RF-Spectrum Opportunities for Cognitive Radio Networks Operating Over GSM Channels

Miguel Luís, *Member, IEEE*, Rodolfo Oliveira, *Senior Member, IEEE*, Rui Dinis, *Senior Member, IEEE*, and Luis Bernardo, *Member, IEEE*

Abstract—In this paper, we characterize the radio frequency spectrum opportunities available in a common global system for mobile communications (GSM) channel to support the operation of a cognitive radio network (CRN). In a first step, we describe the technical details involved to sample the channel using a software defined radio device. Adopting a simple energy-based detector, we identify the two energy regions where the GSM system is active or inactive and evaluate the spectrum sensing accuracy. Based on the output of the detector, we show that the distribution of the durations of busy and idle periods are approximated by geometric distributions. Finally, we validate a theoretical model for the distribution of the service time. The validation results indicate that the service time can be successfully represented by a discrete generalized Pareto distribution, which is confirmed by the Kolmogorov–Smirnov test. Because the throughput of the CRN is represented by the inverse of the service time, the proposed analysis provides an upper bound for the networks' throughput, indicating the maximum throughput that can be attained when a single secondary user transmits over a GSM cellular channel. The results presented in this paper are validated with real data, confirming the accuracy of the proposed service time model.

Index Terms—Cognitive radio networks, service time, performance evaluation and modeling.

I. INTRODUCTION

IN COGNITIVE Radio Networks (CRNs) the transmission channel is licensed to the primary users (PUs), while secondary users (SUs) only access the channel in an opportunistic way when the PUs are inactive, *i.e.,* when the PUs do not use the channel [1], [2]. Because the channel is used by the SUs opportunistically, a SU transmission must be halted whenever a PU becomes active. In a scenario where a SU needs to transmit multiple packets (e.g., in a file transmission), or when a packet may be too long, the amount of time required to finish the SU's service (Service Time) depends on the number

and duration of the PUs' transmissions. By definition, the service time is the interval of time from the instant when the data arrives at the head of the SU transmitting queue (e.g., a packet or a file, depending on the network stack layer), until the instant when its transmission ends. Service time is an important metric in CRNs because it incorporates the level of activity of the PUs. In this work we characterize the service time of a cognitive radio network operating in a GSM channel.

The rest of this section describes related literature and highlights the contributions of our work. Section II describes the RF spectrum sensing procedure, by introducing the experimental setup and describing the steps followed to determine the sensing accuracy. Section III characterizes the duration of the periods when the GSM channel is found idle or busy due to the GSM operation. Based on the statistics of the experimental data, Section IV summarizes the steps involved to numerically compute the distribution of the service time. In Section V we assess the accuracy of the numerical results by comparing them with experimental data. Finally, conclusions are drawn in Section VI.

A. Related Work

Over the last years, with the introduction of the CR concept, a high number of publications have focused on the characterization of transmission opportunities in GSM channels. Recently, the work in [3] evaluated the possibility of using GSM whitespaces in rural areas for dynamic spectrum sharing, therefore supporting the growth of community cellular networks, and consequently, improving the rural access to communications services. Gao *et al.* [4] analyze the practical capacity limits that can be achieved by an opportunistic network when using the 850 MHz GSM uplink frequency. The capacity limits were evaluated under different constraints such as the instantaneous interference power limit to the GSM basestation and the SU transmit power limit. The work in [5] addresses the possibility of exploiting the underutilized GSM channels for secondary communications assuming two distinct communication strategies: i) an interweave communication strategy, where the unlicensed network can only use the channel when it is found vacant, and ii) an underlay strategy, allowing concurrent primary and secondary transmissions. After comparing the performance of both strategies through simulation the authors focus their work on the access strategies to be adopted by the secondary network under

M. Luís is with the Instituto de Telecomunicações, 3810-193 Aveiro, Portugal (e-mail: nmal@av.it.pt).

R. Oliveira, R. Dinis, and L. Bernardo are with the Instituto de Telecomunicações, 1049-001 Lisbon, Portugal, and also with the Departamento de Engenharia Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal (e-mail: rado@fct.unl.pt; rdinis@fct.unl.pt; lflb@fct.unl.pt).

the underlay paradigm. Exploring a different perspective, [6] presented a novel dynamic spectrum access scheme to use the available spectrum vacancies in the GSM band. The work relies on extensive simulation results, showing the advantages of simultaneously adopting TDMA and FDD features to enhance the spectrum access performance in cognitive radio networks.

A few works describe the GSM spectrum availability at specific locations. Reference [7] evaluates the GSM spectrum availability at Jaipur city, India, describing the experimental measurement setup and the analysis procedure followed in the study. The work identifies multiple spectrum spaces which can support potential CR applications. The authors used a spectrum analyzer in the GSM band to determine the average power spectral density. Similarly, [8] also identifies and characterizes the channel opportunities over a GSM network in Aveiro, Portugal. The results have showed that RF-spectrum opportunities over GSM licensed frequencies follow an exponential behavior. Similar results were observed in [9] and [10]. Patil *et al.* [11] characterize the temporal opportunities through detailed measurements obtained in GSM channels, including statistical as well as spectral occupancy details. However, the works in [4], [5], and [8]–[11] focused mostly on characterizing the transmission opportunities over the GSM licensed bands and do not address the theoretical characterization of the service time when using the transmission opportunities.

The characterization of different algorithms capable of identifying GSM signals at specific bands have also attracted attention. Reference [12] proposed an algorithm to identify if GSM signals are present in a given band. Compared with other algorithms in the literature, the proposed algorithm improves the identification performance with short observation intervals at low SNRs. Instead of adopting a specific spectrum sensing module, the spectrum availability in [7] and [12] were determined by the power level measured by a spectrum analyzer, which may penalize the analysis, namely in terms of time-granularity of the spectrum availability. Contrarily, [13] analyzed the spectrum opportunities available in the GSM bands adopting an energy detector to sense the spectrum in real-time. Reference [13] characterized the duration of the available spectral opportunities, quantifying the service time required to transmit different packets. Both [13] and [14] are based on results obtained with Software Defined Radio devices, which are particularly useful to test innovative schemes in realistic radio environments. Reference [14] demonstrates the integration of cognitive radio capabilities, on the fly, such as the spectrum sensing of GSM bands being used by secondary users and its interaction with GSM basestations. SDR devices were also adopted in [15] to sense GSM bands in order to feed artificial neural networks able to predict real world RF power within the GSM 900 band. The main purpose of the work described in [15] is to predictively select channels with the least noise among those that were unused, improving traditional channel selection schemes proposed to GSM channels. However, while the works [7], [12], [14], [15] show the existence of spectrum opportunities in GSM bands, they do not rely on an analytical model capable of describing

(and estimating) the service time when using the available channel opportunities.

Motivated by the successful identification and characterization of spectrum opportunities over GSM channels, recently a few works have characterized the service time in a theoretical way by considering the case when the SUs handoff to a different wireless transmission channel after a PU transmission [16], or the case when the SUs wait for the PU to vacate the channel [17]. In this work we consider that SUs do not handoff to a different channel and wait for the PU to vacate the channel. The purpose of this assumption is to focus our attention in the characterization of the spectrum opportunities available on a single channel.

The average packet service time of CRNs was studied in [18] and [19] without considering the role of the size of the SU's and PU's packets. Tran *et al.* [20] evaluate the average service time of delay-sensitive and delay-insensitive packets showing that the average service time of the packets transmitted by a SU not only depends on the arrival rate and size of SU's packets but also on the arrival rate and size of the PU's packets. In [21] the service time is analyzed for a single channel CRN and the analysis is extended in [22] to characterize the average service response time of elastic data (*i.e.,* variable packet-length) when the PU activity follows an ON-OFF behavior with ON and OFF durations following exponential distributions. While the previous works only characterize the average of the service time, [23] derived expressions for the distribution of the packet service time when the SUs are traffic-saturated and fixed-length packets are transmitted. More recently, [24] proposed a theoretical characterization of the distribution of the service time when both saturated and non-saturated traffic conditions occur, and variable-length packets are transmitted by the SUs. Although the works in [18]–[24] propose different theoretical analysis of the SUs' service time, they rely on unvalidated initial assumptions about the statistics of the licensed radio system (e.g., the probabilities of sensing channel busy/idle periods and the distribution of its duration).

B. Contributions

Mainly motivated by the use of GSM channels by narrow-band devices acting as secondary users, such as the recently proposed NarrowBand IoT[1] devices, this work characterizes the spectrum opportunities available in GSM channels. The contributions of this work are summarized as follows:

- Differently from the works in [18]–[24], in this paper we use real data sampled from a GSM channel to characterize the distribution of the duration of idle and busy GSM channel periods;
- We describe the technical details to sample the GSM channel using a Software Defined Radio (SDR) device. The sampled data[2] was used to detect the activity of the GSM users, showing that the distribution of the duration of busy periods is approximated by a geometric distribution. The duration of the idle periods (occurring when the

---

[1]IoT is the acronym of Internet of Things.

[2]Available at http://tele1.dee.fct.unl.pt/projects/etc/GSM_sample_set.zip.

channel is not used by the GSM system) also follows a geometric distribution;

- We validate the model we have proposed in [24] for the distribution of the service time through the Kolmogorov-Smirnov test. The results presented in the paper are validated with real data, confirming the accuracy of the proposed service time model when used in GSM channels.

We highlight that the service time is an important metric in CRNs because the maximum throughput that the secondary network may achieve is the inverse of the service time. In this way the proposed analysis provides an upper bound on the throughput of the CRNs when a single SU transmits over a typical GSM cellular channel. The service time can also be used as a channel selection metric, when estimates are available for different channels. Similarly, the characterization of the spectrum opportunities available on a single channel can also benefit SUs that handoff to different channels, because the information of the spectrum opportunities can be used to schedule the SUs' transmissions in order to maximize the network performance [25]. In this way, the analysis proposed in this paper can be used to evaluate the comparative performance of a GSM channel, which may support efficient handoff schemes that maximize or improve the throughput of secondary users.

II. SPECTRUM SENSING

A. Experimental Setup

The experimental setup consisted of an Universal Software Radio Peripheral (USRP) model B100 (a Software Defined Radio (SDR) device manufactured by Ettus Research [26]), and a Linux-based laptop running GNUradio (an open-source signal processing tool [27]). The SDR equipment was tuned to the downlink GSM channel 94 (identified according to the Absolute Radio-Frequency Channel Number (ARFCN) list), available on 953.8 MHz. The SDR was configured to sample the radio channel and to use the sampled data to feed a Fast Fourier Transform (FFT) algorithm in real-time. The FFT was computed for a resolution bandwidth of 4 MHz and for 128 discrete points in the frequency domain, leading to a frequency spacing of 31.25 kHz between each FFT point. The sampling rate of the SDR was set to obtain a new FFT value every 32 μs, and the square magnitude of each sample i was computed to obtain its power value ($|x(i)|^2$).

Regarding the time and place where the spectrum data was collected, we opted to select a suburban region that was sampled during the highest period of utilization by the primary users. The data reported in the different results presented in the paper was sampled in the suburbs of Lisbon, Portugal, and similar results were obtained at Nova University Campus, located in Monte de Caparica, Portugal. The radio measurements took place for three consecutive working days (27, 28 and 29 of September, 2016), between 7pm and 8pm. We have decided to choose this time interval due to the increased use of the GSM network during this period. In the other periods of the day the spectrum activity was not so high, and although the same conclusions can be taken (in terms of family of the different distributions analyzed in our work), we decided to only focus on the most active period of spectrum utilization by the primary users, to highlight the spectrum opportunities in the worst case scenario. The sampled frequency is commercially explored by a Portuguese telecommunications operator, and we highlight that the data was collected during different days to increase the statistical diversity of the sample set.

B. Performance Evaluation

The detection of the presence or absence of activity in the GSM channel (caused by the operation of the GSM system) was implemented through an Energy-based (EB) detector [28]. The detector determines the amount of energy available on N_S samples,

$$Y = T_S \sum_{i=1}^{N_S} |x(i)|^2, \tag{1}$$

where $x(i)$ represents the average of the amplitude of the FFT discrete points lying in the 200 kHz GSM channel band, and $T_S = 32\mu s$ represents the channel sampling rate. Y is further compared with a decision threshold γ to decide if the GSM channel is being used by the GSM system.

To distinguish between occupied and vacant spectrum bands, SUs sample the channel and, for each sample k, two hypotheses can be distinguished

$$\begin{aligned} \mathcal{H}_0 &: x(k) = w(k), & k = 1, 2, \ldots, N_S \\ \mathcal{H}_1 &: x(k) = w(k) + s(k), & k = 1, 2, \ldots, N_S, \end{aligned} \tag{2}$$

where $s(k)$ denotes the GSM signal and $w(k)$ represents the noise floor. The performance of the EB detector is evaluated through the probability of miss-detection, denoted by P_m, which occurs when the EB detector classifies the channel as being idle when we have GSM activity, and the probability of false alarm, P_f, that occurs when the detector erroneously classifies the channel as being occupied when there is no GSM activity. These probabilities are formulated as

$$\begin{aligned} P_m &= Pr(Y \leq \gamma | \mathcal{H}_1) \\ P_f &= Pr(Y > \gamma | \mathcal{H}_0), \end{aligned}$$

where Y is the detector output in (1).

After evaluating the energy detector for different values of N_S, we concluded that 2 sensing samples ($N_S = 2$) are enough to select a decision threshold capable of providing an accurate decision of the current channel activity. To justify the adoption of $N_S = 2$, Figure 1 presents the normalized histogram of the energy (Y) collected during the experimental measurements, when $N_S = 2$ sensing samples are considered. As can be seen in Figure 1, the distribution of Y exhibits two distinct regions representing the presence or absence of activity in the GSM channel. The one on the left (until approximately -100 dBJ, which is equivalent to 98.9 pJ[3]) represents the sensing samples collected during the absence of activity in the channel, while the one on the right is due to the activity of the GSM system. From the obtained results, it is important to highlight the importance of the distance between the two regions,

---

[3]dBJ is obtained from $10\log_{10}(E)$, where E denotes the Energy, in Joules.
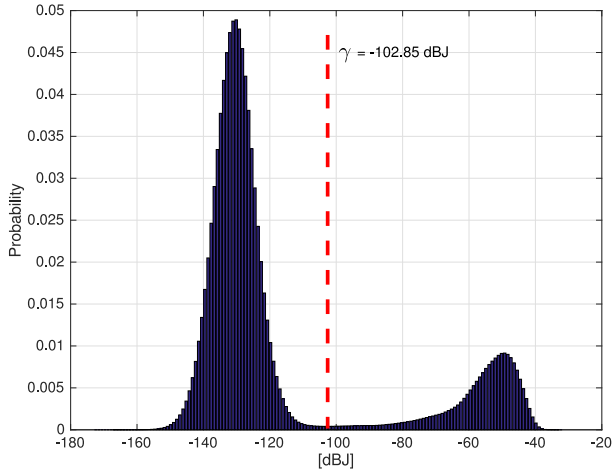
Fig. 1. Normalized histogram of the energy (Y) collected during the experimental measurements $(N_S = 2)$. The red line represents the selected threshold (γ).

because the probabilities of false alarm (incorrectly deciding the presence of GSM activity when the channel is idle) and miss-detection (incorrectly deciding the absence of GSM activity when it effectively occurs) decrease as the regions become more distant. Based on the distribution presented in Figure 1, the energy threshold adopted in the energy detector was set to $\gamma = -102.85$ dBJ (51.9 pJ).

To quantify the miss-detection and false alarm probabilities for $N_S = 2$ and $\gamma = -102.85$ dBJ, we collected sets of 2 samples and computed Y for $N_S = 2$. The different values of Y were then classified in two sets $\mathcal{Y}_{\mathcal{H}_0}$ and $\mathcal{Y}_{\mathcal{H}_1}$ if $Y \leq \gamma$ or $Y > \gamma$, respectively. In this way we obtained the approximate (empirical) cumulative distribution function (CDF) of $\mathcal{Y}_{\mathcal{H}_0}$ and $\mathcal{Y}_{\mathcal{H}_1}$. The empirical CDFs were used in a fit tool to identify the theoretical CDF that could better approximate the empirical data. The sets $\mathcal{Y}_{\mathcal{H}_0}$ and $\mathcal{Y}_{\mathcal{H}_1}$ were evaluated for different distributions (Normal, Skew Normal, Log-Normal, Generalized Extreme Value, Nakagami, Gamma, Shifted Log-Logistic, among others) taking into account the log-likelihood goodness of fit. From the distributions compared in the test, the Normal distribution was the one that better approximated the data in $\mathcal{Y}_{\mathcal{H}_0}$. Figure 2(a) represents the empirical CDF of $\mathcal{Y}_{\mathcal{H}_0}$ ("emp." curve), as well as the theoretical approximation ("approx." curve). Regarding the set $\mathcal{Y}_{\mathcal{H}_1}$, the log-likelihood test identifies the Shifted Log-logistic distribution as the best candidate and the theoretical approximation is represented in Figure 2(b).

To evaluate the error between the theoretical approximations of the data in $\mathcal{Y}_{\mathcal{H}_0}$ and $\mathcal{Y}_{\mathcal{H}_1}$, we applied a goodness-of-fit statistical test over each approximation. The statistical test adopted in the evaluation was the discrete Kolmogorov-Smirnov (KS) test [29]. The KS statistic value indicates the maximum distance between the empirical CDF of each distribution ($F_{\mathcal{Y}_{\mathcal{H}_0}}$ or $F_{\mathcal{Y}_{\mathcal{H}_1}}$, respectively) and the approximated one ($F_{\mathcal{Y}_{\mathcal{H}_0}^*}$ or $F_{\mathcal{Y}_{\mathcal{H}_1}^*}$, respectively), being represented by

$$KS = \sqrt{n} \max_{1 \leq x \leq M} \left| F_{\mathcal{Y}_{\mathcal{H}}}(x) - F_{\mathcal{Y}_{\mathcal{H}}^*}(x) \right|, \qquad (3)$$

where n is the set size and $M = \max_{1 \leq i \leq x} y_i$ is the highest value contained in the real set of sensing samples for the
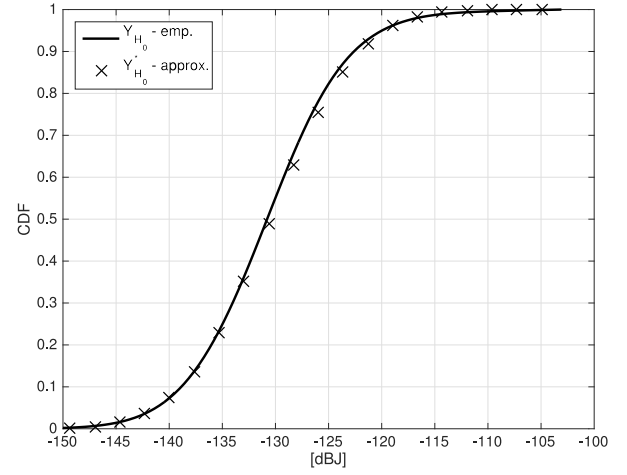
(a)
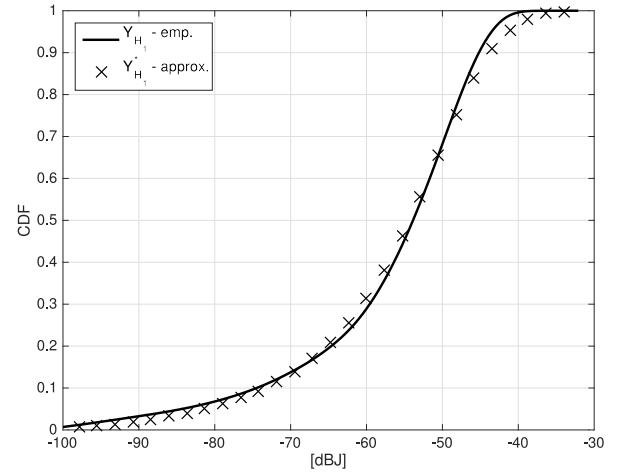
(b)

Fig. 2. Empirical and approximated CDFs of the set of sensing samples: a) $F_{\mathcal{Y}_{\mathcal{H}_0}}$ (*emp.* curve) and $F_{\mathcal{Y}_{\mathcal{H}_0}^*}$ (*approx.* curve); b) $F_{\mathcal{Y}_{\mathcal{H}_1}}$ (*emp.* curve) and $F_{\mathcal{Y}_{\mathcal{H}_1}^*}$ (*approx.* curve).

evaluated distribution. The KS statistic value is used has follows: if KS is higher than the critical value K_α we can reject the null hypothesis with a desired level of significance α; on the other hand, if KS is lower than K_α we can not reject the null hypothesis with the same level of significance. For $\mathcal{Y}_{\mathcal{H}_0}$ the null hypothesis can be expressed as

$$H_{null} : Y_{\mathcal{H}_0} \text{ follows a Normal distribution.}$$

The critical value K_α depends on the set size n, as well on the desired level of significance α, and it can be obtained from the Kolmogorov distribution. For both approximations we have assumed a level of significance $\alpha = 0.05$ and a set size of $n = 30$, resulting in a critical value of $K_\alpha = 0.2417$ [30]. Regarding the approximation $\mathcal{Y}_{\mathcal{H}_0}^*$ we have obtained a statistic value $KS_{\mathcal{Y}_{\mathcal{H}_0}} = 0.1512$, which is smaller than the critical value, meaning that our approximation can not be rejected at a 0.05 level of significance. Similarly, for $\mathcal{Y}_{\mathcal{H}_1}^*$ we have obtained $KS_{\mathcal{Y}_{\mathcal{H}_1}} = 0.2382$, which is again smaller than the critical value.

The KS test is inline with the results presented in Figure 2, where we can see that $\mathcal{Y}^*_{\mathcal{H}_0}$ better approximates $\mathcal{Y}_{\mathcal{H}_0}$ than $\mathcal{Y}^*_{\mathcal{H}_1}$ approximates $\mathcal{Y}_{\mathcal{H}_1}$.

Regarding the probability of miss-detection, it can be approximated by

$$P_m \approx Pr\left(\mathcal{Y}_{\mathcal{H}_1} \leq \gamma\right)$$
$$\approx F_{\mathcal{Y}^*_{\mathcal{H}_1}}(\gamma),$$

where $F_{\mathcal{Y}^*_{\mathcal{H}_1}}$ represents the CDF of the Shifted Log-logistic distribution, given by

$$F_{\mathcal{Y}^*_{\mathcal{H}_1}}(x) \approx \frac{1}{1 + \left(1 + \frac{\xi_{\mathcal{H}_1}\left(x - \mu_{\mathcal{H}_1}\right)}{\sigma_{\mathcal{H}_1}}\right)^{-\frac{1}{\xi_{\mathcal{H}_1}}}}. \tag{4}$$

From the fit tool, we have $\mu_{\mathcal{H}_1} = -48.75$, $\sigma_{\mathcal{H}_1} = 5.1$ and $\xi_{\mathcal{H}_1} = -0.17$, which are respectively the estimated location, scale and shape parameters of the Shifted Log-logistic. In this way, the probability of miss-detection is approximately $P_m \approx 0.0136$.

Regarding the probability of false alarm, it can be approximated by

$$P_f = Pr(\mathcal{Y}_{\mathcal{H}_0} > \gamma)$$
$$\approx 1 - F_{\mathcal{Y}^*_{\mathcal{H}_0}}(\gamma),$$

where $F_{\mathcal{Y}^*_{\mathcal{H}_0}}$ represents the CDF of the Normal distribution, given by

$$F_{\mathcal{Y}^*_{\mathcal{H}_0}}(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x - \mu_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_0}\sqrt{2}}\right)\right], \tag{5}$$

where $\mathrm{erf}(\cdot)$ denotes the error function. From the fit tool we obtained $\mu_{\mathcal{H}_0} = -130.48$ and $\sigma_{\mathcal{H}_0} = 6.55$, representing the location and scale parameters of the Normal distribution used to approximate $\mathcal{Y}_{\mathcal{H}_0}$. In this case, based on the estimated parameters and the selected energy threshold, the probability of false alarm is $P_f \approx 0$. While based in real data, the approximated values of the probabilities of miss-detection and false alarm are effectively close to 0, indicating a high accuracy of the spectrum sensing outcomes. As mentioned before, this fact is also demonstrated in the distribution represented in Figure 1, where we can observe two different energy regions reasonably separated from the γ threshold.

III. CHARACTERIZATION OF THE SPECTRUM OPPORTUNITIES

The GSM channel may be represented by a sequence of busy (PU ON) and idle (PU OFF) periods lasting $k \geq 1$ time units, as illustrated in Figure 3 (in the figure the first PU OFF period lasts $k = 3$ time units). Throughout the rest of the paper the term time unit will be used to denote the duration of the time period required to obtain a new decision provided by the energy detector. Since the energy detector requires $N_S = 2$ consecutive FFTs, and a new FFT is computed every 32 μs, a time unit lasts 64 μs. During the busy period the channel is used by the GSM system. During the idle period the channel
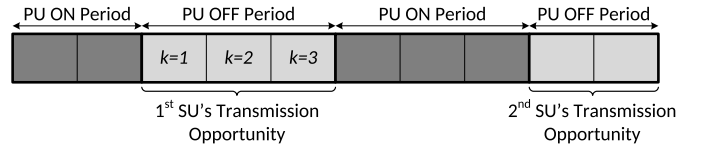

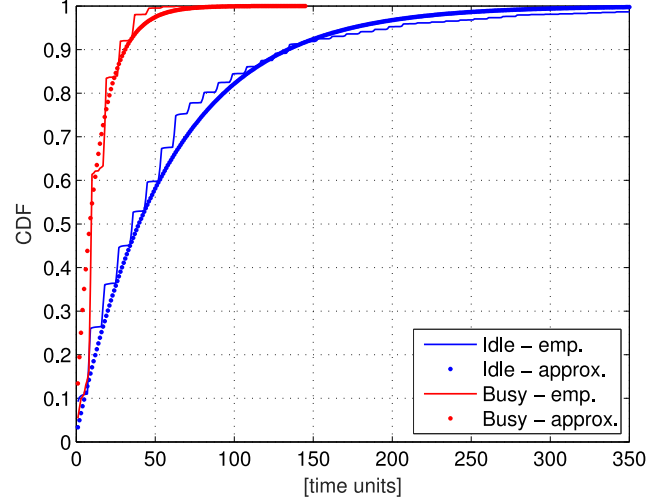Fig. 3. Hypothetical sequence of idle (PU OFF) and busy (PU ON) periods.

Fig. 4. Empirical and theoretical CDFs of the duration of the GSM busy and idle periods.

is available for opportunist access of the secondary users. The GSM channel is used by the GSM system (primary system) during μ_B time units (in average), and not used during μ_I time units.

The first point evaluated in the experimental data was the usage rate of the GSM channel. For each time unit k, the channel was classified as being busy or idle, depending on the energy detector's output. The statistics obtained with the experimental data indicated that the GSM channel was busy during 20% of the sensed time, $P^B_{PU} \approx 0.2$, and idle during approximately 80%, $P^I_{PU} \approx 0.8$. Further, we characterized the duration of each idle and busy period. Figure 4 illustrates the empirical CDFs of the duration of idle and busy periods (identified in the legend by "Idle - emp." and "Busy - emp.", respectively). The empirical CDFs were obtained with the experimental data. The idle periods can last longer than 350 time units, while the duration of the busy periods is almost limited to 100 time units. The average duration of the busy period is $\mu_B \approx 14.39$ time units, while the duration of the idle period is $\mu_I \approx 59.10$ time units. From the empirical CDFs we can observe a quantization-like effect, where the CDF value remains constant for approximately 8 time units. This is because the GSM channel maintains its state (idle or busy) during a GSM time slot, which lasts 577 μs.

Motivated by the fact that the duration of busy and idle periods are assumed to be geometrically distributed in several theoretical works (e.g., [24]), we evaluate the accuracy of such assumption. In this way, the average duration of each period obtained with the experimental data was used to define the single parameter of the geometric distribution, i.e., $p_I = 1/\mu_I$ and $p_B = 1/\mu_B$ for idle and busy periods, respectively. The

TABLE I
EXPERIMENTAL SETUP, SPECTRUM SENSING PARAMETERIZATION AND
GSM CHANNEL CHARACTERIZATION

ARFCN (GSM Band)	94 (900)
Downlink frequency	953.8 MHz
Observable bandwidth (FFT)	4 MHz
FFT size	128
Detector type	Energy detector
Sensing samples (N_S)	2
Decision threshold (γ)	-102.85 dBJ (51.9 pJ)
Prob. of false alarm (P_f)	0
Prob. of miss-detection (P_m)	0.0136
Time unit	64 μs
Idle duration (μ_I)	59.10 time units
Busy duration (μ_B)	14.39 time units
Idle probability (P_{PU}^I)	0.8
Busy probability (P_{PU}^B)	0.2

geometric CDFs are plotted in Figure 4 (identified by "Idle - approx." and "Busy - approx."). As can be observed, the geometric distributions parametrized with p_I and p_B are close to the distributions obtained with the experimental data.

Following the same rationale as in Section II-B, we applied the KS test to evaluate the assumptions performed in the previous paragraphs, *i.e.,* the duration of busy and idle periods can be approximated by geometric distributions. Thus, for the duration of the busy period, the null hypothesis can be expressed as

H_{null} : *the duration of busy periods over a GSM channel follows a geometric distribution.*

Assuming a level of significance $\alpha = 0.05$, the obtained critical value for a set size of $n = 6$ was $K_\alpha = 0.5193$, which is higher than the obtained KS statistic value of $KS_{busy} = 0.5135$. For the duration of the idle period, we have used a set size of $n = 24$ samples, resulting in a critical value of $K_\alpha = 0.2693$. The obtained KS value was $KS_{idle} = 0.2645$. Consequently, for GSM channels, the duration of idle and busy periods may be approximated by geometric distributions.

Table I summarizes the configuration of the experimental setup, the parameters adopted in the energy detector, and the values observed in the analysis of the experimental data.

IV. NUMERICAL COMPUTATION OF THE SERVICE TIME

The packet service time of a secondary user may be computed if the primary users' channel occupancy statistics are known. As observed in Section II-C, the duration of the busy and idle periods, represented by the random variables (RVs) B and I, respectively, are distributed according to geometric distributions with parameters $p_B = 1/\mu_B$ and $p_I = 1/\mu_I$, respectively. The probabilities of the GSM channel staying idle and busy are respectively given by $P_{PU}^I = \mu_I/(\mu_I + \mu_B)$ and $P_{PU}^B = \mu_B/(\mu_I + \mu_B)$.

To compute the service time we assume that a secondary user generates packets to a transmission queue. In what follows, it is always considered that the transmission queue always contains a packet to be transmitted, representing a traffic-saturated condition. The length of the data packet is represented by the RV L, and we assume that L follows a geometric distribution with parameter $p_L = 1/\mu_L$, meaning that

the average packet length is μ_L time units. Without loss of generality we assume that a data packet of length μ_L requires μ_L time units to be transmitted and, consequently, L may also represent the packet duration.

Considering that the idle and busy periods are geometrically distributed, from [24] we known that the distribution of the packet service time of the secondary user may be approximated by a discrete Generalized Pareto (dGP) distribution, with Probability Mass Function (PMF) given by

$$Pr\{X = x\} = \bar{F}(x) - \bar{F}(x+1), \quad (6)$$

and CDF expressed by

$$F(x) = 1 - \bar{F}(x), \quad (7)$$

where $\bar{F}$ is given by

$$\bar{F}(x) = \begin{cases} \left(1 + \dfrac{\xi(x-1)}{\sigma}\right)^{-1/\xi}, & \xi \neq 0 \\ e^{-\frac{x-1}{\sigma}}, & \xi = 0 \end{cases}, \quad (8)$$

and σ and ξ are respectively the scale and shape parameters. σ and ξ are the solutions of the following equation system [24]

$$\begin{cases} \sigma \approx \mathbb{E}[S](1-\xi) - 1 + \xi \\ \xi \approx \frac{1}{2} - \dfrac{(\mathbb{E}[S]-1)^2}{2\text{Var}[S]} \end{cases}, \quad (9)$$

where $\mathbb{E}[S]$ and $\text{Var}[S]$ represent the expectation and variance of the service time obtained through

$$\mathbb{E}[S] = j^{-1}\frac{d}{d\omega}\Phi_S(\omega)\Big|_{\omega=0}, \quad (10)$$

and

$$\text{Var}[S] = \mathbb{E}[S^2] - \big(\mathbb{E}[S]\big)^2, \quad (11)$$

respectively. $\mathbb{E}[S^2]$ represents the second moment of S and is given by $\mathbb{E}[S^2] = j^{-2}\frac{d^2}{d\omega^2}\Phi_S(\omega)|_{\omega=0}$. In (10) and (11) Φ_S represents the Characteristic Function (CF) of the service time when the secondary user always has a packet to transmit, which is given by [24]

$$\Phi_S(\omega) = (1 - p_I) \cdot \sum_{r=1}^{+\infty} Pr\{I_p = r\} \cdot \Phi_L(\omega) \cdot \Phi_B^{r-1}(\omega)$$
$$+ p_I \sum_{r=1}^{+\infty} Pr\{I_p = r\} \cdot \Phi_L(\omega) \cdot \Phi_B^{r}(\omega), \quad (12)$$

where Φ_L represents the CF of the packet length and Φ_B represents the CF of the duration of the busy period (both given by the CF of a geometric RV). In (12) the probability that a data packet is transmitted in exactly r PU idle periods, $Pr\{I_p = r\}$ is represented by [24]

$$Pr\{I_p = r\} = \sum_{t=0}^{+\infty} Pr\{I_r^{Sum} - L = t\}$$
$$- \sum_{t=0}^{+\infty} Pr\{I_{r-1}^{Sum} - L = t\},$$

where

$$Pr\left\{I_r^{Sum} - L = t\right\} = \sum_{k=1}^{+\infty} Pr\{L = k\} \cdot Pr\left\{I_r^{Sum} = k + t\right\}$$

and

$$Pr\left\{I_r^{Sum} = t\right\} = \begin{cases} \binom{t-1}{r-1} p_I^r (1 - p_I)^{t-r}, & t \geq r \\ 0, & t < r \end{cases}.$$

V. VALIDATION AND NUMERICAL RESULTS

This section presents a set of experimental and numerical results to validate the distribution of the packet service time for different data packet lengths transmitted by a secondary user over the GSM channel. For each packet length the Kolmogorov-Smirnov test is used to validate the proposed approximation. Furthermore, we present a set of experimental results showing that we can successfully estimate the service time using a reduced set of real-time samples of the service time.

The results were obtained considering a traffic saturated secondary network formed by a pair of SUs (sender and receiver) that use the GSM channel to transmit packets in an opportunistic way. In this way, the SU transmitter always has a packet to transmit. The SUs are able to identify transmission opportunities through spectrum sensing, as described in Section II-B.

A. Validation of Discrete Generalized Pareto Approximation

This subsection characterizes the packet service time distribution, evaluating the assumption of being distributed according to a discrete Generalized Pareto distribution, as considered in Section III. Figure 5 illustrates the empirical CDF of the packet service time for different data packet lengths, μ_L. Before discussing the comparison results, we point out that the CDFs ploted in the figure are described by discrete RVs. However, for easiness of comparison between the numerical and the experimental results, we have decided to plot the CDF of the numerical results using continuous curves. In Figure 5 the empirical CDF, denoted as $\hat{F}$, is obtained from the experimental data as follows

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \; y_i \leq x, \tag{13}$$

where y_i is the i-th service time value obtained from the experimental data and n is the size of the experimental service time set (that accumulates consecutive experimental service time values). For each scenario, the discrete Generalized Pareto CDF (numerically computed with (7)) is also plotted (indicated as "approx." in the figure). The results presented in the figure show that the empirical CDF obtained with the experimental data is close to the numerical CDF, for the considered data packet lengths. Hence, it can be foreseen that the packet service time over a GSM channel follows a dGP distribution.

B. Goodness of Fit Test for dGP Distribution

In the previous subsection we had visually evaluated the possibility of approximating the service time of the secondary
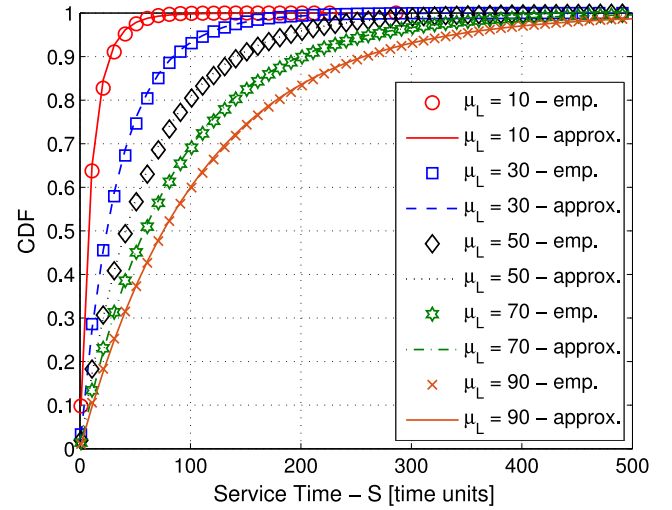

Fig. 5. Empirical CDF of the packet service time and discrete Generalized Pareto approximation for different data packet lengths (μ_L).

TABLE II
EMPIRICAL KS STATISTICS FOR THE DISCRETE GENERALIZED PARETO
DISTRIBUTION FOR DIFFERENT PACKET LENGTHS

μ_L	KS
10	0.1588
20	0.1506
30	0.1271
40	0.1048
50	0.1005
60	0.0913
70	0.0861
80	0.0769
90	0.0719
100	0.0713

network using a discrete Generalized Pareto distribution. Now we evaluate the same assumption using the discrete KS test, as used in Section II-B. In this case the null hypothesis can be expressed as

H_{null} : *the packet service time follows a dGP distribution,*

and, assuming a level of significance $\alpha = 0.05$ and a set size of $n = 70$, the obtained critical value was $K_\alpha = 0.1598$ [30].

Table II presents the KS statistic values for different packet lengths. As we can see, for all the packet lengths the *KS* value is smaller than the critical value, which means that our null hypothesis can not be rejected at a 0.05 level of significance. This means that the service time of a secondary network using a GSM channel can effectively be approximated by a discrete Generalized Pareto distribution.

C. Service Time Estimation

In this subsection we evaluate the possibility of estimating the service time based on real-time observations, *i.e.,* using a set of experimental service time values already experienced by the secondary user. For each set of samples, a SU computes the average and the variance of the packet service time, which are used to compute the scale and the shape parameter of the dGP distribution through (9).

Figure 6 presents the CDF of the packet service time obtained through simulation, as well as their estimates based
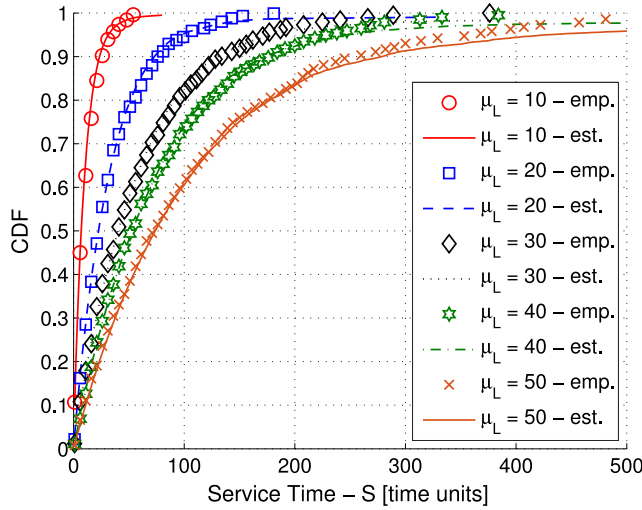
Fig. 6. Empirical CDF of the packet service time and discrete Generalized Pareto approximation based on real-time observations with 10 samples.

TABLE III
THEORETICAL AND ESTIMATED AVERAGE PACKET SERVICE TIME
COMPUTED WITH DIFFERENT SIZES OF REAL TIME ESTIMATIONS. THE
95% CONFIDENCE INTERVAL IS ALSO REPRESENTED

μ_L	set size	$\mathbb{E}[S]$ - model	$\mathbb{E}[S]$ - estimated
10	10	11.415	11.426 ± 0.218
	20		11.433 ± 0.158
	50		11.445 ± 0.110
20	10	22.010	22.101 ± 0.421
	20		22.152 ± 0.306
	50		22.133 ± 0.222
30	10	33.492	33.496 ± 0.745
	20		33.465 ± 0.528
	50		33.516 ± 0.351
40	10	43.074	42.945 ± 0.815
	20		42.861 ± 0.585
	50		42.983 ± 0.385
50	10	58.435	58.445 ± 1.173
	20		58.266 ± 0.830
	50		57.996 ± 0.469
60	10	73.229	73.170 ± 1.402
	20		73.089 ± 0.997
	50		72.528 ± 0.590
70	10	76.311	75.814 ± 1.525
	20		75.566 ± 1.042
	50		75.175 ± 0.632
80	10	91.773	91.925 ± 1.629
	20		92.113 ± 1.176
	50		92.697 ± 0.824
90	10	108.403	108.563 ± 2.223
	20		108.330 ± 1.540
	50		107.798 ± 0.877
100	10	113.992	113.564 ± 2.261
	20		113.481 ± 1.584
	50		113.555 ± 0.966

on a set of 10 samples. For each data packet length μ_L, the CDF indicated as "emp." represents the average of 1000 estimated CDFs based on real-time observations. Figure 6 shows that it is possible to approximate the service time CDF using a small set of 10 samples. We can also observe that, for larger values of data packet durations, 10 service time samples might not be enough to approximate the CDF of the packet service time with the desirable accuracy, which can be justified by the increase of the variance when higher data packet lengths are considered. To better demonstrate this effect we

have characterized the estimation process for different data packet durations and different number of samples used in the estimation process. Table III compares the packet service time estimated from the experimental data with the numerical results computed with (10) (the confidence interval is shown for 95% of confidence level). The different number of samples used in the estimation process are represented in the "set size" column. The results show that the average packet service time estimated from the experimental data closely matches with the average packet service time numerically computed. Moreover, the confidence interval decreases as more samples are used in the estimation process, as expected.

VI. CONCLUSION

In this work we have characterized the spectrum opportunities available in a GSM channel to support the operation of a cognitive radio network. Using a software defined radio device, we implemented the spectrum sensing procedure in order to characterize the periods of time when the channel is used and not used by the GSM radio system. As a result of the channel characterization, we have shown that the distribution of the duration of idle and busy channel periods can be approximated by geometric distributions. Using the experimental data obtained with the software defined radio setup, we validated a theoretical model for the distribution of the packet service time. The validation results, including the Kolmogorov-Sminorv statistic test, have shown that the packet service time can be successfully represented by a discrete Generalized Pareto distribution. Additionally, we have shown that the distribution of the service time can be estimated in real-time by only using a small set of the last service time values experienced by a secondary user. In this way the service time can be accurately estimated in a short period of time, highlighting the practical drive of our work.

REFERENCES

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] S. Hasan *et al.*, "GSM whitespaces: An opportunity for rural cellular service," in *Proc. IEEE Int. Symp. Dyn. Spectrum Access Netw. (DYSPAN)*, McLean, VA, USA, Apr. 2014, pp. 271–282.

[4] J. Gao, H. A. Suraweera, M. Shafi, and M. Faulkner, "Channel capacity of a cognitive radio network in GSM uplink band," in *Proc. Int. Symp. Commun. Inf. Technol.*, Sydney, NSW, Australia, Oct. 2007, pp. 1511–1515.

[5] A. Carniani, L. Giupponi, and R. Verdone, "Evaluation of spectrum opportunities in the GSM band," in *Proc. Eur. Wireless Conf. (EW)*, Lucca, Italy, Apr. 2010, pp. 948–954.

[6] A. H. Chowdhury, Y. Song, and C. Pang, "Accessing the hidden available spectrum in cognitive radio networks under GSM-based primary networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[7] A. Agarwal, A. S. Sengar, R. Gangopadhyay, and S. Debnath, "A real time measurement based spectrum occupancy investigation in northwestern India for cognitive radio applications," in *Proc. Int. Conf. Wireless Commun. Signal Process. Netw. (WiSPNET)*, Chennai, India, Mar. 2016, pp. 2035–2039.

[8] L. Mendes, L. Gonçalves, and A. Gameiro, "GSM downlink spectrum occupancy modeling," in *Proc. IEEE 22nd Int. Symp. Pers. Indoor Mobile Radio Commun.*, Toronto, ON, Canada, Sep. 2011, pp. 546–550.

[9] A. Palaios *et al.*, "Two days of European spectrum: Preliminary analysis of concurrent spectrum use in seven European sites in GSM and ISM bands," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 2666–2671.

[10] M. Wellens, J. Riihijarvi, and P. Mähönen, "Empirical time and frequency domain models of spectrum use," *Phys. Commun.*, vol. 2, nos. 1–2, pp. 10–32, Mar./Jun. 2009.

[11] K. Patil, S. Barge, K. Skouby, and R. Prasad, "Evaluation of spectrum usage for GSM band in indoor and outdoor scenario for dynamic spectrum access," in *Proc. Int. Conf. Adv. Comput. Commun. Informat. (ICACCI)*, Mysore, India, Aug. 2013, pp. 655–660.

[12] Y. A. Eldemerdash, O. A. Dobre, O. Üreten, and T. Yensen, "Fast and robust identification of GSM and LTE signals," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Turin, Italy, May 2017, pp. 1–6.

[13] M. Luís, R. Oliveira, R. Dinis, and L. Bernardo, "RF-spectrum opportunities for cognitive radio networks operating over GSM channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[14] Yuva Kumar S., M. S. Saitwal, M. Z. A. Khan, and U. B. Desai, "Cognitive GSM OpenBTS," in *Proc. IEEE 11th Int. Conf. Mobile Ad Hoc Sensor Syst.*, Philadelphia, PA, USA, Oct. 2014, pp. 529–530.

[15] S. Iliya, E. Goodyer, M. Gongora, J. Shell, and J. Gow, "Optimized artificial neural network using differential evolution for prediction of RF power in VHF/UHF TV and GSM 900 bands for cognitive radio networks," in *Proc. 14th U.K. Workshop Comput. Intell. (UKCI)*, Bradford, U.K., Sep. 2014, pp. 1–6.

[16] L.-C. Wang, C.-W. Wang, and C.-J. Chang, "Modeling and analysis for spectrum handoffs in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 9, pp. 1499–1513, Sep. 2012.

[17] Y. Zhang, "Dynamic spectrum access in cognitive radio wireless networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008, pp. 4927–4932.

[18] I. Suliman and J. Lehtomaki, "Queueing analysis of opportunistic access in cognitive radios," in *Proc. 2nd Int. Workshop Cogn. Radio Adv. Spectrum Manag. (CogART)*, Aalborg, Denmark, May 2009, pp. 153–157.

[19] C. Zhang, X. Wang, and J. Li, "Cooperative cognitive radio with priority queueing analysis," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dresden, Germany, Jun. 2009, pp. 1–5.

[20] H. Tran, T. Q. Duong, and H.-J. Zepernick, "Average waiting time of packets with different priorities in cognitive radio networks," in *Proc. 5th IEEE Int. Symp. Wireless Pervasive Comput. (ISWPC)*, Modena, Italy, May 2010, pp. 122–127.

[21] S. Gunawardena and W. Zhuang, "Service response time of elastic data traffic in cognitive radio networks with SPT service discipline," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, pp. 5602–5607.

[22] S. Gunawardena and W. Zhuang, "Service response time of elastic data traffic in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 559–570, Mar. 2013.

[23] M. Usman, H.-C. Yang, and M.-S. Alouini, "Service time analysis of secondary packet transmission with opportunistic channel access," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Vancouver, BC, Canada, Sep. 2014, pp. 1–5.

[24] M. Luís, R. Oliveira, R. Dinis, and L. Bernardo, "Characterization of the opportunistic service time in cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 2, no. 3, pp. 288–300, Sep. 2016.

[25] L.-C. Wang, C.-W. Wang, and C.-J. Chang, "Optimal target channel sequence design for multiple spectrum handoffs in cognitive radio networks," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2444–2455, Sep. 2012.

[26] M. Ettus. (2016). *Universal Software Radio Peripheral*. [Online]. Available: http://www.ettus.com/

[27] G. Radio. (2016). *Free Software Toolkit for SDR*. [Online]. Available: http://gnuradio.org/

[28] M. Luis, A. Furtado, R. Oliveira, R. Dinis, and L. Bernardo, "Energy sensing parameterization criteria for cognitive radios," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, Aug. 2012, pp. 61–65.

[29] N. Henze, "Empirical-distribution-function goodness-of-fit tests for discrete models," *Can. J. Stat.*, vol. 24, no. 1, pp. 81–93, 1996.

[30] M. A. Stephens, "EDF statistics for goodness of fit and some comparisons," *J. Amer. Stat. Assoc.*, vol. 69, no. 347, pp. 730–737, 1974.

Miguel Luís (S'11–M'17) received the M.Sc. and Ph.D. degrees in electrical and computer engineering from the Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal, in 2009 and 2015, respectively. In 2010, he joined the research center Instituto de Desenvolvimento de Novas Tecnologias, and later the Instituto de Telecomunicações (IT). He is currently a Post-Doctoral Researcher with IT involved in the research project "MobiWise: From Mobile Sensing to Mobility Advising." His main research interests include medium access control schemes for wireless systems, particularly cognitive radio networks, and routing and dissemination strategies for vehicular networks.
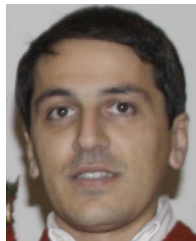
Rodolfo Oliveira (S'04–M'10–SM'15) received the Licenciatura degree in electrical engineering from the Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa (UNL), Lisbon, in 2000, the M.Sc. degree in electrical and computer engineering from the Instituto Superior Técnico, Technical University of Lisbon, in 2003, and the Ph.D. degree in electrical engineering from FCT, UNL, in 2009. From 2007 to 2008, he was a Visiting Researcher with the University of Thessaly and from 2011 to 2012, he was a Visiting Scholar with Carnegie Mellon University. He is currently an Assistant Professor with the Department of Electrical Engineering, FCT, UNL, and is also affiliated as a Researcher with the Instituto de Telecomunicações and UNINOVA, where he researches in the areas of wireless communications and networking.

Rui Dinis (S'96–M'00–SM'14) received the Ph.D. degree from the Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 2001, and the Habilitation degree in telecommunications from the Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa (UNL), in 2010. From 2001 to 2008, he was a Professor with IST. He is currently an Associated Professor with FCT-UNL. From 2003, he was an Invited Professor with Carleton University, Ottawa, Canada.

He was a Researcher with the Centro de Análise e Processamento de Sinal, IST, from 1992 to 2005, and the Instituto de Sistemas e Robótica, from 2005 to 2008. Since 2009, he has been a Researcher with Instituto de Telecomunicações. He has been actively involved in several national and international research projects in the broadband wireless communications area. His research interests include transmission, estimation, and detection techniques.

He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON COMMUNICATIONS (Transmission Systems—Frequency-Domain Processing and Equalization), and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was also a Guest Editor for *Physical Communication* (Elsevier) (Special Issue on Broadband Single-Carrier Transmission Techniques).

Luis Bernardo (M'03) received the Ph.D. degree from the Instituto Superior Técnico, Technical University of Lisbon, Portugal, in 2002, and the Habilitation degree in telecommunications from the Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa (UNL), in 2013.

Since 1999, he has been an Assistant Professor with the FCT, UNL. Since 2000, he has been a Researcher with the Instituto de Desenvolvimento de Novas Tecnologias, Departamento de Engenharia Electrotécnica, FCT, UNL. Since 2013, he has been a Researcher with the Instituto de Telecomunicações, Lisbon. His research interests include medium access control protocols for wireless communications, cross-layer optimization of wireless systems, routing protocols, and network modeling.