max}^{\widehat{\Psi}})$, with $\varepsilon_{r,\max}^{\widehat{\Psi}}$ given by (3.27). As it can be appreciated in this example, an insufficient sample size ($N = 100$) can lead to estimation errors of up to ± 5 Mbit/s in the expected data rate, which for $\Psi = 0.5$ (where the true available data rate is 20 Mbit/s) represents an error of 25%. On the other hand, with a sufficiently large sample size ($N = 2000$), the estimation error is less than 1 Mbit/s (5% error for $\Psi = 0.5$), which can be made arbitrarily low by further increasing the sample size. This simple numerical example illustrates the potential impact that inaccurate primary activity statistics can have in the performance of DSA/CR systems and highlights the practical importance of the analytical results obtained in this work to determine the sample size required for an accurate estimation of the primary activity statistics.

Another aspect of practical interest is for how long (in absolute time units) the DSA/CR system needs to observe a PU channel before an accurate estimation of the PU statistics is available, and how this compares to the operation time scale of the DSA/CR system. All mathematical results for the considered PU activity statistics are provided in terms of the number of periods that need to be observed (N), which can be easily tracked by the DSA/CR system in a practical implementation (increasing a counter every time a new PU period is observed until the required sample size is reached). The total observation time in absolute time units, if it needs to be known, can

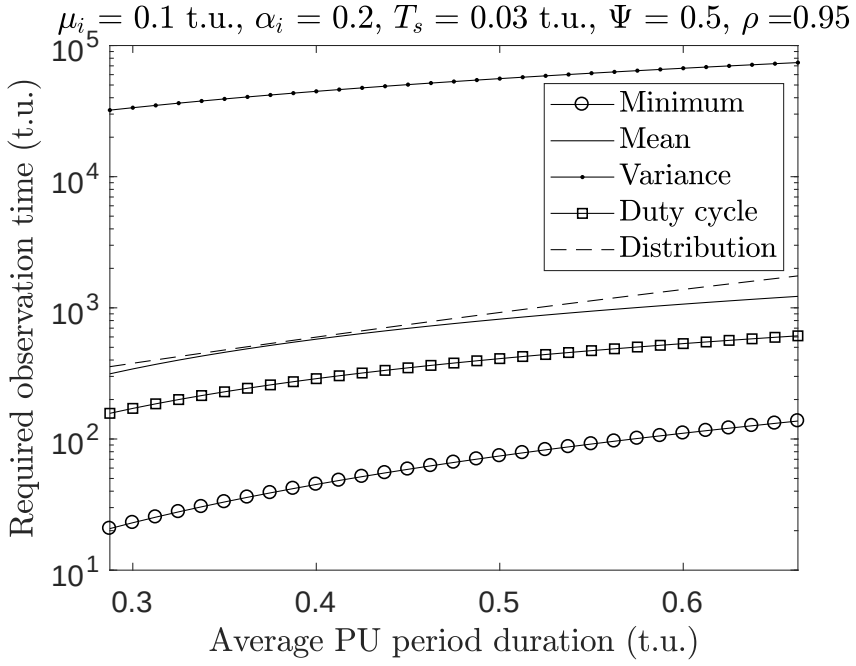


FIGURE 3.8: Required observation time for different PU activity statistics as a function of the average PU period duration (target estimation error of 0.05).

be readily obtained as $N \cdot \mathbb{E}(T_i)$, where $\mathbb{E}(T_i)$ is the average PU period duration (which would also need to be estimated as described in Section 3.4). Notice that the required observation time for a particular PU channel depends on the PU channel itself and its specific PU occupancy, quantified through $\mathbb{E}(T_i)$. This is illustrated in Fig. 3.8, which shows the required observation time for different PU activity statistics as a function of the average PU period duration. In most practical cases, the required observation time can be expected to be greater than the typical operation time scale of the DSA/CR network as a result of the need to observe hundreds/thousands of periods for an accurate estimation (as illustrated in the results obtained in Section 3.8). This is compatible with the notion that the estimated PU activity statistics will normally be exploited in the long term and once they are estimated for the first time they only need to be updated sporadically, which can be done while the DSA/CR system is in normal operation.

It is also worth mentioning that an accurate estimation of the PU activity statistics requires a careful consideration not only of the sample size N, which is the aspect of interest investigated in this work, but also the sensing period T_s. In the estimation of the moments and related metrics (mean, variance, duty cycle) the impact of the sensing period can be removed by introducing appropriate correction factors as discussed in Section 3.4. In the estimation of the distribution the problem is more challenging since the accuracy of the estimated distribution depends directly on the accuracy of the estimated minimum period (as shown in detail in [22]), which in turn can only be estimated accurately if the sensing period is an integer submultiple of the true (and often unknown) minimum period. To enable a fair evaluation of the individual impact of the sample size, which is the aspect of interest in this work, on the accuracy of the

estimated statistics (in particular, the distribution), the sensing period has been selected as an integer submultiple of the true minimum period. In a practical implementation it may be necessary to use more sophisticated methods for an accurate estimation of the minimum period such as those proposed in [120].

3.10 Summary

DSA/CR systems can benefit from the knowledge of the primary traffic statistics, which can be exploited by DSA/CR users to achieve a more efficient utilisation of the free primary spectrum. This information can be obtained from spectrum sensing by estimating the duration of individual idle/busy periods (from the sequence of spectrum sensing decisions) and then processing a sufficiently large number of observed periods (referred to in this chapter as the observation sample size) to calculate relevant statistics on the primary activity such as the minimum period duration, the mean and variance of the observed periods, the channel duty cycle or the underlying distribution. An important practical question is how many periods need to be observed in order to guarantee that the estimated statistics will meet a predefined level of accuracy. In this context, this chapter has performed a detailed mathematical analysis on the observation sample size required for an accurate estimation of each of the above mentioned primary activity statistics, and provided closed-form expressions for the estimation error as a function of the sample size as well as the required sample size as a function of the desired estimation error. The obtained analytical results have been compared to both simulation and experimental results, showing an excellent agreement in all cases. The expressions provided in this chapter can be used in practical DSA/CR systems to guarantee that PU activity statistics are estimated to the desired level of accuracy. Moreover, an iterative stopping algorithm has been proposed to enable SU perform a real-time calculation of the required sample size, which has been shown to be very accurate.

Appendix: Correction of Sample Moments

As discussed in Section 3.4, the sample mean needs no correction and the sample variance needs to be corrected as shown in (3.14). A similar analysis can be carried out for the third and fourth central moments.

For the third central moment of the estimated periods:

$$\begin{aligned}\mathbb{M}_3(\widehat{T}_i) &= \mathbb{E}([\widehat{T}_i - \mathbb{E}(\widehat{T}_i)]^3) = \mathbb{E}([T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e)]^3) \\ &= \mathbb{E}([T_i - \mathbb{E}(T_i)]^3) + \mathbb{E}([T_e - \mathbb{E}(T_e)]^3) \\ &= \mathbb{M}_3(T_i) + \mathbb{M}_3(T_e) \end{aligned} \tag{3.46}$$

Thus the third sample central moment needs no correction since $\mathbb{M}_3(T_e) = 0$ for a symmetric triangular distribution.

For the fourth central moment of the estimated periods:

$$\begin{aligned}\mathbb{M}_4(\widehat{T}_i) &= \mathbb{E}([\widehat{T}_i - \mathbb{E}(\widehat{T}_i)]^4) = \mathbb{E}([T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e)]^4) \\ &= \mathbb{E}([T_i - \mathbb{E}(T_i)]^4) + 6\,\mathbb{E}([T_i - \mathbb{E}(T_i)]^2) \cdot \\ &\quad \cdot \mathbb{E}([T_e - \mathbb{E}(T_e)]^2) + \mathbb{E}([T_e - \mathbb{E}(T_e)]^4) \\ &= \mathbb{M}_4(T_i) + 6\,\mathbb{V}(T_i)\mathbb{V}(T_e) + \mathbb{M}_4(T_e) \end{aligned} \tag{3.47}$$

where $\mathbb{V}(T_e) = T_s^2/6$ and $\mathbb{M}_4(T_e) = T_s^4/15$ for a symmetric triangular distribution. Solving (3.47) for $\mathbb{M}_4(T_i)$ it can be shown that a sample estimate $\widehat{c}_{4,i}$ of the fourth central moment needs to be corrected as $\widetilde{c}_{4,i} = \widehat{c}_{4,i} - \widetilde{v}_i T_s^2 - T_s^4/15$.

Chapter 4

Estimation of Primary Activity Statistics under Imperfect Spectrum Sensing

4.1 Introduction

CR systems can benefit from accurate the knowledge of the PU statistics to increase spectrum efficiency as discussed in previous chapters. CRs utilise spectrum sensing techniques to estimate the PU statistical information. Unfortunately, spectrum sensing is imperfect resulting in an inaccurate estimation of PU statistics, in particular under low SNR conditions when sensing errors may occur. Chapter 3 considered the problem of studying and overcoming the impact of finite spectrum sensing period and proposed methods to improve the estimation of PU statistics. On the other hand, Chapter 4, was concerned with the effect of a limited number of period observations and the resulting PU statistics. Another source of imperfection for spectrum sensing is sensing errors. Sensing errors have a significant impact on the performance of communication systems (both PU and SU). In the case of a false alarm, the SU transmitter will be silent when it should be transmitting, leading to a low CR transmission rate. In the case of a missed detection, interference with the PU signal will occur. Inaccurate detection leads to inaccurate estimation for PU activity statistics as well.

Multiple methods in the literature have been proposed to increase the estimation accuracy of PU activity statistics and reduce the effect of imperfect spectrum sensing (ISS) such as [10, 80, 81, 121]. In this chapter sensing errors and their effects on the estimated PU distribution are investigated. New algorithms to correctly estimate the imperfectly sensed periods are proposed. The methods in [81] are the closest ones to the proposed methods and the obtained results will be compared to them. Computational cost/complexity analysis for all the considered methods is provided as well.

Moreover, the PU activity distribution is affected by multiple factors including for example the time when the PU channels are observed (e.g., patterns may be different

during day and night times [122]), the location (indoor/outdoor and rural/urban areas), application type (e.g., voice data is modeled differently from regular internet data) [123] and sensing duration [64]. This results in multiple characterisations for PU statistics of idle/busy periods. Results have shown that the generalised Pareto distribution provides the most accurate fit for low time resolution observations (i.e., long sensing periods) while for high time resolution (i.e., short sensing periods) the best distribution fit varies depending on the PU radio technology [8]. Even though in the literature multiple methods to improve the estimation of PU statistics have been proposed, their accuracy for different PU distributions has not been assessed (typically exponential distribution is considered to model busy/idle periods or the study is limited to a single PU distribution only). The accuracy of estimation for best performing methods is tested under different PU distributions. The mathematical analysis of the low SNR scenario with sensing errors requires a significantly more complex study and it is feasibility is unclear, thus left as future work.

This chapter considers the problem of accurate estimation of the primary activity distribution in the presence of spectrum sensing errors (false alarms and missed detections). This problem is different from the detection problem. The probability of false alarms and missed detections are performance metrics for spectrum sensing. Using different detection methods is irrelevant to this work as it only changes the values of false alarms and missed detections and would not affect the functionality of the proposed methods.

The main contributions of this chapter are outlined as follows:

1. Three novel algorithms are proposed to palliate the effects of sensing errors on the estimation of the primary traffic distribution.

2. The impact of different PU occupancy models is investigated under both PSS and ISS. The aim is to determine whether algorithms that have been designed and evaluated for a particular PU distribution can still achieve a high accuracy in statistics prediction when the PU pattern follows a different distribution.

4.2 System Model

A CR senses the channel at a finite rate with duration T_s as shown in Fig. 4.1. In every sensing event (T_s), a binary decision is made to result in either idle (H_0) or busy (H_1) state of the channel. A perfect spectrum sensing (PSS) performance is obtained when a CR receiver operates in a high PU transmission power. However under low SNR imperfect spectrum sensing (ISS) may occur with two types of errors: false alarm ($\breve{H} = H_1|H_0$) where the signal is not present but announced as present because of the high noise level, and missed detection ($\breve{H} = H_0|H_1$) where the signal is present but with power lower than the energy detection threshold.

There are multiple factors that affect the estimation of the PU activity pattern, including the probability of false alarm (P_{fa}), the probability of missed detection (P_{md})

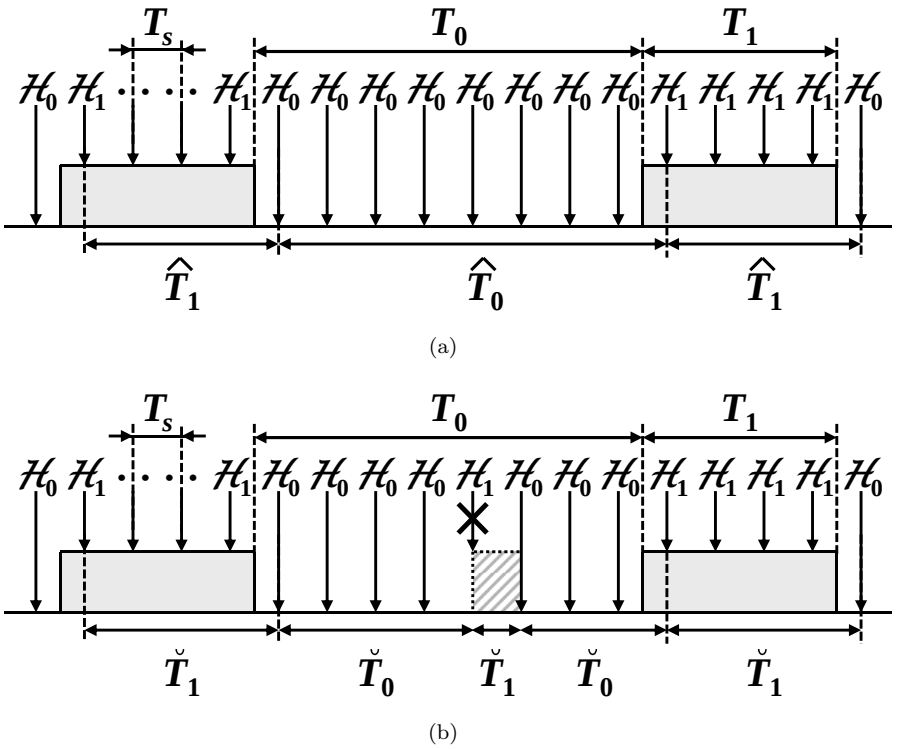


FIGURE 4.1: Estimation of period durations from spectrum sensing decisions: (a) under perfect spectrum sensing, (b) under imperfect spectrum sensing [81]. $\widehat{T}_i$ represents the estimated period under perfect spectrum sensing and $\breve{T}_i$ represents the estimated period under imperfect spectrum sensing.

and how the errors are distributed. Note that the sensing errors will affect the lengths of periods depending on their location as it can be appreciated from Fig. 4.2 where three types of errors can be identified: missed detection, false alarm and originally correct periods but with missed detection/false alarm that occurs at its beginning, middle or end, which results in making its length shorter than the minimum PU activity time μ_i and therefore detectable if the value of μ_i is known. According to this, multiple errors could occur in a long period dividing it into multiple short periods which would have a significant effect on the calculation of PU statistics. The objective of this chapter is to develop appropriate methods to overcome the impact of these errors on the estimated PU statistics.

4.3 Methods Proposed to Overcome the Effect of Spectrum Sensing Errors on the Estimated Statistics

As described above, in some cases CR requires to work at low SNR (close to the noise floor) and this will introduce sensing errors. The presence of errors makes the calculation of PU activity statistics inaccurate. In order to improve the estimated statistics, three methods aimed at reconstructing the original periods are presented. These methods require some knowledge of the PU signal. In this chapter, it is assumed that the PU

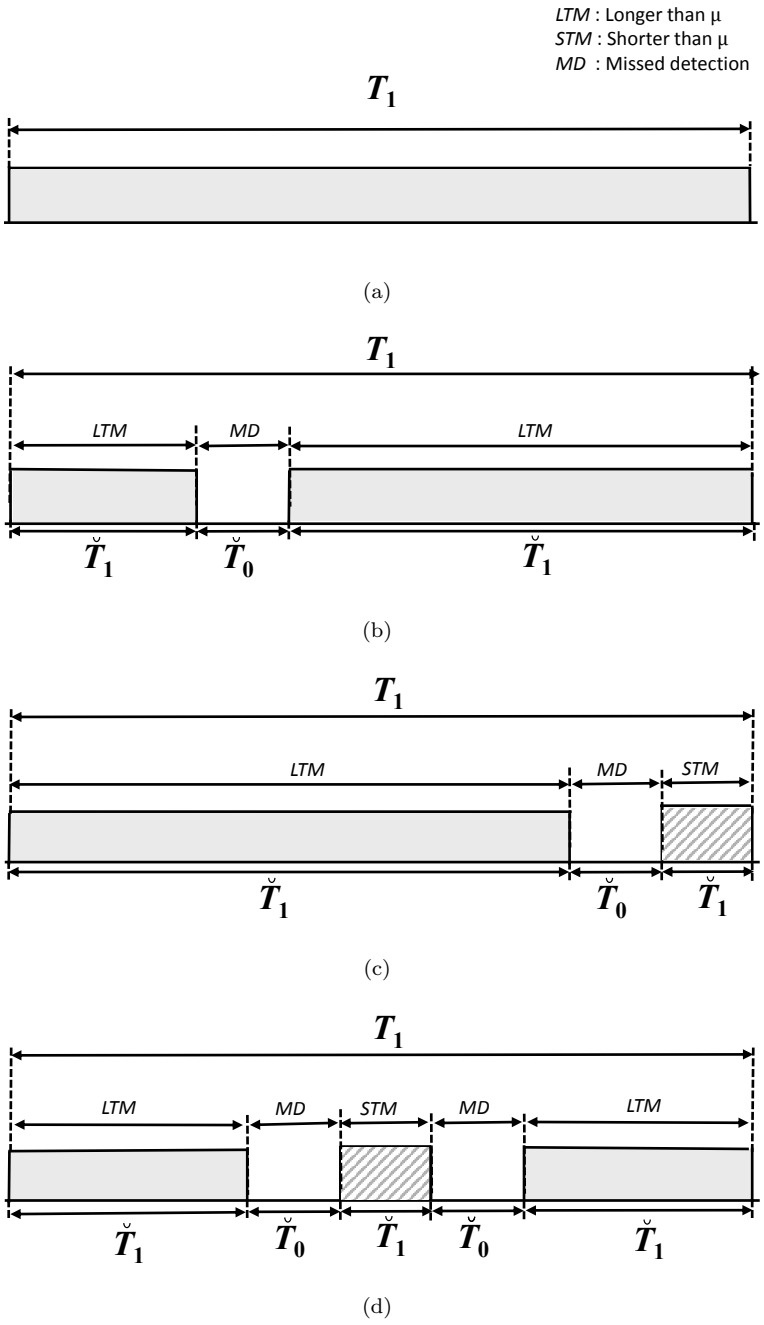


FIGURE 4.2: Sensing errors according to their location: (a) Original period, (b) Single detectable incorrect period, (c) Two detectable incorrect periods, (d) Multiple detectable incorrect periods.

minimum activity time μ_i is known at the CR receiver. There are three methods to obtain the minimum activity duration:

1. Regional beacon signals with real-time information about minimum activity time for PU present in the geographical area. The main drawback of this method is the requirement to modify the primary network [98].

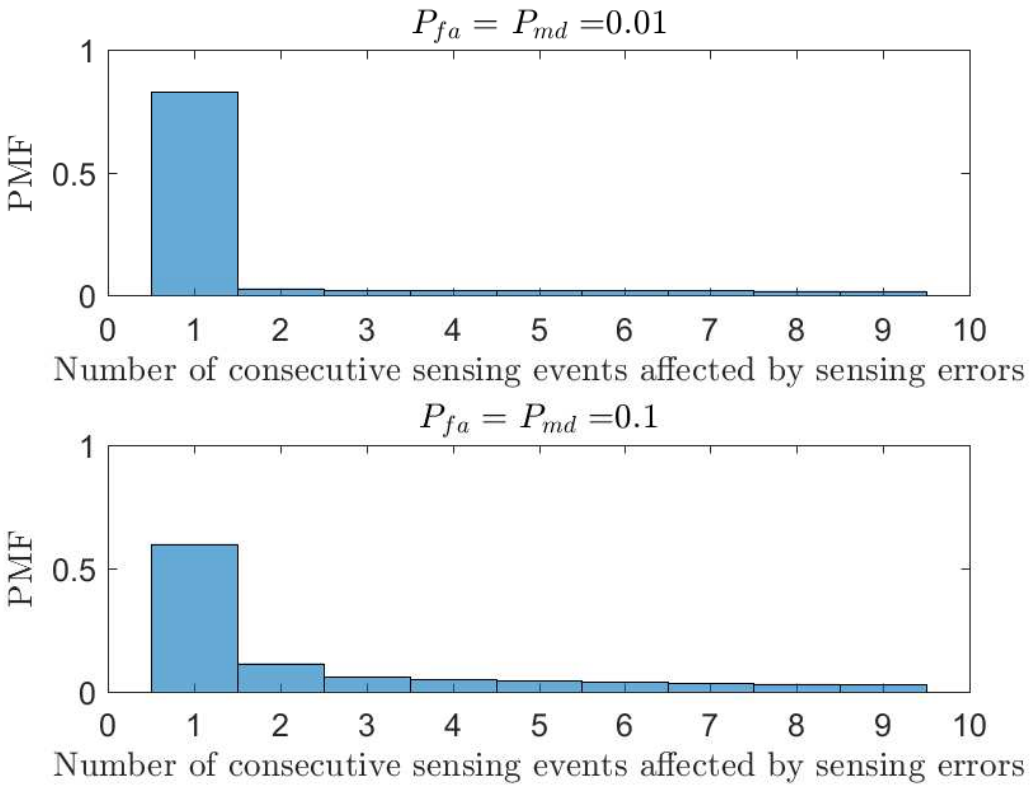


FIGURE 4.3: PMF of the number of consecutive sensing events affected by sensing errors (duty cycle (Ψ) = 0.8, μ_i = 10 time unit (t.u.), $E\{T_i\}$ = 50 t.u.).

2. Offline method as in [97], where the SU utilize blind estimation methods in order to determine the minimum PU activity time.

3. Offline methods with prior knowledge on the operating standards. These methods can be applied in slotted systems where the frame duration is known (e.g., GSM [124]).

Based on the knowledge of μ_i, three novel algorithms are proposed to palliate the effects of spectrum sensing errors.

4.3.1 Method 1

Fig. 4.3 shows the probability mass function (PMF) of the number of consecutive sensing events (T_s) affected by sensing errors (obtained from simulations). As it can be appreciated, in most cases sensing errors occur in individual isolated sensing events; only when the probability of error (P_{fa}/P_{md}) is higher (bottom of Fig. 4.3) bursts of two or more consecutive erroneous sensing events can occur. This observation can be exploited to identify the occurrence of sensing errors and reconstruct the original periods as follows. First, a threshold βT_s (with $\beta \in \mathbb{N}^+$) is defined, which can be tuned based on Fig. 4.3. Starting from an initial estimated period $\breve{T}_{i,n}$ which has a duration less than the threshold and the minimum ($\breve{T}_{i,n} < \beta T_s$ and $\breve{T}_{i,n} < \mu_i$), where $n \in \mathbb{N}^+$ represents the

Algorithm 2: Method 1

for *each* $\breve{T}_{i,n} < \mu_i$ *and* $\breve{T}_{i,n} < \beta T_s$ **do**
  W = 1
  while $\breve{T}_{1-i,n+W} < \beta T_s$ **do**
    W=W+1
  end
  $\widetilde{T}_{i,n} = \breve{T}_{i,n} + \ldots + \breve{T}_{i,n+W-1}$
end
for *each* $\widetilde{T}_{i,n}$ **do**
  if $\widetilde{T}_{i,n} < \mu_i$ **then**
    discard $\widetilde{T}_{i,n}$
  else
    keep $\widetilde{T}_{i,n}$
  end
end

index in the sequence of observed periods, the durations of all subsequent periods (both busy and idle) are checked until a period of the opposite type with a duration greater than βT_s (i.e., $\breve{T}_{1-i,n+W} > \beta T_s$) is found. A reconstructed period of the same type (idle/busy) as $\breve{T}_{i,n}$ is then estimated by adding the durations $\widetilde{T}_{i,n} = \breve{T}_{i,n} + \ldots + \breve{T}_{i,n+W-1}$ where $\widetilde{T}_{i,n}$ is the new reconstructed period of the original. The period $\breve{T}_{1-i,n+W}$ (which is of different type than the previously reconstructed period) is then taken as the starting point for a new reconstruction based on the same principle. This process is repeated over the sequence of estimated periods so that all periods shorter than βT_s (which based on Fig. 4.3 can be assumed to be short periods resulting from sensing errors) will be added in an attempt to reconstruct the original sequence of busy/idle periods. After applying the method above, some of the reconstructed periods were observed to be shorter than μ_i, thus indicating the presence of a few incorrectly reconstructed periods. Therefore a second step is performed after the reconstruction process above where all the reconstructed periods $\widetilde{T}_{i,n}$ shorter than μ_i are discarded and not included in the computation of the PU activity statistics.

4.3.2 Method 2

In this method, whenever a period shorter than the minimum PU activity time is observed (i.e., $\breve{T}_{i,n} < \mu_i$), a window of K sensing events with a total duration KT_s ($K \in \mathbb{N}^+$) is defined centred around that period. The sensing events within the window that are observed in busy (idle) state are given a weight of $+1$ (-1) respectively. The weights are then added and the original period is reconstructed as follows:

- If the sign of the sum is different from the sign of the weight associated with $\breve{T}_{i,n}$, then the state (idle/busy) of $\breve{T}_{i,n}$ is reversed and a reconstructed period is estimated by adding the durations $\widetilde{T}_{i,n} = \breve{T}_{i,n-1} + \breve{T}_{i,n} + \breve{T}_{i,n+1}$.

Algorithm 3: Method 2

for *each* $\breve{T}_{i,n} < \mu_i$ **do**
  Define a window of size KT_s centered at $\breve{T}_{i,n}$
  Give weight of $+1$ to busy sensing events in window
  Give weight of -1 to idle sensing events in window
  Calculate the sum S of the weights in window
  if *sign*(S) == *sign*(*weight of* $\breve{T}_{i,n}$) **then**
    $\widetilde{T}_{i,n} = \breve{T}_{i,n} + \breve{T}_{i,n+1}$
  else
    $\widetilde{T}_{i,n} = \breve{T}_{i,n-1} + \breve{T}_{i,n} + \breve{T}_{i,n+1}$
  end
end

- If the sign of the sum is the same as the sign of the weight associated with $\breve{T}_{i,n}$, then the state of $\breve{T}_{i,n+1}$ is reversed and added to its preceding period $\breve{T}_{i,n}$ to produce a reconstructed period of the same type (idle/busy) as $\breve{T}_{i,n}$ with duration $\widetilde{T}_{i,n} = \breve{T}_{i,n} + \breve{T}_{i,n+1}$.

Note that when a period shorter than μ_i is found, a majority rule is used to determine the most likely state of the channel around that period and determine how the original period should be reconstructed according to that most likely case. This is the main idea this method is based on.

4.3.3 Method 3

As observed in Fig. 4.2, the occurrence of multiple sensing errors can lead to a pattern where two periods that are longer than μ_i (and therefore can be assumed to be observed in their original idle/busy state) contain a number of other periods shorter than μ_i (which cannot be classified as either correct or incorrect idle/busy observations). This method aims at reconstructing the original period by adding the durations of all the periods $\breve{T}_{i,n} + \ldots + \breve{T}_{i,n+W}$ between two consecutive periods longer than or equal to the minimum PU transmission time (i.e, $\breve{T}_{i,n-1} \geq \mu_i$, $\breve{T}_{i,n+W+1} \geq \mu_i$ and $\breve{T}_{i,n}, \ldots, \breve{T}_{i,n+W} < \mu_i$) to the next or previous period depending on the period types.

If the starting period $\breve{T}_{i,n-1}$ and ending period $\breve{T}_{i,n+W+1}$ are of the same type (busy/idle), then the reconstructed period will be of that same type. However, if $\breve{T}_{i,n-1}$ and $\breve{T}_{i,n+W+1}$ are of different types, then it is not possible to determine unambiguously the original period type since it depends on the particular order in which sensing errors occurred and it is not possible to determine which of the periods shorter than μ_i are correct/incorrect idle/busy observations. In this other case, the period type is randomly decided as busy with probability $\widehat{\Psi}$ and idle with probability $1-\widehat{\Psi}$, where $\widehat{\Psi}$ is an estimation of the PU channel duty cycle Ψ obtained from past spectrum sensing observations. If the instantaneous average busy duration is $\mathbb{E}\left\{\widehat{T}_1\right\}_N$ and the average instantaneous

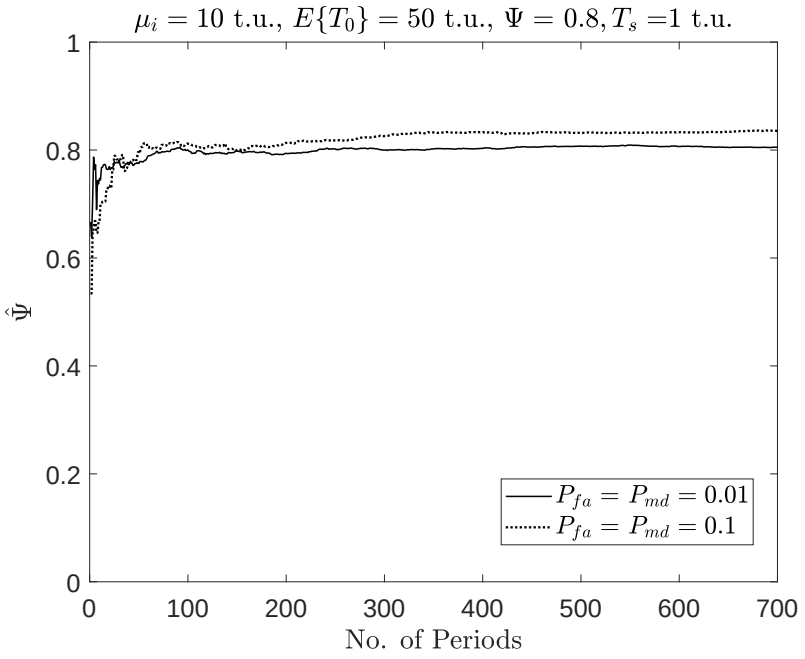


FIGURE 4.4: Estimated PU channel duty cycle $\widehat{\Psi}$ as a function of the number of periods used in the estimation ($\Psi = 0.8$, $\mu_i = 10$ t.u., $E\{T_i\} = 50$ t.u.).

idle time is $\mathbb{E}\left\{\widehat{T}_0\right\}_N$, then the instantaneous duty cycle ($\widetilde{\Psi}_N$) can be estimated as:

$$\widehat{\Psi}_N = \frac{\mathbb{E}\left\{\widehat{T}_1\right\}_N}{\mathbb{E}\left\{\widehat{T}_1\right\}_N + \mathbb{E}\left\{\widehat{T}_0\right\}_N}, \tag{4.1}$$

where N is the sample size. The instantaneous average period duration is found as $\mathbb{E}\left\{\widehat{T}_i\right\}_N = \frac{1}{N}\sum_{n=1}^{N}\left\{\widehat{T}_{i,n}\right\}$. Fig. 4.4 shows the estimated duty cycle $\widehat{\Psi}_N$ quickly converges to the real value Ψ, meaning that the algorithm needs a short transition time before it can operate correctly.

Method 3 is different from method 2 in terms how they aim to reconstruct the observed to the original. Method 2 is based on MAJORITY rule. This method centres a window of KT_s sensed events around the period that is shorter than the minimum $\breve{T}_{i,n}$. If the majority of periods are of the same type of $\breve{T}_{i,n}$ then $\breve{T}_{i,n+1}$ type is changed to the same type of $\breve{T}_{i,n}$. If the majority of periods are of different type of $\breve{T}_{i,n}$ then $\breve{T}_{i,n}$ type is changed to the same of the majority. While method 3 takes all short periods between two long periods. In case the two long periods are of the same type (i.e., busy-busy or idle-idle) then all short perids are selected to be of the same type of the two long periods. In case the two long periods are of different type (i.e., busy-idle or idle-busy) then based on the instantaneous duty cycle the short periods types are selected.

4.3.4 Other related Methods

The proposed three methods will be compared with the ones presented in [81], which are the closest ones to this research. The methods proposed in [81] are summarised below:

Algorithm 4: Method 3

for *each* $\breve{T}_{i,n} < \mu_i$ **do**
  W = 0
  while $\breve{T}_{i,n+W+1} < \mu_i$ **do**
    W=W+1
  end
  if *type of* $\breve{T}_{i,n-1}$ == *type of* $\breve{T}_{i,n+W+1}$ **then**
    $\widetilde{T}_{i,n-1} = \breve{T}_{i,n-1} + \breve{T}_{i,n} + \ldots + \breve{T}_{i,n+W} + \breve{T}_{i,n+W+1}$
  else
    Estimate duty cycle $\widehat{\Psi}$ based on equation (3.24)
    Generate uniform random number $\zeta \in (0, 1)$
    if $\zeta \leq \widehat{\Psi}$ **then**
      if $\widetilde{T}_{i,n-1}$ *busy and* $\widetilde{T}_{i,n+W+1}$ *idle* **then**
        $\widetilde{T}_{i,n-1} = \breve{T}_{i,n-1} + \ldots + \breve{T}_{i,n+W}$
      else
        $\widetilde{T}_{i,n} = \breve{T}_{i,n} + \ldots + \breve{T}_{i,n+W+1}$
      end
    else
      if $\widetilde{T}_{i,n-1}$ *idle and* $\widehat{T}_{i,n+W+1}$ *busy* **then**
        $\widetilde{T}_{i,n-1} = \breve{T}_{i,n-1} + \ldots + \breve{T}_{i,n+W}$
      else
        $\widetilde{T}_{i,n} = \breve{T}_{i,n} + \ldots + \breve{T}_{i,n+W+1}$
      end
    end
  end
end

- Method 4: Every observed period with a length shorter than the minimum period is discarded.
- Method 5: Every observed period with a length shorter than the minimum period is discarded along with the preceding and succeeding periods.
- Method 6: For every observed period with a length shorter than the minimum a reconstruction attempt is made by adding it with the preceding and succeeding periods.

4.4 Simulation Results

The performance of the six methods considered in this chapter is evaluated by means of simulations, which can be summarized as follows:

1. Generate idle/busy periods' lengths T_i following a generalized Pareto distribution, which has been proven to provide the best fit to empirical spectrum data [64].

2. Perform idle/busy sensing decisions H_0/H_1 on the generated sequence in step 1 every T_s time units (t.u.).

3. Add random errors (with $P_{fa} > 0$ and $P_{md} > 0$) in the sequence resulting from step 2.

4. Using the new H_0/H_1 sequence from step 3, calculate the period lengths $\breve{T}_i$ that would be estimated under ISS.

5. Process the sequence of period lengths resulting from step 4 in order to reconstruct the original periods by making use of one of the six methods considered in this chapter.

6. Compute the cumulative distribution function (CDF) of the idle/busy lengths obtained in steps 4 & 5, and compare with the CDF of the original lengths in step 1.

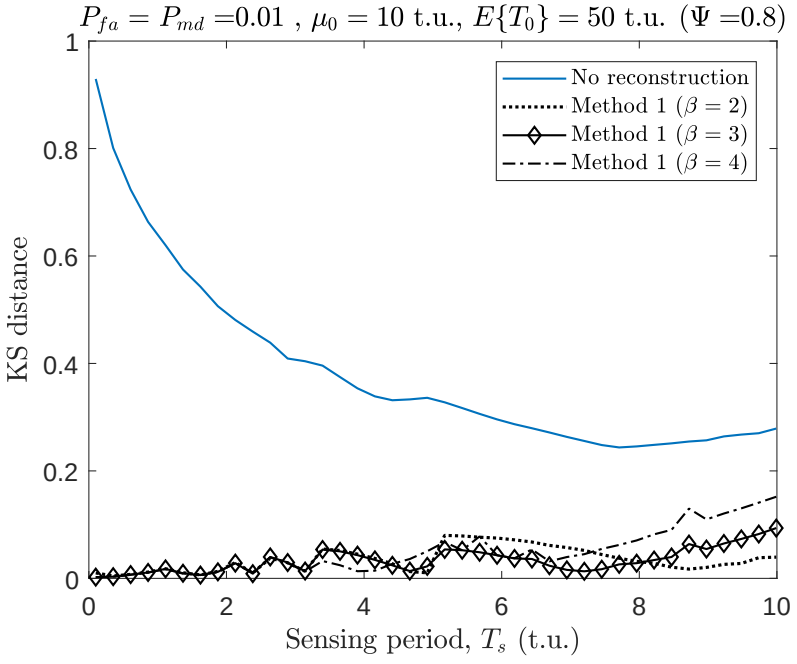
The comparison of step 6 quantifies the accuracy of the PU statistics estimated with and without the proposed methods and therefore determines whether accuracy improvements can be obtained with the proposed methods. The comparison between the estimated and original distributions is performed using the classic Kolmogorov-Smirnov (KS) distance [96].

$$D_{KS} = \sup_{T_i} \left| F_{T_i}(T_i) - F_{\breve{T}_i}(T_i) \right|. \tag{4.2}$$

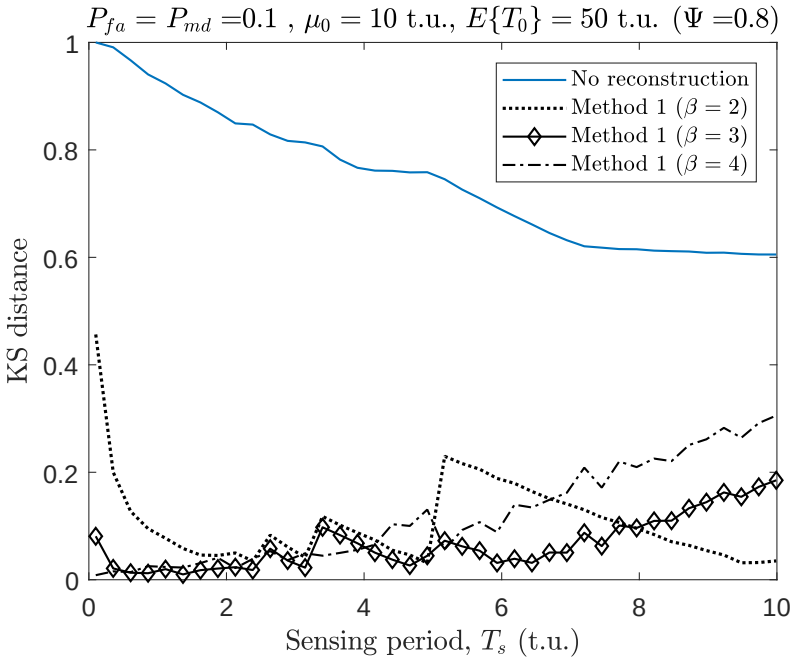
Note that the sequence of idle/busy lengths T_i generated in step 1 contains real positive values and its CDF $F_{T_i}(T_i)$ has a continuous domain, while the sequences resulting from steps 4 and 5 contain values $\breve{T}_i$ that are integer multiples of the sensing period ($\breve{T}_i = NT_s$, $N \in \mathbb{N}^+$) and their CDFs $F_{\breve{T}_i}(\breve{T}_i)$ have a discrete domain. Since it is not possible to compare continuous and discrete sets, the discrete set $F_{\breve{T}_i}(\breve{T}_i)$ is interpolated to produce $F_{\breve{T}_i}(T_i)$, which is compared in (4.2) with the distribution of the original lengths $F_{T_i}(T_i)$.

Fig. 4.5, Fig. 4.6, Fig. 4.7, show the accuracy of the estimated CDF of PU idle periods in terms of the KS distance (with respect to the original periods of step 1) as a function of the sensing period, when no reconstruction method is used (i.e., after step 4) and when a reconstruction method is used (i.e., after step 5). Results are shown for method 1 (Fig. 4.5), method 2 (Fig. 4.6) and method 3 (Fig. 4.7), for $P_{fa} = P_{md} = 0.01$ (top row) and $P_{fa} = P_{md} = 0.1$ (bottom row). Simulations were performed for duty cycle values of $\Psi = 0.8$ (heavy PU channel load).

As it can be appreciated from Fig. 4.5, Fig. 4.6, Fig. 4.7, the accuracy in all cases is noticeably better when the probability of error is lower (top row). However, all the methods proposed in this work can provide significant accuracy improvements with respect to the case where no reconstruction is performed (i.e., the PU statistics are computed based on the raw period lengths observed under ISS). In some particular cases the estimation is nearly perfect (i.e., $D_{KS} \approx 0$).

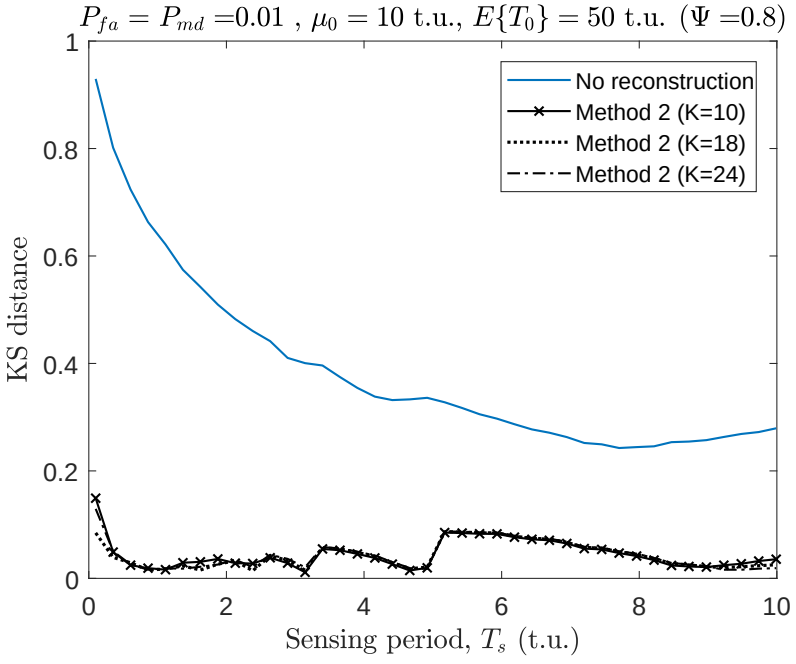


(a)

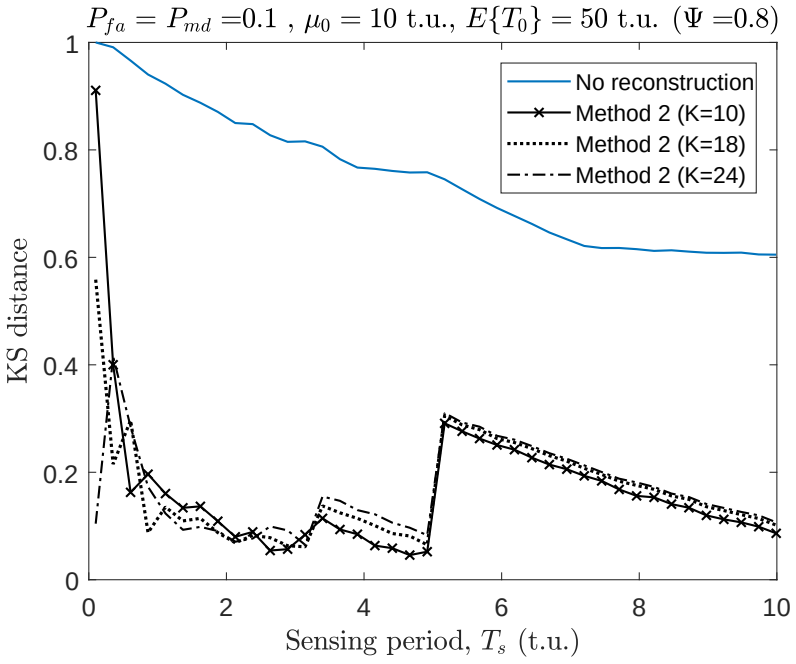


(b)

FIGURE 4.5: Performance of method 1.



(a)



(b)

FIGURE 4.6: Performance of method 2.

For method 1 (Fig. 4.5), there is a wide range of (low) values of the sensing period T_s that can provide a nearly perfect estimation ($D_{KS} \approx 0$). When the probability of sensing errors is lower ($P_{fa} = P_{md} = 0.01$), this is true for all the considered values of the parameter β. However when the probability of sensing errors is higher ($P_{fa} = P_{md} = 0.1$) this may not be true for low values of β (e.g., for the case $\beta = 2$ in Fig. 4.5(b)). This can be explained based on Fig. 4.3: when the probability of sensing errors is lower, most sensing errors occur in individual isolated sensing events and values of β as low as $\beta = 2$ are sufficient to correctly reconstruct the original period lengths; however when the probability of sensing errors is higher, there are some cases where two or a few more consecutive sensing events can be affected by sensing errors and the value of the threshold β needs to be increased (e.g., $\beta = 3$ or $\beta = 4$) in order to reconstruct the original period lengths. In any case, with an adequate configuration of the β parameter and the employed sensing period T_s, method 1 can provide a nearly perfect estimation of the PU statistics.

Method 2 (Fig. 4.6), can provide levels of accuracy comparable to those of method 1. However in this case the best accuracy seems to be attained when the sensing period is approximately half of the minimum PU transmission time (i.e., $T_s \approx \mu_i/2$). Moreover, the accuracy obtained in that region is very sensitive to the selection of the value of T_s (i.e., small variations of the value of T_s around the optimum point can result in a noticeable degradation of the estimation accuracy), which makes the configuration of this method more complex than method 1. As observed in Fig. 4.6, a window size of $K = 10$ sensing events tends to provide the best accuracy but there are no significant differences for other values of the parameter K.

Method 3 can also provide an equally significant accuracy improvement as shown in Fig. 4.7. Notice that there is no significant difference in the level of accuracy obtained when the method's decisions are based on the real-time duty cycle estimated from past sensing observations ($\widehat{\Psi}$) or the true long-term PU channel duty cycle (Ψ). This means that the method is implementable in practice since the true PU channel duty cycle does not need to be known beforehand and the estimation obtained from past sensing observations is accurate and converges quickly to the correct value (see Fig. 4.4). As observed in Fig. 4.7, method 3 tends to provide the best estimation accuracy when the sensing period T_s is similar to the minimum PU transmission time μ_i (notice that T_s should not be greater than μ_i since some short idle/busy periods would then be missed).

Finally, Fig. 4.8 compares the methods proposed in this chapter (methods 1-3) with the methods proposed in [81] (methods 4-6). The proposed methods are capable to attain (with an adequate configuration of their parameters) a nearly perfect estimation ($D_{KS} \approx 0$) of the PU channel activity statistics, which cannot be achieved by methods 4-6. It is worth noting that the performance results shown in Figs. 7 and 8 of [81] for methods 4-6 show that there are some values of the sensing period T_s for which the KS distance D_{KS} is very close to zero. However, this is due to the fact that KS distance in [81] was computed by comparing the statistics estimated under ISS with those estimated

TABLE 4.1: Computational cost (complexity) analysis for the considered methods ($\Psi = 0.8$, $\mu_i = 10$ *t.u.*, $T_s = 9.9$ *t.u.*, $P_{fa} = P_{md} = 0.01$).

Method	**Average computation time (s)**
Method 1 ($\beta = 3$)	0.0662
Method 2 ($K = 10$)	0.1252
Method 3 (real-time est. $\widehat{\Psi} = 0.8$)	0.3811
Method 4	0.0004
Method 5	0.0011
Method 6	0.0008

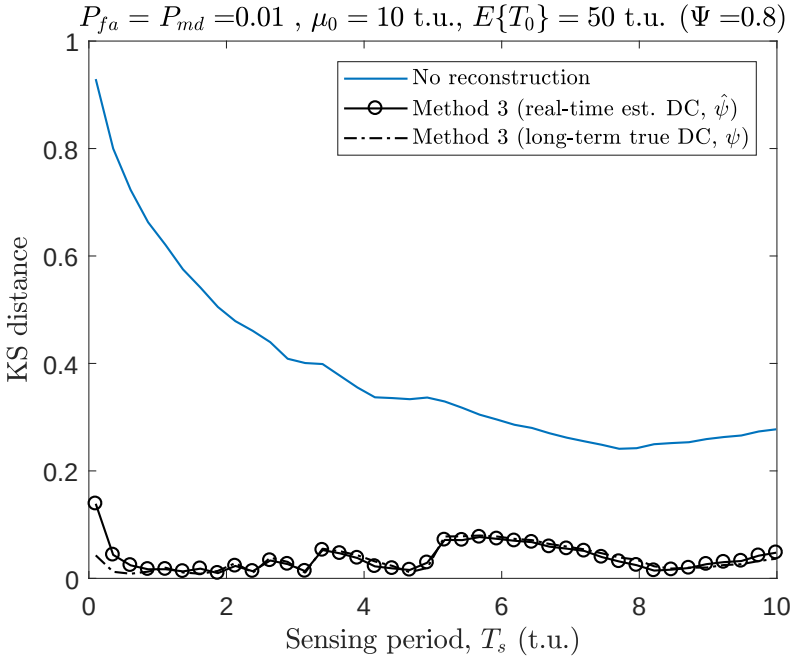
under PSS, instead of comparing with the original period lengths as it is done in this chapter. This second type comparison, which provides a more realistic evaluation of the estimation accuracy, shows that the proposed methods can attain a significantly improved accuracy in the estimated PU channel activity statistics as shown in Fig. 4.8.

Table 4.1 shows the computational cost (complexity) of the considered methods when applied to 10,000 pairs of periods (with 20 repetitions) based on a workstation with an Intel Xeon processor (E5-1620v3 @ 3.50GHz). Methods 4-6 provide simplicity and fast execution at the expense of accuracy, while methods 1-3 are slower but provide a nearly perfect estimation. It is worth mentioning that method 3 is the most demanding algorithm as it requires to compute real-time estimated DC ($\widehat{\Psi} = 0.8$) which adds a significant calculation burden.

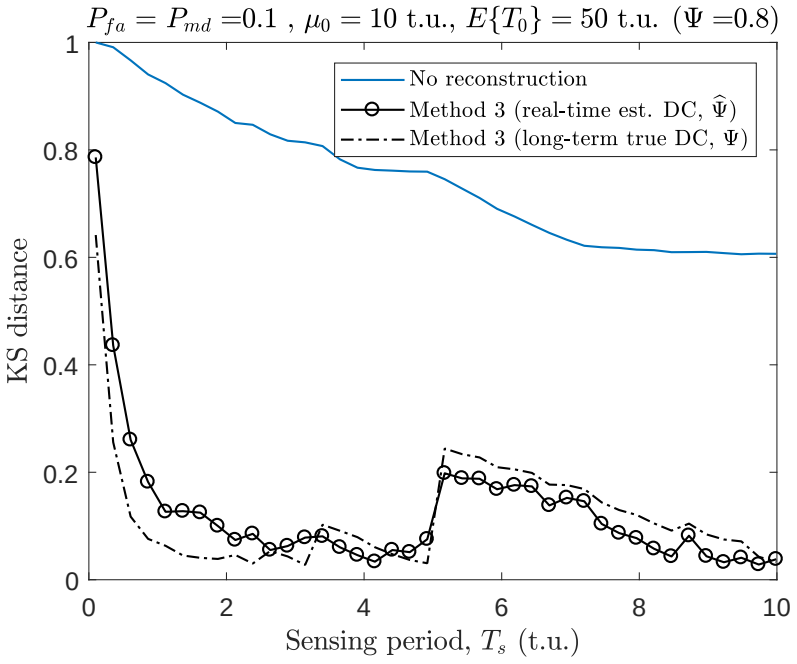
4.5 Configuration of the Proposed Methods

A relevant practical aspect of the proposed methods is how the parameters should be configured to ensure an optimum operation point that can provide a (nearly) perfect estimation of the PU channel activity statistics.

The system designer can essentially control two parameters, namely the *operating point* of the employed spectrum sensing method and the employed sensing period T_s. The operating point of the employed spectrum sensing method determines its performance (i.e., the values of P_{fa} and P_{md}) and can be tuned by modifying the parameters of the sensing algorithm; for example, in the case of energy detection the main design parameter is the detection/decision threshold. As suggested by the results of Fig. 4.5, Fig. 4.6, Fig. 4.7, the sensing algorithm should ideally be configured to operate in a point where P_{fa} and P_{md} are as low as possible since for lower values of P_{fa} and P_{md} there is a relatively wider range of values of the sensing period T_s where the error is zero or very close to zero, which makes it relatively easy to provide a (nearly) perfect estimation of the PU activity statistics (i.e., $D_{KS} \approx 0$). If the operating point of the sensing algorithm cannot be modified (e.g., it is configured to provide specific performance targets required by a particular service) and the resulting P_{fa} and P_{md} are high, the proposed methods can still provide a (nearly) perfect estimation of the PU activity statistics, however in this case the value of the sensing period T_s should be selected carefully. In particular, Fig.



(a)



(b)

FIGURE 4.7: Performance of method 3.

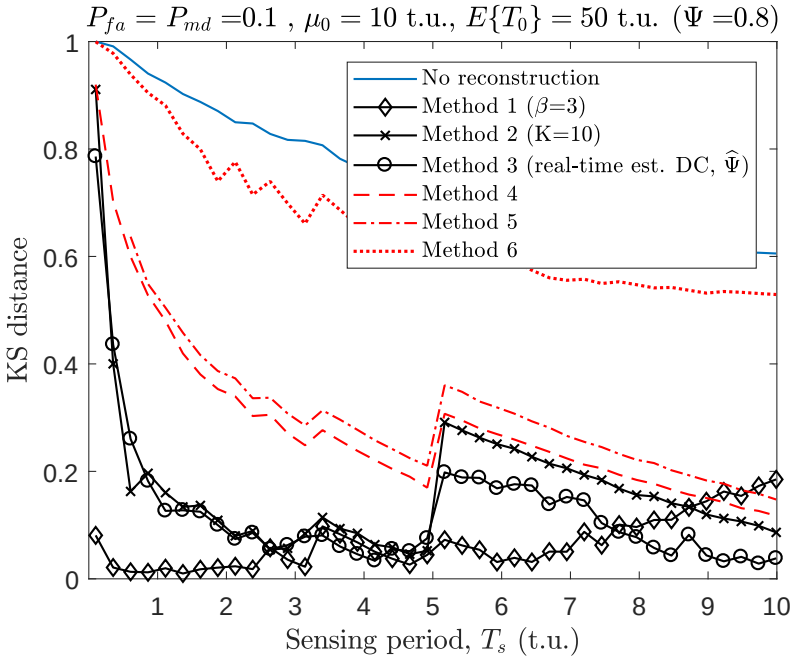


FIGURE 4.8: Performance comparison of the methods considered in this chapter.

4.5 shows that for method 1 it should take low values (ideally close to zero, $T_s \approx 0$), for method 2 it should be $T_s \approx \mu_i/2$, and for method 3 it should take high values (ideally close to the minimum PU transmission time, $T_s \approx \mu_i$). Moreover, if the operating point of the sensing method and the sensing period are both constrained, the only degree of freedom for the system designer would be the selection of one of the three proposed methods. In such a case, the decision could be based on which method provides the best accuracy for the employed sensing period T_s (e.g., method 1 when $T_s \approx 0$, method 2 when $T_s \approx \mu_i/2$, and method 3 when $T_s \approx \mu_i$). In any case, the set of methods proposed in this work can provide a very accurate (nearly perfect) estimation of the PU activity statistics under imperfect sensing, even with a high probability of sensing errors, over a wide range of operating conditions.

4.6 Impact of the Primary User Activity Pattern

In the previous section, the PU is assumed to follow a GP distribution. Nevertheless, the modelling of PU occupancy depends on multiple factors including the radio technology (e.g., amateur systems, paging systems and cellular mobile communication systems) and the time scale resolution of sensing device (the duration required to perform spectrum sensing, T_s) as long time scales would result in different distribution patterns than short time scales, which in turn depends on the sensed radio technology.

It is interesting to note that most researches assume an exponential distribution to model the binary PU holding states (idle/busy) for both large/small T_s. This can be related to the mathematical simplicity of exponential distribution. In this section, the impact of the PU activity pattern/distribution on the performance of the proposed

methods is investigated. The considered distributions are: exponential [125], generalised exponential [126], Pareto [127], generalised Pareto [128], gamma [125] and Weibull [125]. Table 4.2 shows the cumulative distribution function (CDF) for the considered models of PU binary occupancy patterns. In order to maintain a fair comparison between the different distributions, their parameters (λ_i and α_i) are configured as shown in Table 4.3 to provide the same mean period duration ($\mathbb{E}\{T_i\} = 50$ t.u.) with the same minimum activity time ($\mu_i = 10$ t.u.). The same parameters are considered for both busy and idle periods, which leads to a channel duty cycle (DC) of 50% ($\Psi = 0.5$).

Fig 4.9 shows the CDFs of the considered distributions for original generated periods (T_i). As it can be appreciated, some distributions show similar trends. In particular, three different groups of distributions can be distinguished. The first group is generalised Pareto and exponential distributions, the second group contains gamma, generalised exponential and Weibull distributions, while the last group contains only the Pareto distribution which behaves differently from other distributions. This observation will serve to explain later on the differences and similarities observed in the obtained results.

Fig. 4.10 shows the performance of PSS in comparison with the original periods. In this case, the only source of error is the use of a finite sensing period T_s, which results in all the observed periods being integer multiples of T_s (i.e., $\widehat{T}_i = kT_s$, $k \in \mathbb{N}^+$). As appreciated, the estimation error can be decreased by reducing T_s (i.e., increasing the time resolution improves the accuracy). Moreover it can also be noted that there are some differences in the accuracy obtained for different distributions, however these are mainly minor differences, except for the Pareto distribution, which shows a slightly different trend as T_s increases (note in Fig. 4.9 that the CDF of the Pareto distribution behaves differently from all other distributions).

Fig. 4.11, Fig. 4.12 and Fig. 4.13 show the estimation accuracy for the CDF of the PU busy/idle periods under ISS in terms of the KS distance (with respect to the original periods T_i obtained from of step 1) as a function of the sensing period. The results are shown for the following cases: no reconstruction methods are used and the PU statistics are estimated from the periods $\breve{T}_i$ observed under ISS (Fig. 4.11), the periods $\breve{T}_i$ are processed according to reconstruction method 4 (Fig. 4.12), and the periods $\breve{T}_i$ are processed according to reconstruction method 1 (Fig. 4.13). The top row for Fig. 4.11, Fig. 4.12 and Fig. 4.13 shows the results obtained for $P_{fa} = P_{md} = 0.01$ while the bottom row show the results obtained for $P_{fa} = P_{md} = 0.1$. Simulations are performed for $\Psi = 0.5$ (medium load).

As it can be appreciated from Fig. 4.11, Fig. 4.12 and Fig. 4.13, the accuracy is in general better when having a lower probability of error. Nevertheless, the considered reconstruction methods can provide better accuracy compared to the case where no reconstruction is considered (i.e., when the PU statistics are computed straight from the periods observed under ISS). As opposed to the PSS case shown in Fig. 4.10, the accuracy under ISS degrades as the sensing period T_s decreases. This is due to the fact that reducing T_s increases the number of sensing events per PU idle/busy period

TABLE 4.2: Considered probability distribution models for PU idle/busy period durations. Distribution names: E (exponential), GE (generalised exponential), P (Pareto), GP (generalised Pareto), G (gamma), and W (Weibull). Distribution parameters: μ_i (location), λ_i (scale), and α_i (shape). T_i represents the period length. $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$ represent the mean and the variance of the distribution, respectively. $\psi(\cdot)$ is the digamma function [129, 6.3.1] and $\psi'(\cdot)$ is its derivative. $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function [129, 6.5.2] and $\Gamma(\cdot)$ is the (complete) gamma function [129, 6.1.1]. (Reproduced from [64]).

Distribution function	**Parameters**	**Moments**
$F_E(T_i;\mu_i,\lambda_i) = 1 - e^{-\lambda_i(T_i-\mu_i)}$	$T_i \geq \mu_i > 0$ $\lambda_i > 0$	$\mathbb{E}\{T_i\} = \mu_i + \frac{1}{\lambda_i}$ $\mathbb{V}\{T_i\} = \frac{1}{\lambda_i^2}$
$F_{GE}(T_i;\mu_i,\lambda_i,\alpha_i) = \left[1 - e^{-\lambda_i(T_i-\mu_i)}\right]^{\alpha_i}$	$T_i \geq \mu_i > 0$ $\lambda_i > 0$ $\alpha_i > 0$	$\mathbb{E}\{T_i\} = \mu_i + \frac{\psi(\alpha_i+1)-\psi(1)}{\lambda_i}$ $\mathbb{V}\{T_i\} = \frac{\psi'(1)-\psi'(\alpha_i+1)}{\lambda_i^2}$
$F_P(T_i;\lambda_i,\alpha_i) = 1 - \left(\frac{\lambda_i}{T_i}\right)^{\alpha_i}$	$T_i \geq \lambda_i$ $\lambda_i > 0$ $\alpha_i > 2$	$\mathbb{E}\{T_i\} = \frac{\alpha_i\lambda_i}{\alpha_i-1}$ $\mathbb{V}\{T_i\} = \frac{\alpha_i\lambda_i^2}{(\alpha_i-1)^2(\alpha_i-2)}$
$F_{GP}(T_i;\mu_i,\lambda_i,\alpha_i) = 1 - \left[1 + \frac{\alpha_i(T_i-\mu_i)}{\lambda_i}\right]^{-1/\alpha_i}$	$T_i \geq \mu_i$ $(\alpha_i \geq 0)$ $T_i \in [\mu_i, \mu_i - \frac{\lambda_i}{\alpha_i}]$ $(\alpha_i < 0)$ $\mu_i, \lambda_i > 0,\ \alpha_i < 1/2$	$\mathbb{E}\{T_i\} = \mu_i + \frac{\lambda_i}{1-\alpha_i}$ $\mathbb{V}\{T_i\} = \frac{\lambda_i^2}{(1-\alpha_i)^2(1-2\alpha_i)}$
$F_G(T_i;\mu_i,\lambda_i,\alpha_i) = \frac{\gamma\left(\alpha_i, \frac{T_i-\mu_i}{\lambda_i}\right)}{\Gamma(\alpha_i)}$	$T_i \geq \mu_i > 0$ $\lambda_i > 0$ $\alpha_i > 0$	$\mathbb{E}\{T_i\} = \mu_i + \lambda_i\alpha_i$ $\mathbb{V}\{T_i\} = \lambda_i^2\alpha_i$
$F_W(T_i;\mu_i,\lambda_i,\alpha_i) = 1 - \exp\left[-\left(\frac{T_i-\mu_i}{\lambda_i}\right)^{\alpha_i}\right]$	$T_i \geq \mu_i > 0$ $\lambda_i > 0$ $\alpha_i > 0$	$\mathbb{E}\{T_i\} = \mu_i + \lambda_i\Gamma\left(1 + \frac{1}{\alpha_i}\right)$ $\mathbb{V}\{T_i\} = \lambda_i^2\left[\Gamma\left(1 + \frac{2}{\alpha_i}\right) - \Gamma^2\left(1 + \frac{1}{\alpha_i}\right)\right]$

TABLE 4.3: Distribution parameters.

Distribution	**Shape** (α_i)	**Scale** (λ_i)
E	-	0.025
GE	29.179	0.0993
P	2.001	25.355
GP	0.25	30
G	9.499	4.215
W	3.167	44.7721

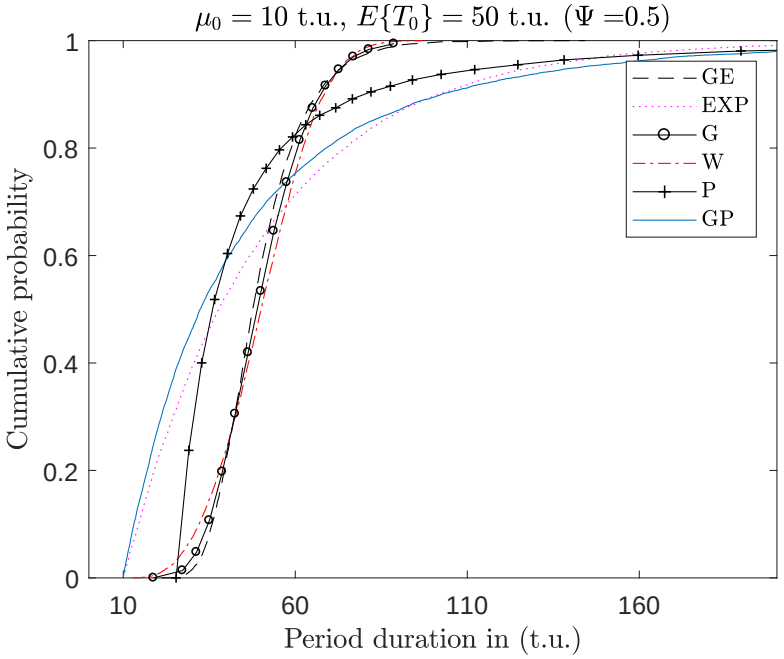


FIGURE 4.9: CDF of idle/busy periods ($\mu_i = 10$ t.u., $E\{T_i\} = 50$ t.u., $\Psi = 0.5$).

and therefore the probability of having sensing errors within each of the PU idle/busy periods.

For the case where no reconstruction methods are applied 4.11, the accuracy of the estimated PU statistics depends on the considered distributions. Notice that similar distributions lead to similar levels of accuracy. Both generalised Pareto and exponential distributions, which are similar as shown in Fig. 4.9, lead to the highest accuracy and immunity against sensing errors. On the other hand, the generalised exponential, gamma and Weibull distributions, which are also similar among them, provide relatively similar estimation accuracies, while the Pareto distribution results in the lowest accuracy (this observation holds for both cases of sensing errors, i.e., low $P_{fa} = P_{md} = 0.01$ and high $P_{fa} = P_{md} = 0.1$ probabilities).

For method 4 ([81]) Fig. 4.12, similar comments as above can be made. However, in this case some improvement in the accuracy of the estimated PU statistics is obtained for all the distributions as a result of the reconstruction operation performed by this method. It is interesting to note that for the case of low probability of sensing errors ($P_{fa} = P_{md} = 0.01$) there are some cases where the accuracy of the estimated statistics

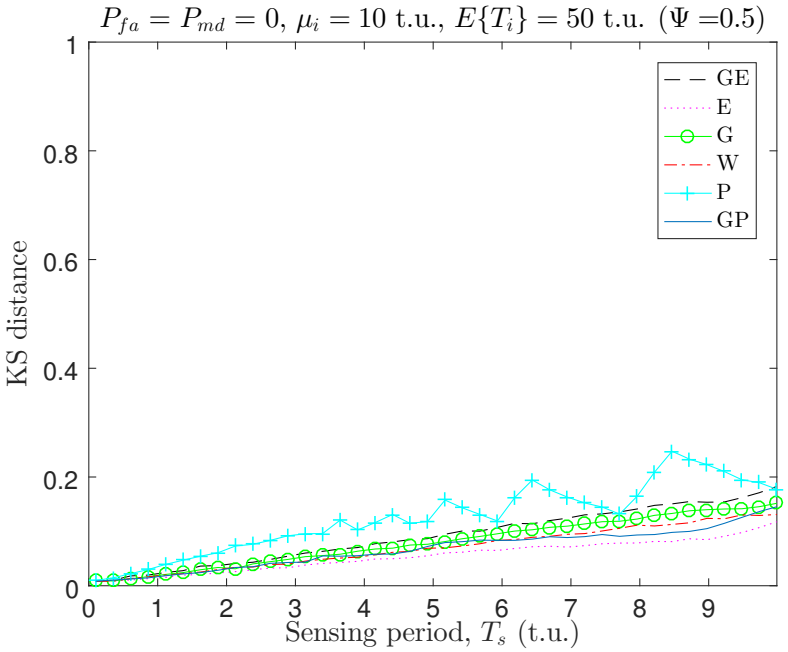
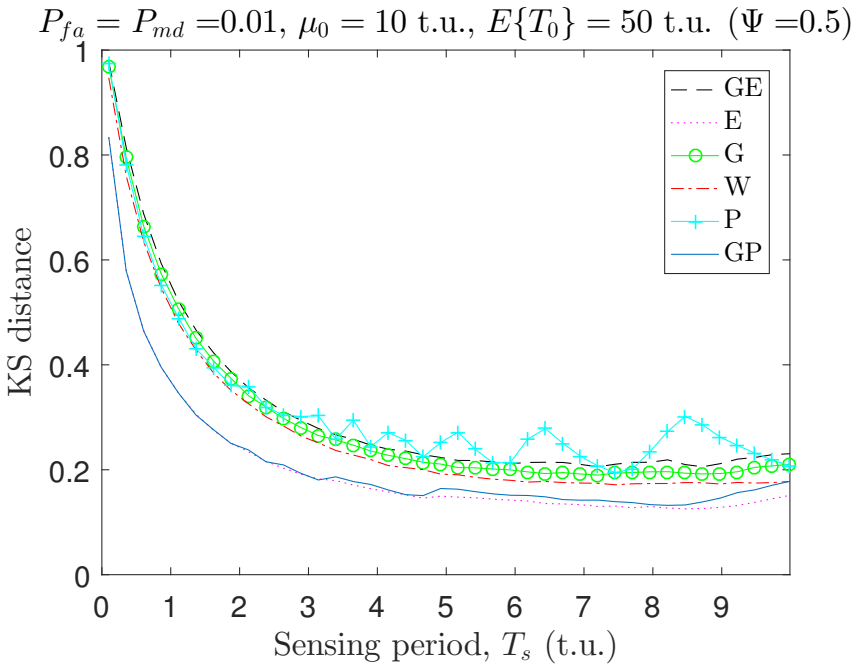


FIGURE 4.10: Estimation accuracy for the considered distributions under PSS.

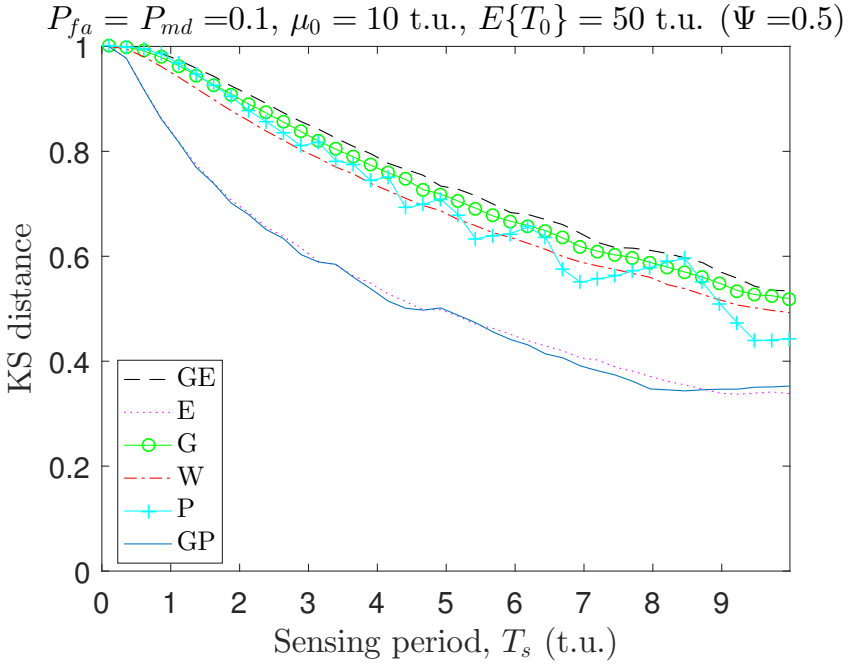
is nearly perfect ($D_{KS} \approx 0$). As shown in Fig. 4.12, this occurs for the generalised Pareto and exponential distributions when the sensing period is close to the minimum PU activity time ($T_s \approx \mu_i$) but not for any other distributions. This may incorrectly lead to the conclusion that certain methods can provide a nearly perfect estimation of the PU statistics when this actually depends on the particular PU distribution considered (i.e., it may be true for some PU systems but not for some others and the efficacy of a certain method may actually depend on the PU system over which it is applied).

For method 1 (proposed method) Fig. 4.13, the threshold is set to $\beta = 3$ as it provided the best results as can be seen from Fig. 4.5. Note that this method is more complex than method 4 and the level of improvement is notably higher. Notice that in this case there is a wide range of (low) values of the sensing period T_s that can provide a nearly perfect estimation ($D_{KS} \approx 0$). When the probability of sensing errors is low ($P_{fa} = P_{md} = 0.01$) this observation holds for all the distributions (even though for the Pareto distribution the range is slightly shorter). However for higher probability of sensing errors ($P_{fa} = P_{md} = 0.1$) this observation only holds for the generalised Pareto and exponential distributions, where both can achieve nearly perfect accuracy over a wide range (for other distributions there is a lower bound below which the error cannot decrease). In contrast with method 4, the results for method 1 are not so sensitive to the particular PU occupancy pattern (distribution).

Finally, as it can be appreciated from Fig. 4.11, Fig. 4.12 and Fig. 4.13, the efficiency of the considered methods may depend to different extents on the PU activity statistics (i.e., some methods may be more effective for certain PU distributions than for others while some other methods may be less sensitive to the particular PU distribution). If this

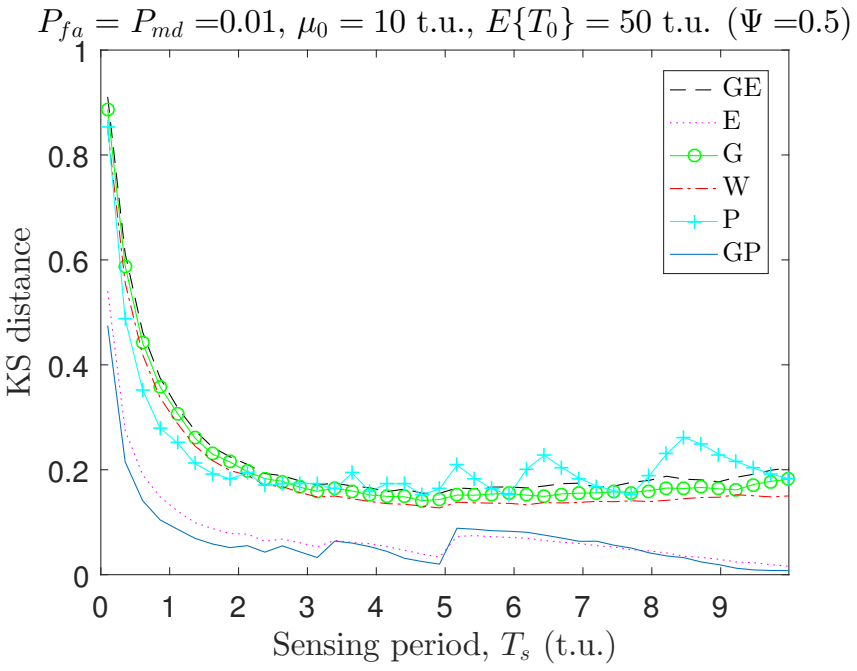


(a)

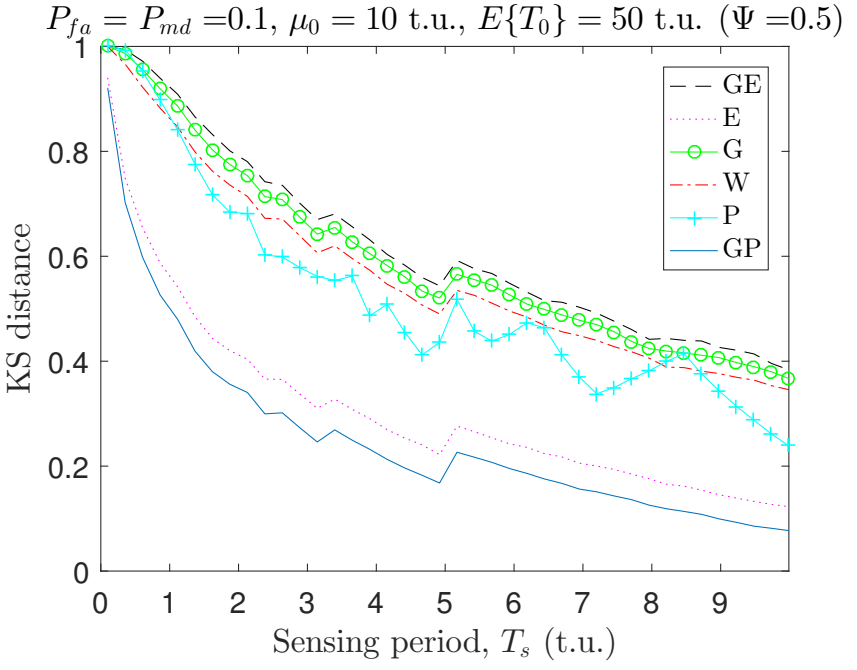


(b)

FIGURE 4.11: Estimation accuracy with no reconstruction under ISS.

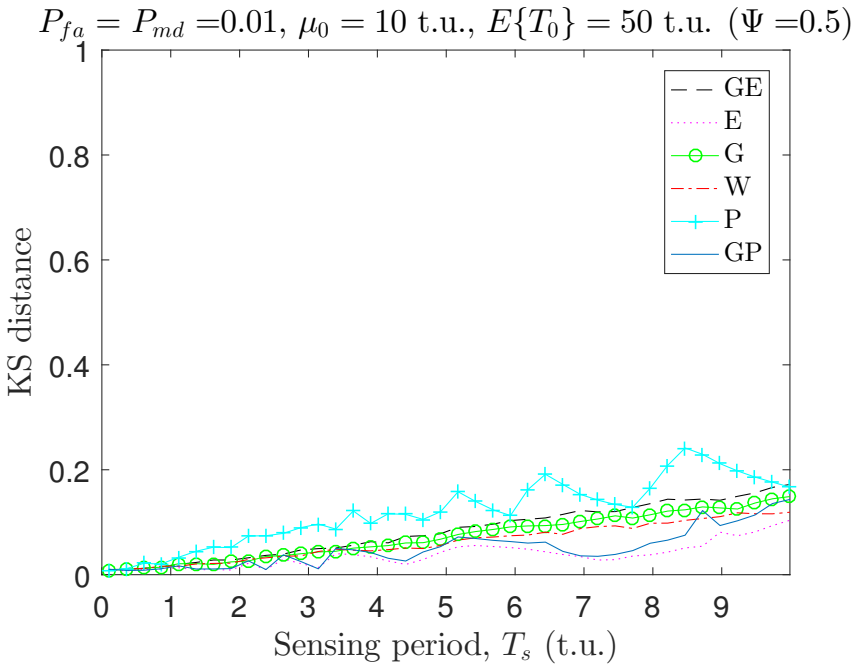


(a)

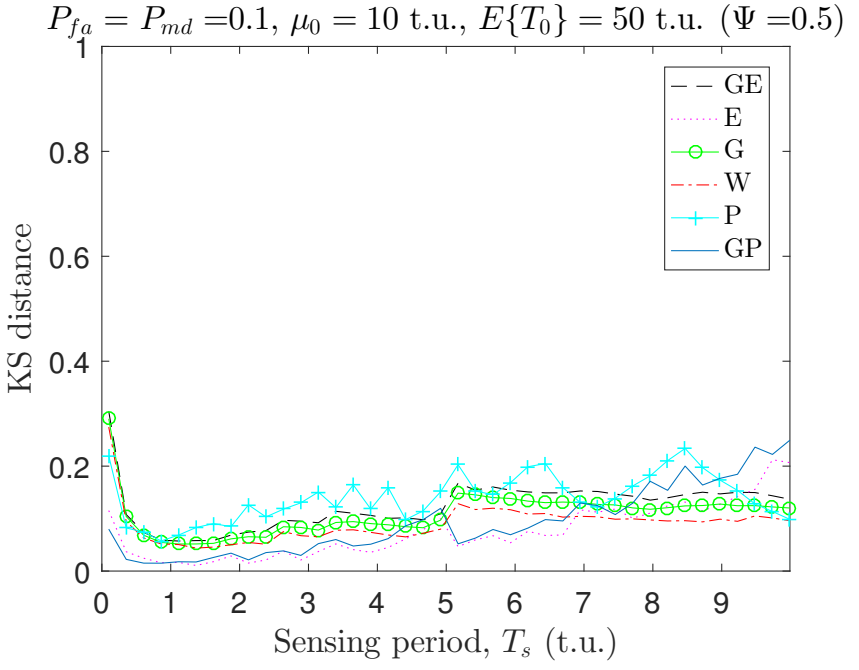


(b)

FIGURE 4.12: Estimation accuracy with method 4 under ISS.



(a)



(b)

FIGURE 4.13: Estimation accuracy with method 1 under ISS.

is not adequately considered then the performance of a real system design could be under or overestimated, thus leading to unexpected behaviours/performances. This important observation should be carefully taken into account in future studies on DSA/CR systems, where traditionally only one distribution model is considered for the PU idle/busy period durations (typically the exponential distribution).

4.7 Summary

CR systems can benefit from the knowledge of the PU spectrum activity statistics. The estimation of such statistics, however, can be very inaccurate when based on sensing observations due to the practical limitations of spectrum sensing and the presence of sensing errors.

In this chapter, two problems have been addressed. First, the problem of accurately estimating the PU activity statistics under ISS by proposing three simple but effective methods to overcome the degrading effects of spectrum sensing errors on the estimated statistics. The performance has been evaluated by means of computer simulations and compared to other methods previously proposed in the literature. The obtained results show that the methods proposed in this work are able not only to provide significant accuracy improvements with respect to the existing methods but also, and more importantly, an accurate estimation of the PU activity statistics. This chapter has also discussed how the proposed methods should be configured in a practical system design in order to achieve the best attainable accuracy.

Second, multiple models to imitate PU occupancy pattern have been considered and their impact on the accuracy of methods to estimate the PU activity statistics has been investigated by means of simulations under both PSS and ISS. The obtained results have demonstrated that the efficiency of different methods may depend on the particular PU activity statistics. As a matter of fact, some methods may be less sensitive than others to the particular PU distribution and exhibit more similar or dissimilar levels of accuracy under different operating conditions. Future studies on DSA/CR systems in general, and on methods for the estimation of PU activity statistics in particular, should not be constrained to a single model for the distribution of PU idle/busy periods (as it has usually been the case in the past) but consider a sufficiently broad range of models for a more accurate and comprehensive performance evaluation.

Chapter 5

Cooperative Estimation of Primary Activity Statistics under Imperfect Spectrum Sensing

5.1 Introduction

As discussed earlier, spectrum sensing is a key enabling technology for CR operation, as it allows SUs to detect the presence/absence of PU traffic which is essential to reduce the interference [11, 14]. An essential requirement for SU is to work in a fast and accurate manner while identifying empty slots in the primary channel.

While the cooperative estimation improves the operation of spectrum sensing by taking advantage of spatial diversity at every receiving SU. The improvement in performance is hindered by both the increase of cooperation overhead (as every SU is required to report local sensing decision to FC) and security issues (The CR network can is penetrated by MUs) with MUs sending false reports to FC. This chapter provides solutions to these problems by providing a power efficient and secure reporting mechanisms. The main contributions of this chapter can be summarised as follows:

1. Study the cooperative estimation of PU traffic statistics under both sensing errors and finite sensing period with experimental validation.

2. Propose a new reporting mechanism (differential reporting) to reduce the overhead in the reporting channel and increase the spectrum and energy efficiency.

3. Study the estimation of primary distribution under both sensing errors and SSDF attacks and propose a new algorithm to counter the effect of such attacks on the estimation of PU traffic statistics. While both aspects have received some attention in the literature separately, they have not been considered simultaneously along with their combined effects on the cooperative estimation of primary traffic statistics.

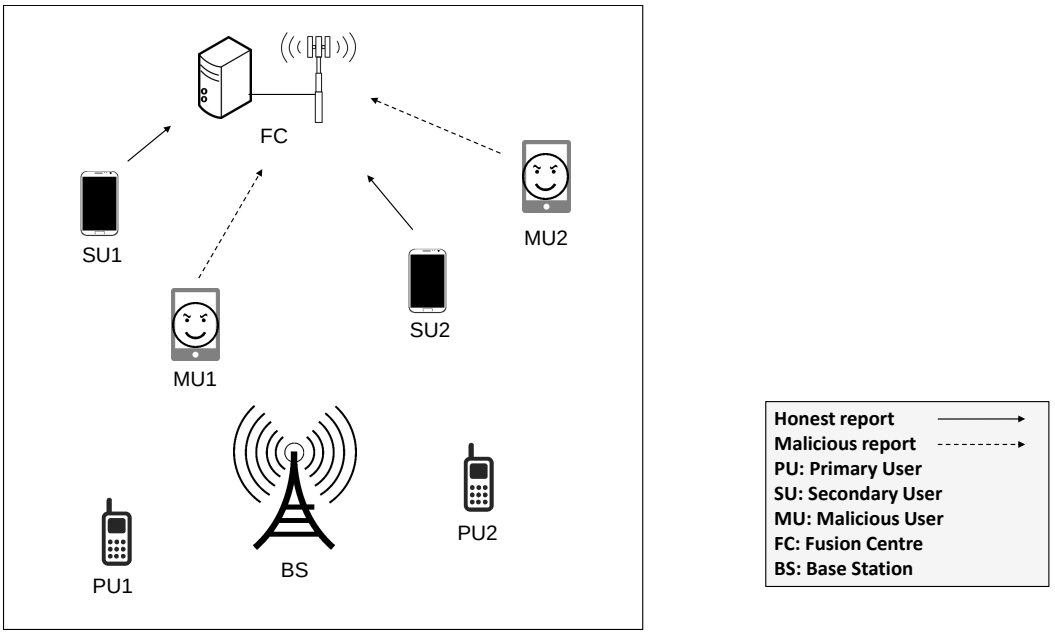


FIGURE 5.1: System model for cooperative primary traffic estimation with malicious users.

5.2 System Model and Problem Formulation

In this chapter, a single PU channel is considered for the sake of simplicity. The PU state holding times (T_0 for idle periods and T_1 for busy periods) are random variables assumed to be independent and exponentially distributed. The exponential distribution is the most common model used to describe the periods of the on/off states in the literature [71, 80, 112–114] even though it has been proven not to be the most accurate since other distributions provide better fit for real scenarios such as the generalized Pareto, Gamma or even more complicated distributions [64]. We use the exponential distribution because it is a special case of the generalized Pareto distribution with a simpler mathematical form. As for cooperative network side, K SUs with a specialised FC are considered along with MUs. The FC is in charge of making the final decision of PU channel state through one of the decision rules (only hard decision rules are considered, soft decision is out of the scope for this chapter) and then exploit the sequence of reported idle/busy channel states to estimate the durations of the channel holding times T_0/T_1 and the statistics (i.e., distribution). The considered system model is shown in in Fig. 5.1.

The cooperative estimation can be sub-characterised into four indispensable stages. Starting with the sensing stage, every SU performs spectrum sensing on a regular basis to estimate the primary channel availability. Second stage is the local hard decision, every SU utilises a detection algorithm to generate the binary channel state decisions (0 for idle/absence of PU and 1 for busy/presence of PU). The decisions of all SUs are assumed to be independent. The third stage is the reporting phase, where the local decisions of every SU are reported to the central FC through a dedicated reporting channel for the final global decision, where the FC (CR base station) is in charge of

the final global decisions while the SUs function as cooperative sensing nodes. At every sensing event (performed with a sensing period of T_s time units), the FC makes the global decisions regarding the presence/busy (H_1) or absence/idle (H_0) of a PU. The decision rules considered in this chapter are the most popular ones (AND, OR, MAJORITY) [130].

1. AND-rule: The FC decides that a PU is present only if all cooperative SUs report with PU present (i.e., all SUs report with 1).

2. OR-rule: The FC decides that a PU is present when at least one cooperative SU reports with PU present (i.e., at least one SU reports with 1).

3. MAJORITY-rule: The FC decides that a PU is present when half or more of the cooperative SUs decide the presence of a PU (i.e., $K/2$ or more SUs report with 1).

Based on one of these three hard decision rules the FC makes a decision on the PU channel state and then exploits the sequence of reported idle/busy channel states to estimate the durations of the channel holding times $\widetilde{T}_i$ ($i = 0$ for idle periods, $i = 1$ for busy periods) of the original primary busy/idle periods T_i. Note that the estimated periods are integer multiples of the employed sensing period (i.e., $\widetilde{T}_i = mT_s$, $m \in \mathbb{N}^+$) and as a result the estimated periods will differ from the true original periods, which can in general be assumed to have a continuous domain (i.e., $T_i \in \mathbb{R}^+$).

In practise, SUs can work under both low and high SNR conditions. Under low SNR, SUs suffer from sensing errors (on local decisions, as every SU contributes in the final decision). ISS occurs in two types of errors: false alarm ($\widetilde{H} = H_1|H_0$) which is characterised by the probability of false alarm (P_{fa}), where the PU signal is not present but announced as present because of the high noise level present at SU's receiver, and missed detection ($\widetilde{H} = H_0|H_1$) which is characterised by the probability of missed detection (P_{md}), where the PU signal is present but with power lower than the receiver's threshold because of fading and shadowing.

Sensing errors have a significant impact on the performance of cognitive network systems (both PUs and SUs) and on the estimation of PU traffic statistical information as well. Inaccurate detection leads to inaccurate estimation for PU traffic activity statistics as the estimated durations can be longer or shorter than the original values. Another source of error is MUs who report with fake channel states to confuse the FC and lead it to announce wrong global decisions, thus missing the opportunity of transmission and leading to inaccurate PU traffic estimations.

The main objective of this chapter is to study the cooperative estimation of the primary statistics (distribution of period durations) under spectrum sensing errors and SSDF attacks, and propose methods that can provide an accurate estimation of the PU traffic statistics under such challenging conditions.

5.3 Cooperative Estimation of the Distribution of Primary Channel Holding Times

Two methods are considered in this chapter for the estimation of the distribution of primary idle/busy periods, the Direct Estimation Method (DEM) and the Method of Moments (MoM).

5.3.1 Direct Estimation Method (DEM)

The direct estimation of the distribution is based on the empirical cumulative distribution function (ecdf in MATLAB), where the Kaplan-Meier estimation is obtained utilising the ecdf function for the given samples. The main advantage of this method is that it requires no prior knowledge about the primary distribution. The main drawback of this method is that the estimated distribution is a discrete version of the original continuous distribution as the estimated periods are discrete (integer) multiples of the sensing period T_s. Moreover, this method can not achieve high accuracy for all sensing periods, which can not be improved even by increasing the number of SUs as it will be seen in the results section. This motivates the consideration of the following method.

5.3.2 Method of Moments (MoM)

To overcome the limitations of the DEM, a solution based on the MoM is considered. For the MoM, the distribution of the primary periods has to be known or assumed to be known. The distribution parameters are then estimated from the sample moments. The probability density function (PDF) and cumulative density function (CDF) for the exponential distribution are given by [94]:

$$f_{T_i}(t) = \begin{cases} 0 & t < \mu_i \\ \lambda_i e^{-\lambda_i (t-\mu_i)} & t \geq \mu_i, \end{cases} \tag{5.1}$$

$$F_{T_i}(t) = \begin{cases} 0 & t < \mu_i \\ 1 - e^{-\lambda_i (t-\mu_i)} & t \geq \mu_i, \end{cases} \tag{5.2}$$

where $\lambda_i \geq 0$ is the scale parameter of the distribution and $\mu_i > 0$ is the location parameter (also the smallest value for the PU activity period. i.e., $T_i \geq \mu_i$) .

The distribution parameters can be estimated following three approaches:

5.3.2.1 Direct estimation of minimum

The minimum period $\widetilde{\mu}_i^{dem}$ can be estimated as:

$$\widetilde{\mu}_i^{dem} = \min\left(\{\widetilde{T}_i\}_{n=1}^{N}\right) = T_s, \tag{5.3}$$

where $\{\widetilde{T}_{i,n}\}_{n=1}^{N}$ is a set of N observed periods and its minimum value under ISS is given by T_s as discussed in [81]. The value of λ_i can be inferred from:

$$\begin{aligned}\widetilde{\mathbb{V}}(T_i) &= \frac{1}{\lambda_i^2} \\ &\approx \frac{1}{N-1}\sum_{n=1}^{N}\left[\widetilde{T}_{i,n} - \mathbb{E}(\widetilde{T}_i)\right]^2, \end{aligned} \tag{5.4}$$

where $\widetilde{\mathbb{V}}(T_i)$ is the variance of the observed PU periods and $\mathbb{E}(\widetilde{T}_i)$ is the mean which is given by (5.5).

5.3.2.2 Minimum based on MoM

In general, the higher the number of SUs for the cooperative estimation of mean and variance, the higher the accuracy of the estimation. This observation can be utilised to estimate μ_i as follows:

$$\begin{aligned}\mathbb{E}(\widetilde{T}_i) &= \widetilde{\mu}_i + \frac{1}{\lambda_i} \\ &\approx \frac{1}{N}\sum_{n=1}^{N}\widetilde{T}_{i,n}, \end{aligned} \tag{5.5}$$

$$\widetilde{\mu}_i = \mathbb{E}(\widetilde{T}_i) - \sqrt{\widetilde{\mathbb{V}}(T_i)}, \tag{5.6}$$

where $\mathbb{E}(\widetilde{T}_i)$ is the mean of the observed PU periods.

5.3.2.3 Minimum based on modified MoM

A similar procedure as above is utilised, but with a correction factor to reduce the effects of finite spectrum sensing period. The estimation of $\widetilde{\mathbb{V}}(T_i)$ is given by:

$$\begin{aligned}\widetilde{\mathbb{V}}(T_i) &= \frac{1}{\lambda_i^2} - \frac{T_s^2}{6} \\ &\approx \frac{1}{N-1}\sum_{n=1}^{N}\left[\widetilde{T}_{i,n} - \mathbb{E}(\widetilde{T}_i)\right]^2 - \frac{T_s^2}{6}, \end{aligned} \tag{5.7}$$

where $T_s^2/6$ is the correction factor introduced in Chapter 2 to remove the effect of the finite sensing period T_s.

5.4 Local State Reporting methods and Overhead

Cooperation can improve the estimation of both the instantaneous channel state and the primary traffic statistics, however the cooperative process introduces signalling overhead, which reduces the spectrum and energy efficiencies. Reporting in every sensing event

Algorithm 5: Periodic reporting

Input : $\lambda \in \mathbb{R}^+$ ▷ Energy decision threshold
$N_s \in \mathbb{N}^+$ ▷ Number of signal samples
Output: $R_{ch,i} \in \{0, 1\}$ ▷ Channel state report

```
for each sensing event i do
    Y_i ← Energy of N_s samples                ▷ Energy detection
    if Y_i ≥ λ then
        R_{ch,i} ← 1                           ▷ Flag channel as busy
    else
        R_{ch,i} ← 0                           ▷ Flag channel as idle
    end
    SU sends R_{ch,i} to FC
end
```

is in general necessary in the case of cooperative spectrum sensing but is not essential in the case of cooperative PU traffic estimation considered in this chapter. A possible increase in spectrum and energy efficiency can be achieved by reducing the amount of channel reports required at each sensing stage. In this section, first the original reporting mechanism is described followed by the On/Off reporting method proposed in [131], then a new method (differential reporting) is proposed. Fig. 5.2 shows the required number of reports for each reporting mechanism.

5.4.1 Periodic Reporting Mechanism

In the default periodic reporting mechanism, every SU transmits a report containing the local decision (at every sensing event) during the reporting stage to the central FC. Each report is sent through a dedicated report channel for every SU. The periodic reporting is summarised in Algorithm 5. The main drawback with periodic reporting is the high number of reports as every SU sends reports to the FC with local decisions via its own dedicated reporting channel in every single sensing event.

5.4.2 On/Off Reporting Mechanism

In this method, which is proposed in [131], all SUs report the local states back to the FC only during busy periods and remain silent during idle periods. This way the reporting overhead would be reduced from the periodic reporting, especially at low channel usage (i.e., low duty cycle). An alternative approach is to report the local states back to the FC only during idle periods and remain silent otherwise. This way the reporting overhead would be reduced from the periodic reporting under high channel usage (i.e., high duty cycle). The reporting option that provides the lowest number of reports depends on whether the duty cycle is lower than 0.5 (reporting during every busy periods) or greater (reporting during every idle periods). If the primary channel duty cycle is around 0.5, then both options are equivalent. In practice, SUs target primary channels with limited primary usage. As a result, only the first case for the On/Off reporting mechanism

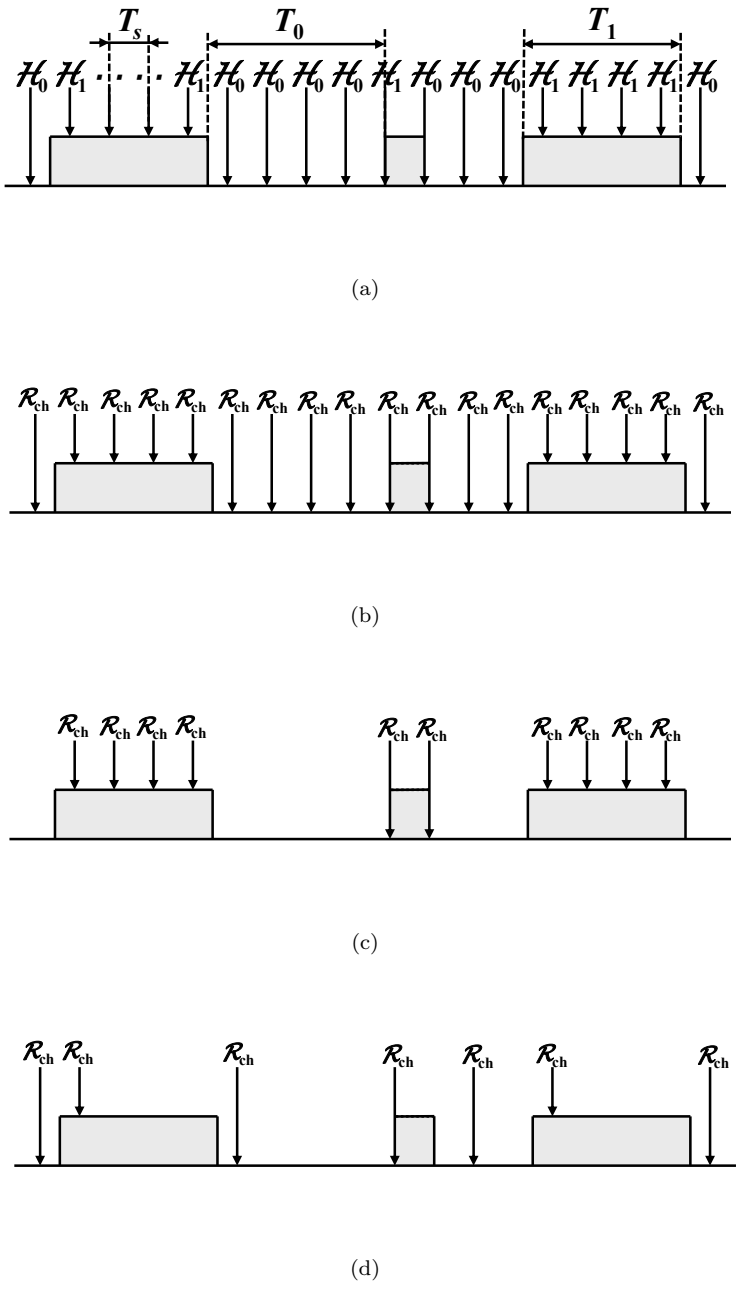


FIGURE 5.2: The operation for the considered reporting mechanisms. (a) The sensing stage at SU (it shows the required number of sensing events per SU. (b) Periodic reporting mechanism. (c) On/Off reporting mechanism (report for busy periods case). (d) Differential reporting mechanism.

(i.e., reporting during busy periods) will be considered for comparison purposes in this chapter. The considered On/Off reporting is summarised in Algorithm 6.

5.4.3 Proposed Differential Reporting Mechanism

A differential reporting method is proposed where in contrast to periodic reporting, SUs report their local decisions only when there is a change in the locally detected PU state (i.e., bit 1 is sent when the local decision goes from idle to busy and bit 0 is sent when

Algorithm 6: On/Off reporting

Input : $\lambda \in \mathbb{R}^+$ ▷ Energy decision threshold
$N_s \in \mathbb{N}^+$ ▷ Number of signal samples
Output: $R_{ch,i} \in \{0, 1\}$ ▷ Channel state report

```
for each sensing event i do
    Y_i ← Energy of N_s samples                    ▷ Energy detection
    if Y_i ≥ λ then
        R_{ch,i} ← 1                               ▷ Flag channel as busy
        SU sends R_{ch,i} to FC
    else
        R_{ch,i} ← 0                               ▷ Flag channel as idle
        SU remains silent
    end
end
```

Algorithm 7: Differential reporting

Input : $\lambda \in \mathbb{R}^+$ ▷ Energy decision threshold
$N_s \in \mathbb{N}^+$ ▷ Number of signal samples
Output: $R_{ch,i} \in \{0, 1\}$ ▷ Channel state report

```
for each sensing event i do
    Y_i ← Energy of N_s samples                    ▷ Energy detection
    if Y_i ≥ λ then
        R_{ch,i} ← 1                               ▷ Flag channel as busy
    else
        R_{ch,i} ← 0                               ▷ Flag channel as idle
    end
    if R_{ch,i} = R_{ch,i−1} (i.e., same as previous state) then
        SU remains silent
    else
        SU sends R_{ch,i} to FC
    end
end
```

the local decision goes from busy to idle). When SUs remain silent, the FC assumes that the new detected state is the same as the last reported state. The differential reporting mechanism is summarised in Algorithm 7. For differential reporting, the FC needs to keep a copy of every SU last state (for comparison with new sensed states) to estimate the PU period durations.

The differential reporting mechanism is expected to have a significant impact on reporting overhead by reducing the amount of required reports and therefore increase the total system efficiency. This will be discussed in detail in Section 5.7.

5.4.4 Analysis of the Required Number of Reports

Closed form expressions for the expected number of reports for the periodic, On/Off and differential reporting mechanisms are derived for two scenarios: first, under perfect

spectrum sensing ($P_{fa} = P_{md} = 0$), second, under imperfect spectrum sensing (P_{fa}, $P_{md} > 0$).

First, the expected number of reports n_p for the periodic reporting mechanism, at both high SNRs (Perfect Spectrum Sensing, PSS) and low SNRs (Imperfect Spectrum Sensing, ISS) scenarios, is given by:

$$E\{n_p\} = \frac{E\{T_1\}}{T_s}\frac{N}{2} + \frac{E\{T_0\}}{T_s}\frac{N}{2}, \tag{5.8}$$

where $N \in \mathbb{N}^+$ is the total number of idle and busy periods in the observed set $\{\widetilde{T}_{i,n}\}_{n=1}^{N}$, $E\{T_0\}$ and $E\{T_1\}$ are the expected durations of idle and busy periods, respectively, and T_s is the sensing period. Notice that P_{fa} and P_{md} do not affect the total amount of reports since in the periodic reporting case a report is always sent in every sensing event.

Second, for the On/Off reporting mechanism, the expected number of reports n_{of} under PSS is given by:

$$E\{n_{of}\} = \frac{E\{T_1\}}{T_s}\frac{N}{2}, \tag{5.9}$$

while it can be easily seen that for ISS the expected number of reports is given by:

$$E\{n_{of}\} = \frac{E\{T_1\}}{T_s}\frac{N}{2}(1 - P_{md}) + \frac{E\{T_0\}}{T_s}\frac{N}{2}P_{fa}. \tag{5.10}$$

Lastly, for the differential reporting mechanism, the expected number of reports n_d under PSS is given by:

$$E\{n_d\} = N, \tag{5.11}$$

since under high SNR, the total number of reports sent to the FC is the same as the total number of periods, as one report is sent for every new observed period. On the other hand, the upper bound for the expected number of reports for differential reporting under ISS is found as follows:

$$E\{n_d\} = N + N\left[\frac{E\{T_1\}}{T_s}P_{md} + \frac{E\{T_0\}}{T_s}P_{fa}\right], \tag{5.12}$$

notice that one error (either false alarm or missed detection) will result in two reports. The upper bound in (5.12) is loose and can be approximated by taking into consideration the effect of the sensing error position within the period. For instance, consecutive sensing errors within the same period or sensing errors occurring at beginning or ending of the period result in a single report, then the following expression is obtained:

$$
\begin{aligned}
E\{n_d\} =& N + N\left[\frac{E\{T_1\}}{T_s}P_{md} + \frac{E\{T_0\}}{T_s}P_{fa}\right] - \\
&- NP_{md} - \sum_{k=2}^{\lfloor E\{T_1\}/T_s\rfloor} \frac{N}{2}\frac{E\{T_1\}}{T_s}P_{md}^{k} - \\
&- NP_{fa} - \sum_{k=2}^{\lfloor E\{T_0\}/T_s\rfloor} \frac{N}{2}\frac{E\{T_0\}}{T_s}P_{fa}^{k}.
\end{aligned} \tag{5.13}
$$

The previous analytical results are for a single CR and can be easily scaled up by multiplying by the number of cooperative SUs K.

5.5 Spectrum Sensing Data Falsification

In previous sections, all cooperative users are assumed to be honest. Unfortunately, given the openness nature of wireless communications, cognitive networks and advances in software defined radios have made the system vulnerable to data falsification attacks carried out by malicious or greedy nodes disguised [132]. MUs will send falsified reports. This type of attack is known as SSDF [133]. MUs have two main objectives for attacks [134]: first is to interfere with the primary system by having MUs report with idle states at busy primary channels, second is to report with busy states when local sensing decisions provide an idle state and as a result, the FC falsely declares the primary channel as busy so that legitimate SUs have to wait for another sensing event. Meanwhile, MUs can access the idle channel exclusively. This attack strategy is typically utilised by greedy MUs to maximise their data rate. In this chapter, the main focus is on the later scenario.

5.5.1 Spectrum Sensing Data Falsification Attacks

In this chapter, the considered SSDF attacks are similar to the ones described in [135–137], in which the MUs attack with a predefined attack strategy. Intelligent attacks such as those described in [138], where MUs are capable altering their attacks adaptively (based on the CR network) to maximise their own utilities are out of scope of this chapter and hence is left to future work. The considered SSDF attacks are:

1. Blind attack: The attackers report with busy state in every sensing event [133].
2. Random attack: The MUs attack (i.e., report an idle channel as busy) with a given probability of attack $P_a < 1$ [139].

The blind attack would have a devastating effect on the resulting global detection if it succeeds, however its detection is straightforward. Notice that under periodic and On/Off reporting, the MU would report a busy PU channel in 100% of the submitted reports, while under differential reporting the MU would indicate the channel as busy

in the initial report and then would not report anymore, implying that the channel still remains busy. These extreme cases would be very easy to detect by the FC by simply counting the number of reports and states sent by each user and comparing with the rest of users (taking into account the employed reporting mechanism). As a result, a modified version of the pure blind attack is here considered (see Algorithm 8), which is more sophisticated and therefore increases the chances of this type of attack to succeed. Notice that this modified blind attack requires MUs to sense the PU channel before sending a report to the FC, while the pure blind attack would not require any sensing at all. The random attack (Algorithm 9) also requires MUs to sense the PU channel before sending a report (regardless of the reporting mechanism employed) since the actual states of the PU channel need to be known in order to meet the desired probability of attack (P_a). Therefore, in both types of attack (blind and random) MUs need to sense the channel before sending the report to the FC.

For the case of differential reporting, in order to be able to apply SSDF attacks successfully, MUs need to follow the reporting rules imposed by the FC. Not following the reporting rules would lead to anomalous sequences of reports, with much higher/lower number of reports than the average, which would make the attack process susceptible of being detected by the FC. Thus, it is essential for MUs to follow the same reporting procedure imposed by the FC.

Finally, it is also worth mentioning that MUs may attack not only during the idle periods of the PU channel (by sending a busy report), but also during the busy periods of the PU channel (by sending an idle report), or a combination of both. While attacks during PU busy periods may be possible, in this case the MU does not obtain an individual benefit from leading the FC to believe that the channel is idle when it is actually busy and therefore the MU does not have a strong incentive to carry out such attack. On the other hand, leading the FC to believe that the channel is busy when it is actually free allows the MU to prevent other SUs from transmitting and hence use the PU channel idle times for its own transmissions. Therefore the MU does have a strong incentive to attack during idle periods (by sending a busy report), which is not the case during busy periods. Notice that the algorithms and analyses presented in this chapter can be readily adapted to the either type of attack by simply reverting idle/busy periods (both in the algorithms and analysis of results). However, in order to simplify the subsequent analysis and discussion, we restrict ourselves, without loss of generality, to the case where MUs attack during idle periods only.

5.5.2 Proposed Algorithm

To eliminate the effects of SSDF attacks, a secure and efficient data fusion is essential, which in turn requires a reliable defence reference to identify MUs [136]. However in practical scenarios, a reliable reference is not always available. Eventually honest reports are mixed with malicious ones. In this context, we propose a novel algorithm to identify contrived MUs reports without the requirement of a previous reference. The key idea

Algorithm 8: Modified blind attack (with differential reporting)

Input : $\lambda \in \mathbb{R}^+$ ▷ Energy decision threshold
$N_s \in \mathbb{N}^+$ ▷ Number of signal samples
Output: $R_{ch,i} \in \{0,1\}$ ▷ Channel state report

for *each sensing event* i **do**
  $Y_i \leftarrow$ Energy of N_s samples ▷ Energy detection
  if $Y_i \geq \lambda$ **then**
    $R_{ch,i} \leftarrow 1$ ▷ Flag channel as busy
  else
    $R_{ch,i} \leftarrow 0$ ▷ Flag channel as idle
  end
  if $R_{ch,i} = R_{ch,i-1}$ *(i.e., same as previous state)* **then**
    MU remains silent
  else
    MU sends $R_{ch,i} = 1$ to FC
  end
end

Algorithm 9: Random attack (with differential reporting)

Input : $\lambda \in \mathbb{R}^+$ ▷ Energy decision threshold
$N_s \in \mathbb{N}^+$ ▷ Number of signal samples
Output: $R_{ch,i} \in \{0,1\}$ ▷ Channel state report

for *each sensing event* i **do**
  $Y_i \leftarrow$ Energy of N_s samples ▷ Energy detection
  if $Y_i \geq \lambda$ **then**
    $R_{ch,i} \leftarrow 1$ ▷ Flag channel as busy
  else
    $R_{ch,i} \leftarrow 0$ ▷ Flag channel as idle
  end
  if $R_{ch,i} = R_{ch,i-1}$ *(i.e., same as previous state)* **then**
    MU remains silent
  else
    MU generates a radom number $Z \sim U(0,1)$
    if $Z < P_a$ **then**
      MU sends $R_{ch,i} = 1$ to FC
    else
      MU sends $R_{ch,i}$ to FC
    end
  end
end

of Algorithm 10 is based on the differential reporting mechanism. Whenever a report is available at the FC from a specific SU, a comparison is made with the previous report from the same SU. If the report contains information of same state as the previous report, then the report is discarded and the decision rule is applied based on the reports from the other $K-1$ SUs. Furthermore, the proposed algorithm can almost function in real-time without the need for a comparison with statistical characteristics for sensors as

Algorithm 10: Defence against attackers (with differential reporting)

Input : Reports from sensing nodes
Output: Decision
for *each report* $R_{k,i}$ *from* SU_k *in sensing event* i **do**
  if $R_{k,i} = R_{k,i-1}$ **then**
    $R_{k,i}$ is discarded
    Apply MAJORITY rule to $K-1$ SUs
  else
    Apply MAJORITY rule to K SUs
  end
end

the operation of obtaining accurate statistical information requires a significant sample size [140]. The proposed algorithm differs from the literature in that it is much simpler and does not require any pre-defined trusted nodes nor sophisticated rules at the FC.

5.6 Simulation and Experimental Methodology

The performance of the considered methods was evaluated both with simulations and hardware experiments. Simulations were performed in MATLAB by generating several sequences with a sufficiently large number of interleaved on/busy and off/idle periods from an exponential distribution. The simulation procedure can be summarised as follows:

1. Generate idle/busy periods' lengths T_i following an exponential distribution with predefined location (μ_i) and scale (λ_i) parameters.
2. Perform idle/busy sensing decisions H_0/H_1 on the generated sequence in step 1 every T_s time units (t.u.).
3. Calculate the idle/busy lengths estimated under PSS.
4. Add random errors (with $P_{fa} > 0$ and $P_{md} > 0$) in the sequence resulting from step 2.
5. Using the new H_0/H_1 sequence from step 4, calculate the period lengths $\widetilde{T}_i$ that would be estimated under ISS.
6. MUs will fake H_0 to H_1 with a given attack probability of $P_a > 0$.
7. FC computes the CDF of the idle/busy lengths obtained in steps 5 & 6 by applying a hard decision rule and compares with the CDF of the original periods.

The hardware experiments were conducted using the same PECAS platform described in Section 2.4.4. In general, the energy detection threshold can be selected through one of the following criteria:

- To meet a specific probability of false alarm (P_{fa}). This method requires knowledge of the SU noise power. In practice, this can be achieved by keeping the receiver function on an empty frequency channel for a sufficient time (several minutes in PECAS [103]) then setting the threshold to maintain the desired P_{fa} [141].
- To meet a specific probability of missed detection. This method requires knowledge of the received primary SNR in addition to the device noise power [142].
- To minimise the combined error from P_{fa} and P_{md}. This method also requires the knowledge of both the device noise power and primary signal SNR [143].

A more detailed description of these methods can be found in [144]. Since it is difficult to set accurately the energy detection threshold to result in a specific P_{fa} and P_{md} with the RTL-SDR [103], the errors are introduced through emulations to the on/off periods received by the RTL-SDR.

Even though the original PECAS is designed for a single CR scenario, the experiments are repeated for the required number of SUs to produce different streams for every SU and emulate a cooperative estimation scenario.

5.7 Simulation and Experimental Results

In this section, the analysis and validation of the proposed methods are provided. The value considered for each parameter is shown in the title of each figure in terms of generic time units (t.u.). In the case of experimental results, where a particular time unit needs to be selected according to the real-time capabilities of the employed hardware platform, the reference unit is the second (i.e., 1 t.u. = 1 second). First, different decision rules will be assessed, followed by different methods to assess the estimated primary distribution accuracy. The comparison between the estimated and original distributions is performed using the classic Kolmogorov-Smirnov (KS) distance [96], defined as:

$$D_{KS} = \sup_{T_i} \left| F_{T_i}(T_i) - F_{\widetilde{T}_i}(T_i) \right|, \tag{5.14}$$

where $F_{T_i}(T_i)$ and $F_{\widetilde{T}_i}(T_i)$ represent the CDFs of the original and estimated periods, respectively.

Fig. 5.3 compares the estimation accuracy of the considered hard decision rules (AND, OR, MAJORITY) when the cooperative SUs use periodic reporting and the FC uses the DEM to estimate the CDF of busy periods. For comparison purposes, the case of single SU is also included in Fig. 5.3. The duty cycle is set to 0.5 ($\Psi = 0.5$), where both busy and idle periods will have similar parameters. The MAJORITY rule outperforms the other rules in the estimation of the primary statistics (4 cooperative SUs can estimate accurately the primary statistics under $P_{fa} = P_{md} = 0.01$, while 12 SUs are required to estimate the primary statistics under $P_{fa} = P_{md} = 0.1$). As for the AND and the OR rules, both of them fail to provide an accurate estimation of the primary

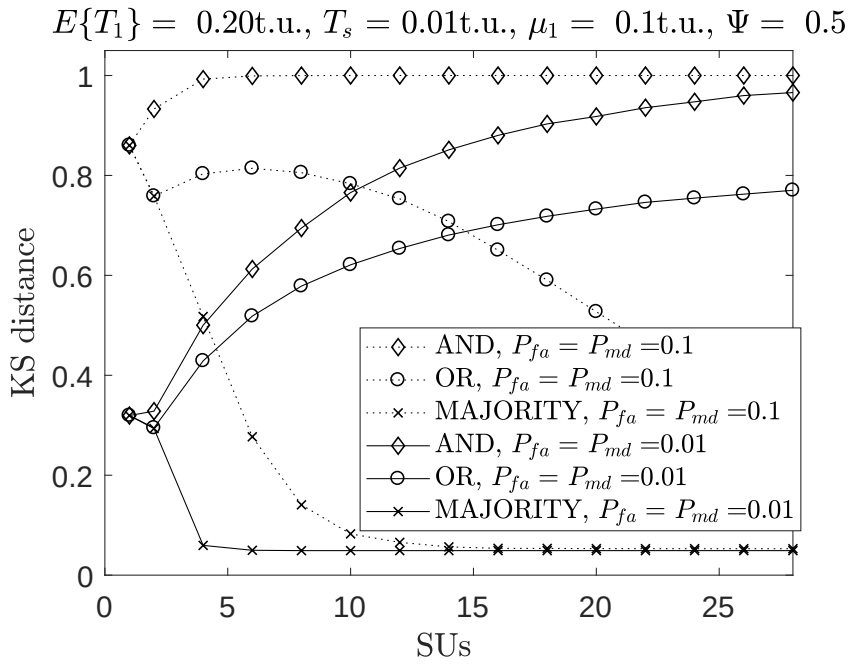
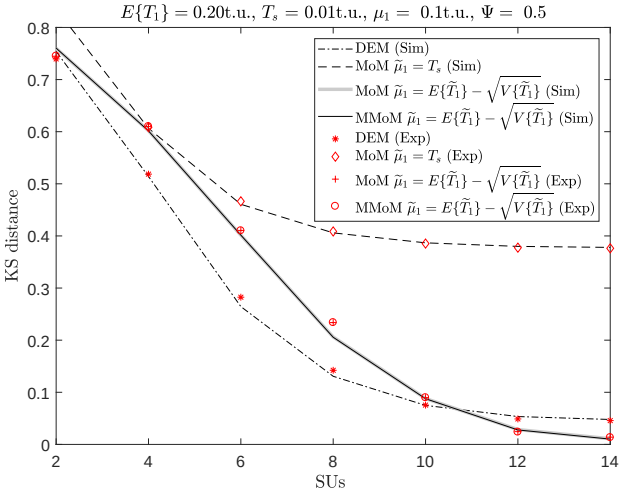


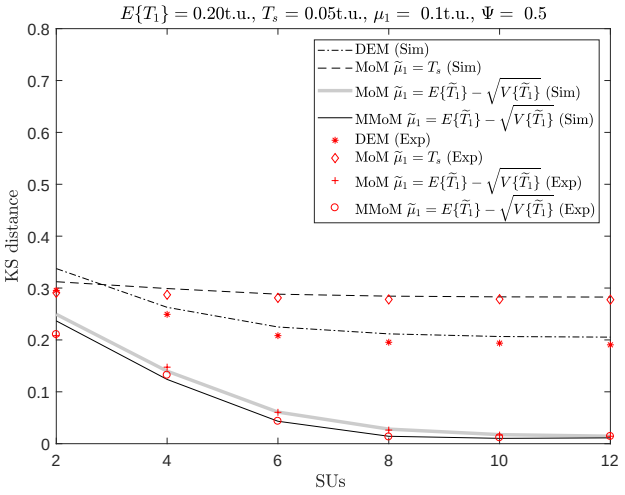
FIGURE 5.3: Accuracy of the estimated distribution for different fusion rules and periodic reporting under sensing errors.

distribution (low KS distance value) for both scenarios of high and low sensing error probabilities. For the OR rule, the obtained results can be explained as any CR reports with a busy period will result in the FC announcing the channel as busy and as the number of cooperative SUs increases the probability of having false alarms increases as well. It is interesting to notice that the direct estimation of the CDF (i.e., DEM, which is considered in all cases in Fig. 5.3) never reaches a perfect accuracy ($D_{KS} = 0$) regardless of the number of SUs and the fusion rule. This is a result of the finite sensing period T_s. Based on these results, further numerical results will only consider the MAJORITY fusion rule.

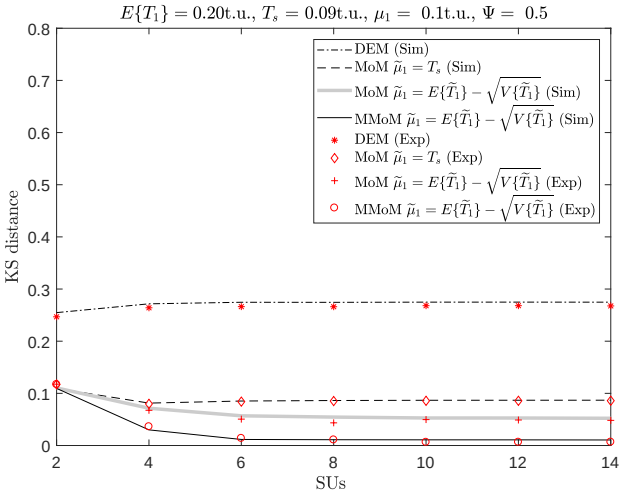
Fig. 5.4 shows the accuracy of the considered methods to estimate the distribution of the primary traffic for different sensing periods. Experimental results are considered only here due to the significant amount of time required to run experiments for cooperative SU scenarios using a single SU hardware platform. As it can be observed, the experimental results (with PECAS) provide a perfect fit with simulations. For small sensing periods (Fig. 5.4(a)), the DEM performs better than the MoM and its modified version in Section 5.3.2.3 (MMoM), but for high number of SUs, MoM and MMoM can provide a more accurate estimation. For higher sensing durations (Fig. 5.4(b) and Fig. 5.4(c)) MoM and MMoM provide better accuracy in the estimation of the primary traffic over the whole range of the number of cooperative SUs. The minimum period obtained from MMoM gives better estimation than the minimum obtained through the original MoM, except for the case where T_s has a small duration (i.e., multiple sensing events occur in a single period) where both minimums provide a similar KS distance. The direct estimated minimum with MoM provides results with significant inaccuracy regardless of



(a)



(b)



(c)

FIGURE 5.4: Different methods to estimate the distribution under periodic reporting: (a) $T_s = 0.01$ t.u., (b) $T_s = 0.05$ t.u., (c) $T_s = 0.09$ t.u.

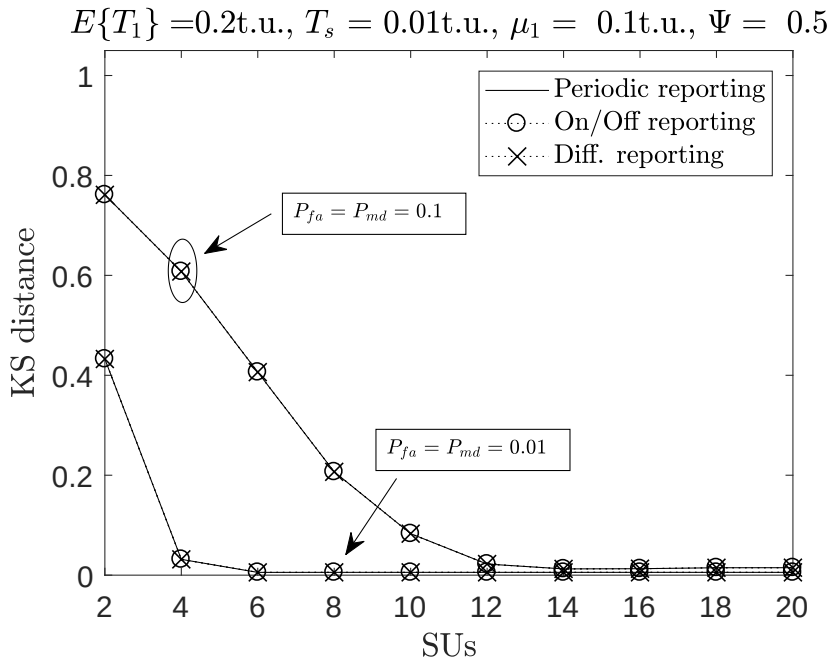


FIGURE 5.5: Accuracy of the estimated distribution for different reporting mechanisms.

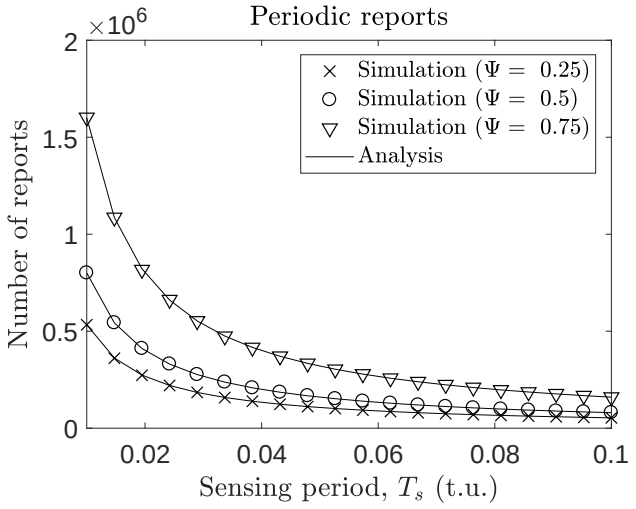
the sensing period or the number of cooperative SUs. Since the MMoM performs better than the rest of methods, it will be the only method considered in the remainder of this section.

The performance of the periodic reporting, On/Off reporting and the proposed differential reporting mechanisms for cooperative estimation will be discussed based on the MAJORITY fusion rule with MMoM distribution estimation. As it can be appreciated in Fig. 5.5, the three considered methods have a similar performance under sensing errors, however the differential reporting mechanism provides higher efficiency and security advantages in comparison with the other methods as discussed below.

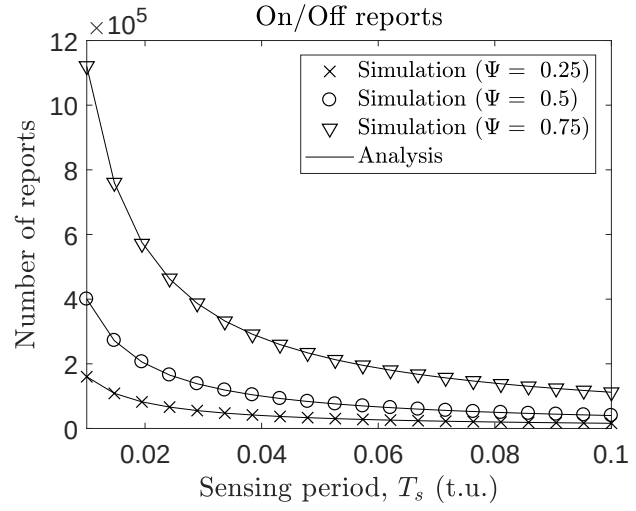
Fig. 5.6 shows the required number of channel reports for 20,000 periods for the three considered reporting mechanisms (periodic, On/Off and differential) under different primary loads (high $\Psi = 0.75$, moderate $\Psi = 0.5$ and low $\Psi = 0.25$). As it can be appreciated, the derived analytical expressions provide a perfect match for the periodic and On/Off reporting methods, while the result of (5.13) provides a tight upper bound for the required number of reports in the case of the differential reporting mechanism. The reduction in the amount of reports transmitted using the On/Off and differential reporting mechanisms with respect to the periodic reporting mechanism can be quantified, respectively, as:

$$B_{of} = \frac{E\{n_{of}\}}{E\{n_p\}}, \tag{5.15}$$

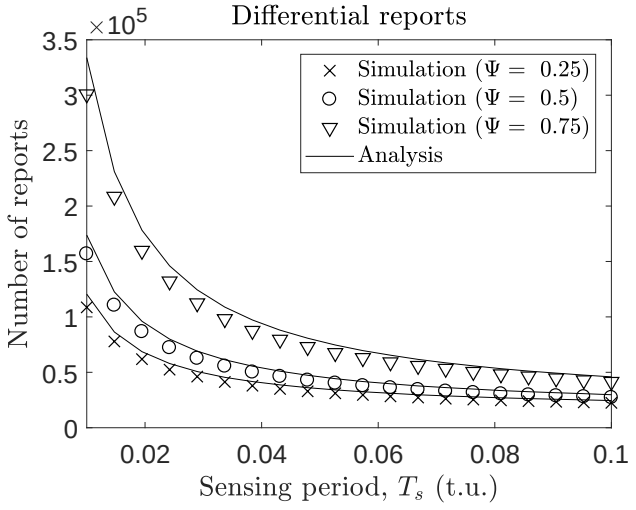
$$B_d = \frac{E\{n_d\}}{E\{n_p\}}, \tag{5.16}$$



(a)



(b)



(c)

FIGURE 5.6: Required number of reports under sensing errors ($P_{fa} = P_{md} = 0.1$) for: (a) Periodic reporting, (b) On/Off reporting, (c) Differential reporting.

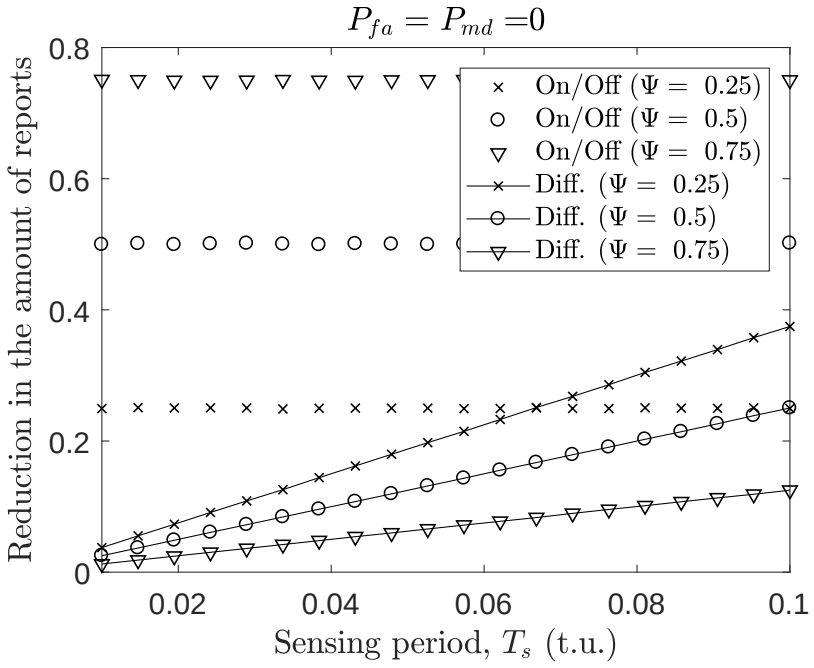


FIGURE 5.7: Reduction in the number of reports under PSS ($P_{fa} = P_{md} = 0$).

where B_{of} and B_d are the reduction in the amount of reports for On/Off and differential reporting mechanisms respectively. The smaller the value of B_{of}/B_d , the lower the amount of reporting overhead required for feedback and therefore the higher the efficiency. Figs. 5.7 and 5.8 show the reduction in the amount of reports for On/Off and differential reporting mechanisms with respect to the periodic reporting mechanism under perfect and imperfect spectrum sensing scenarios, respectively. The scenario of perfect spectrum sensing is considered to give an idea on the reduction in the case of high primary signal power present at the SU. As it can be concluded from both figures, the differential reporting mechanism outperforms the On/Off in nearly every channel load, except for small duty cycles ($\Psi = 0.25$) and large sensing periods ($T_s > \frac{\mu_i}{2}$) as at low duty cycles the SUs will remain idle for most of the time due to the absence of PU traffic. As it can be observed, the best estimation accuracy obtained for smaller sensing periods. In practice, the duty cycle of PU is unknown and the differential reporting mechanism provides higher efficiency. As it can be appreciated, the proposed mechanism reduces significantly the amount of required reports for all scenarios.

The accuracy of the estimation of primary traffic statistics under random attacks with different fusion rules is shown in Fig. 5.9. As it can be appreciated, the MAJORITY rule outperforms the AND/OR rules in the presence of attacks. The OR rule is ineffective against attacks because only one busy report is required to declare the channel as busy, therefore a single MU would be able to prevent the whole SU network from transmitting. The AND rule would be effective in an ideal case of perfect spectrum sensing, since a single honest SU who reports an idle channel as idle would be enough to make any attack fail, regardless of the number of MUs; however, in a realistic ISS scenario, the presence of sensing errors means that an idle channel may be reported as

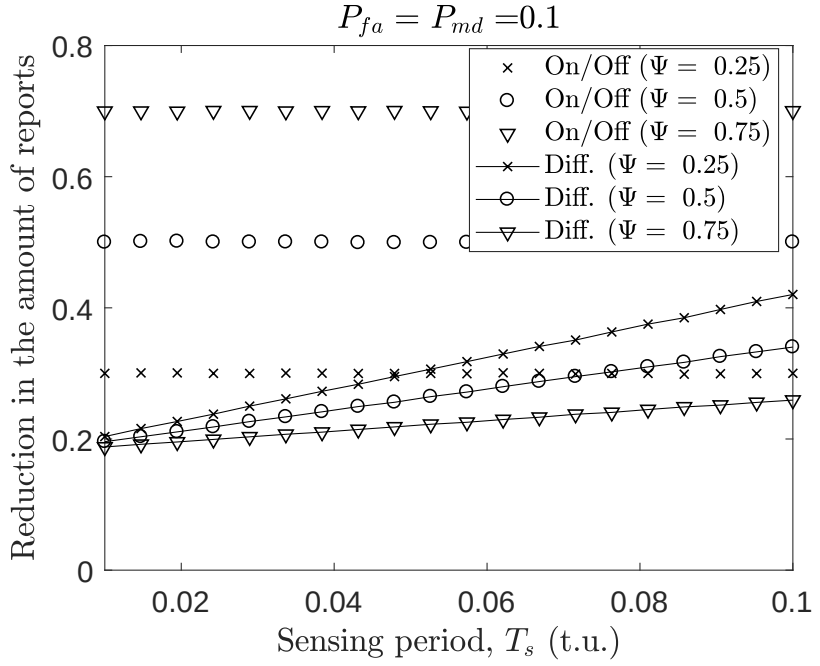


FIGURE 5.8: Reduction in the number of reports under ISS ($P_{fa} = P_{md} = 0.1$).

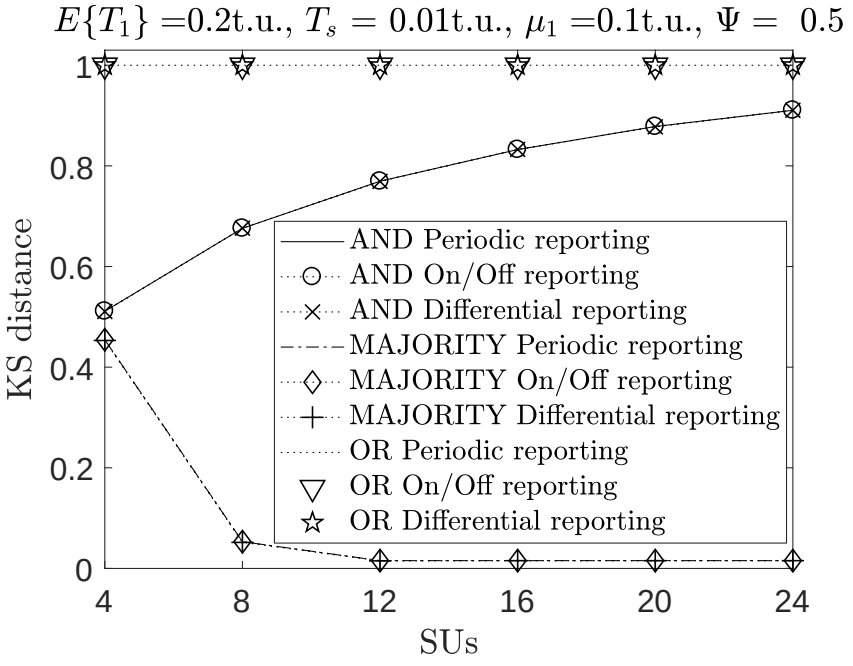


FIGURE 5.9: Accuracy of the estimated distribution for different fusion rules under both ISS ($P_{fa} = P_{md} = 0.01$) and random attacks (MUs = $K/4$, $P_a = 0.75$).

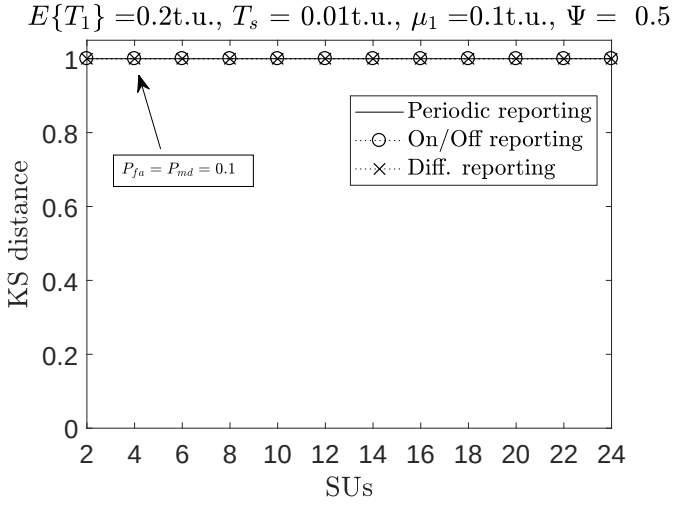
busy (false alarm) and vice versa (missed detection) even by honest SUs. Overall, the MAJORITY rule provides the best balance between malicious and erroneous reports, and therefore leads to the best estimation accuracy as observed in Fig. 5.9. By increasing the number of SUs, the MAJORITY rule enables an accurate estimation even under SSDF attacks. Comparing Figs. 5.3 and 5.9, it can be observed that the presence of MUs (Fig. 5.9) increases the total number of required SUs in order to achieve

an accurate estimation of the distribution with the MAJORITY rule with respect to the case of no MUs (Fig. 5.3), however the MAJORITY rule still provides the best estimation accuracy. Similar conclusions are obtained in the case of blind attacks (not shown here for brevity). Therefore, the MAJORITY rule provides the best estimation accuracy, even in the presence of SSDF attacks. The subsequent performance analysis under SSDF attacks will consider the MAJORITY fusion rule only.

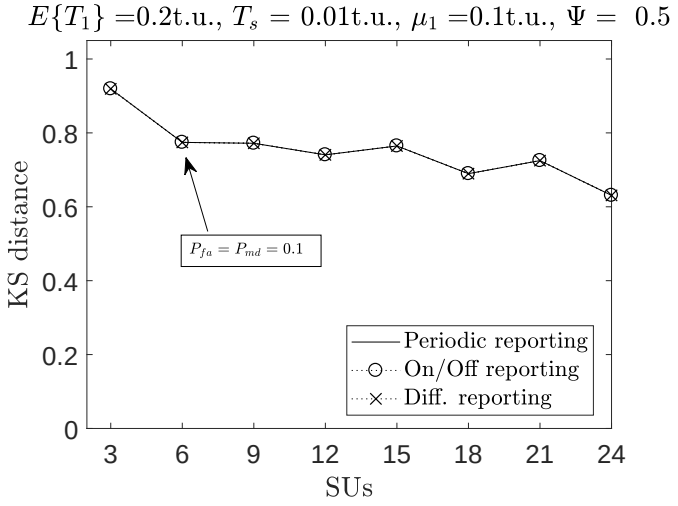
The accuracy of the estimation of primary traffic statistics under blind and random attacks is shown in Figs. 5.10 and 5.11 respectively. As it can be appreciated, the blind attack has the same level of degradation on the estimation of the PU distribution for the three reporting mechanisms (periodic, On/Off and differential). Moreover, when the population of attackers becomes half of the SUs (Fig. 5.10(a)), the FC will be overwhelmed with wrong reports and produce false global decisions regardless of the probability of missed detection and false alarm. For smaller MUs population (Fig. 5.10(c)), a large number of SUs is required to produce an accurate estimation of the primary statistics. The random attack has less severe effects on the estimation of the statistics in comparison with the blind attack. In fact, the blind attack is a special case of the random attack with attack probability $P_a = 1$. In general, the differential reporting mechanism performs better than the periodic and On/Off counterparts regardless of the P_a value. Nevertheless, all methods fail to provide an accurate estimation of the PU statistics except for small P_a (Fig. 5.11(a)), where a high number of SUs are essential to have a relatively acceptable estimation (SUs > 20). As it can be appreciated from Fig. 5.12, the proposed defense algorithm can significantly improve the estimation of primary statistics while mitigating the effects of MUs by discarding the contrived reports and keeping the correct ones for the cooperative estimation. Moreover, the proposed method provides accurate results regardless of the attack type or the population of MUs.

5.8 Summary

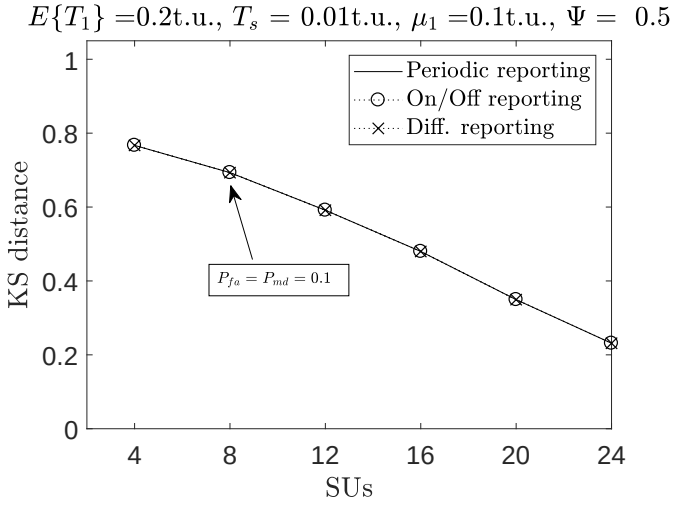
CR systems can benefit from the knowledge of PU activity statistics, which can be exploited to prevent interference and access the spectrum more efficiently. This information can be obtained individually by each CR user based on its local spectrum sensing observations, however a cooperative estimation approach can provide significant benefits both in terms of accuracy (overcoming the degrading effects of sensing errors) and reliability (overcoming the degrading effects of malicious users). In this context, this chapter has provided a detailed study on the cooperative estimation of the PU activity statistics (in particular, the distribution of the channel holding times) under both spectrum sensing errors and SSDF attacks. This chapter has evaluated the impact on the accuracy of the estimated statistics that several aspects may have, such as the hard decision rule used for cooperative sensing-based estimation (the MAJORITY rule was observed to provide the best performance) and the method employed to estimate the distribution (the MMoM approach proposed in this work has been proven to provide the



(a)

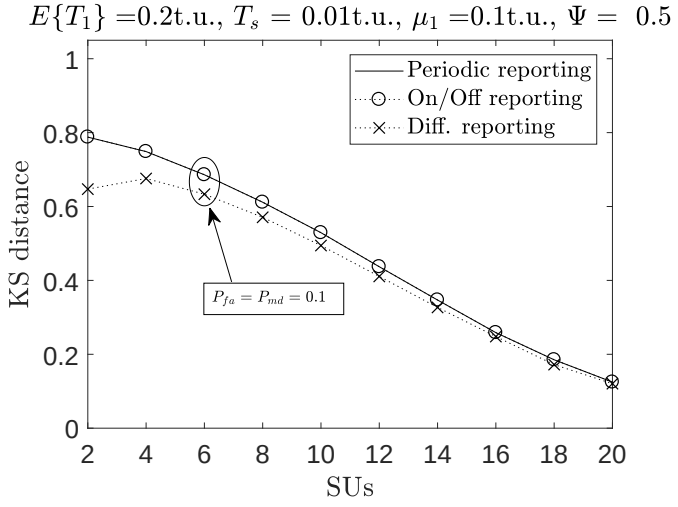


(b)

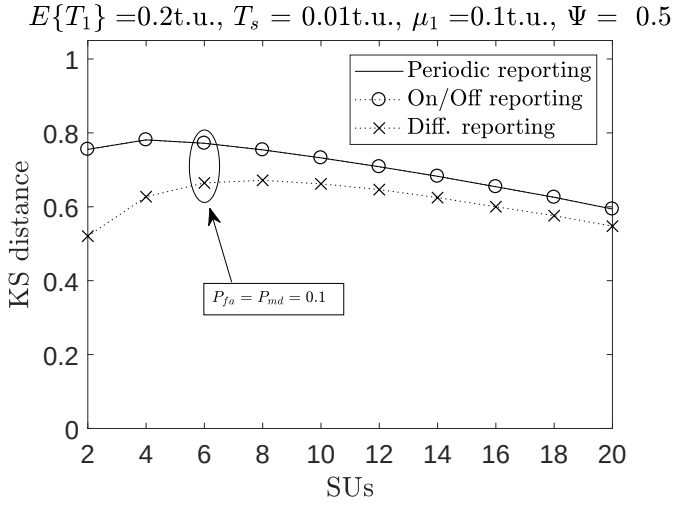


(c)

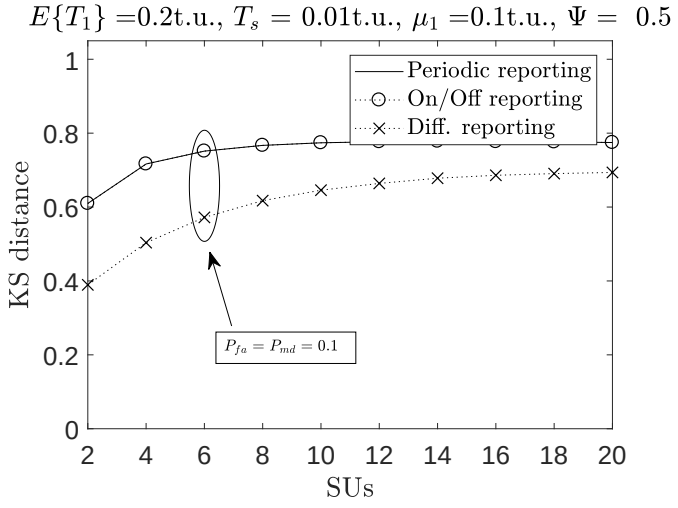
FIGURE 5.10: Accuracy of the estimated distribution under blind attacks: (a) MUs = $K/2$, (b) MUs = $K/3$, (c) MUs = $K/4$.



(a)

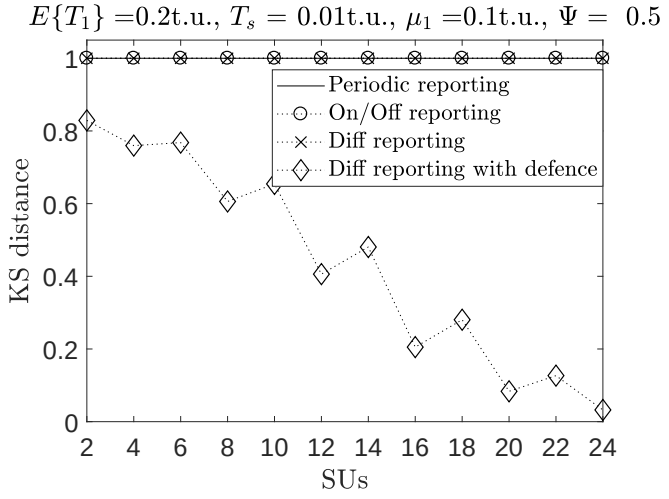


(b)

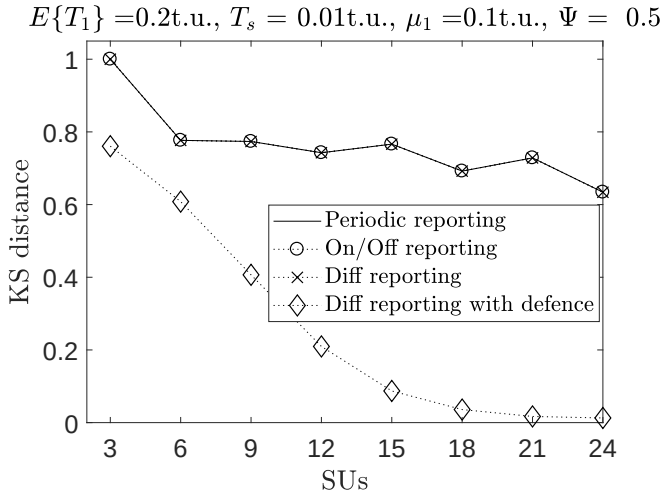


(c)

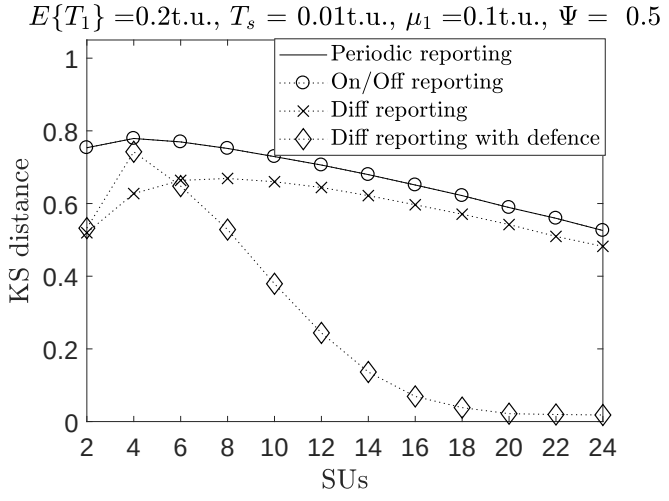
FIGURE 5.11: Accuracy of the estimated distribution under smart attacks: (a) MUs $= K/2$ and $P_a = 0.25$, (b) MUs $= K/2$ and $P_a = 0.5$, (c) MUs $= K/2$ and $P_a = 0.75$.



(a)



(b)



(c)

FIGURE 5.12: Accuracy of the estimated distribution under SSDF attacks with the proposed defence method: (a) MUs = $K/2$ and $P_a = 1$ (Blind attack), (b) MUs = $K/3$ and $P_a = 1$ (Blind attack), (c) MUs = $K/2$ and $P_a = 0.5$ (Smart attack).

most accurate estimation). While cooperative estimation can improve the estimation accuracy, it also increases the amount of signalling in the system (associated with the reporting overhead) and introduces security threads (from MUs deliberately sending incorrect reports). Both issues have been successfully addressed in this work by proposing a differential reporting mechanism that can decrease significantly the signalling overhead as well as a defence mechanism that can effectively remove both blind and random SSDF attacks. The obtained simulation and experimental results demonstrate that the methods proposed in this chapter enable a more accurate estimation of the PU activity statistics with a reduced level of signalling overhead and a high level of security against SSDF attacks.

Chapter 6

Conclusions and Future Work

The advances and growing demands for wireless networks and devices, lead to the problem of spectrum scarcity. Novel solutions to enhance the efficiency of wireless spectrum utilization are essential to solve the scarcity of available transmission bandwidth. DSA/CR is a promising solution to increase spectrum efficiency by accessing white spectrum holes in an opportunistic and non-interfering manner. The opportunistic access type to the PU channel implies that the DSA/CR system depends on their functionality on the PU activity time. Having accurate knowledge of PU traffic statistics can greatly enhance the performance of the CR network.

6.1 Conclusions

Spectrum sensing is one of the enabling techniques to detect PU channel availability and estimate traffic statistics. Unfortunately, spectrum sensing is hindered by several practical limitations when utilised to estimate PU statistics (such as the minimum transmission time, mean, variance, duty cycle, and distribution). The main objective of this thesis is to investigate the problem of improving the estimation of primary user traffic statistics under spectrum sensing. In this context, the main conclusions of the thesis are:

1. **Chapter 2**. The problem of finite sensing period is considered, which imposes limitations on the measured duration of busy/idle periods and hence the time resolution to which the resulting distribution for PU activity can be estimated. First, the analytical perspective of how this limitation affects the estimation of PU distribution under finite sensing periods is considered. Second, methods have been proposed to enable SU to obtain an accurate estimation of the distribution of PU activity periods. A solution is proposed based on the method of moments namely a modified version of the Method of Moments to improve the primary distribution estimation. Simulation results have shown that the proposed method outperforms the conventional approach based on the direct estimation by means of empirical CDF calculation as well as the approach based on the standard Method

of Moments. It has been found out that an accurate estimation or knowledge of the minimum PU activity time is essential to achieve an accurate estimation of the distribution of PU activity periods. Provided that the minimum PU activity time is known or can be estimated to a sufficient degree of accuracy, the proposed MMoM method constitutes an ideal solution to provide an accurate (nearly perfect) estimation of the PU activity statistics based on spectrum sensing observations.

2. **Chapter 3**. This chapter aims to answer the question of how many periods need to be observed in order to guarantee that the estimated statistics will meet a predefined level of accuracy for SNR scenarios. Closed form expressions are provided for the required sample size to attain the desired estimation error (for the minimum, mean, variance, duty cycle and distribution). The obtained analytical results have been compared to both simulation and experimental results, showing an excellent agreement in all cases. Moreover, an iterative stopping algorithm has been proposed to enable SU to perform a real-time calculation of the required sample size, which has been shown to be very accurate.

3. **Chapter 4**. In this chapter, two problems have been addressed. First, the problem of accurately estimating the PU activity statistics under ISS (low SNR scenario) three methods are proposed to overcome the degrading effects of spectrum sensing errors on the estimated statistics. The performance has been evaluated by means of computer simulations and compared to other methods previously proposed in the literature. The obtained results showed that the proposed methods outperformed the literature methods and provided an accurate estimation of the PU activity statistics. Moreover, multiple PU occupancy distributions have been considered and their impact on the accuracy of methods to estimate the PU activity statistics has been investigated by means of simulations under both PSS and ISS. The obtained results have demonstrated that the efficiency of different methods may depend on the particular PU activity statistics.

4. **Chapter 5**. Cooperative estimation of the PU activity statistics (in particular, the distribution of the channel holding times) under spectrum sensing errors and SSDF attacks is studied in detail. A novel reporting mechanism is proposed to both reduce the reporting overhead and the effects of SSDF attacks on the distribution estimation. This part of the thesis has evaluated the impact on the accuracy of the estimated statistics. While cooperative estimation can improve the estimation accuracy, it also increases the amount of signalling in the system (associated with the reporting overhead) and introduces security threads (from MUs deliberately sending incorrect reports). Both issues have been successfully addressed in this chapter by proposing a differential reporting mechanism that can decrease significantly the signalling overhead as well as a defence mechanism that can effectively remove both blind and random SSDF attacks. The obtained simulation and experimental results demonstrate that the methods proposed in this chapter enable

a more accurate estimation of the PU activity statistics with a reduced level of signalling overhead and a high level of security against SSDF attacks.

6.2 Future Work

This work has briefly discussed the estimation of primary traffic statistics. Future work for improving the estimation of primary user in cognitive radios could include the following issues. Some of them are already being addressed by the author. Nevertheless, extensions to this work can be pursued, including but not limited to:

1. **Closed form expressions to characterise the expected primary user traffic statistics under imperfect spectrum sensing**
   The work reported in Chapters 2 and 3 model the estimation of the primary statistics under both finite sensing length and a limited number of observations for SUs with high SNR (high power received at SU receiver from the PU transmitter) scenarios and for single SU. It would be both interesting and important to provide closed-form expressions to for the PU statistics in case of low SNR environment (sensing errors). In this case, several points are needed to be taken into account including the probabilities of errors (both P_{fa} and P_{md}). Also, more complex aspects such as the (random) number of errors and their (random) relative locations within each PU period, since this determines how the original PU periods would be split into shorter periods. Moreover, for the case of cooperative estimation, the number of SUs is needed to be taken into account. These expressions are essential for the design and deployment for cognitive radio systems.

2. **The design and implementation of an affordable test-bed for the cooperative estimation of the PU activity statistics**
   It might be interesting to provide a cooperative system test-bed for the cooperative PU traffic estimation followed by their limitations and constraints, which can provide a functional tool for proof-of-concept, validation, optimisation of algorithms and designs. The current work in this thesis is based on a single SU prototype and repeated multiple times to simulate the problem of cooperative estimation.

3. **Other approaches for the sample size analysis**
   The sample size analysis carried out in this work is based on intuitive and simple estimation methods for the considered activity statistics, the investigation of other estimation approaches based on more complex techniques that might potentially lead to accurate estimations with lower sample sizes would be a plausible extension to this work as well. In this context, machine learning techniques [145–150] have already demonstrated their potential benefits in other areas of wireless networks and their application to the problem here considered is also suggested as further future work.

4. **Differential reporting under intelligent attacks**
   Chapter 5 presented MUs performing SSDF attacks with defence algorithm based on differential reporting to minimise the attacks effects on the estimation of PU statistics in real-time. Future extension to the presented in Chapter 5, is to pair the differential reporting mechanism with other defence algorithms to tackle more sophisticated and intelligent attacks performed by MUs.

Bibliography

[1] J. S. Belrose, "Reginald aubrey fessenden and the birth of wireless telephony," *IEEE Antennas and Propagation Magazine*, vol. 44, no. 2, pp. 38–47, April 2002.

[2] PwC, "Number of network connected devices per person around the world from 2003 to 2020," 2015, accessed = 2018-07-01. [Online]. Available: https://www.statista.com/statistics/678739/forecast-on-connected-devices-per-person/

[3] R. W. Lucky, "The precious radio spectrum [reflections]," *IEEE Spectrum*, vol. 38, no. 9, pp. 90–90, Sept 2001.

[4] S. Pagadarai and A. M. Wyglinski, "A quantitative assessment of wireless spectrum measurements for dynamic spectrum access," in *2009 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, June 2009, pp. 1–5.

[5] K. Patil, R. Prasad, and K. Skouby, "A survey of worldwide spectrum occupancy measurement campaigns for cognitive radio," in *2011 International Conference on Devices and Communications (ICDeCom)*, Feb 2011, pp. 1–5.

[6] J. Xue, Z. Feng, and K. Chen, "Beijing spectrum survey for cognitive radio applications," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, Sept 2013, pp. 1–5.

[7] M. Hyhty, A. Mmmel, M. Eskola, M. Matinmikko, J. Kalliovaara, J. Ojaniemi, J. Suutala, R. Ekman, R. Bacchus, and D. Roberson, "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2386–2414, Fourthquarter 2016.

[8] M. López-Benítez, "Spectrum usage models for the analysis, design and simulation of cognitive radio networks," Ph.D. dissertation, Universitat Politècnica de Catalunya, 2011.

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

[10] C. H. Liu, P. Pawelczak, and D. Cabric, "Primary user traffic classification in dynamic spectrum access networks," *IEEE Journal on Selected Areas in Comms.*, vol. 32, no. 11, pp. 2237–2251, November 2014.

[11] M. López-Benítez and F. Casadevall, *Spectrum usage models for the analysis, design and simulation of cognitive radio networks.* Dordrecht: Springer Netherlands, 2012, ch. 2, pp. 27–73.

[12] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/-dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[13] J. Mitola, *Cognitive radio architecture: The engineering foundations of radio XML.* Hoboken, N.J. : Wiley-Interscience, 2006., 2006.

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

[15] Y. C. Liang, K. C. Chen, G. Y. Li, and P. Mahonen, "Cognitive radio networking and communications: An overview," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3386–3407, Sept 2011.

[16] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov 2005, pp. 328–337.

[17] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[18] J. 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*, vol. 61, no. 5, pp. 2271–2293, Jun 2012.

[19] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey, "IEEE 802.11af: a standard for TV white space spectrum sharing," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 92–100, October 2013.

[20] C. S. Sum, L. Lu, M. T. Zhou, F. Kojima, and H. Harada, "Design considerations of IEEE 802.15.4m low-rate WPAN in TV white space," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 74–82, April 2013.

[21] A. A. Franklin, J. Pak, H. Jung, S. Kim, S. You, J. Um, S. Lim, G. Ko, S. Hwang, B. Jeong, M. Song, and C. Kim, "Cognitive radio test-bed based on ECMA-392 international standard," in *2010 7th International Symposium on Wireless Communication Systems*, Sept 2010, pp. 1026–1030.

[22] J. Wang, M. S. Song, S. Santhiveeran, K. Lim, G. Ko, K. Kim, S. H. Hwang, M. Ghosh, V. Gaddam, and K. Challapali, "First cognitive radio networking standard for personal/portable devices in TV white spaces," in *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, April 2010, pp. 1–12.

[23] V. Osa, C. Herranz, J. F. Monserrat, and X. Gelabert, "Implementing opportunistic spectrum access in LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 99, Mar 2012. [Online]. Available: https://doi.org/10.1186/1687-1499-2012-99

[24] X. Hong, J. Wang, C. X. Wang, and J. Shi, "Cognitive radio in 5g: a perspective on energy-spectral efficiency trade-off," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 46–53, July 2014.

[25] K. A. Yau, J. Qadir, C. Wu, M. A. Imran, and M. H. Ling, "Cognition-inspired 5g cellular networks: A review and the road ahead," *IEEE Access*, vol. 6, pp. 35 072–35 090, 2018.

[26] F. Tseng, L. Chou, H. Chao, and J. Wang, "Ultra-dense small cell planning using cognitive radio network toward 5G," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 76–83, December 2015.

[27] A. Mukherjee, J. Cheng, S. Falahati, H. Koorapaty, D. H. Kang, R. Karaki, L. Falconetti, and D. Larsson, "Licensed-assisted access LTE: coexistence with IEEE 802.11 and the evolution toward 5G," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 50–57, June 2016.

[28] Q. Chen, G. Yu, and Z. Ding, "Optimizing unlicensed spectrum sharing for LTE-U and WiFi network coexistence," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2562–2574, Oct 2016.

[29] Q. Zhao, "A survey of dynamic spectrum access: signal processing, networking, and regulatory policy," in *in IEEE Signal Processing Magazine*, 2007, pp. 79–89.

[30] T. A. Le and K. Navaie, "On the interference tolerance of the primary system in cognitive radio networks," *IEEE Wireless Communications Letters*, vol. 4, no. 3, pp. 281–284, June 2015.

[31] M. G. Khoshkholgh, K. Navaie, and H. Yanikomeroglu, "Access strategies for spectrum sharing in fading environment: Overlay, underlay, and mixed," *IEEE Transactions on Mobile Computing*, vol. 9, no. 12, pp. 1780–1793, Dec 2010.

[32] M. E. Tanab and W. Hamouda, "Resource allocation for underlay cognitive radio networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 1249–1276, Secondquarter 2017.

[33] C. T. Charles, "Formalizing the interference temperature model," *Wireless Communications and Mobile Computing*, vol. 7, no. 9, pp. 1077–1086. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcm.482

[34] A. Alsharoa, H. Ghazzai, E. Yaacoub, and M. Alouini, "Bandwidth and power allocation for two-way relaying in overlay cognitive radio systems," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 911–916.

[35] H. Sun, A. Nallanathan, C. X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: A survey," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 74–81, April 2013.

[36] 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, First 2009.

[37] A. Ali and W. Hamouda, "Advances on spectrum sensing for cognitive radio networks: Theory and applications," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 1277–1304, Secondquarter 2017.

[38] E. Hossain, D. Niyato, and D. I. Kim, "Evolution and future trends of research in cognitive radio: a contemporary survey," *Wireless Communications and Mobile Computing*, vol. 15, no. 11, pp. 1530–1564. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcm.2443

[39] S. K. Sharma, T. E. Bogale, S. Chatzinotas, B. Ottersten, L. B. Le, and X. Wang, "Cognitive radio techniques under practical imperfections: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 1858–1884, Fourthquarter 2015.

[40] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, April 1967.

[41] A. Mariani, A. Giorgetti, and M. Chiani, "Effects of noise power estimation on energy detection for cognitive radio applications," *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3410–3420, December 2011.

[42] M. López-Benítez and F. Casadevall, "Improved energy detection spectrum sensing for cognitive radio," *IET Communications*, vol. 6, no. 8, pp. 785–796, May 2012.

[43] L. Claudino and T. Abrão, "Spectrum sensing methods for cognitive radio networks: A review," *Wirel. Pers. Commun.*, vol. 95, no. 4, pp. 5003–5037, Aug. 2017. [Online]. Available: https://doi.org/10.1007/s11277-017-4143-1

[44] F. Salahdine, H. E. Ghazi, N. Kaabouch, and W. F. Fihri, "Matched filter detection with dynamic threshold for cognitive radio networks," in *2015 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Oct 2015, pp. 1–6.

[45] G. Huang and J. K. Tugnait, "On cyclostationarity based spectrum sensing under uncertain gaussian noise," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2042–2054, April 2013.

[46] A. V. Dandawate and G. B. Giannakis, "Statistical tests for presence of cyclostationarity," *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2355–2369, Sept 1994.

[47] E. Biglieri, A. Goldsmith, L. Greenstein, N. Mandayam, and H. Poor, *Principles of Cognitive Radio*. Cambridge University Press, 2012. [Online]. Available: https://books.google.co.uk/books?id=KwkgAwAAQBAJ

[48] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 32–39, April 2008.

[49] A. F. Molisch, L. J. Greenstein, and M. Shafi, "Propagation issues for cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 787–804, May 2009.

[50] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov 2005, pp. 131–136.

[51] G. Taricco, "Optimization of linear cooperative spectrum sensing for cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 77–86, Feb 2011.

[52] Z. Quan, S. Cui, and A. H. 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, Feb 2008.

[53] J. Lai, E. Dutkiewicz, R. P. Liu, and R. Vesilo, "Comparison of cooperative spectrum sensing strategies in distributed cognitive radio networks," in *2012 IEEE Global Communications Conference (GLOBECOM)*, Dec 2012, pp. 1513–1518.

[54] J. Unnikrishnan and V. V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 18–27, Feb 2008.

[55] Y. Saleem and M. H. Rehmani, "Primary radio user activity models for cognitive radio networks: A survey," *Journal of Network and Computer Applications*, vol. 43, pp. 1 – 16, 2014.

[56] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework, year=2007, volume=25, number=3, pages=589-600, issn=0733-8716, month=April,," *IEEE Journal on Selected Areas in Communications*.

[57] D. Bertsekas and J. Tsitsiklis, *Introduction to Probability*, ser. Athena Scientific books. Athena Scientific, 2002. [Online]. Available: https://books.google.co.uk/books?id=bcHaAAAAMAAJ

[58] D. Datla, R. Rajbanshi, A. M. Wyglinski, and G. J. Minden, "Parametric adaptive spectrum sensing framework for dynamic spectrum access networks," in *2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, April 2007, pp. 482–485.

[59] Z. Wang, Y. H. Chew, and C. Yuen, "On discretizing the exponential on-off primary radio activities in simulations," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2011, pp. 556–560.

[60] M. H. Rehmani, A. Carneiro Viana, H. Khalife, and S. Fdida, "Activity Pattern Impact of Primary Radio Nodes on Channel Selection Strategies," in *CogART 2011 - 4th International Workshop on Cognitive Radio and Advanced Spectrum Management*. Barcelona, Spain: ACM, Oct. 2011, pp. 36:1–36:5. [Online]. Available: https://hal.inria.fr/inria-00630230

[61] S. Wang, J. Zhang, and L. Tong, "A characterization of delay performance of cognitive medium access," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 800–809, February 2012.

[62] L. Jiao, E. Song, V. Pla, and F. Y. Li, "Capacity upper bound of channel assembling in cognitive radio networks with quasistationary primary user activities," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1849–1855, May 2013.

[63] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in WLAN channels: Empirical model and its stochastic analysis," in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum*, ser. TAPAS '06. New York, NY, USA: ACM, 2006. [Online]. Available: http://doi.acm.org/10.1145/1234388.1234402

[64] M. López-Benítez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2091–2104, Jun 2013.

[65] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive radios for dynamic spectrum access - dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, May 2007.

[66] M. Wellens, A. D. Baynast, and P. Mahonen, "Performance of dynamic spectrum access based on spectrum occupancy statistics," *IET Communications*, vol. 2, no. 6, pp. 772–782, July 2008.

[67] K. Umebayashi, Y. Suzuki, and J. J. Lehtomäki, "Dynamic selection of CWmin in cognitive radio networks for protecting IEEE 802.11 primary users," in *Proc. 2011 6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, June 2011, pp. 266–270.

[68] M. Wellens, J. Riihijrvi, and P. Mähönen, "Empirical time and frequency domain models of spectrum use," *Physical Communication*, vol. 2, no. 1, pp. 10 – 32, 2009.

[69] X. Liu, B. Krishnamachari, and H. Liu, "Channel selection in multi-channel opportunistic spectrum access networks with perfect sensing," in *Proc. 2010 IEEE Int'l. Symp. Dyn. Spect. Access Networks (DySPAN 2010)*, Apr. 2010, pp. 1–8.

[70] A. Kaur, S. Sharma, and A. Mishra, "Sensing period adaptation for multiobjective optimisation in cognitive radio using Jaya algorithm," *Electronics Letters*, vol. 53, no. 19, pp. 1335–1336, 2017.

[71] H. Kim and K. G. Shin, "Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 533–545, May 2008.

[72] K. Umebayashi, Y. Suzuki, and J. J. Lehtomki, "Dynamic selection of cwmin in cognitive radio networks for protecting ieee 802.11 primary users," in *2011 6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, June 2011, pp. 266–270.

[73] K. Umebayashi, K. Hayashi, and J. J. Lehtomäki, "Threshold-setting for spectrum sensing based on statistical information," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1585–1588, July 2017.

[74] C. H. Liu, W. Gabran, and D. Cabric, "Prediction of exponentially distributed primary user traffic for dynamic spectrum access," in *Proc. 2012 IEEE Global Communications Conference (GLOBECOM)*, Dec 2012, pp. 1441–1446.

[75] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y. D. Yao, and L. Hanzo, "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2017.

[76] V. Southivong and T. Fujii, "Primary and secondary identification using energy detection with statistical primary information," in *2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, March 2015, pp. 164–169.

[77] M. Ghaznavi and A. Jamshidi, "Defence against primary user emulation attack using statistical properties of the cognitive radio received power," *IET Communications*, vol. 11, no. 9, pp. 1535–1542, 2017.

[78] M. López-Benítez, *Cognitive radio*. Cambridge University Press, 2013, ch. 13, pp. 383–425.

[79] J. J. Lehtomäki, M. López-Benítez, K. Umebayashi, and M. Juntti, "Improved channel occupancy rate estimation," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 643–654, March 2015.

[80] W. Gabran, C. H. Liu, P. Pawelczak, and D. Cabric, "Primary user traffic estimation for dynamic spectrum access," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 544–558, March 2013.

[81] M. López-Benítez, "Can primary activity statistics in cognitive radio be estimated under imperfect spectrum sensing?" in *Proc. 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 750–755.

[82] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40 – 62, 2011.

[83] A. Ghasemi and E. S. Sousa, "Opportunistic spectrum access in fading channels through collaborative sensing," vol. 2, no. 2, pp. 71–82, Mar 2007.

[84] K. Cicho, A. Kliks, and H. Bogucka, "Energy-efficient cooperative spectrum sensing: A survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1861–1886, thirdquarter 2016.

[85] J. Imtiaz and D. Kim, "Energy-efficient management of cognitive radio terminals with quality-based activation," *IEEE Communications Letters*, vol. 21, no. 5, pp. 1171–1174, May 2017.

[86] J. R. Long, W. Wu, Y. Dong, Y. Zhao, M. A. T. Sanduleanu, J. F. M. Gerrits, and G. van Veenendaal, "Energy-efficient wireless front-end concepts for ultra lower power radio," in *2008 IEEE Custom Integrated Circuits Conference*, Sept 2008, pp. 587–590.

[87] Z. Khan, J. Lehtomaki, K. Umebayashi, and J. Vartiainen, "On the selection of the best detection performance sensors for cognitive radio networks," *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 359–362, April 2010.

[88] K. Umebayashi, J. J. Lehtomäki, T. Yazawa, and Y. Suzuki, "Efficient decision fusion for cooperative spectrum sensing based on OR-rule," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2585–2595, July 2012.

[89] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-noy, and T. Brown, "A survey of sensor selection schemes in wireless sensor networks," in *In SPIE Defense*

and Security Symposium Conference on Unattended Ground, Sea, and Air Sensor Technologies and Applications IX, 2007.

[90] T. Zhao and Y. Zhao, "A new cooperative detection technique with malicious user suppression," in *2009 IEEE International Conference on Communications*, June 2009, pp. 1–5.

[91] Y. Al-Mathehaji, S. Boussakta, M. Johnston, and H. Fakhrey, "Defeating SSDF attacks with trusted nodes assistance in cognitive radio networks," *IEEE Sensors Letters*, vol. 1, no. 4, pp. 1–4, Aug 2017.

[92] J. Wu, X. Li, T. Song, L. Zhang, M. Liu, and J. Hu, "Two-stage credit threshold on cooperative spectrum sensing to exclude malicious users in mobile cognitive radio networks," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, June 2017, pp. 1–6.

[93] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive radios for dynamic spectrum access - dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, May 2007.

[94] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, ser. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley & Sons, 1995, no. v. 2.

[95] R. J. Larsen and M. L. Marx, *An introduction to mathematical statistics and its applications; 4th ed.* Boston, MA: Prentice Hall, 2006.

[96] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes: The art of scientific computing*, 3rd ed. Cambridge University Press, 2007.

[97] M. López-Benítez, F. Casadevall, A. Umbert, J. Perez-Romero, R. Hachemani, J. Palicot, and C. Moy, "Spectral occupation measurements and blind standard recognition sensor for cognitive radio networks," in *2009 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, June 2009, pp. 1–9.

[98] M. López-Benítez, "Cognitive radio," in *Heterogeneous cellular networks: Theory, simulation and deployment.* Cambridge University Press, 2013, ch. 13.

[99] K. A. Stroud and D. J. Booth, *Engineering mathematics.* Basingstoke : Palgrave Macmillan, 2013., 2013.

[100] MATLAB, *version 7.10.0 (R2016b).* Natick, Massachusetts: The MathWorks Inc., 2016.

[101] D. Ruppert, *Statistics and Data Analysis for Financial Engineering (Springer Texts in Statistics)*, 1st ed. Springer, Berlin, 2010. [Online]. Available: http://www.amazon.de/Statistics-Analysis-Financial-Engineering-Springer/dp/1441977864

[102] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, 1997.

[103] M. López-Benítez, A. Al-Tahmeesschi, K. Umebayashi, and J. Lehtomäki, "PECAS: A low-cost prototype for the estimation of channel activity statistics in cognitive radio," in *Proc. 2017 IEEE Wireless Comms. & Networking Conf. (WCNC)*, March 2017, pp. 1–6.

[104] M. López-Benítez and J. Lehtomäki, "On the sensing sample size for the estimation of primary channel occupancy rate in cognitive radio," in *Proc. 2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.

[105] A. Al-Tahmeesschi, M. López-Benítez, K. Umebayashi, and J. Lehtomäki, "Analytical study on the estimation of primary activity distribution based on spectrum sensing," in *Proc. 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2017), Workshop on Cognitive Radio and Innovative Spectrum Sharing Paradigms for Future Networks (CRAFT 2017)*, Montreal, Quebec, Canada, 8-13 Oct. 2017, pp. 1–5.

[106] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki, and K. Umebayashi, "Accurate estimation of primary user traffic based on periodic spectrum sensing," in *Proc. 2018 IEEE Wireless Comms. & Networking Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[107] A. Mood, F. Graybill, and D. Boes, *Introduction to the Theory of Statistics.* McGraw-Hill, 1974.

[108] I. R. Savage, "Probability inequalities of the Tchebycheff type," *Journal of Research of the National Bureau of Standards-B. Mathematics and Mathematical Physics B*, vol. 65, no. 3, pp. 211–222, 1961.

[109] H. N. Phien, "On the computation of sample central moments," *Int'l. J. Mathematical Education in Science and Tech.*, vol. 19, no. 3, pp. 403–412, May 1988.

[110] I. Hughes and T. P. A. Hase, *Measurements and their uncertainties: A practical guide to modern error analysis.* Oxford University Press, 2010.

[111] H. H. Ku, "Notes on the use of propagation of error formulas," *Journal of Research of the National Bureau of Standards. Section C: Engineering and Instrumentation*, vol. 70C, no. 4, pp. 263–273, Oct. 1966.

[112] Q. Liang, M. Liu, and D. Yuan, "Channel estimation for opportunistic spectrum access: uniform and random sensing," *IEEE Trans. Mobile Computing*, vol. 11, no. 8, pp. 1304–1316, Aug. 2012.

[113] P. Tehrani, L. Tong, and Q. Zhao, "Asymptotically efficient multichannel estimation for opportunistic spectrum access," *IEEE Trans. Signal Processing*, vol. 60, no. 10, pp. 5347–5360, Oct. 2012.

[114] W. Saad, Z. Han, H. V. Poor, T. Basar, and J. B. Song, "A cooperative bayesian nonparametric framework for primary user activity monitoring in cognitive radio networks," *IEEE J. Sel. Areas Comms.*, vol. 30, no. 9, pp. 1815–1822, Oct. 2012.

[115] S. Geirhofer, L. Tong, and B. M. Sadler, "A measurement-based model for dynamic spectrum access in WLAN channels," in *Proc. IEEE Military Comms. Conf. (MILCOM 2006)*, Oct. 2006, pp. 1–7.

[116] L. Stabellini, "Quantifying and modeling spectrum opportunities in a real wireless environment," in *Proc. IEEE Wireless Comms. and Networking Conf. (WCNC 2010)*, Apr. 2010, pp. 1–6.

[117] M. Wellens, J. Riihijärvi, and P. Mähönen, "Empirical time and frequency domain models of spectrum use," *Physical Comm.*, vol. 2, no. 1-2, pp. 10–32, Mar. 2009.

[118] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, 2nd ed. Wiley, Nov. 1994, vol. 1.

[119] L. Zhang, "Sample mean and sample variance: Their covariance and their (in)dependence," *The American Statistician*, vol. 61, no. 2, pp. 159–160, 2007.

[120] M. López-Benítez, A. Al-Tahmeesschi, and D. Patel, "Accurate estimation of the minimum primary channel activity time in cognitive radio based on periodic spectrum sensing observations," in *Proc. 24th European Wireless Conference (EW 2018)*, May 2018, pp. 131–136.

[121] J. Lehtomäki, M. López-Benítez, K. Umebayashi, and M. Juntti, "Improved channel occupancy rate estimation," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 643–654, March 2015.

[122] M. López-Benítez and F. Casadevall, "Empirical time-dimension model of spectrum use based on a discrete-time markov chain with deterministic and stochastic duty cycle models," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 6, pp. 2519–2533, July 2011.

[123] K. Sithamparanathan and A. Giorgetti, *Cognitive radio techniques*. Artech House, 2012.

[124] K. Cafe, "Gsm timeslot and frequency specifications - rf cafe," 2016. [Online]. Available: http://www.rfcafe.com/references/electrical/gsm-specs.htm

[125] J. F. Lawless, *Statistical models and methods for lifetime data*. Wiley, 1982.

[126] R. D. Gupta and D. Kundu, "Generalized exponential distributions," *Australian and New Zealand Journal of Statistics*, vol. 41, no. 2, pp. 173–188, Jun. 1999.

[127] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, 3rd ed. John Wiley & Sons, 2003.

[128] T. Öztekin, "Comparison of parameter estimation methods for the three-parameter generalized Pareto distribution," *Turkish Journal of Agriculture and Forestry*, vol. 29, no. 6, pp. 419–428, Dec. 2005.

[129] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, 10th ed. New York: Dover, 1972.

[130] S. Kyperountas, N. Correal, and Q. Shi, "A comparison of fusion rules for cooperative spectrum sensing in fading channels," In Proceedings of the Wireless Symposium and Summer School, Blacksburg, VA, USA, 24 June 2010.

[131] S. Bae and H. Kim, "Robust cooperative sensing with ON/OFF signaling over imperfect reporting channels," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2196–2205, Dec 2016.

[132] V. Selis and A. Marshall, "A classification-based algorithm to detect forged embedded machines in iot environments," *IEEE Systems Journal*, pp. 1–11, 2018.

[133] L. Zhang, G. Ding, Q. Wu, Y. Zou, Z. Han, and J. Wang, "Byzantine attack and defense in cognitive radio networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 3, pp. 1342–1363, thirdquarter 2015.

[134] O. Fatemieh, R. Chandra, and C. A. Gunter, "Secure collaborative sensing for crowd sourcing spectrum data in white space networks," in *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, April 2010, pp. 1–12.

[135] N. Marchang, A. Taggu, and A. K. Patra, "Detecting byzantine attack in cognitive radio networks by exploiting frequency and ordering properties," *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, 2018.

[136] L. Zhang, G. Ding, Q. Wu, and F. Song, "Defending against byzantine attack in cooperative spectrum sensing: Defense reference and performance analysis," *IEEE Access*, vol. 4, pp. 4011–4024, 2016.

[137] L. Zhang, Q. Wu, G. Ding, S. Feng, and J. Wang, "Performance analysis of probabilistic soft SSDF attack in cooperative spectrum sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 81, May 2014. [Online]. Available: https://doi.org/10.1186/1687-6180-2014-81

[138] W. Wang, L. Chen, K. G. Shin, and L. Duan, "Thwarting intelligent malicious behaviors in cooperative spectrum sensing," *IEEE Transactions on Mobile Computing*, vol. 14, no. 11, pp. 2392–2405, Nov 2015.

[139] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 774–786, Feb 2011.

[140] M. López-Benítez and J. Lehtomäki, "On the sensing sample size for the estimation of primary channel occupancy rate in cognitive radio," in *Proc. IEEE Wireless Comms. and Networking Conf. (WCNC 2016)*, Apr. 2016, pp. 2599–2604.

[141] S. Atapattu, C. Tellambura, and H. Jiang, "Spectrum sensing via energy detector in low SNR," in *2011 IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–5.

[142] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Comms.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[143] W. Zhang, R. K. Mallik, and K. B. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5761–5766, December 2009.

[144] M. López-Benítez and J. Lehtomäki, "Energy detection based estimation of primary channel occupancy rate in cognitive radio," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.

[145] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," Available: https://arxiv.org/abs/1710.02913, Oct. 2017.

[146] O. Simeone, "A brief introduction to machine learning for engineers," Available: https://arxiv.org/abs/1709.02840, Sep. 2017.

[147] A. Yamada, T. Nishio, M. Morikura, and K. Yamamoto, "Machine learning-based primary exclusive region update for database-driven spectrum sharing," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, June 2017, pp. 1–5.

[148] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 2, pp. 257–265, June 2018.

[149] Y. Lu, P. Zhu, D. Wang, and M. Fattouche, "Machine learning techniques with probability vector for cooperative spectrum sensing in cognitive radio networks," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.

[150] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2209–2221, November 2013.