

# Estimation of Primary Channel Activity Statistics in Cognitive Radio Based on Periodic Spectrum Sensing Observations

Miguel López-Benítez, *Senior Member, IEEE*, Ahmed Al-Tahmeesschi, *Student Member, IEEE*, Dhaval K. Patel, *Member, IEEE*, Janne Lehtomäki, *Member, IEEE*, and Kenta Umebayashi, *Member, IEEE*

**Abstract**—Primary channel activity statistics (such as the minimum idle/busy periods of the channel, the moments or the underlying distributions) can be exploited by Cognitive Radio (CR) systems to adapt their operation and improve their performance. Such statistics can be directly estimated from periodic observations of the instantaneous idle/busy state of the primary channel (i.e., periodic spectrum sensing). However, the periodicity of such observations (i.e., the sensing period) imposes a fundamental limit on the time resolution to which idle/busy periods can be observed and consequently on the accuracy of any subsequent estimated statistics. In this context, this work provides a comprehensive analysis on the estimation of the primary activity statistics based on periodic channel state observations performed with a finite sensing period. In particular, this work provides a comprehensive set of closed-form expressions for the estimated statistics as a function of the true primary activity statistics and the employed sensing period. These expressions can find a wide range of applications in the analysis, design and simulation of CR systems. Moreover, several methods to minimise the estimation errors and improve the accuracy are proposed and validated both with simulations and hardware experiments.

**Index Terms**—Cognitive radio, dynamic spectrum access, spectrum sensing, primary activity statistics.

Manuscript received March 16, 2018; revised September 26, 2018; accepted November 30, 2018. Date of publication Month XX, 201X; date of current version December 13, 2018. This work was supported by the British Council under UKIERI DST Thematic Partnerships 2016-17 (grant ref. DST-198/2017). The work of J. Lehtomäki was financially supported by Academy of Finland 6Genesis Flagship (grant 318927) and by Infotech Oulu. The work of K. Umebayashi was supported by the European Commission in the framework of the H2020-EUJ-02-2018 project 5G-Enhance (grant agreement no. 815056) and the Ministry of Internal Affairs and Communications (MIC) of Japan and the JSPS KAKENHI grant numbers JP18K04124 and JP18KK0109. This work was presented in part at the 2018 IEEE Wireless Communications and Networking Conference [34] and the 2018 European Wireless Conference [37]. The associate editor coordinating the review of this paper and approving it for publication was L. DaSilva. (*Corresponding author: M. López-Benítez*)

M. López-Benítez and A. Al-Tahmeesschi are with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, United Kingdom (email: m.lopez-benitez@liverpool.ac.uk; ahmed.al-tahmeesschi@liverpool.ac.uk).

M. López-Benítez is also with the ARIES Research Centre, Antonio de Nebrija University, 28040 Madrid, Spain.

D. K. Patel is with the School of Engineering and Applied Science, Ahmedabad University, Ahmedabad 380009, India (e-mail: dhaval.patel@ahduni.edu.in).

J. Lehtomäki is with the Centre for Wireless Communications, University of Oulu, FI-90014 Oulu, Finland (e-mail: jannel@ee.oulu.fi).

K. Umebayashi is with the Graduate School of Engineering, Tokyo University of Agriculture and Technology, Koganei 184-8588, Japan (e-mail: ume\_k@cc.tuat.ac.jp).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2018.XXXXXXX

## I. INTRODUCTION

DESPISE being a much broader concept [1]–[3], Dynamic Spectrum Access/Cognitive Radio (DSA/CR) is commonly understood as an opportunistic spectrum access method whereby unlicensed (secondary) users are allowed to access the spectrum bands assigned to licensed (primary) systems during the inactivity periods of the primary users [4], [5]. This spectrum sharing approach is commonly known in the literature as *interweave*, which has received considerable more attention than the alternative *underlay* and *overlay* approaches [5]. In underlay and overlay, secondary transmissions concurrent to the primary transmissions are allowed under predefined interference constraints and therefore the number of users and power generated by the primary and secondary systems during their joint activity periods is an important aspect to take into account. In interweave, which constitutes the focus of this work, the DSA/CR system is not allowed to transmit during the activity periods of the primary system and therefore features such as the aggregated power generated by the primary system during its activity periods or number of primary users are not relevant. In interweave, the DSA/CR system is mainly concerned with the binary on/off occupancy pattern of the primary channel and the associated durations of its activity/inactivity periods. Owing to the opportunistic nature of the interweave operation principle, the behaviour and performance of DSA/CR systems depend on the spectrum occupancy patterns of the primary systems. An accurate knowledge of the primary activity patterns and their statistical properties can be useful to DSA/CR systems. Statistical information such as the duration of past idle/busy periods, their minimum, moments (mean, variance, etc.) or the underlying distribution, can be exploited to predict potential future trends in spectrum occupancy [6]–[8], schedule spectrum sensing [9], select the most appropriate band and channel of operation for the DSA/CR system [10]–[14], and take appropriate actions to minimise interference, optimise the system performance and improve the spectrum efficiency [15]–[20].

The activity statistics of a primary channel are initially unknown to the DSA/CR system but can be estimated using spectrum sensing decisions. DSA/CR systems periodically sense the instantaneous state of primary channels and in every sensing event a binary (idle/busy) decision is made based on a spectrum sensing (signal detection) method [21], [22]. While the main purpose of spectrum sensing is the

instantaneous detection of transmission opportunities [23], the sequence of spectrum sensing decisions can also be used to estimate the durations of the idle and busy periods and, with an appropriate processing of these values, obtain relevant primary activity statistics. The periodicity of spectrum sensing observations (i.e., the sensing period) imposes a fundamental limit on the time resolution to which idle/busy periods can be observed and consequently on the accuracy of any subsequent estimated statistics. While being an important issue of practical relevance in DSA/CR systems, the impact of this particular practical limitation on the estimated statistics still represents a problem that, to the best of the authors' knowledge, has not received a rigorous and formal treatment in the literature. Previous related work in this area (see for example [24]–[27]) suffers from several limitations: 1) the focus has been mainly on the estimation of the channel duty cycle or the mean idle/busy times, neglecting other relevant statistics of practical interest; 2) idle/busy times have commonly been assumed to be exponentially distributed, an assumption that facilitates the analytical tractability of the problem but has been proven by experimental studies to be unrealistic [28]–[32]; and 3) the conclusions of analytical studies have rarely been validated with experimental results (only in very few cases hardware experiments have been considered in addition to simulations [33]).

This work overcomes the above mentioned limitations by providing a comprehensive analysis on the estimation of primary activity statistics based on periodic channel state observations performed with a finite sensing period [34]. This work considers a broad range of relevant primary activity statistics, including the minimum idle/busy times, moments (not only the mean idle/busy times but also the variance, skewness and kurtosis) and the underlying distributions (first without assuming any distribution model and then considering more realistic models than the exponential distribution). The interest of this work is in analysing the impact of the employed sensing period on the accuracy of the estimated primary activity statistics. To this end, this work explores the relation between the true primary activity statistics and those estimated by a DSA/CR system as a function of the employed sensing period. In particular, this work answers (for each primary activity statistic) the following key questions, which constitute the main contributions:

- **Q1:** *What are the statistics estimated from periodic channel observations?* This work provides a comprehensive set of closed-form expressions for the *estimated* statistics as a function of the *true* primary statistics and the employed sensing period. These expressions can find a wide range of applications in the analysis, design and simulation of CR systems.
- **Q2:** *What is the deviation or error with respect to the true primary activity statistics?* Based on the above mentioned expressions, the error of the estimated statistics is quantified. The resulting expressions are useful in system designs (e.g., to determine the maximum sensing period that can guarantee a certain level of accuracy for each primary activity statistic).

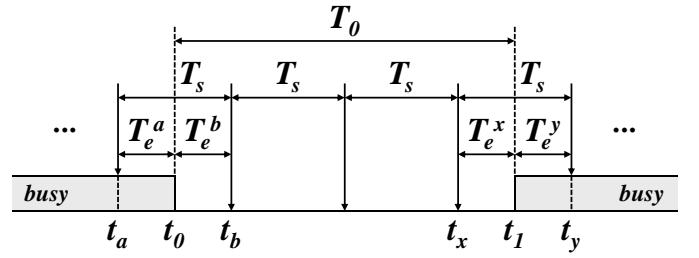


Fig. 1. Considered model (the estimation of an idle period is illustrated).

TABLE I  
PARAMETERS INVOLVED IN THE MODEL OF FIG. 1.

Parameter	Description
$t_0$	Instant of transition from busy to idle state
$t_1$	Instant of transition from idle to busy state
$t_a$	Instant of last sensed busy before switching to idle
$t_b$	Instant of first sensed idle after switching to idle
$t_x$	Instant of last sensed idle before switching to busy
$t_y$	Instant of first sensed busy after switching to busy
$T_0 = t_1 - t_0$	Real duration of period being estimated ( $T_i$ , $i = 0$ )
$T_s = t_b - t_a$ $= t_y - t_x$	Time interval between consecutive channel state observations (sensing period)
$T_e^a = t_0 - t_a$	Differences (errors) between the instants of channel state transitions and the nearest sensing instants
$T_e^b = t_b - t_0$	
$T_e^x = t_1 - t_x$	
$T_e^y = t_y - t_1$	

• **Q3:** *What can be done to minimise the estimation error?*

This work also proposes several methods to minimise the estimation errors and improve the accuracy of the estimated statistics, which are validated by both simulations and hardware experiments.

The rest of this work is organised as follows. First, Section II analyses the estimation of individual period durations from spectrum sensing decisions. The findings are used in Section III to analyse the estimation of several scalar statistics (minimum idle/busy period, moments of the estimated periods, and estimated channel duty cycle) and in Section IV to analyse the estimation of the underlying distribution. Section V validates the analytical results and assesses the performance of the methods proposed to reduce the estimation error, both with simulations and hardware experiments. Finally, Section VI summarises and concludes this work.

## II. ESTIMATION OF INDIVIDUAL PERIODS

Primary activity statistics (including both scalar statistics and the distribution of period durations) are estimated based on a sufficiently large number of observed idle/busy periods in the primary channel. Before the analysis of such statistics can be performed, a detailed study on the estimation of individual periods is first required, which is presented in this section.

DSA/CR systems sense periodically the instantaneous idle/busy state of a primary channel in order to detect transmission opportunities (idle periods) and determine when the

channel needs to be vacated to avoid interference (busy periods). The sequence of idle/busy decisions provided by the employed spectrum sensing method can be exploited to estimate the duration of the idle/busy periods of the primary channel as illustrated in Fig. 1, which shows the estimation of the duration of an idle period (the involved parameters are described in Table I). The primary channel is assumed to be sensed periodically with a spectrum sensing period of  $T_s$  time units (t.u.). The DSA/CR system estimates the duration of the real idle/busy period duration  $T_i$  ( $i = 0$  for idle periods,  $i = 1$  for busy periods) based on the time difference between the sensing events observed around the channel state transitions ( $t_a, t_b, t_x, t_y$ ). Based on the channel states observed at these sensing events, the DSA/CR system can make an estimation of the period duration, denoted as  $\hat{T}_i$ , following three different methods or Estimation Strategies (ESs):

- ES1:  $\hat{T}_i = t_x - t_b = T_i - T_e^b - T_e^x$ , which results in an underestimation of  $T_i$ .
- ES2:  $\hat{T}_i = t_y - t_a = T_i + T_e^a + T_e^y$ , which results in an overestimation of  $T_i$ .
- ES3:  $\hat{T}_i = \frac{(t_x - t_b) + (t_y - t_a)}{2} = t_x - t_a = t_y - t_b = T_i + T_e^a - T_e^x = T_i - T_e^b + T_e^y$ , which provides a more accurate estimation of  $T_i$  by averaging the estimations of ES1/ES2.

The true primary period  $T_i$  would be estimated accurately if the channel were sensed exactly when the channel state changes ( $t_0$  and  $t_1$ ). However, since spectrum sensing is in practice desynchronised with the primary activity, the channel is in general sensed at arbitrary time instants around the channel state transitions ( $t_a, t_b, t_x, t_y$ ) and, as a result, the estimated periods  $\hat{T}_i$  are affected by certain estimation errors ( $T_e^a, T_e^b, T_e^x, T_e^y$ ). For the three ESs, the estimated period duration can be expressed as (answer to Q1 for the estimation of individual periods):

$$\hat{T}_i = T_i + T_e \quad (1)$$

where  $T_e$  represents the error component, which is in turn composed of the sum of two other error subcomponents as shown above for each ES.

It can be shown as follows that the error (sub)components are directly related with the sensing period  $T_s$ . Note that  $T_e^a, T_e^b, T_e^x$  and  $T_e^y$  can take any value within the interval  $[0, T_s]$  and, as a matter of fact, can be assumed to be uniformly distributed within such interval, i.e.,  $T_e^a, T_e^b, T_e^x, T_e^y \sim U(0, T_s)$ , where  $U(a, b)$  denotes a uniform distribution between  $a$  and  $b$ . Since the error component  $T_e$  for all ESs is composed of the sum of two uniform random variables in  $[0, T_s]$ , assumed independent, then  $T_e$  can be modelled as a random variable whose Probability Density Function (PDF) can be obtained as the convolution of two uniform distributions with width  $T_s$  – either  $U(0, T_s)$  or  $U(-T_s, 0)$  depending on the sign of the error subcomponent – which leads to a triangular distribution with width  $2T_s$  [35, theorem 3.8.3]. More concretely, calculating such convolution for each ES, the PDF of  $T_e$  is found to be:

$$f_{T_e}^{\text{ES1}}(T) = f_{T_e}^{\text{ES3}}(T + T_s) \quad (2a)$$

$$f_{T_e}^{\text{ES2}}(T) = f_{T_e}^{\text{ES3}}(T - T_s) \quad (2b)$$

$$f_{T_e}^{\text{ES3}}(T) = \begin{cases} \frac{T_s + T}{T_s^2}, & -T_s \leq T \leq 0 \\ \frac{T_s - T}{T_s^2}, & 0 \leq T \leq T_s \\ 0, & \text{otherwise} \end{cases} \quad (2c)$$

$$f_{T_e}^{\text{ES3}}(T) = \begin{cases} \frac{T_s + T}{T_s^2}, & -T_s \leq T \leq 0 \\ \frac{T_s - T}{T_s^2}, & 0 \leq T \leq T_s \\ 0, & \text{otherwise} \end{cases} \quad (2c)$$

$$f_{T_e}^{\text{ES3}}(T) = \begin{cases} \frac{T_s + T}{T_s^2}, & -T_s \leq T \leq 0 \\ \frac{T_s - T}{T_s^2}, & 0 \leq T \leq T_s \\ 0, & \text{otherwise} \end{cases} \quad (2c)$$

Denoting by  $\Delta(a, b)$  a symmetric triangular distribution between  $a$  and  $b$ , then  $T_e \sim \Delta(-2T_s, 0)$  for ES1,  $T_e \sim \Delta(0, 2T_s)$  for ES2, and  $T_e \sim \Delta(-T_s, T_s)$  for ES3.

As it can be appreciated, the estimation error  $T_e$  of individual periods can be modelled as a symmetric triangular random variable with width  $2T_s$  for all the considered ESs. The maximum deviation of  $\hat{T}_i$  with respect to  $T_i$ , i.e., the maximum absolute error is given by (answer to Q2):

$$\varepsilon_{a,\max}^{\hat{T}_i} = \max(|T_i - \hat{T}_i|) = \max(|T_e|) = \begin{cases} 2T_s, & \text{for ES1} \\ 2T_s, & \text{for ES2} \\ T_s, & \text{for ES3} \end{cases} \quad (3a)$$

$$\varepsilon_{a,\max}^{\hat{T}_i} = \max(|T_i - \hat{T}_i|) = \max(|T_e|) = \begin{cases} 2T_s, & \text{for ES1} \\ 2T_s, & \text{for ES2} \\ T_s, & \text{for ES3} \end{cases} \quad (3b)$$

$$\varepsilon_{a,\max}^{\hat{T}_i} = \max(|T_i - \hat{T}_i|) = \max(|T_e|) = \begin{cases} 2T_s, & \text{for ES1} \\ 2T_s, & \text{for ES2} \\ T_s, & \text{for ES3} \end{cases} \quad (3c)$$

This result not only is consistent with the intuitive notion that reducing the sensing period  $T_s$  improves the accuracy of the estimated periods but, more interestingly, indicates that given an irreducible sensing period (the minimum value of  $T_s$  may be constrained by hardware limitations or other system design aspects [36]) the estimation error of individual periods can be minimised by employing ES3 (answer to Q3). In the remainder of this work it will be assumed (unless otherwise stated) that individual idle/busy periods are estimated based on ES3.

While the model in (1)–(2) can describe the statistical properties of the estimated individual periods, it is unable to accurately characterise their discrete nature. Note that the estimated periods are integer multiples of the sensing period (i.e.,  $\hat{T}_i = kT_s, k \in \mathbb{N}^+$ ). To account for this, a second model is proposed. It can be shown that the estimated periods under ES3 can also be expressed as (answer to Q1):

$$\hat{T}_i = \left( \left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s \quad (4)$$

where  $\lfloor \cdot \rfloor$  denotes the floor operator and  $\xi \in \{0, 1\}$  is a Bernoulli random variable introduced to model the fact that the same original period  $T_i$  can lead to two possible estimated periods, either  $\hat{T}_i = \lfloor T_i/T_s \rfloor T_s$  or  $\hat{T}_i = (\lfloor T_i/T_s \rfloor + 1)T_s = \lceil T_i/T_s \rceil T_s$  (where  $\lceil \cdot \rceil$  denotes the ceil operator), depending on the relative (random) position of the sensing events with respect to the beginning/end of the original period  $T_i$ . The Bernoulli probabilities  $P(\xi = 0)$  and  $P(\xi = 1)$  can be calculated noting that the true value of the original period  $T_i$  is within the interval  $[\lfloor T_i/T_s \rfloor T_s, \lceil T_i/T_s \rceil T_s]$ , and the estimation  $\hat{T}_i = \lfloor T_i/T_s \rfloor T_s$  (i.e.,  $\xi = 0$ ) will be more likely when  $T_i$  is closer to the lower extreme of the interval, while the estimation  $\hat{T}_i = \lceil T_i/T_s \rceil T_s$  (i.e.,  $\xi = 1$ ) will be more likely when  $T_i$  is closer to the upper extreme of the interval, hence:

$$P(\xi = 0) = P\left(\hat{T}_i = \left\lfloor \frac{T_i}{T_s} \right\rfloor T_s\right) = \left\lfloor \frac{T_i}{T_s} \right\rfloor - \frac{T_i}{T_s} \quad (5a)$$

$$P(\xi = 1) = P\left(\hat{T}_i = \left\lceil \frac{T_i}{T_s} \right\rceil T_s\right) = \frac{T_i}{T_s} - \left\lceil \frac{T_i}{T_s} \right\rceil \quad (5b)$$

which completes the characterisation of the model in (4).

The estimation error of individual periods can also be inferred from (4) as follows. As mentioned above, when

$T_i \in [\lfloor T_i/T_s \rfloor T_s, \lceil T_i/T_s \rceil T_s]$  the estimated period  $\widehat{T}_i$  will be one of the two extremes of the interval. The maximum estimation error will occur when  $T_i$  is close to one of the extremes of the interval and the estimated period  $\widehat{T}_i$  is the other extreme. Hence, the maximum error (answer to Q2) is equal to the difference between the extremes of the interval:

$$\varepsilon_{a,\max}^{\widehat{T}_i} = \max(|T_i - \widehat{T}_i|) = \left| \left\lceil \frac{T_i}{T_s} \right\rceil T_s - \left\lfloor \frac{T_i}{T_s} \right\rfloor T_s \right| = T_s \quad (6)$$

which leads to the same conclusion that was obtained in (3c) for ES3 with the model in (1)–(2).

### III. ESTIMATION OF SCALAR ACTIVITY STATISTICS

DSA/CR systems can calculate primary activity statistics based on a set of idle/busy periods observed from spectrum sensing as illustrated in Fig. 1. Let  $\{\widehat{T}_{i,n}\}_{n=1}^N$  be a set of  $N$  estimated periods  $\widehat{T}_i$  of type  $i$  ( $i = 0$  for idle periods,  $i = 1$  for busy periods) based on which the primary activity statistics are calculated. This section analyses the estimation of scalar primary activity statistics based on such set (namely the minimum period duration, the moments of the estimated periods, and the estimated channel duty cycle) and determines the impact of the sensing period  $T_s$  on the resulting accuracy. The impact of the sample size  $N$  on the accuracy of the estimated statistics is out of the scope of this work and therefore it will be assumed to be *sufficiently large* to provide the best attainable accuracy, so that all observed estimation errors will be the result of the employed sensing period  $T_s$ , which is the aspect of interest in this work.

#### A. Estimation of the Minimum Period Duration

The minimum period duration, denoted as  $\mu_i$ , is an important parameter since it determines the minimum period of time for which the primary channel will be available for opportunistic transmission ( $\mu_0$ ) or the minimum amount of time the DSA/CR system will have to wait before the channel becomes available ( $\mu_1$ ). The *minimum period duration* is a relative concept that can be considered at both *short* and *long* times scales. The value of  $\mu_i$  at short time scales is typically known since it is determined by the physical layer features of the radio technology employed by the primary system. For example, the value of  $\mu_i$  at short time scales is typically the slot duration (e.g., 417  $\mu$ s in DECT, 577  $\mu$ s in GSM, 14.167 ms in TETRA), the transmission time interval (e.g., 1 ms in LTE), or similar (25  $\mu$ s PIFS in IEEE 802.11a). However, such value is not usually very useful to DSA/CR systems since it may be challenging to exploit spectrum opportunities (idle periods) at such short time scale. On the other hand, idle/busy periods at longer time scales resulting from the primary traffic pattern at higher layers (e.g., user sessions) are more suitable for opportunistic transmissions. In such another case, the value of  $\mu_i$  at long time scales may be known or may need to be estimated from spectrum sensing [37].

The minimum period duration can be estimated from a set  $\{\widehat{T}_{i,n}\}_{n=1}^N$  of  $N$  estimated periods as  $\widehat{\mu}_i = \min_n(\{\widehat{T}_{i,n}\}_{n=1}^N)$ . The relation between the estimated minimum period  $\widehat{\mu}_i$  and

the true minimum period  $\mu_i$  (answer to Q1) can be obtained from (4) by noting that  $\min(T_i) = \mu_i$  and  $\min(\xi) = 0$ :

$$\widehat{\mu}_i = \min(\widehat{T}_i) = \min \left[ \left( \left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s \right] = \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s \quad (7)$$

The absolute error in (6) for individual periods is also applicable to the minimum estimated period in (7), hence  $\varepsilon_{a,\max}^{\widehat{\mu}_i} = T_s$  (answer to Q2).

Since the floor function is not invertible, it is not possible to uniquely identify the value of the true minimum  $\mu_i$  based on a single pair of employed (known)  $T_s$  and observed  $\widehat{\mu}_i$ . However, the result in (7) suggests a possible method to accurately estimate the true minimum  $\mu_i$  (answer to Q3). Notice that  $\widehat{\mu}_i = \mu_i$  when  $T_s = \mu_i/k$  with  $k \in \mathbb{N}^+$  (i.e., the estimated minimum is equal to the true minimum if the sensing period is an integer submultiple of the true minimum), and  $\widehat{\mu}_i < \mu_i$  otherwise. Based on this observation, the DSA/CR system could increase/decrease progressively the employed  $T_s$  and observe the estimated values of  $\widehat{\mu}_i$  until a maximum is found by trial-and-error, which would be an accurate estimation of the real  $\mu_i$ . However, an exhaustive search based on testing a large number of values for  $T_s$  may not be practical or feasible in real DSA/CR implementations. To address this problem, a method based on curve-fitting is proposed, which is aimed at providing an accurate estimation of the true minimum  $\mu_i$  based on a reduced set of tested  $T_s$  and observed  $\widehat{\mu}_i$  values. Let  $\{T_{s,l}\}_{l=1}^L$  be a set of  $L$  values of the sensing period tested by the DSA/CR system and  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$  the corresponding set of estimated minimum periods for each sensing period (i.e.,  $\widehat{\mu}_{i,l} = \lfloor \mu_i/T_{s,l} \rfloor T_{s,l}$ ). The proposed method consists in using curve-fitting to find the value of  $\mu_i$  that provides the best fit of (7) to the experimentally observed sets  $\{T_{s,l}\}_{l=1}^L$  and  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$ . The discontinuities of the floor function in (7) are problematic for curve-fitting methods, which are typically based on gradient search. This can be easily solved by rewriting the floor function in (7) in terms of its (continuous) Fourier series [38, eq. (2.1.7)], which yields:

$$\widehat{\mu}_i = \mu_i - \frac{T_s}{2} + T_s \sum_{m=1}^{\infty} \frac{\sin\left(\frac{2\pi m \mu_i}{T_s}\right)}{\pi m} \quad (8)$$

The expression in (8) can be readily fitted to  $\{T_{s,l}\}_{l=1}^L$  and  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$  using conventional curve-fitting methods. This provides an estimation of the true minimum as the value of  $\mu_i$  in (8) that offers the best fit to the sets  $\{T_{s,l}\}_{l=1}^L$  and  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$ . While this method needs several values of  $T_s$  as an input, this does not necessarily require the DSA/CR system to actually modify the employed sensing period, which might not be practical in some cases. Different sensing periods can be emulated by selectively discarding some sensing decisions (e.g., discarding every other sensing decision would be equivalent to employing a sensing period  $2T_s$ ). An additional method to estimate the minimum, based on the assumption of a certain distribution for  $T_i$ , will be discussed in Section IV-B. The accuracy of these approaches will be assessed in Section V.

#### B. Estimation of the Moments of Period Durations

The moments of the idle/busy period durations can be used by DSA/CR systems to characterise the primary channel

activity pattern in a simple manner. The most commonly used moments in statistics are the mean (first raw moment), variance (second central moment), skewness (third normalised moment) and kurtosis (fourth normalised moment). Given a set  $\{\widehat{T}_{i,n}\}_{n=1}^N$  of  $N$  estimated periods, the mean  $\mathbb{E}(\widehat{T}_i)$ , variance  $\mathbb{V}(\widehat{T}_i)$ , skewness  $\mathbb{S}(\widehat{T}_i)$  and kurtosis  $\mathbb{K}(\widehat{T}_i)$  of the observed periods can be estimated based on the corresponding (unbiased) sample moment estimators ( $\widehat{m}_i$ ,  $\widehat{v}_i$ ,  $\widehat{s}_i$  and  $\widehat{k}_i$ , respectively)<sup>1</sup>:

$$\mathbb{E}(\widehat{T}_i) \approx \widehat{m}_i = \frac{1}{N} \sum_{n=1}^N \widehat{T}_{i,n} \quad (9)$$

$$\mathbb{V}(\widehat{T}_i) \approx \widehat{v}_i = \frac{1}{N-1} \sum_{n=1}^N (\widehat{T}_{i,n} - \widehat{m}_i)^2 \quad (10)$$

$$\mathbb{S}(\widehat{T}_i) \approx \widehat{s}_i = \frac{\sqrt{N(N-1)}}{N-2} \frac{\frac{1}{N} \sum_{n=1}^N (\widehat{T}_{i,n} - \widehat{m}_i)^3}{\left( \sqrt{\frac{1}{N} \sum_{n=1}^N (\widehat{T}_{i,n} - \widehat{m}_i)^2} \right)^3} \quad (11)$$

$$\begin{aligned} \mathbb{K}(\widehat{T}_i) \approx \widehat{k}_i &= \frac{N-1}{(N-2)(N-3)} \left[ (N+1) \times \right. \\ &\times \left. \frac{\frac{1}{N} \sum_{n=1}^N (\widehat{T}_{i,n} - \widehat{m}_i)^4}{\left( \sqrt{\frac{1}{N} \sum_{n=1}^N (\widehat{T}_{i,n} - \widehat{m}_i)^2} \right)^2} - 3(N-1) \right] + 3 \end{aligned} \quad (12)$$

The relation between the estimated moments in (9)–(12) and the true moments  $\mathbb{E}(T_i)$ ,  $\mathbb{V}(T_i)$ ,  $\mathbb{S}(T_i)$  and  $\mathbb{K}(T_i)$  as a function of the employed sensing period  $T_s$  (answer to Q1) can be obtained based on the model in (1):

$$\mathbb{E}(\widehat{T}_i) = \mathbb{E}(T_i) + \mathbb{E}(T_e) = \begin{cases} \mathbb{E}(T_i) - T_s, & \text{for ES1} \\ \mathbb{E}(T_i) + T_s, & \text{for ES2} \\ \mathbb{E}(T_i), & \text{for ES3} \end{cases} \quad (13a)$$

$$\mathbb{V}(\widehat{T}_i) = \mathbb{V}(T_i) + \mathbb{V}(T_e) = \mathbb{V}(T_i) + \frac{T_s^2}{6}, \quad \text{for all ESs} \quad (14)$$

$$\begin{aligned} \mathbb{S}(\widehat{T}_i) &= \frac{\mathbb{E}([\widehat{T}_i - \mathbb{E}(\widehat{T}_i)]^3)}{[\mathbb{V}(\widehat{T}_i)]^{3/2}} = \frac{\mathbb{E}([T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e)]^3)}{[\mathbb{V}(T_i) + \mathbb{V}(T_e)]^{3/2}} \\ &= \frac{\mathbb{E}([T_i - \mathbb{E}(T_i)]^3) + \mathbb{E}([T_e - \mathbb{E}(T_e)]^3)}{[\mathbb{V}(T_i) + \mathbb{V}(T_e)]^{3/2}} \\ &= \mathbb{S}(T_i) \left[ 1 + \frac{\mathbb{V}(T_e)}{\mathbb{V}(T_i)} \right]^{-\frac{3}{2}} + \mathbb{S}(T_e) \left[ 1 + \frac{\mathbb{V}(T_i)}{\mathbb{V}(T_e)} \right]^{-\frac{3}{2}} \\ &= \mathbb{S}(T_i) \left[ 1 + \frac{T_s^2}{6\mathbb{V}(T_i)} \right]^{-\frac{3}{2}}, \quad \text{for all ESs} \end{aligned} \quad (15)$$

<sup>1</sup>The expressions in (9)–(12) may impose high storage/memory requirements in DSA/CR systems, in particular if the sample size  $N$  is large. This problem can be solved in practical implementations by employing equivalent recurrence formulae, which only require two (instead of  $N$ ) values to be processed (see [39] for details).

$$\begin{aligned} \mathbb{K}(\widehat{T}_i) &= \frac{\mathbb{E}([\widehat{T}_i - \mathbb{E}(\widehat{T}_i)]^4)}{[\mathbb{V}(\widehat{T}_i)]^2} = \frac{\mathbb{E}([T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e)]^4)}{[\mathbb{V}(T_i) + \mathbb{V}(T_e)]^2} \\ &= \frac{\mathbb{E}([T_i - \mathbb{E}(T_i)]^4) + \mathbb{E}([T_e - \mathbb{E}(T_e)]^4)}{[\mathbb{V}(T_i) + \mathbb{V}(T_e)]^2} + \\ &\quad + \frac{6\mathbb{E}([T_i - \mathbb{E}(T_i)]^2)\mathbb{E}([T_e - \mathbb{E}(T_e)]^2)}{[\mathbb{V}(T_i) + \mathbb{V}(T_e)]^2} \\ &= \frac{\mathbb{K}(T_i)[\mathbb{V}(T_i)]^2 + 6\mathbb{V}(T_i)\mathbb{V}(T_e) + \mathbb{K}(T_e)[\mathbb{V}(T_e)]^2}{[\mathbb{V}(T_i) + \mathbb{V}(T_e)]^2} \\ &= \frac{\mathbb{K}(T_i)[\mathbb{V}(T_i)]^2 + \mathbb{V}(T_i)T_s^2 + \frac{T_s^4}{15}}{\left[ \mathbb{V}(T_i) + \frac{T_s^2}{6} \right]^2}, \quad \text{for all ESs} \end{aligned} \quad (16)$$

where  $T_i$  and  $T_e$  have been assumed to be independent, and  $\mathbb{E}(T_e) \in \{0, \pm T_s\}$ ,  $\mathbb{V}(T_e) = T_s^2/6$ ,  $\mathbb{S}(T_e) = 0$  and  $\mathbb{K}(T_e) = 12/5$  are respectively the mean, variance, skewness and kurtosis of the symmetric triangular distribution  $f_{T_e}(T)$  in (2) [40, ch. 40].

Based on (13)–(16), the error of the estimated moments is given by (answer to Q2):

$$\begin{aligned} \varepsilon_{\text{exact}}^{\mathbb{E}(\widehat{T}_i)} &= \mathbb{E}(\widehat{T}_i) - \mathbb{E}(T_i) = \mathbb{E}(T_e) = \begin{cases} -T_s, & \text{for ES1} \\ T_s, & \text{for ES2} \\ 0, & \text{for ES3} \end{cases} \end{aligned} \quad (17a)$$

$$(17b)$$

$$(17c)$$

$$\varepsilon_{\text{exact}}^{\mathbb{V}(\widehat{T}_i)} = \mathbb{V}(\widehat{T}_i) - \mathbb{V}(T_i) = \mathbb{V}(T_e) = \frac{T_s^2}{6}, \quad \text{for all ESs} \quad (18)$$

$$\varepsilon_{\text{exact}}^{\mathbb{S}(\widehat{T}_i)} = \frac{\mathbb{S}(\widehat{T}_i) - \mathbb{S}(T_i)}{\mathbb{S}(T_i)} = \left[ 1 + \frac{T_s^2}{6\mathbb{V}(T_i)} \right]^{-\frac{3}{2}} - 1, \quad \text{for all ESs} \quad (19)$$

$$\begin{aligned} \varepsilon_{\text{exact}}^{\mathbb{K}(\widehat{T}_i)} &= \frac{\mathbb{K}(\widehat{T}_i) - \mathbb{K}(T_i)}{\mathbb{K}(T_i)} \\ &= \frac{\left[ \mathbb{V}(T_i)T_s^2 + \frac{T_s^4}{15} \right] \left[ \frac{1}{\mathbb{K}(T_i)} - \frac{1}{3} \right] - \frac{T_s^4}{180}}{\left[ \mathbb{V}(T_i) + \frac{T_s^2}{6} \right]^2}, \quad \text{for all ESs} \end{aligned} \quad (20)$$

Note that (17)–(20) establish unique exact relations between true and estimated moments based on  $T_s$  and other known or accurately estimable parameters. This indicates that corrected versions ( $\widetilde{m}_i$ ,  $\widetilde{v}_i$ ,  $\widetilde{s}_i$ ,  $\widetilde{k}_i$ ) of the sample moments ( $\widehat{m}_i$ ,  $\widehat{v}_i$ ,  $\widehat{s}_i$ ,  $\widehat{k}_i$ ) can be used as accurate estimations of the true moments ( $\mathbb{E}(T_i)$ ,  $\mathbb{V}(T_i)$ ,  $\mathbb{S}(T_i)$ ,  $\mathbb{K}(T_i)$ ) as follows (answer to Q3):

$$\mathbb{E}(T_i) \approx \widetilde{m}_i = \begin{cases} \widehat{m}_i + T_s, & \text{for ES1} \\ \widehat{m}_i - T_s, & \text{for ES2} \\ \widehat{m}_i, & \text{for ES3} \end{cases} \quad (21a)$$

$$(21b)$$

$$(21c)$$

$$\mathbb{V}(T_i) \approx \widetilde{v}_i = \widehat{v}_i - \frac{T_s^2}{6}, \quad \text{for all ESs} \quad (22)$$

$$\mathbb{S}(T_i) \approx \widetilde{s}_i = \widehat{s}_i \left[ 1 + \frac{T_s^2}{6\widetilde{v}_i} \right]^{\frac{3}{2}}, \quad \text{for all ESs} \quad (23)$$

$$\mathbb{K}(T_i) \approx \tilde{k}_i = \frac{1}{\tilde{v}_i^2} \left( \hat{k}_i \left[ \tilde{v}_i + \frac{T_s^2}{6} \right]^2 - \tilde{v}_i T_s^2 - \frac{T_s^4}{15} \right), \quad \text{for all ESs} \quad (24)$$

which allows an accurate estimation of the real moments of  $T_i$  for any employed sensing period.

### C. Estimation of the Channel Duty Cycle

The Duty Cycle (DC) is commonly used in DSA/CR to quantify the overall occupancy level (load) of primary channels, which can be useful in many contexts (e.g., channel selection). This parameter is defined as the probability (probabilistic definition) or fraction of time (empirical definition) that the channel is busy (i.e., in use by the primary system). The true DC of a primary channel, denoted as  $\Psi$ , can be expressed as:

$$\Psi = \frac{\mathbb{E}(T_1)}{\mathbb{E}(T_0) + \mathbb{E}(T_1)} \quad (25)$$

The analysis of Section III-B demonstrated that the terms  $\mathbb{E}(T_i)$  in (25) can be accurately estimated by applying a certain correction factor to the sample means as shown in (21). Therefore, the estimation of the DC obtained from sample means,  $\tilde{\Psi} \approx \tilde{m}_1 / (\tilde{m}_0 + \tilde{m}_1)$ , is an accurate estimation of the true DC,  $\Psi$ , regardless of the employed sensing period (answer to Q1) with no estimation error (answer to Q2) provided that the estimation is based on the corrected sample means  $\tilde{m}_i$  (answer to Q3).

## IV. ESTIMATION OF THE DISTRIBUTION

The distribution of the idle/busy period durations provides a complete characterisation of the primary channel activity. A common assumption widely employed in the existing literature is that idle/busy periods are exponentially distributed (see [24]–[27] for some examples and [41, Section II] for a detailed discussion). While this assumption simplifies analytical studies, field measurements have demonstrated that the exponential model is unrealistic [28]–[32]. A more realistic model over a broad range of frequency bands is the Generalised Pareto (GP) distribution [42]. The Probability Density Function (PDF),  $f_{T_i}(T)$ , and the Cumulative Distribution Function (CDF),  $F_{T_i}(T)$ , of the GP distribution are respectively given by [43, ch. 20]:

$$f_{T_i}(T) = \frac{1}{\lambda_i} \left[ 1 + \frac{\alpha_i(T - \mu_i)}{\lambda_i} \right]^{-1/\alpha_i - 1}, \quad T \geq \mu_i \quad (26a)$$

$$F_{T_i}(T) = 1 - \left[ 1 + \frac{\alpha_i(T - \mu_i)}{\lambda_i} \right]^{-1/\alpha_i}, \quad T \geq \mu_i \quad (26b)$$

where  $\mu_i > 0$ ,  $\lambda_i > 0$ ,  $\alpha_i \in \mathbb{R}$  are the location, scale and shape parameters, respectively. The analysis of empirical data indicates that  $\alpha_i \in [0, 0.25]$  in most practical cases [42]; in such a case (for  $\alpha_i \geq 0$ ) the support of the distribution in (26) is  $T \geq \mu_i$ . This work assumes GP-distributed periods. It can be shown, by applying the change of variable  $x = \alpha_i(T - \mu_i) / \lambda_i$  and using  $\lim_{x \rightarrow 0} (1 + x)^{1/x} = e$ , that the limit of (26) as  $\alpha_i$  tends towards zero is an exponential distribution with location parameter  $\mu_i$  and rate parameter  $1/\lambda_i$ . The same applies for

the moments [43, eq. (20.156)] when  $\alpha_i = 0$ . Therefore, the results obtained in this work can be particularised to the exponential case by deriving the limit as  $\alpha_i$  tends towards zero (the results can be found in [44]).

### A. Direct Estimation

The Direct Estimation Method (DEM) makes an estimation of the distribution  $F_{T_i}(T)$ , denoted as  $F_{\hat{T}_i}(\hat{T})$ , based on a set  $\hat{\mathcal{T}}_i = \{\hat{T}_{i,n}\}_{n=1}^N$  of  $N$  observed period durations as follows:

$$F_{\hat{T}_i}(\hat{T}) = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\hat{T}_i(\hat{T})} \{\hat{T}_{i,n}\} = \frac{|\hat{\mathcal{T}}_i(\hat{T})|}{N} \quad (27)$$

where  $|\hat{\mathcal{T}}_i(\hat{T})|$  indicates the cardinality (number of elements) of  $\hat{\mathcal{T}}_i(\hat{T}) = \{\hat{T}_{i,n} : \hat{T}_{i,n} \leq \hat{T}, n = 1, \dots, N\}$  (the subset of period durations lower than or equal to  $\hat{T}$ ), and  $\mathbf{1}_A\{x\}$  is the indicator function of subset  $A$ , which is equal to one for the elements  $x \in A$  and zero otherwise. Note that this method can be used without any assumptions on  $F_{T_i}(T)$  and the estimated distribution  $F_{\hat{T}_i}(\hat{T})$  (usually referred to as *empirical CDF*) has a discrete domain  $\hat{T} = kT_s$  ( $k \in \mathbb{N}^+$ ), even though the original distribution  $F_{T_i}(T)$  is in general continuous, since the periods of the set  $\hat{\mathcal{T}}_i$  are integer multiples of the sensing period ( $\hat{T}_{i,n} = kT_s$ ,  $k \in \mathbb{N}^+$ ).

The relation between the estimated distribution  $F_{\hat{T}_i}(\hat{T})$  in (27) and the true distribution  $F_{T_i}(T)$  in (26b) as a function of the employed sensing period  $T_s$  (answer to Q1) can be obtained based on the model in (4). A period with duration  $\hat{T} = kT_s$  ( $k \in \mathbb{N}^+$ ) can be observed when  $T_i \in [\hat{T}, \hat{T} + T_s]$  and  $\xi = 0$ , or when  $T_i \in [\hat{T} - T_s, \hat{T}]$  and  $\xi = 1$ . Hence, the Probability Mass Function (PMF) of the estimated periods,  $f_{\hat{T}_i}(\hat{T}) = P(\hat{T}_i = \hat{T})$ , can be obtained as:

$$f_{\hat{T}_i}(\hat{T}) = P(\hat{T} \leq T_i \leq \hat{T} + T_s) \cdot \mathbb{E}(P(\xi = 0)) + \\ + P(\hat{T} - T_s \leq T_i \leq \hat{T}) \cdot \mathbb{E}(P(\xi = 1)) \quad (28)$$

which leads to the result shown in (29).

Note that two particular cases in (28)–(29) need special consideration, concretely the values of  $\hat{T}$  around the minimum period, namely  $\hat{T} = \lfloor \mu_i/T_s \rfloor T_s$  and  $\hat{T} = \lceil \mu_i/T_s \rceil T_s$ . Note 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, in the interval around the minimum period  $\mu_i$  the range of valid values is  $T_i \in [\mu_i, \lceil \mu_i/T_s \rceil T_s]$  (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, the probability  $P(\xi = 0)$  needs to be scaled accordingly by the factor:

$$\chi_0 = \frac{\left[ \frac{\mu_i}{T_s} \right] T_s - \mu_i}{T_s} = \left[ \frac{\mu_i}{T_s} \right] - \frac{\mu_i}{T_s} \quad (32)$$

It is worth noting that when  $T_s = \mu_i/k$ ,  $k \in \mathbb{N}^+$  (i.e., the sensing period is an integer submultiple of the true minimum) then  $\lfloor \mu_i/T_s \rfloor T_s = \lceil \mu_i/T_s \rceil T_s$ ; in this case (29b) must be used.

$$\begin{aligned}
f_{\widehat{T}_i}(\widehat{T}) &= \begin{cases} F_{T_i}(\widehat{T} + T_s) \cdot \mathbb{E}(P(\xi = 0)) \cdot \chi_0, \\ \left[ F_{T_i}(\widehat{T} + T_s) - F_{T_i}(\widehat{T}) \right] \cdot \mathbb{E}(P(\xi = 0)) + F_{T_i}(\widehat{T}) \cdot [1 - \mathbb{E}(P(\xi = 0)) \cdot \chi_0], \\ \left[ F_{T_i}(\widehat{T} + T_s) - F_{T_i}(\widehat{T}) \right] \cdot \mathbb{E}(P(\xi = 0)) + \left[ F_{T_i}(\widehat{T}) - F_{T_i}(\widehat{T} - T_s) \right] \cdot \mathbb{E}(P(\xi = 1)), \end{cases} \quad (29a) \\
&\quad \widehat{T} = \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s \quad (29b) \\
&\quad \widehat{T} = \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s \quad (29c) \\
g_{\widehat{T}_i}(T) &= \begin{cases} 0, & T \leq \mu_i - T_s \\ \omega(T + T_s) - \frac{\lambda_i}{(1-\alpha_i)T_s^2} + \frac{T+T_s-\mu_i}{T_s^2}, & \mu_i - T_s < T \leq \mu_i \\ \omega(T + T_s) - 2\omega(T) + \frac{\lambda_i}{(1-\alpha_i)T_s^2} - \frac{T-T_s-\mu_i}{T_s^2}, & \mu_i < T \leq \mu_i + T_s \\ \omega(T + T_s) - 2\omega(T) + \omega(T - T_s), & T > \mu_i + T_s \end{cases} \quad (30a) \\
&\quad \mu_i - T_s < T \leq \mu_i \quad (30b) \\
&\quad \mu_i < T \leq \mu_i + T_s \quad (30c) \\
&\quad T > \mu_i + T_s \quad (30d) \\
G_{\widehat{T}_i}(T) &= \begin{cases} 0, & T \leq \mu_i - T_s \\ \Omega(T + T_s) + \frac{\lambda_i^2}{(1-\alpha_i)(1-2\alpha_i)T_s^2} + \frac{T^2-(\mu_i-T_s)^2}{2T_s^2} + \left(T_s - \mu_i - \frac{\lambda_i}{1-\alpha_i}\right) \frac{T+T_s-\mu_i}{T_s^2}, & \mu_i - T_s < T \leq \mu_i \\ \Omega(T + T_s) - 2\Omega(T) - \frac{\lambda_i^2}{(1-\alpha_i)(1-2\alpha_i)T_s^2} + \frac{\lambda_i}{1-\alpha_i} \frac{T-T_s-\mu_i}{T_s^2} - \frac{T^2-T_s^2-\mu_i^2}{2T_s^2} + \frac{(T-\mu_i)(T_s+\mu_i)}{T_s^2}, & \mu_i < T \leq \mu_i + T_s \\ 1 + \Omega(T + T_s) - 2\Omega(T) + \Omega(T - T_s), & T > \mu_i + T_s \end{cases} \quad (31a) \\
&\quad \mu_i - T_s < T \leq \mu_i \quad (31b) \\
&\quad \mu_i < T \leq \mu_i + T_s \quad (31c) \\
&\quad T > \mu_i + T_s \quad (31d)
\end{aligned}$$

The Bernoulli probabilities in (28)-(29) are obtained as the expected value of those in (5) as:

$$\begin{aligned}
\mathbb{E}(P(\xi = 0)) &= \int_T P(\xi = 0) f_{T_i}(T) dT \\
&= \sum_{m=0}^{\infty} (m+1) [F_{T_i}((m+1)T_s) - F_{T_i}(mT_s)] - \frac{\mathbb{E}(T_i)}{T_s} \quad (33)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(P(\xi = 1)) &= \int_T P(\xi = 1) f_{T_i}(T) dT \\
&= \frac{\mathbb{E}(T_i)}{T_s} - \sum_{m=0}^{\infty} m [F_{T_i}((m+1)T_s) - F_{T_i}(mT_s)] \quad (34)
\end{aligned}$$

In practice,  $\mathbb{E}(P(\xi = 0)) \approx \mathbb{E}(P(\xi = 1)) \approx 1/2$  is a valid approximation in many cases.

The corresponding CDF can then be obtained as:

$$F_{\widehat{T}_i}(\widehat{T}) = P(\widehat{T}_i \leq \widehat{T}) = \sum_{k=0}^{\widehat{T}/T_s} f_{\widehat{T}_i}(kT_s) \quad (35)$$

Note that (28), (29), (32), (33), (34) and (35) do not make any assumptions on the distribution of the true periods  $T_i$ .

An alternative expression for the estimated distribution  $F_{\widehat{T}_i}(\widehat{T})$  in (27) as a function of the true distribution  $F_{T_i}(T)$  and  $T_s$  (answer to Q1) can be obtained based on the model in (1). Assuming  $T_i$  and  $T_e$  independent, a (continuous) PDF associated with the estimated periods,  $g_{\widehat{T}_i}(T)$ , can be obtained as  $g_{\widehat{T}_i}(T) = f_{T_i}(T) * f_{T_e}(T)$  [35, theorem 3.8.3], where  $f_{T_i}(T)$  is given by (26a) (assuming GP-distributed periods),  $f_{T_e}(T)$  is given by (2) and the operator  $*$  indicates convolution. The resulting expression for  $g_{\widehat{T}_i}(T)$  is shown in (30) where  $\omega(\cdot)$  is given by (36), and its direct integration yields the CDF,  $G_{\widehat{T}_i}(T) = \int_{-\infty}^T g_{\widehat{T}_i}(\tau) d\tau$ , shown in (31) where  $\Omega(\cdot)$  is given by (37)<sup>2</sup>.

$$\omega(\tau) = \frac{[\lambda_i + \alpha_i(\tau - \mu_i)]^2}{(1 - \alpha_i)T_s^2} f_{T_i}(\tau) \quad (36)$$

<sup>2</sup>The results in (30)–(31) assume ES3. The expressions for ES1 and ES2 can be obtained by replacing  $T$  with  $T + T_s$  and  $T - T_s$ , respectively.

$$\Omega(\tau) = \frac{[\lambda_i + \alpha_i(\tau - \mu_i)]^2}{(\alpha_i - 1)(1 - 2\alpha_i)T_s^2} [1 - F_{T_i}(\tau)] \quad (37)$$

An expression for the (discrete) distribution estimated with DEM can be obtained by evaluating (30)–(31) at the discrete values  $T = (k + 1/2)T_s = \widehat{T} + T_s/2$ , where a continuity correction factor 1/2 is necessary as a discrete distribution is being represented by a continuous expression:

$$f_{\widehat{T}_i}(\widehat{T}) = g_{\widehat{T}_i}\left(\widehat{T} + \frac{T_s}{2}\right) \quad (38)$$

$$F_{\widehat{T}_i}(\widehat{T}) = G_{\widehat{T}_i}\left(\widehat{T} + \frac{T_s}{2}\right) \quad (39)$$

The estimation error or deviation of the estimated distribution  $F_{\widehat{T}_i}(\widehat{T})$  in (27) with respect to the true distribution  $F_{T_i}(T)$  in (26b) (answer to Q2) cannot be quantified in a straightforward manner since the former has a discrete domain ( $\widehat{T} = kT_s$ ,  $k \in \mathbb{N}^+$ ) while the latter is in general continuous ( $T \in \mathbb{R}^+$ ). The difference between two distributions is typically quantified in terms of the Kolmogorov-Smirnov (KS) distance, which is defined as the maximum absolute difference between the CDFs [45, eq. (14.3.17)]. The KS distance, however, is defined in the context of continuous distributions. For the purposes of this analysis, a modified version of the KS distance can be defined as the maximum absolute difference between the discrete estimated CDF  $F_{\widehat{T}_i}(\widehat{T})$  and the continuous original CDF  $F_{T_i}(T)$  evaluated at the discrete points  $T = \widehat{T} = kT_s$  ( $k \in \mathbb{N}^+$ ) for which  $F_{\widehat{T}_i}(\widehat{T})$  is defined:

$$D_{KS,i} = \sup_{\widehat{T}} |F_{\widehat{T}_i}(\widehat{T}) - F_{T_i}(\widehat{T})| \quad (40)$$

This KS distance can be evaluated analytically by determining the period duration  $\widehat{T}^*$  for which the absolute difference of (40) is maximum and then computing its value at  $\widehat{T} = \widehat{T}^*$ . In order to find the discrete value  $\widehat{T}^*$ , let first consider the continuous value  $T^*$  that maximises the absolute difference between the continuous CDF  $G_{\widehat{T}_i}(T)$ , which is related with

the estimated distribution as shown in (39), and the true CDF  $F_{T_i}(T)$ . Solving  $\frac{\partial}{\partial T}|F_{T_i}(T) - G_{\widehat{T}_i}(T)|$  for  $T$  yields  $T^* = \mu_i$ , and based on (4) it can be seen that the sought  $\widehat{T}^*$  can be either  $\widehat{T}^* = \lfloor \mu_i/T_s \rfloor T_s$  or  $\widehat{T}^* = \lceil \mu_i/T_s \rceil T_s$ . The maximum absolute difference is provided by either value of  $\widehat{T}^*$ , hence:

$$\begin{aligned} D_{\text{KS},i} &= \max \left\{ F_{\widehat{T}_i} \left( \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s \right) - F_{T_i} \left( \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s \right), \right. \\ &\quad \left. F_{\widehat{T}_i} \left( \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s \right) - F_{T_i} \left( \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s \right) \right\} \\ &= \max \left\{ F_{\widehat{T}_i} \left( \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s \right), \right. \\ &\quad \left. F_{\widehat{T}_i} \left( \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s \right) - F_{T_i} \left( \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s \right) \right\} \quad (41) \end{aligned}$$

Recall that the estimated distribution  $F_{\widehat{T}_i}(\cdot)$  can be evaluated based on (35) (regardless of the true distribution of  $T_i$ ) or (39) (assuming GP-distributed periods), while  $F_{T_i}(\cdot)$  represents the true distribution of  $T_i$ , which in the case of GP-distributed periods is given by (26b).

The result in (41) is useful to evaluate the impact of the employed sensing period on the accuracy of the estimated distribution. A detailed inspection of (41) reveals that the only way to improve the accuracy with DEM is to decrease the sensing period. To overcome this limitation, an alternative estimation method is proposed below (answer to Q3).

### B. Moment-based Estimation

The main drawback of the DEM approach is that the estimated distribution is discrete (this constrains the resulting estimation accuracy since the original distribution is in general continuous). To overcome this limitation, an alternative approach is here proposed based on the estimation of the distribution's parameters (namely,  $\mu_i$ ,  $\lambda_i$  and  $\alpha_i$ , assuming the GP model in (26) as discussed above) by means of statistical inference methods. In contrast with DEM, the resulting distribution is continuous, thus offering the possibility to provide a more accurate estimation.

Various parameter-inference techniques have been proposed for the GP distribution, including solutions based on the Method of Moments (MoM), Probability-Weighted Moments (PWM), Maximum Likelihood Estimation (MLE), Principle of Maximum Entropy (PoME), and Least Squares Estimation (LSE) [43, ch. 20]. The solution here proposed is based on MoM, which estimates the distribution parameters  $\mu_i$ ,  $\lambda_i$  and  $\alpha_i$  based on sample moments. This approach is here preferred for three reasons: 1) MoM has been reported to be superior for the GP distribution [46]; 2) the population moments can be accurately estimated from sample moments by applying appropriate corrections as discussed in Section III-B; and 3) the sample moments can be computed based on recursion equations that only require the last sample [39] while other inference methods in general need the whole history of past observed period durations, which leads to significantly higher memory and computation requirements in practical implementations.

The parameters  $\lambda_i$  and  $\alpha_i$  of the GP distribution can be estimated as follows [43, ch. 20]:

$$\lambda_i \approx \tilde{\lambda}_i = \frac{1}{2} \left( 1 + \frac{(\tilde{m}_i - \mu_i)^2}{\tilde{v}_i} \right) (\tilde{m}_i - \mu_i) \quad (42a)$$

$$\alpha_i \approx \tilde{\alpha}_i = \frac{1}{2} \left( 1 - \frac{(\tilde{m}_i - \mu_i)^2}{\tilde{v}_i} \right) \quad (42b)$$

where  $\tilde{\lambda}_i$  and  $\tilde{\alpha}_i$  are the estimations obtained based on the corrected sample mean  $\tilde{m}_i$  in (21) and the corrected sample variance  $\tilde{v}_i$  in (22). The expressions in (42) assume that the value of the minimum period  $\mu_i$  is known. If it is unknown, it can be estimated with the methods discussed in Section III-A, which do not make any assumptions on the distribution of the true periods  $T_i$ . If the true periods  $T_i$  are assumed to be GP-distributed, then the minimum period  $\mu_i$  can also be estimated based on sample moments as follows. Combining the expressions for the mean and variance of the GP distribution [43, ch. 20], the minimum period can be estimated as:

$$\tilde{\mu}_i = \tilde{m}_i - \sqrt{\tilde{v}_i(1 - 2\tilde{\alpha}_i)} \quad (43)$$

where  $\tilde{\alpha}_i$  can be estimated from the corrected sample skewness  $\tilde{s}_i$  in (23) based on [43, ch. 20]:

$$\tilde{s}_i = \frac{2(1 + \tilde{\alpha}_i)\sqrt{1 - 2\tilde{\alpha}_i}}{1 - 3\tilde{\alpha}_i} \quad (44)$$

Solving (44) for  $\tilde{\alpha}_i$  leads to a cubic equation with the following solution:

$$\tilde{\alpha}_i = -\frac{1}{24} \left( \frac{\Phi}{\zeta\Upsilon} + \zeta\Upsilon + 9\tilde{s}_i^2 + 12 \right) \quad (45a)$$

where

$$\Phi = 9(9\tilde{s}_i^4 + 40\tilde{s}_i^2 + 16) \quad (45b)$$

$$\zeta = -\frac{1}{2} + \frac{1}{2}\sqrt{3}i \quad (45c)$$

$$\Upsilon = \left( \frac{\Gamma - \theta i}{2} \right)^{\frac{1}{3}} \quad (45d)$$

$$\Gamma = 54(27\tilde{s}_i^6 + 180\tilde{s}_i^4 + 272\tilde{s}_i^2 - 64) \quad (45e)$$

$$\theta = 3456\tilde{s}_i(\tilde{s}_i^2 + 4) \quad (45f)$$

Despite the presence of the imaginary number  $i = \sqrt{-1}$  in (45), the resulting  $\tilde{\alpha}_i$  is always real.

The introduction of the estimated  $\tilde{\mu}_i$ ,  $\tilde{\lambda}_i$  and  $\tilde{\alpha}_i$  in (26) provides a (continuous) estimated distribution  $F_{\widehat{T}_i}(T)$ . Notice that  $F_{\widehat{T}_i}(T)$  is estimated based on corrected versions of the sample moments that, as shown in Section III-B, are not affected by the employed sensing period. Therefore, the estimated distribution  $F_{\widehat{T}_i}(T)$  and the real distribution  $F_{T_i}(T)$  can be assumed to be identical provided that the corrected sample moments are calculated based on a sufficiently large set of observed periods.

## V. NUMERICAL, SIMULATION AND EXPERIMENTAL RESULTS

This section validates the developed analytical results and assesses the accuracy of the proposed estimation methods

with both simulation and experimental results. Simulations are conducted in Matlab by generating a sufficiently large number ( $\approx 10^6\text{-}10^7$ ) of GP-distributed random periods  $T_i$  with known pre-defined parameters ( $\mu_i = 100$  t.u.,  $\lambda_i = 300$  t.u. and  $\alpha_i = 0.2$ , unless otherwise stated), then sensing the sequence of periods  $T_i$  with a given sensing period  $T_s$  in order to calculate the corresponding sequence of estimated periods  $\hat{T}_i$  that would be observed by a DSA/CR system, and finally calculating the relevant statistics resulting from the observed periods  $\hat{T}_i$ , which are compared with the true statistics of the generated periods  $T_i$ . This simulation approach is similar to that employed in [47]. Hardware experiments are conducted with a Prototype for the Estimation of Channel Activity Statistics (PECAS) (see [48] for details<sup>3</sup>). The (primary) transmitter sends a sequence of  $10^5$  pairs of GP-distributed idle/busy periods (with the same parameters as in the simulations) using a 433 MHz ON-OFF Keying (OOK) modulator with an output power of 2 dBm. The (secondary DSA/CR) receiver, placed 1 metre apart, uses a Software-Defined Radio (SDR) with a gain of 50 dB to monitor the idle/busy activity at 433 MHz. The SDR senses the channel every  $T_s = 10$  ms (longer sensing periods in steps of 10 ms up to 90 ms are emulated by selectively discarding sensing events). The maximum absolute timing error (jitter) is 50  $\mu$ s. In every sensing event, signal samples are captured for 3 ms at a sample rate of  $10^6$  samples per second; this provides 3000 signal samples, which are processed to decide the instantaneous idle/busy state of the channel using energy detection. The outcomes of the energy detection decisions are used to estimate the durations of the observed idle/busy periods (using ES3 only) and compute the primary activity statistics. By comparing the parameters configured in the transmitter with the estimations provided by the receiver, the correctness of the developed analytical results can be verified and the accuracy of the proposed estimation methods can be assessed under realistic evaluation conditions, including the channel degrading effects and practical limitations of real transmitters and receivers. Note that the analytical expressions derived in this work are independent of the particular units of the time variable and therefore the results are shown in generic *time units* (t.u.). In hardware experiments, where a specific unit needs to be selected, the reference time unit is the millisecond (i.e., 1 t.u. = 1 ms).

It is worth mentioning that the sample sizes (number of random periods) considered in simulations and experiments were constrained by practical limitations: in simulations by the capabilities of the employed computer (mainly the amount of memory available), and in hardware experiments by the shortest time scales allowed by the hardware components as well as the need to keep the length of experiments within practical limits. While the use of a finite sample size may affect the accuracy of the simulation and experimental results in some specific points/cases, the considered sample sizes provide in general a high level of accuracy, sufficient to validate the correctness and accuracy of the analytical results

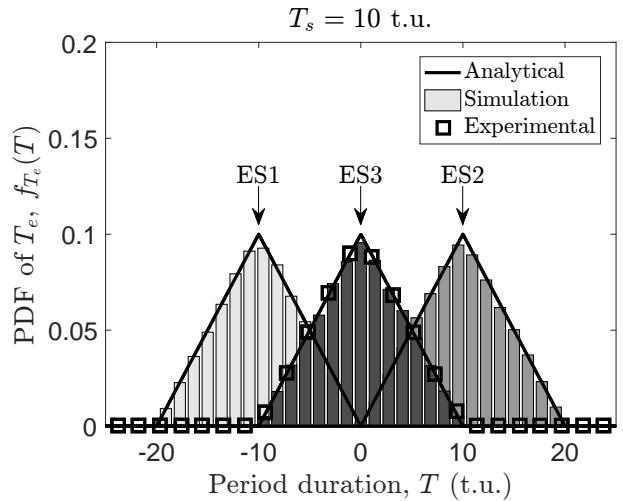


Fig. 2. PDF of the error component  $T_e$  of the model for  $\hat{T}_i$  in (1) for a sensing period  $T_s = 10$  t.u.

obtained in this work. The sequences of random periods were generated both as independent random numbers as well as inducing several correlation features using the method described in [49]. In both cases the same results were obtained, thus indicating that the presence of correlations in the primary traffic does not affect the estimation of the considered primary activity statistics and therefore the obtained analytical results are valid regardless of whether the primary traffic is correlated or uncorrelated.

Let consider first the models proposed in Section II for the estimation of individual periods. Fig. 2 compares the expression in (2) for the PDF of the error component  $T_e$  of the first model proposed in (1) with simulation and experimental results (experimental results are shown for ES3 only). A sensing period  $T_s = 10$  t.u. was considered and the absolute differences between each original period  $T_i$  and the corresponding estimated period  $\hat{T}_i$  were computed to determine the estimation errors  $T_e$ , which were then used to calculate a histogram (empirical PDF). As it can be appreciated, there is a perfect match, which validates the model in (1)–(2). Fig. 3 compares the expressions in (5) for the Bernoulli probabilities of the second model proposed in (4) with simulation and experimental results obtained by sensing a period  $T_i = 100$  t.u. with  $T_s \in (0, T_i]$ . For each simulated value of  $T_s$  the sensing process was repeated 1000 times starting at different time instants so that the sensing events could take place at different relative positions with respect to the original period  $T_i$ . The first sensing event was shifted from the beginning of the original period  $T_i$  by a random amount uniformly distributed in  $[0, T_s]$ . The probability  $P(\xi = 0)$  was estimated as the proportion of cases where the estimated period was  $\hat{T}_i = \lfloor T_i/T_s \rfloor T_s$ , while  $P(\xi = 1)$  was obtained as the proportion of cases where  $\hat{T}_i = \lceil T_i/T_s \rceil T_s$ . The experiments were conducted by transmitting  $10^5$  pairs of idle/busy periods with duration  $T_i = 100$  ms and computing the probabilities  $P(\xi = 0)$  and  $P(\xi = 1)$  as the proportion of cases where the observed periods were  $\hat{T}_i = \lfloor T_i/T_s \rfloor T_s$  and  $\hat{T}_i = \lceil T_i/T_s \rceil T_s$ , respectively. As it can be appreciated in Fig. 3, all results show

<sup>3</sup>The software component of the PECAS prototype is based on free open source code available at: <http://www.lopezbenitez.es/misc/PECAS.zip>

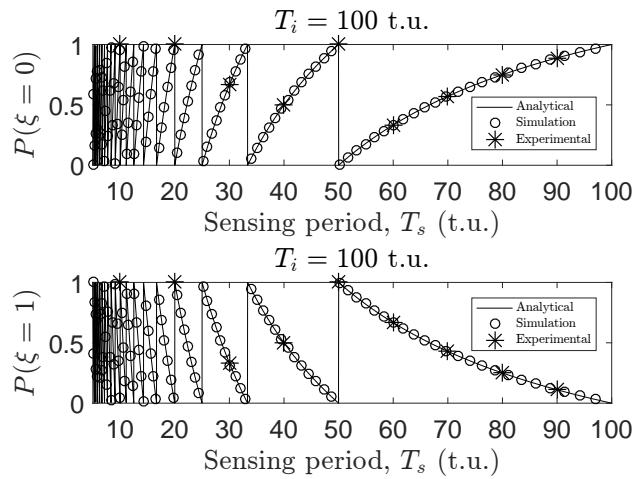


Fig. 3. Bernoulli probabilities of the model for  $\widehat{T}_i$  in (4) as a function of the sensing period  $T_s$  for a period  $T_i = 100$  t.u.

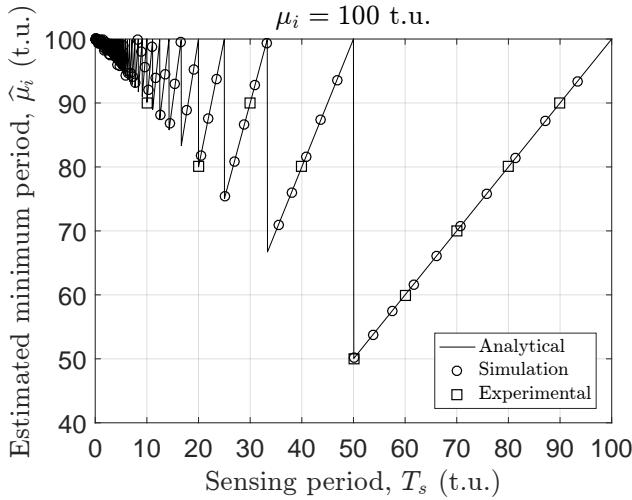


Fig. 4. Estimated minimum period  $\widehat{\mu}_i$  as a function of the sensing period  $T_s$  for a true minimum period  $\mu_i = 100$  t.u.

a perfect agreement, which validates the model in (4)–(5).

Fig. 4 shows the expression in (7) for the estimated minimum period  $\widehat{\mu}_i$  as a function of the sensing period  $T_s$  when the true minimum period is  $\mu_i = 100$  t.u., which provides a perfect match with simulation and experimental results. It is worth noting, as pointed out in Section III-A, that the estimated minimum is equal to the true minimum if the sensing period is an integer submultiple of the true minimum. However, as appreciated, the estimated minimum period can in general differ significantly from the true value. An exhaustive search method for  $T_s$  based on (7) was discussed in Section III-A along with an alternative method based on fitting (8) to a reduced set of tested sensing periods  $\{T_{s,l}\}_{l=1}^L$  and observed minimum periods  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$ . The performance of the latter is illustrated in Fig. 5, which shows the estimated minimum  $\widehat{\mu}_i$  as a function of the number of terms considered in the sum of (8). These results were obtained by fitting (8) to four different pairs of sets  $\{T_{s,l}\}_{l=1}^L$  and  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$ . For each of the four considered

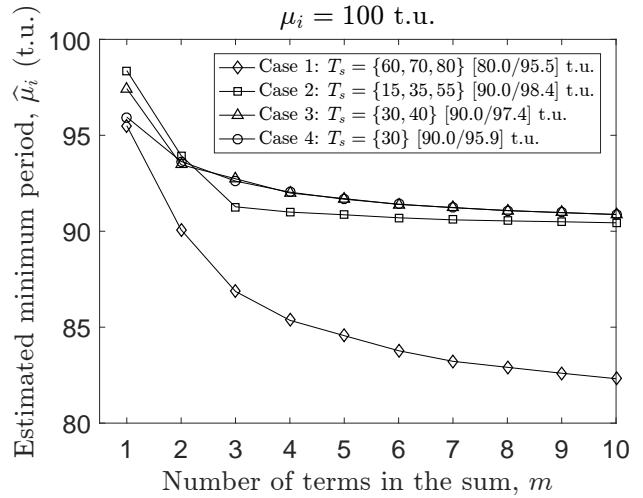


Fig. 5. Curve-fitting based estimated minimum period  $\widehat{\mu}_i$  as a function of the number of terms considered in (8) for a true minimum period  $\mu_i = 100$  t.u.

cases, the legend of Fig. 5 shows the set of tested sensing periods  $\{T_{s,l}\}_{l=1}^L$  and two additional values in the format  $[A/B]$ , where  $A$  is the best estimation provided individually by any of the tested sensing periods in the set  $\{T_{s,l}\}_{l=1}^L$  and  $B$  is the estimation obtained by means of a non-linear least squares fit of (8) to  $\{T_{s,l}\}_{l=1}^L$  and  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$ . As it can be appreciated, the proposed curve-fitting approach can provide significant accuracy improvements in the estimation of the minimum period duration. For example, in Case 1 the proposed approach can reduce the relative error of the estimated minimum from  $(100 - 80)/100 = 20\%$  to  $(100 - 95.5)/100 = 4.5\%$ , while in the other three cases the relative error is reduced from 10% to 1.6% (Case 2), 2.6% (Case 3), and 4.1% (Case 4). Note that these accurate estimations can be obtained by testing just a few sensing periods, which can be as few as  $L = 3$  (cases 1 and 2) or  $L = 2$  (case 3), and in some cases only  $L = 1$  (case 4). This represents a significant reduction compared with the exhaustive search approach discussed in Section III-A. It is worth noting that the best estimation is (typically) obtained, as observed in Fig. 5, when the fitted expression in (8) includes only one term ( $m = 1$ ). This is because as the number of terms considered in (8) increases, the expression reproduces more accurately the discontinuities of the floor function in (7) and this leads to a poorer performance of curve-fitting methods based on gradient search. It is also worth noting that this approach was observed to provide the best results when the initial value of  $\mu_i$  for the curve-fitting algorithm is the highest value observed in the sample set  $\{\widehat{\mu}_{i,l}\}_{l=1}^L$ .

Fig. 6 compares the analytical expressions for the estimated moments, including both the sample moments  $\widehat{x}_i$  in (9)–(12) and their corrected versions  $\tilde{x}_i$  in (21)–(24), with the simulation and experimental counterparts. As it can be appreciated, there is a perfect agreement, which validates the analytical results. Note that the default sample moments  $\widehat{x}_i$  provide an accurate estimation of the true moments (shown on top of each subfigure) only for very low values of the sensing period ( $T_s \approx 0$ ) and the accuracy rapidly degrades as  $T_s$  increases. However, the corrected moments  $\tilde{x}_i$ , calculated

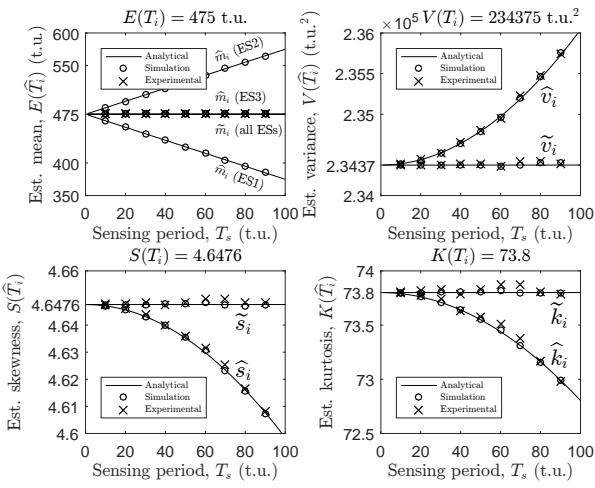


Fig. 6. Estimated moments ( $\hat{v}_i$ ) and their corrected versions ( $\tilde{v}_i$ ) as a function of the sensing period  $T_s$  ( $\mu_i = 100$  t.u.,  $\lambda_i = 300$  t.u.,  $\alpha_i = 0.2$ ).

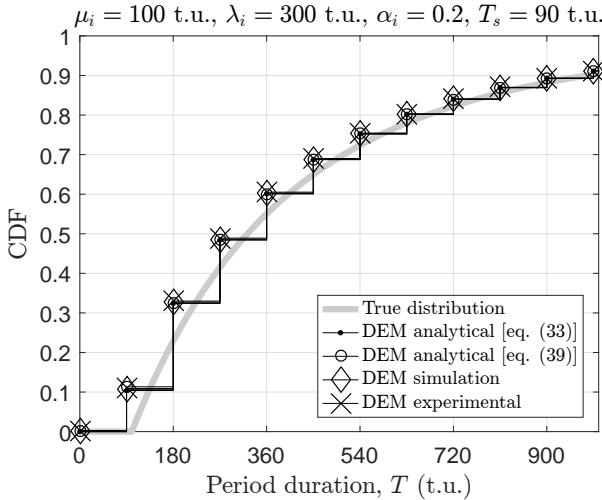


Fig. 7. Original and DEM-estimated CDFs ( $\mu_i = 100$  t.u.,  $\lambda_i = 300$  t.u.,  $\alpha_i = 0.2$ ,  $T_s = 90$  t.u.).

as shown in (21)–(24), remove the impact of  $T_s$  and provide accurate estimations of the true moments regardless of the employed sensing period.

Fig. 7 compares the CDF of GP-distributed periods (with  $\mu_i = 100$  t.u.,  $\lambda_i = 300$  t.u., and  $\alpha_i = 0.2$ ) with the CDF estimated by DEM when a sensing period  $T_s = 90$  t.u. is employed. The estimated distribution is evaluated analytically using both (35) and (39). Simulation and experimental results are also included to validate the analytical expressions. As it can be appreciated, the distribution estimated with DEM can diverge significantly from the true distribution. For example, the value of the estimated CDF at  $\hat{T} = T_s = 90$  t.u. is approximately 0.11, while the value of the true distribution at the same point is zero (estimation error of  $D_{KS,i} = 0.11$ ). Similarly, the value of the estimated CDF at  $\hat{T} = 2T_s = 180$  t.u. is approximately 0.33 while the value of the true distribution at the same point is 0.23 (estimation error of  $D_{KS,i} = 0.10$ ). This estimation error can obviously be decreased by reducing

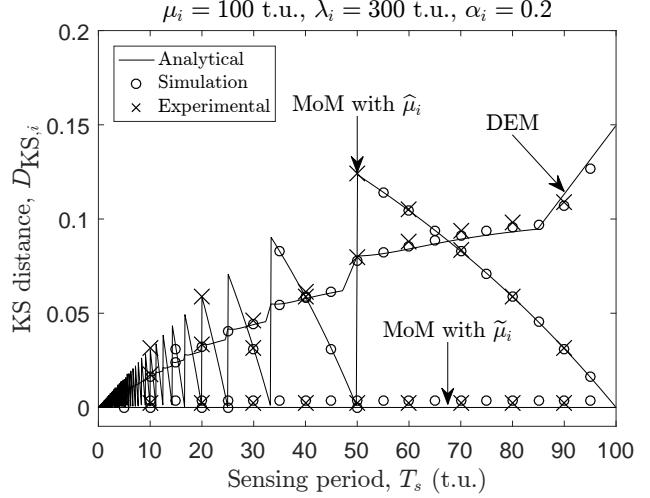


Fig. 8. Accuracy of DEM and MoM as a function of the sensing period ( $\mu_i = 100$  t.u.,  $\lambda_i = 300$  t.u.,  $\alpha_i = 0.2$ ).

the sensing period  $T_s$ , however this may not be possible in practice since the value of  $T_s$  is typically selected to achieve a certain signal detection performance under specified hardware constraints rather than an accurate estimation of the primary activity statistics. To overcome this limitation, the MoM approach is proposed.

Fig. 8 compares the estimation accuracy (in terms of the KS distance) of DEM and MoM as a function of the employed sensing period. The performance of MoM is shown for two cases: when the minimum period duration  $\hat{\mu}_i$  is estimated as the minimum value of the sequence of observed periods,  $\hat{\mu}_i = \min_n(\{\hat{T}_{i,n}\}_{n=1}^N)$ , and when the minimum period duration  $\tilde{\mu}_i$  is estimated based on corrected sample moments as shown in (43)–(45). The analytical curve for DEM is obtained by evaluating (41), while the counterparts for MoM are obtained by evaluating the maximum absolute difference between (26b) with the true parameters and (26b) with the MoM-estimated parameters. Simulation and experimental results are also included to validate the analytical expressions. As it can be appreciated, DEM can provide an accurate estimation of the true distribution only if the employed sensing period is sufficiently short. The MoM approach can provide accurate estimations provided that the estimated minimum period is sufficiently accurate. The minimum period duration estimated based on (7) can be significantly inaccurate depending on the employed sensing period and this has a direct impact on the overall accuracy of the corresponding MoM-estimated distribution. It is worth noting, as discussed above, that when the sensing period is an integer submultiple of the true minimum, then the minimum period estimated with (7) is accurate and in such a case the estimated distribution is accurate as well (i.e., the points in Fig. 8 where MoM with  $\hat{\mu}_i$  leads to  $D_{KS,i} = 0$ ). However, the minimum period estimated based on (7) will in general differ significantly from the true value and as a result the estimated distribution can be significantly inaccurate (in some cases the estimation error is even higher than that of DEM as observed in Fig. 8). On the other hand, the minimum

period estimated based on (43)-(45) is accurate regardless of the employed sensing period and so is the resulting estimated distribution (MoM with  $\tilde{\mu}_i$  leads to  $D_{KS,i} \approx 0$  for all sensing periods). These results demonstrate that the MoM approach based on the proposed moment estimation methods can provide a highly accurate (virtually perfect) estimation of the distribution of primary periods regardless of the employed sensing period.

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 estimating the opportunistic data rate that a DSA/CR system can expect from a primary channel. The opportunistic data rate can be useful for example in channel selection decisions or to determine the optimum schedule for the task of sensing a number of primary channels by a set of secondary users as investigated in detail in [9]. The expected opportunistic data rate ( $\hat{R}_b$ ) can be expressed as a function of the estimated duty cycle ( $\hat{\Psi}$ ) as  $\hat{R}_b = (1 - \hat{\Psi})W\eta$ , where  $W$  is the primary channel bandwidth and  $\eta$  is the spectrum efficiency associated to the modulation and coding schemes used by the DSA/CR system. Fig. 9 shows the estimated available data rate (with  $W = 20$  MHz and  $\eta = 2$  bit/s/Hz) as a function of the estimated channel duty cycle for the three estimation strategies described in Section II, with and without the corrections derived in (21) from the analysis carried out in Section III-B. As it can be appreciated, ES1 and ES2 can lead to inaccurate predictions if the impact of the sensing period is not removed by applying appropriate corrections. For ES1, the maximum error is observed at  $\hat{\Psi} = 0.33$ , where the actual data rate is 26.91 Mbit/s but the resulting prediction is 32.84 Mbit/s, which represents an overestimation error of 22.05% (this would result in a significant outage for the DSA/CR system since the actual available capacity would be significantly lower than expected). For ES2, the maximum error is observed at  $\hat{\Psi} = 0.22$ , where the actual data rate is 31.03 Mbit/s but the resulting prediction is 28.38 Mbit/s, which represents an underestimation error of 8.55% (this would result in a significant waste of free capacity since the actual available capacity would be significantly higher than expected). On the other hand, by applying the corrections proposed for ES1/ES2 or using ES3, an accurate estimation would be obtained and the spectrum resources would be exploited efficiently. This simple 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 and methods that this work has provided for an accurate estimation of the primary activity statistics.

## VI. CONCLUSIONS

DSA/CR users can monitor periodically the idle/busy state of primary channels to estimate the primary activity statistics and exploit this information to improve the system performance. However, the unavoidable use of a finite sensing period imposes fundamental limits on the temporal resolution to which the duration of idle and busy periods can be observed. This work has carried out a comprehensive

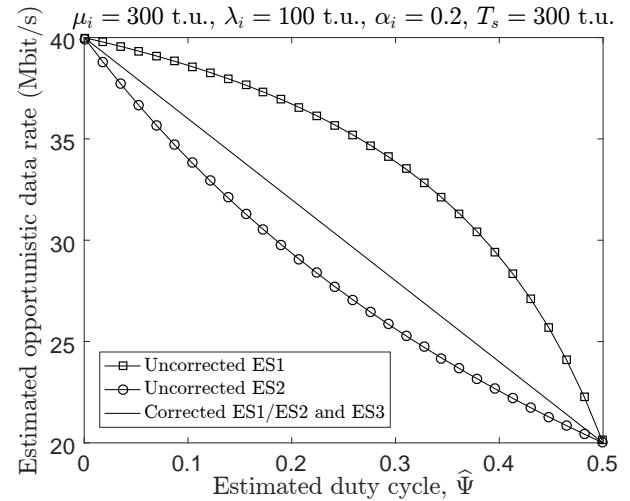


Fig. 9. Estimated opportunistic data rate as a function of the estimated duty cycle ( $\mu_i = 300$  t.u.,  $\lambda_i = 100$  t.u.,  $\alpha_i = 0.2$ ,  $T_s = 300$  t.u.,  $W = 20$  MHz,  $\eta = 2$  bit/s/Hz).

analytical study on how this limitation affects the estimation of individual idle/busy periods, their minimum, their moments (mean, variance, skewness and kurtosis) and the underlying distribution. Closed-form expressions have been developed to determine the relation between the estimated and the true statistics as a function of the employed sensing period as well as the resulting estimation error. The obtained analytical results can be useful in the analysis, design and simulation of DSA/CR systems. Various estimation strategies and methods have also been proposed to minimise the estimation error resulting from the use of a finite sensing period. Simulation and experimental results have demonstrated that the proposed methods can minimise the impact of the sensing period, thus providing a highly accurate (virtually perfect) estimation of all the considered primary activity statistics regardless of the employed sensing period.

While this work has focused on the impact of the sensing period, another important aspect is the impact of the sample size (i.e., how the accuracy of the estimated statistics depends on the number of observed periods). Moreover, this work implicitly focuses on a link-level scenario where a single secondary receiver is considered. The extension to a network-level scenario where multiple secondary nodes cooperate to produce a more accurate estimation is another interesting problem worth investigating. These problems will be addressed in our future work.

## REFERENCES

- [1] Q. Zhao and A. Swami, "A survey of dynamic spectrum access: signal processing and networking perspectives," in *Proc. IEEE Int'l. Conf. Acoustics, Speech and Signal Process. (ICASSP 2007)*, vol. 4, Apr. 2007, pp. IV/1349–IV/1352.
- [2] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.
- [3] M. M. Buddhikot, "Understanding dynamic spectrum access: Taxonomy, models and challenges," in *Proc. 2nd IEEE Int'l. Symp. Dyn. Spect. Access Networks (DySPAN 2007)*, Apr. 2007, pp. 649–663.
- [4] Y.-C. Liang, K.-C. Chen, G. Y. Li, and P. Mähönen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Vehic. Tech.*, vol. 60, no. 7, pp. 3386–3407, Sep. 2011.

- [5] M. López-Benítez, "Cognitive radio," in *Heterogeneous cellular networks: Theory, simulation and deployment*. Cambridge University Press, 2013, ch. 13, pp. 383–425.
- [6] S. Yarkan and H. Arslan, "Binary time series approach to spectrum prediction for cognitive radio," in *Proc. IEEE 66th Vehic. Tech. Conf. (VTC 2007 Fall)*, Sep. 2007, pp. 1563–1567.
- [7] V. K. Tumuluru, P. Wang, and D. Niyato, "A neural network based spectrum prediction scheme for cognitive radio," in *Proc. 2010 IEEE Int'l. Conf. Comms. (ICC 2010)*, May 2010, pp. 1–5.
- [8] Y. Chen and H.-S. Oh, "Spectrum measurement modelling and prediction based on wavelets," *IET Communications*, vol. 10, no. 16, pp. 2192–2198, Oct. 2016.
- [9] C.-H. Liu, A. Azarfar, J.-F. Frigon, B. Sansò, and D. Cabric, "Robust cooperative spectrum sensing scheduling optimization in multi-channel dynamic spectrum access networks," *IEEE Trans. Mobile Comp.*, vol. 15, no. 8, pp. 2094–2108, Aug. 2016.
- [10] 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.
- [11] J. Virtanen, M. Höyhtyä, J. Lehtomäki, and T. Bräysy, "Priority channel selection based on detection history database," in *Proc. Fifth Int'l. Conf. Cognitive Radio Oriented Wireless Networks & Comms. (CROWNCOM 2010)*, Jun. 2010, pp. 1–5.
- [12] M. Höyhtyä, S. Pollin, and A. Mämmelä, "Classification-based predictive channel selection for cognitive radios," in *Proc. 2010 IEEE Int'l. Conf. Comms. (ICC 2010)*, May 2010, pp. 1–6.
- [13] F. Bouali, O. Sallent, J. Pérez-Romero, and R. Agustí, "A novel spectrum selection strategy for matching multi-service secondary traffic to heterogeneous primary spectrum opportunities," in *Proc. IEEE 22nd Int'l. Symp. Pers. Indoor and Mobile Radio Comms. (PIMRC 2011)*, Sep. 2011, pp. 417–422.
- [14] ———, "Strengthening radio environment maps with primary-user statistical patterns for enhancing cognitive radio operation," in *Proc. Sixth Int'l. ICST Conf. Cognitive Radio Oriented Wireless Networks and Comms. (CROWNCOM 2011)*, Jun. 2011, pp. 256–260.
- [15] E. Jung and X. Liu, "Opportunistic spectrum access in multiple-primary-user environments under the packet collision constraint," *IEEE/ACM Trans. Networking*, vol. 20, no. 2, pp. 501–514, Apr. 2012.
- [16] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive medium access: constraining interference based on experimental models," *IEEE J. Sel. Areas Comms.*, vol. 26, no. 1, pp. 95–105, Jan. 2008.
- [17] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Comms.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.
- [18] H. Su and X. Zhang, "Cross-layer-based opportunistic MAC protocols for QoS provisioning over cognitive radio wireless networks," *IEEE J. Sel. Areas Comms.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.
- [19] K. W. Sung, S.-L. Kim, and J. Zander, "Temporal spectrum sharing based on primary user activity prediction," *IEEE Trans. Wireless Comms.*, vol. 9, no. 12, pp. 3848–3855, Dec. 2010.
- [20] A. Anandkumar, N. Michael, A. K. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE J. Sel. Areas Comms.*, vol. 29, no. 4, pp. 731–745, Apr. 2011.
- [21] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Comms. Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130, First Quarter 2009.
- [22] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing techniques for cognitive radio: Challenges and solutions," *EURASIP J. Advances Signal Process.*, vol. 2010, pp. 1–15, Jan. 2010.
- [23] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 785–796, Feb. 2008.
- [24] W. Gabran, C.-H. Liu, P. Pawelczak, and D. Cabric, "Primary user traffic estimation for dynamic spectrum access," *IEEE J. Sel. Areas Comms.*, vol. 31, no. 3, pp. 544–548, Mar. 2013.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] ———, "Dynamic spectrum access in WLAN channels: Empirical model and its stochastic analysis," in *Proc. First Int'l. Works. Tech. and Policy for Accessing Spectr. (TAPAS 2006)*, Aug. 2006, pp. 1–10.
- [30] ———, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Comms. Mag.*, vol. 45, no. 5, pp. 66–72, May 2007.
- [31] 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.
- [32] 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.
- [33] J. J. Lehtomäki, M. López-Benítez, K. Umebayashi, and M. Juntti, "Improved channel occupancy rate estimation," *IEEE Trans. Comms.*, vol. 63, no. 3, pp. 643–654, Mar. 2015.
- [34] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki, and K. Umebayashi, "Accurate estimation of primary user traffic based on periodic spectrum sensing," in *Proc. IEEE Wireless Comms. and Networking Conf. (WCNC 2018)*, Barcelona, Spain, Apr. 2018, pp. 1–6.
- [35] R. J. Larsen and M. L. Marx, *An introduction to mathematical statistics and its applications*, 5th ed. Prentice Hall, 2012.
- [36] 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.
- [37] 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 Conf. (EW 2018)*, May 2018, pp. 131–136.
- [38] E. C. Titchmarsh and D. R. Heath-Brown, *The theory of the Riemann zeta-function*, 2nd ed. Oxford, 1986.
- [39] 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.
- [40] M. Evans, N. Hastings, and B. Peacock, *Statistical distributions*, 3rd ed. Wiley, 2000.
- [41] M. López-Benítez and F. Casadevall, "Empirical time-dimension model of spectrum use based on discrete-time Markov chain with deterministic and stochastic duty cycle models," *IEEE Trans. Vehic. Tech.*, vol. 60, no. 6, pp. 2519–2533, Jul. 2011.
- [42] ———, "Time-dimension models of spectrum usage for the analysis, design and simulation of cognitive radio networks," *IEEE Trans. Vehic. Tech.*, vol. 62, no. 5, pp. 2091–2104, Jun. 2013.
- [43] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, 2nd ed. Wiley, Nov. 1994, vol. 1.
- [44] 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. IEEE 28th Int'l. Symp. Pers. Indoor and Mobile Radio Comms. (PIMRC 2017 workshops)*, Mar. 2017, pp. 1–5.
- [45] 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.
- [46] T. Öztekin, "Comparison of parameter estimation methods for the three-parameter generalized Pareto distribution," *Turkish J. Agriculture and Forestry*, vol. 29, no. 6, pp. 419–428, Dec. 2005.
- [47] M. López-Benítez, "Can primary activity statistics in cognitive radio be estimated under imperfect spectrum sensing?" in *Proc. IEEE 24th Int'l. Symp. Pers. Indoor and Mobile Radio Comms. (PIMRC 2013)*, Sep. 2013, pp. 750–755.
- [48] 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. IEEE Wireless Comms. and Networking Conf. (WCNC 2017)*, Mar. 2017, pp. 1–6.
- [49] M. López-Benítez and F. Casadevall, "Modeling and simulation of time-correlation properties of spectrum use in cognitive radio," in *Proc. 6th Int'l. Conf. Cogn. Radio Oriented Wireless Networks and Comms. (CrownCom 2011)*, Jun. 2011, pp. 326–330.



**Miguel López-Benítez** (S'08, M'12, SM'17) received the BSc and MSc degrees in Telecommunication Engineering (both with Distinction) from Miguel Hernández University, Elche, Spain in 2003 and 2006, respectively, and the PhD degree in Telecommunication Engineering (with Highest Distinction, *summa cum laude*) from the Technical University of Catalonia, Barcelona, Spain in 2011. He was a Research Fellow in the Centre for Communication Systems Research of the University of Surrey, UK from 2011 to 2013. In 2013 he became a Lecturer (Assistant Professor) in the Department of Electrical Engineering and Electronics of the University of Liverpool, UK and was promoted to Senior Lecturer (Associate Professor) in 2018, which is his current position.

His research interests are in the field of wireless communications and networking, with special emphasis on mobile communications and dynamic spectrum access in cognitive radio systems. He is/has been the principal investigator or co-investigator of research projects funded by the EPSRC, British Council and Royal Society, and has been involved in the European-funded projects AROMA, NEWCOM++, FARAMIR, QoS MOS and CoRaSat. He is Associate Editor of IEEE Access, IET Communications, and Wireless Communications and Mobile Computing, and has been a member of the Organising Committee for the IEEE WCNC International Workshop on Smart Spectrum (IWSS 2015-18). Please visit <http://www.lopezbenitez.es> for details.

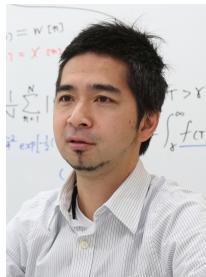


**Janne Lehtomäki** (S'03, M'06) received the Ph.D. degree from the University of Oulu, Oulu, Finland, in 2005. He was with Georgia Tech, Atlanta, GA, USA, as a Visiting Scholar, in 2013. He is currently an Adjunct Professor with the Centre for Wireless Communications, University of Oulu. His current research interest focuses on terahertz band wireless communication. Dr. Lehtomäki was a co-recipient of the Best Paper Award of IEEE WCNC 2012. He is an Editorial Board member of Physical Communication. He was a General Co-Chair of the IEEE WCNC

2017 & 2018 International Workshop on Smart Spectrum, a TPC Co-Chair of the IEEE WCNC 2015 and 2016 International Workshop on Smart Spectrum, and a Publicity/Publications Co-Chair of ACM NANOCOM in 2015-2018. He has served as a Guest Associate Editor for two IEICE Transactions on Communications Special Sections and as a Managing Guest Editor of the Nano Communication Networks Special Issue in 2016.



**Ahmed Al-Tahmeesschi** (S'16) received the BSc and MSc degrees in Communications Engineering from the University of Technology, Baghdad, Iraq, in 2009 and 2011, respectively. He is currently pursuing a PhD degree in the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK. His research interests include cognitive radio networks, dynamic spectrum access techniques, algorithm design and machine learning.



**Kenta Umebayashi** (S'00, M'04) received the LL.B. degree from Ritsumeikan University, Japan in 1996 and the B.E., M.E., and Ph.D. degrees from the Yokohama National University, Japan in 1999, 2001, and 2004, respectively. From 2004 to 2006, he was a Research Scientist at the Centre for Wireless Communications, University of Oulu, Finland. He is currently an Associate Professor at the Tokyo University of Agriculture and Technology, Japan. He was an Associate Editor of IEICE Transactions on Communications from May 2015 to May 2017. He

was a principal investigator in the four Grants-in-Aid for Scientific Research projects and three Strategic Information and Communications R&D Promotion Programme projects. His research interests lie in the areas of signal detection and estimation theories for wireless communication, signal processing for multiple antenna systems, cognitive radio networks, and terahertz band wireless communication. He received the Best Paper Award at IEEE WCNC 2012 for a paper he authored, and the Best Paper Award at IEEE WCNC workshop IWSS 2015 for a paper he co-authored.



**Dhaval K. Patel** has been an Assistant Professor at the School of Engineering and Applied Science, Ahmedabad University, India since 2014. He was also a visiting faculty at Franklin W. Olin College of Engineering, Massachusetts, USA. He worked as Junior Research Fellow in the Post Graduate Lab for Communication Systems at Nirma University, India from 2011 to 2014. He received the B.E. and M.E. degrees in Communication Systems Engineering (both with Distinction/First-Class) from Gujarat University in 2003 and 2010, respectively,

and a Ph.D. degree in Electronics and Communications from the Institute of Technology, Nirma University, India in 2014. His research areas of interest include vehicular cyber-physical systems, 5G wireless networks, non-parametric statistics and physical layer security. He is the principal investigator of research projects funded by Department of Science and Technology (DST), UK-India Education and Research Initiative (UKIERI), Association of South-East Asian Nations (ASEAN)-India Collaborative R&D Project and Gujarat Council on Science and Technology (GUJCOST).