

Modeling and Validation of Channel Idleness and Spectrum Availability for Cognitive Networks

C. Ghosh, *Member, IEEE*, S. Roy, *Fellow, IEEE*, and M. B. Rao

Abstract—The potential of successful cognitive radio networks operating in TV White Spaces (and other future bands re-allocated for unlicensed operation) has led to significant upsurge of interest in their design optimization - particularly those that are cross-layer in nature, involving both MAC protocols as well as physical layer aspects such as channel sensing. Typically, these seek to optimize a network-level metric (notably, aggregate throughput) of secondary (cognitive) network subject to interference constraints on the primary. In turn, this requires suitable sensing by cognitive users to detect availability of primary channels (currently unused by the protected incumbents) for opportunistic usage.

To date, most studies have used largely *hypothetical* assumptions regarding channel idleness and resulting spectrum availability due to primary user dynamics. For example, idleness of channels over any spectrum are typically assumed to be an independent and identically distributed Bernoulli sequence. In contrast, nearly all real-time measurements suggest that channel idleness is *frequency dependent*, *i.e.*, the probability that a channel is idle depends on the channel location¹. Cognitive radio research thus increasingly needs more realistic and validated models for channel idleness as the foundation of credible cross-layer analysis; this is the primary contribution of our work.

We use *two sets of real-time measurements conducted in disparate geographic locations* over four distinct time intervals to show that channel idleness is appropriately modeled as independent but *non-identical* (i.n.i.d.) Bernoulli variables characterized by p_i , the probability of idleness for the i -th channel. We validate that Beta distribution can be used for modeling the variations in channel idleness probabilities; the Beta distribution parameters are estimated from the data to produce the best model fit. Based on the validated i.n.i.d. model, we build a predictive model by computing the *availability probability* of k channels, *i.e.*, $P\{N_{idle} = k\}$, where N_{idle} denotes the number of idle channels over the spectrum of N channels. However, the combinatorial complexity inherent in the computation of $P\{N_{idle} = k\}$ suggests the need for efficient approximations. We accomplish this by *classifying* idleness of channels based on the magnitude of p_i , and propose a novel *Poisson-normal* approximation for computing $P\{N_{idle} = k\}$. For validation, the distribution obtained from our technique is compared with the exact distribution and normal approximation using the approximation error criterion.

Index Terms—channel idleness, spectrum availability, Poisson-normal approximation, Beta distribution, KS test.

Manuscript received 5 January 2012; revised 15 May 2012. Parts of this work were presented at ACM CoRoNet Workshop, September 2010.

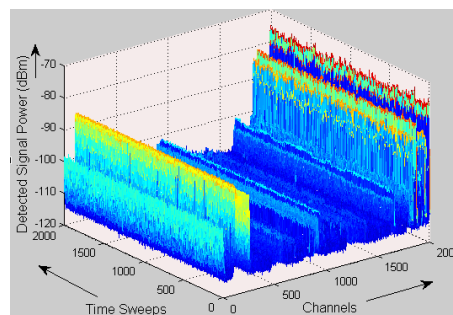
C. Ghosh is with the Nokia Research Center, Berkeley, CA (email: chittabrata.ghosh@nokia.com).

S. Roy is with the Electrical Engineering Department, University of Washington, Seattle, WA (email: sroy@u.washington.edu).

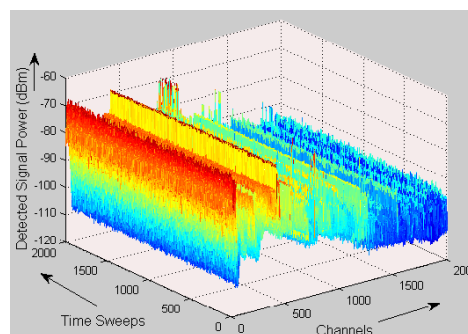
M. B. Rao is with the College of Medicine, University of Cincinnati, Cincinnati, OH (email: marepalli.rao@uc.edu).

Digital Object Identifier 10.1109/JSAC.2012.121117.

¹Such frequency dependence of idleness becomes more pronounced as the operating spectral bandwidth increases, relative to the resolution bandwidth for sensing.



Detected signal power from 8:00 pm to 9:00 pm over 2000 channels (1500-1900 MHz) in Maastricht, Netherlands



Detected signal power from 7:00 am to 8:00 am over 2000 channels (20-420 MHz) in Aachen, Germany

Fig. 1. Dynamic spectral occupancy in a spectrum of 400 MHz conducted at two disparate geographic locations over a stretch of an hour on a weekday

I. INTRODUCTION

THE INCREASING popularity of mobile devices (laptops, cellular phone, and smart phones) and demands for multimedia services via wireless access have resulted in exponentiating demands on present-day 3G networks. One approach (currently favored by regulatory authorities such as the Federal Communications Commission) is to expand the availability of new *unlicensed* spectrum - such as TV White Spaces made available by the transition to digital TV broadcasting - for use by *cognitive radio* networks [1], [2], where secondary users enabled by local sensing and spectrum mapping capabilities are able to detect and use channels that are temporarily idle due to licensed (primary) user inactivity, subject to non-interference with primary reception.

Clearly, allowing such opportunistic secondary usage leads to improved overall spectrum utilization (as measured by achieved aggregate throughput per unit Hz per unit area). However, optimizing the operation of such cognitive networks requires credible models for the occupancy patterns of primary users. To date, only a few efforts [3] - [6] have been reported

that capture the requisite temporal and frequency-dependent dynamics of primary user occupancy. This is the primary motivation for our work, anchored by available spectrum utilization measurements from experiments conducted [3] at Aachen, Germany and at Maastricht, Netherlands over 20-1500 MHz and 1500-3000 MHz spectra, respectively. Spectrum occupancy patterns in 1500–1900 MHz band at Maastricht between 8 - 9 pm and in 20-420 MHz band at Aachen between 7 - 8 am are shown in Figure 1.

It is already known and established using real-time measurements that spectrum availability possesses temporal and spatial variations. In absence of efficient planning on spectrum allocation and operation, network performance may be sacrificed significantly from reduced throughput and increased collisions to unlicensed secondary networks and inadmissible interference to licensed incumbents. Till date, no predictive models have been developed to facilitate spectrum scheduling at a centralized location. In this paper, we utilize publicly available and accredited data collected over several spectra in order to build a predictive model on spectrum availability. In the due process of model building, first, based on the detection threshold, we compute probability of idleness for all the channels under consideration. Then, using real time measurement results, we made an endeavor to validate that these channel idleness probabilities can be modeled by a Beta distribution with appropriate α and β values.

Most work to date within the academic cognitive radio community [7] - [10] assume hypothetical channel idleness models, notably as a set of Bernoulli, independent, and *identically distributed* (i.i.d.) variables for channel idleness probabilities (CIPs). The above measurements directly contradict this precept, and suggest that *independent but not identically distributed* (i.n.i.d.) Bernoulli random variables may be more apropos. Based on the above i.n.i.d. model for CIPs, we next compute *spectrum availability probability* (SAP), $P\{N_{idle} = k\}$, where N_{idle} represents the number of idle channels in a spectrum of N channels at any given instant of time. Our work is centered around 1st-order statistical characterization of N_{idle} . Simultaneously, we compute SAPs over multiple such spectral bands, since knowledge of these SAPs allows *spectrum selection*, i.e., choosing the “best” spectrum region (e.g., that with highest value of N_{idle} and in turn, highest probability of idleness)². However, traditional approaches for computing the probability distribution of N_{idle} has a combinatorial complexity with i.n.i.d. CIPs. Therefore, we propose a novel *Poisson-normal* approximation technique with reduced computational complexity while maintaining acceptable modeling fidelity.

In summary, the main contributions of the paper are as follows:

- Validate the i.n.i.d. model for CIPs over two sets of real-time measurements at two disparate geographic locations using McNemar’s and Pearson’s Chi-square tests;
- Validate the use of Beta distribution for non-identical CIPs by a) using data-driven estimates of its parameters α , β and b) using Kolmogorov-Smirnov (KS) and Chi-square goodness-of-fit tests over the same set of mea-

surements;

- Develop a Poisson-normal approximation technique based on the i.n.i.d. model for CIPs in order to compute probability distribution of N_{idle} and validate its accuracy and efficiency with reference to the exact distribution and a normal approximation.

This centralized approximation model, based on i.n.i.d. channel idleness probabilities, may potentially serve as an input to future cognitive medium access control protocol architectures for *adaptive spectrum selection* at any instant of time, where spectrum allocation can be planned based on the *spectrum availability* computations over several blocks of non-contiguous spectra. Moreover, *adaptive spectrum selection* can lead to better network performance in terms of aggregate throughput. However, the major focus of this paper is limited to developing models for channel idleness and spectrum availability and validation of network performance based on aggregated throughput and latency based on this predictive model will be left for future work.

The rest of the paper is organized as follows. Section II discusses the related work on spectrum measurements and modeling of primary users’ occupancy patterns for better appreciation of modeling channel idleness. Section III illustrates the i.n.i.d. model for the CIPs. In Section IV, we present the novel Poisson-normal approximation method for computation of the distribution of N_{idle} , a measure of spectrum availability and its impact on network performance. Section V discusses the process of validating the channel availability model using statistical testing methods. Further, we evaluate the accuracy and efficiency of our proposed Poisson-normal approximation technique with respect to the conventional normal approximation and exact distribution. Finally, Section VI draws the conclusion of our research.

II. RELATED WORK

Study of spectrum occupancy begins with spectrum measurement data as input, leading to models for channel idleness and spectrum availability as output. McHenry *et al.* [11] - [12] monitored spectrum occupancy for different channels at multiple geographic locations. They deployed a high dynamic range spectrum measurement system for spectrum monitoring ranging from hours to days. Sanders *et al.* [13] used the Radio Spectrum Measurement System to collect observations periodically on channels in the 108 MHz- 10 GHz range providing a vast trove of occupancy data. Roberson *et al.* [14] used passive monitoring over 30 MHz - 3 GHz in order to categorize the degree of utilization of the channels into three different classes: (i) channels seldom used, (ii) channels used during specific intervals of time, and (iii) channels heavily used.

In our previous work [6], we have proposed the first validated probabilistic model for channel occupancy using real-time measurements in the paging band (928-948 MHz). There, we captured the characteristics of primary user occupancies based on variations in transmission powers, selection of center frequencies, and time durations of operations. This leads to estimates of durations for channel idleness and occupancy; however, overall spectrum availability was not explored, which

²This is a building block for any future *adaptive sensing strategy*.

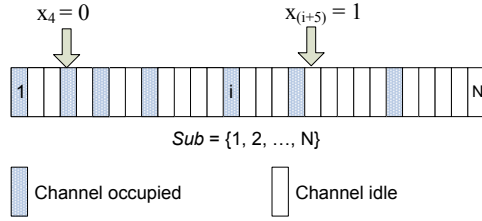


Fig. 2. Spectrum occupancy of N channels by primary users at time t

is undertaken in the present work. Our channel idleness model is inspired by [5] who first proposed the use of Beta distribution to match channel occupancy seen in real-time measurements. Initial model validation was pursued in [4]; in this paper, we have conducted more extensive validations using the two measurement data sets described above.

III. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we first describe a probabilistic model for channel idleness. The spectrum of interest is divided into N non-overlapping primary channels, denoted by $Sub = \{1, 2, \dots, N\}$. We note that typical real-time measurements use a resolution bandwidth (for channel sensing) that is much smaller compared to that of typical primary services. This resolution bandwidth is limited by the available hardware for the measurement set-up; in our case, 30 of 200 KHz (corresponding to the resolution bandwidth) sub-channels span one 6 MHz primary channel. The primary occupancy patterns exhibit temporal dependencies, *i.e.*, $A_t \subset Sub$, the subset of occupied channels at time t , varies as a function of time index t , as shown in Figure 2.

Let $\{x_1, x_2, \dots, x_N\}$ represent Bernoulli variables of (primary) channel status, whereby $x_i = 0$ if i^{th} channel is occupied by a primary user and $x_i = 1$ if it is idle, $i = 1, 2, \dots, N$. The CIP, *i.e.*, $p_i = Pr(x_i = 1)$ can be estimated empirically ([11] - [14]) by observing status of the i^{th} channel at the same time instant, averaged over several days.

Both the measurement data support the following observations:

- The CIP can be small, moderate, or large depending upon the time of the day and its usage; and
- The random variables, x_i 's, are statistically independent, but unlike the assumption in majority of existing research, are not identically distributed.

Therefore, x_i 's are modeled as independent, non-identical Bernoulli random variables, *i.e.*, $Pr(x_i = 1)$ and $Pr(x_j = 1)$ are different.

The Beta distribution is suggested in this paper in order to model non-identically distributed CIPs over any operating spectrum. This is reasonable in view of the fact that a Beta distribution has the capability to approximate a wide range of probability distributions on $(0, 1)$ [15]. Additionally, it has been observed from measurements that (i) majority of TV broadcast channels being *on-air* leads to very low idleness of such channels, (ii) channels allocated for space navigation, telemetry, and governmental activities are rarely used (high

probability of idleness), and (iii) channels designated for FM radio, paging, and text messaging have low-to-moderate idleness.

The Beta probability density function, with parameters $\alpha, \beta > 0$ that largely determine the distribution shape, is expressed as:

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, 0 < x < 1, \quad (1)$$

where

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \quad (2)$$

is a normalization constant.

To determine *spectrum availability*, we now define a (discrete) random variable N_{idle} representing the total number of idle channels in the N -set. Clearly, $N_{idle} = \sum_{i=1}^N x_i$. Efficient computation of the p.m.f of N_{idle} in an N -set swath is the major contribution of our research and will be discussed in the next section. Table I lists all the variables used in this work.

IV. COMPUTATION OF $Pr(N_{idle} = k)$

In this section, we introduce the method of computing spectrum availability using exact distribution and other approximation methods. Later on, we briefly discuss about the interrelation between spectrum availability and overall network performance. We initiate this discussion with the following Lemma that summarizes the combinatorial complexity in computation of exact distribution of spectrum availability.

Lemma 1: When p_i 's are not equal, the complexity involved in computing $Pr(N_{idle} = k)$ in an N -set is given by:

$$\left[\binom{N}{k} \times N \right] - 1, k = 0, 1, 2, \dots, N. \quad (3)$$

Proof: Let A represent a generic symbol for any subset of Sub representing idle channels and A' its complement. Then,

$$Pr(N_{idle} = k) = \sum_A \left(\prod_{i \in A} p_i \right) \left(\prod_{j \in A'} (1 - p_j) \right), \quad (4)$$

where the summation is taken over all subsets $A \subset Sub$ with cardinality k . Each expression inside the summation symbol of Eq. (4) is a product of N terms and the summation involves a total of $\binom{N}{k}$ summands. Therefore, the computational complexity of Eq. (4) is given by $\left(\binom{N}{k} \times N \right) - 1$. \square

The calculation of the probability of $N_{idle} = k$ is thus extremely high per Lemma 1. For example, the probability of exactly 15 idle channels, *i.e.*, $Pr(N_{idle} = 15)$, when $N = 30$, incurs $\binom{30}{15} \approx 155$ million possible scenarios, where in each scenario 30 numbers need to be multiplied. The complexity in computing the full exact distribution of N_{idle} is given by

$$\sum_{k=0}^N \left[\left(\binom{N}{k} \times N \right) - 1 \right] = 2^N \times N - (N + 1). \quad (5)$$

This exhaustive amount of computation is a potential memory constraint on a mobile device when fast, real-time spectrum mapping is desired. Hence, an efficient technique is indispensable for real-time calculation of the distribution of N_{idle} . We

TABLE I
NOTATION

N	Number of channels in operating spectrum	Sub	Set of channels
x_i	Binary random variable indicating status of i^{th} channel	p_i	i^{th} CIP
A_t	Subset of channels occupied at time instant t	A	Subset of idle channels in Sub
N_{idle}	Total number of idle channels	$Pr(N_{idle} = k)$	Exact probability of having $N_{idle} = k$
$\epsilon(k)$	Approximation error	ϵ	Overall approximation error
X	Any discrete random variable	\mathcal{S}_X	Essential support of X
θ	Decision threshold of \mathcal{S}_X	$PrNormal(N_{idle} = k)$	Normal approximation of $N_{idle} = k$
\bar{N}_{idle}	Mean of the normal approximation	C_N	Variance of the normal approximation
P_{th1}	Threshold for selecting small p_i 's	P_{th2}	Threshold for selecting large p_i 's
Sub_{small}	Set of channels with $0 < p_i \leq P_{th1}$	m	Size of Sub_{small}
$N_{idle_{small}}$	Number of idle channels in Sub_{small}	Sub_{mod}	Set of channels with $P_{th1} < p_i < P_{th2}$
n	Size of Sub_{mod}	$N_{idle_{mod}}$	Number of idle channels in Sub_{mod}
Sub_{large}	Set of channels with $p_i \geq P_{th2}$	$N_{idle_{large}}$	Number of idle channels in Sub_{large}
Z	Random variable with Poisson distribution	λ_s	Mean value of $N_{idle_{small}}$
X_i	i^{th} Bernoulli random variable with p_i	S	Sum of all X_i
$Pr_{Poi-Normal}(N_{idle} = k)$	Poisson-normal approximated probability of $N_{idle} = k$	λ_l	Mean value of $N_{idle_{large}}$
$Pr_{Poi}(N_{idle_{small}} = k)$	Probability of $N_{idle_{small}} = k$ using Poisson approximation		

propose a suitable approximation technique and determine the resulting approximation error.

Definition 1 (Approximation error $\epsilon(k)$): For any approximation, $Pr_{app}(N_{idle} = k)$, the corresponding pointwise error $\epsilon(k)$ is given by

$$\epsilon(k) = |Pr(N_{idle} = k) - Pr_{app}(N_{idle} = k)|. \quad (6)$$

We define $\epsilon = \max_{0 \leq k \leq N} \epsilon(k)$ - the *over-all approximation error* - as the figure of merit of the approximation method.

Definition 2 (Essential support: \mathcal{S}_X): For any non-negative, discrete variable X , the essential support \mathcal{S}_X is defined to be the set $\mathcal{S}_X = \{k; Pr(X = k) \geq \theta\}$ for small values of θ .

With this provision, we re-define the *effective over-all approximation error* to be

$$\epsilon = \max_{k \in \mathcal{S}_{N_{idle}}} \epsilon(k). \quad (7)$$

A. Approximation with Normal Distribution

The Central Limit Theorem [15] provides a well-known baseline approach for approximating the limiting distribution of a sequence of variables. Let $\bar{N}_{idle} = \sum_{i=1}^N p_i$ be the mean and $C_N = Var(N_{idle}) = \sum_{i=1}^N p_i(1-p_i)$ be the variance of N_{idle} . For N and C_N sufficiently large, N_{idle} can be approximated by $Normal(\bar{N}_{idle}, C_N)$. Hence the p.m.f. of the discrete variable N_{idle} may be approximated by

$$\begin{aligned} Pr(N_{idle} = k) &\approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{\sqrt{2\pi C_N}} e^{-\left(\frac{x-\bar{N}_{idle}}{\sqrt{2C_N}}\right)^2} dx \\ &= Pr_{Normal}(N_{idle} = k), \end{aligned} \quad (8)$$

where $k = 0, 1, \dots, N$.

The experimental results shown later in Section V reveal that the normal approximation is not acceptable in terms of accuracy, although it is computationally efficient. To improve the accuracy, we propose a novel method by merging the Law of Rare Events, Very High Frequency Events, and Moderate Events, as described next.

B. Poisson-Normal Approximation

As noted, the idleness probabilities p_i can be classified as small, moderate, or large, by defining two thresholds: a lower threshold P_{th1} and an upper threshold P_{th2} ($0 < P_{th1} < P_{th2} < 1$). All p_i 's within the range $0 < p_i \leq P_{th1}$ relate to channels with small CIPs. For p_i 's $\geq P_{th2}$, channels are classified into a group with large CIPs. Otherwise, channels within the range $P_{th1} < p_i < P_{th2}$ are categorized into a group with moderate CIPs. With this, we have the following definitions.

Definition 3 (Sub_{small}): This is a set of all channels having $0 < p_i \leq P_{th1}$. Let m be the size of Sub_{small} and $N_{idle_{small}}$ be the number of idle channels in Sub_{small} .

Definition 4 (Sub_{mod}): This is a set of all channels having $P_{th1} < p_i < P_{th2}$. Let n be the size of Sub_{mod} and $N_{idle_{mod}}$ be the number of idle channels in Sub_{mod} .

Definition 5 (Sub_{large}): This is a set of all channels having $p_i \geq P_{th2}$. The size of Sub_{large} is $(N - m - n)$. Let $N_{idle_{large}}$ be the number of idle channels in Sub_{large} .

Note that P_{th1} is typically close to zero and P_{th2} is close to one. In order to find the distribution of N_{idle} , we first compute the approximate distributions of $N_{idle_{small}}$, $N_{idle_{mod}}$, and $N_{idle_{large}}$. The distribution of N_{idle} is computed by using the relation

$$N_{idle} = N_{idle_{small}} + N_{idle_{mod}} + N_{idle_{large}}. \quad (9)$$

As observed from the real-time measurements, there are an appreciable number $N_{idle_{small}}$ of channels with small CIPs (and so the expectation $N_{idle_{small}} \sum_{i \in Sub_{small}} p_i$ is of intermediate magnitude). Then, the distribution of $N_{idle_{small}}$ channels may be approximated by the Poisson distribution [16], also called *law of rare events*. Similarly, there are an appreciable number $N_{idle_{large}}$ of channels with large CIPs. For such channels, $(1 - p_i)$ is very small for the i^{th} channel. This is true for all $N_{idle_{large}}$ channels. Therefore, distribution of $N_{idle_{large}}$ channels can also be approximated by Poisson distribution. Finally, with moderate probabilities, the distribution of $N_{idle_{mod}}$ in Sub_{mod} can be approximated by a normal distribution. Hence, with N channels comprising of $N_{idle_{small}}$, $N_{idle_{mod}}$, and $N_{idle_{large}}$, we combine Poisson and normal approximations in order to compute spectrum availability. We term this approximation

method as Poisson-normal approximation and the details of this method is illustrated below.

1) *Approximate Distribution of $N_{idle_{small}}$* : The distribution of $N_{idle_{small}}$ can be approximated by a Poisson distribution and the probability that there are k idle channels is

$$Pr(N_{idle_{small}} = k) \approx \frac{\lambda_s^k e^{-\lambda_s}}{k!} = Pr_{Poi}(N_{idle_{small}} = k), \quad (10)$$

where $\lambda_s = \sum_{i \in Sub_{small}} p_i$. This approximation follows the so-called Law of Rare Events. The following Lemma gives an upper bound on the resulting approximation error.

Lemma 2: Let Z be a random variable that has a Poisson distribution with parameter λ_s . We have

$$|Pr(N_{idle_{small}} = k) - Pr(Z = k)| \leq \frac{1 - e^{-\lambda_s}}{\lambda_s} \times \sum_{i \in Sub_{small}} p_i^2, \quad (11)$$

where $k=0, 1, 2, \dots$

Proof: This follows the Law of Rare Events [16]. Suppose that X_1, \dots, X_s are independent Bernoulli random variables with success probabilities p_1, \dots, p_s . If all the p_i 's are sufficiently small, then the Law of Rare Events asserts that the sum $S = \sum_1^s X_i$ is approximately Poisson distributed. More precisely, if Z is a Poisson random variable with mean λ , then

$$|Pr(S = k) - Pr(Z = k)| \leq \frac{1 - e^{-\lambda}}{\lambda} \times \sum_{i=1}^s p_i^2, \quad k = 0, 1, \dots \quad (12)$$

Here, $S = N_{idle_{small}} = \sum_{i \in Sub_{small}} x_i$ and $\lambda = \lambda_s$ and (11) follows directly from (12). \square

From the upper bound, if $N = 10$, each $p_i \leq 0.03$, then $\sum_{i=1}^{10} p_i^2 \leq 0.009$. Then, the exact probability $Pr(N_{idle_{small}} = k)$ and the approximate Poisson probability $Pr(Z = k)$ agree in the first two decimal places, *i.e.*, the approximation error is less than 0.001. Therefore, in our analysis we have taken $P_{th1} = 0.03$ as the threshold value for small CIPs.

2) *Approximate Distribution of $N_{idle_{mod}}$* : The distribution of $N_{idle_{mod}}$ in Sub_{mod} can be approximated by a normal distribution and the probability that there are k idle channels is computed as follows:

$$\begin{aligned} Pr(N_{idle_{mod}} = k) &\approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{\sqrt{2\pi C_n}} e^{-\left(\frac{x-\bar{N}_{mod}}{\sqrt{2C_n}}\right)^2} dx \\ &= Pr_{Normal}(N_{idle_{mod}} = k), \end{aligned} \quad (13)$$

where n is the size of Sub_{mod} , $k = 0, 1, \dots, n$, $\bar{N}_{mod} = E[N_{idle_{mod}}] = \sum_{i \in Sub_{mod}} p_i$, and $C_n = \sum_{i \in Sub_{mod}} p_i(1 - p_i)$ represents the variance of $N_{idle_{mod}}$.

3) *Approximate Distribution of $N_{idle_{large}}$* : The approximation of the distribution for $N_{idle_{large}}$ essentially follows the method for $N_{idle_{small}}$. Since $(1 - p_i)$ is small for $i \in Sub_{large}$, using the Law of Rare Events, the distribution of $N_{idle_{large}}$ can also be approximated by a Poisson distribution and the probability that there are k idle channels is expressed as:

$$Pr(N_{idle_{large}} = k) \approx \frac{e^{-\lambda_l} \lambda_l^{(N-m-n-k)}}{(N-m-n-k)!} = Pr_{Poi}(N_{idle_{large}} = k), \quad (14)$$

where $\lambda_l = \sum_{i \in Sub_{large}} (1 - p_i)$, $k = 0, 1, \dots, (N - m - n)$.

Now, we illustrate the Poisson-normal approximation procedure in order to compute the distribution of N_{idle} using $N_{idle_{small}}$, $N_{idle_{large}}$, and $N_{idle_{mod}}$. To achieve computational efficiency, only the essential supports, $\mathcal{S}_{N_{idle_{small}}}$ and $\mathcal{S}_{N_{idle_{large}}}$ of $N_{idle_{small}}$ and $N_{idle_{large}}$, respectively, are used in the computations. From Eq. (9), we have $N_{idle} = N_{idle_{small}} + N_{idle_{mod}} + N_{idle_{large}}$. As an illustration, we consider two cases: $N_{idle} = 0$ and $N_{idle} = 1$. Since $N_{idle} = 0$ implies no idle channels over the entire band, by Eqs. (9), (10), (13), and (14):

$$\begin{aligned} Pr(N_{idle} = 0) &= Pr(N_{idle_{small}} = 0, N_{idle_{large}} = 0, N_{idle_{mod}} = 0) \\ &\approx e^{-\lambda_s} \frac{e^{-\lambda_l} \lambda_l^{(N-m-n)}}{(N-m-n)!} Pr\left(-\frac{1}{2} < N_{idle_{mod}} < \frac{1}{2}\right) \\ &\approx e^{-\lambda_s} e^{-\lambda_l} \frac{(\lambda_l)^{(N-m-n)}}{(N-m-n)!} \int_{-\frac{1}{2}}^{\frac{1}{2}} N(\bar{N}_{mod}, C_n) dx. \end{aligned}$$

Similarly, $N_{idle} = 1$ means that there is only one idle channel in the N -set. This idle channel can be in one of Sub_{small} , Sub_{mod} , or Sub_{large} . By Eqs. (9), (10), (13), and (14):

$$\begin{aligned} Pr(N_{idle} = 1) &\approx (e^{-\lambda_s} \lambda_s) \left(\frac{e^{-\lambda_l} \lambda_l^{(N-m-n)}}{(N-m-n)!} \right) \int_{-\frac{1}{2}}^{\frac{1}{2}} N(\bar{N}_{mod}, C_n) dx + \\ &\quad (e^{-\lambda_s}) \left(\frac{e^{-\lambda_l} \lambda_l^{(N-m-n-1)}}{(N-m-n-1)!} \right) \int_{-\frac{1}{2}}^{\frac{1}{2}} N(\bar{N}_{mod}, C_n) dx + \\ &\quad (e^{-\lambda_s}) \left(\frac{e^{-\lambda_l} \lambda_l^{(N-m-n)}}{(N-m-n)!} \right) \int_{\frac{1}{2}}^{\frac{3}{2}} N(\bar{N}_{mod}, C_n) dx. \end{aligned}$$

Following the *modus operandi* in the illustration presented above, we now present a general expression for the probability of k idle channels, where k_1 of these idle channels are in Sub_{small} , k_2 in Sub_{mod} , and k_3 in Sub_{large} such that $k_1 + k_2 + k_3 = k$.

$$\begin{aligned} &= \sum Pr(N_{idle_{small}} = k_1, N_{idle_{mod}} = k_2, N_{idle_{large}} = k_3) \\ &= \sum Pr(N_{idle_{small}} = k_1) Pr(N_{idle_{mod}} = k_2) Pr(N_{idle_{large}} = k_3) \\ &\approx Pr_{Poi}(N_{idle_{small}} = k_1) Pr_{Normal}(N_{idle_{mod}} = k_2) Pr_{Poi}(N_{idle_{large}} = k_3) \\ &= Pr_{Poi-Normal}(N_{idle} = k), \end{aligned} \quad (15)$$

where the summation is taken over all $k_1 \geq 0$, $k_2 \geq 0$, and $k_3 \geq 0$ with $k_1 + k_2 + k_3 = k$. By focusing on $k_1 \in \mathcal{S}_{N_{idle_{small}}}$ and $k_3 \in \mathcal{S}_{N_{idle_{large}}}$ in Eq. (15), the amount of calculations can be reduced considerably. In any practical situation, the sets $\mathcal{S}_{N_{idle_{small}}}$ and $\mathcal{S}_{N_{idle_{large}}}$ are determined first, and then Eq. (15) is implemented in obtaining an approximate distribution of N_{idle} .

C. Network Performance in terms of Spectrum Availability

In this sub-section, we evaluate the various cognitive radio-based sensing parameters like probability of detection, P_d and probability of false alarm, P_{fa} [17].

1) *Detection Probabilities*: Detection probability is defined as the probability of detecting at least one idle channel during the scanning period T_s , provided an idle channel is truly available. Assuming a constant hold time τ , let us define two events as follows:

E_1 : The secondary user detects an idle channel during hold time τ ;

E_2 : There is an idle channel available during the sensing period T_s .

Then, for constant τ during the sensing period, P_d is simplified as $P_d = \frac{E_1}{E_2}$, since $E_1 \subset E_2$. Now, occurrence of event E_1 is expressed as $P(E_1) = Pr_{Poi-Normal}[N_{idle}(\delta t + \tau) \geq 1]$ and similarly, probability of E_2 is deduced as $P(E_2) = \prod_{i=1}^N CIP_i(T_s)$ [17] considering N channels. Therefore, from Eq. 15 and replacing expressions for E_1 and E_2 , P_d for constant hold time τ and sensing threshold ϵ is expressed as:

$$P_d(\tau, \epsilon) = \frac{\epsilon C_{Poi-Normal}(\tau)}{\prod_{i=1}^N CIP_i(T_s)}, \quad (16)$$

where $C_{Poi-Normal}(\tau)$ is the cumulative distribution function (CDF) of the proposed Poisson-normal distribution.

For a target P_d expressed in Eq. 16, P_{fa} is expressed [18] in terms of P_d as follow:

$$P_{fa}(\epsilon, \tau) = Q\left(\sqrt{2\gamma + 1}Q^{-1}(P_d(\epsilon, \tau)) + \sqrt{\tau f_s \gamma}\right), \quad (17)$$

where γ is the signal-to-noise ratio at the secondary receiver and f_s is the sampling frequency for sensing N channels under consideration.

2) *Aggregate Throughput*: For a spectrum, let us define $P(H_0)$ as the probability when an available channel is detected and $P(H_1)$ as the probability for which an idle channel is not detected. Therefore, we have $P(H_0) + P(H_1) = 1$. The secondary users can operate in such a spectrum under two scenarios:

- S_1 : An idle channel is detected and there is no false alarm in sensing; throughput achieved on such a channel is $\frac{(T-T_s)}{T}C_0$, where T is time slot reserved for a secondary user and C_0 is the data rate obtained in such a scenario.
- S_2 : An idle channel is mis-detected although the channel is truly not available; throughput on such a channel is expressed as $\frac{(T-T_s)}{T}C_1$.

The probabilities [18] for scenarios S_1 and S_2 are $(1 - P_{fa}(\tau, \epsilon))P(H_0)$ and $(1 - P_d(\tau, \epsilon))P(H_1)$, respectively. Defining achieved throughput for S_1 and S_2 as R_0 and R_1 , respectively, they are expressed as:

$$\begin{aligned} R_0(\epsilon, \tau) &= \frac{T - T_s}{T} C_0 (1 - P_{fa}(\epsilon, \tau)) P(H_0), \\ R_1(\epsilon, \tau) &= \frac{T - T_s}{T} C_1 (1 - P_d(\epsilon, \tau)) P(H_1). \end{aligned} \quad (18)$$

The average throughput of the secondary network is then expressed as:

$$\begin{aligned} R(\tau) &= \frac{T - T_s}{T} C_0 (1 - P_{fa}(\epsilon, \tau)) P(H_0) + \\ &\quad \frac{T - T_s}{T} C_1 (1 - P_d(\epsilon, \tau)) P(H_1). \end{aligned} \quad (19)$$

Replacing expressions for P_d and P_{fa} from Eqs. 16 and 17, system throughput can be obtained in terms of the spectrum availability and CIPs.

V. VALIDATION OF CHANNEL IDLENESS MODEL AND COMPARATIVE EVALUATION

In this section, we first validate the model for channel idleness discussed in Section III. Then, we provide a comparative study of the exact distribution, normal, and Poisson-normal approximations for *spectrum availability* using results obtained from the experimental measurements. For validation purposes,

we have considered real-time measurements conducted on two different bands:

(i) *Measurement Set 1*: 20-1500 MHz band with center frequency 770 MHz inside a modern office building (Latitude: 50°47'24.01" North and Longitude: 6°3'47.42" East) at Aachen [3] and

(ii) *Measurement Set 2*: 1500-3000 MHz band [3] with center frequency 2250 MHz on a rooftop location ((Latitude: 50°50'23.34" North and Longitude: 5°43'14.93" East)) in a residential area at Maastricht. These measurements were conducted over 8192 channels during December 2006 and January 2007 over a simultaneous period of seven days. The resolution bandwidth for each channel is fixed at 200 KHz. Average sweep time is considered to be 1.8 s. The measurement set-up is discussed in [3].

A. Model Validation

1) *Spectrum Idleness Model Validation*: In this section, we validate the i.n.i.d. observation for CIPs using the Pearson's chi-square statistic and McNemar's test statistic.

Independence Validation:

Our intention is to validate that the CIPs of two adjacent channels are statistically independent. Pearson's chi-square statistic, χ_{Ind}^2 [19], is used to assess independence. We compute χ_{Ind}^2 from the experimental results of *Measurement Set 1* for random two adjacent channels 2500 (equivalently 520 MHz) and 2501 (520.2 MHz) during the 7-8 am interval using the 2×2 contingency table for them provided in Table II. Due to space limitations, we have used same Table II for *Measurement Set 2* in order to report frequencies of two adjacent channels 1700 (1840 MHz) and 1701 (1840.2 MHz) during the morning (4-5 am) time interval.

Each observation $O_{i,j}$ consists of the values of two outcomes in the (i, j) cell and the null hypothesis is that the occurrence of these outcomes is statistically independent. Each observation is allocated to one cell of the contingency table, according to the values of the two outcomes. For instance, out of 2000 observations, $O_{i,j} = 68$ (180) when i represents channel 2500 (1700) being busy and j represents channel 2501 (1701) being simultaneously busy. The theoretical frequency $E_{i,j}$ for a cell, given the hypothesis of independence, is expressed as:

$$E_{i,j} = \frac{\sum_{k=1}^c O_{i,k} \sum_{k=1}^r O_{k,j}}{N}, \quad (20)$$

where r is the number of rows, c is the number of columns in the table, and degrees of freedom is $(r - 1)(c - 1)$.

The value of the test statistic is given as:

$$\chi_{Ind}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}. \quad (21)$$

For the test of independence, a p-value of less than or equal to 0.05 is commonly interpreted as justification for rejecting the null hypothesis.

Based on the results presented in Table II obtained from *Measurement Set 1*, the value of χ_{Ind}^2 from Eq. 21 is 1.8125. The chi-squared statistic χ_{Ind}^2 can then be used to calculate a p-value by comparing its value to a chi-squared distribution with specific D_f . The corresponding p-value is 0.1782 for

TABLE II
CONTINGENCY TABLE FOR CHANNELS 2500(1700) AND 2501(1701)

	Observations		
	Ch. 2501 (1701) Busy	Ch. 2501 (1701) Idle	Row total
Ch. 2500 (1700) Busy	68 (180)	994 (852)	1062 (1032)
Ch. 2500 (1700) Idle	46 (167)	892 (801)	938 (968)
Column total	114 (347)	1886 (1653)	2000

Spectrum 1 data set. Since the p -value is greater than 0.05, we accept the hypothesis of statistical independence of idleness between channels 2500 and 2501. The value of χ^2_{Ind} for the data set in *Spectrum 2* during the 4-5 am and 4-5 pm intervals are 0.0028 and 0.0014, respectively. The corresponding p -values are 0.9578 and 0.9698. Since both these values are higher than 0.05, we infer that the independence assumption of channel idleness between channels 1700 and 1701 is valid for the Maastricht measurements.

Similarly, we perform the same analysis with randomly selected adjacent channels 562 and 563, 1819 and 1820, 3155 and 3156, 3519 and 3520, and 6435 and 6436 from both the spectra. We need to emphasize here that independence test proved to be successful in majority of the adjacent channels we have randomly chosen from the Maastricht data set, especially in the Global System for Mobile Communication (GSM) 1800 MHz cellular band. However, we must admit that there are a few examples of adjacent channels like 3519 and 3520 and 1000 and 1001, which proved to have dependence in terms of channel availability implying occupancy by the same primary user.

Non-Identical Distribution Validation: For the validation of the claim for non-identical distribution of idleness between adjacent channels, we again refer to the experimental measurements and apply a non-parametric method called the McNemar's test [19]. The test is applied to the same contingency table as in Table II, which tabulates the outcomes of frequencies on adjacent channels 2500 (1700) and 2501 (1701) for the Aachen (Maastricht) measurements. Here, the null hypothesis is the marginal homogeneity between these two adjacent channels, *i.e.*, probability of channel 2500 (1700) being occupied and channel 2501 (1701) being idle (p_{bi}) subsequently is identical to the probability of channel 2500 (1700) being idle and channel 2501 (1701) being occupied (p_{ib}) subsequently. In the sequel, p_{bi} , *etc.* are the theoretical probability of occurrences in cells with the corresponding label.

The McNemar test statistic with Yates' correction for continuity is given by:

$$\chi^2_{Iden} = \frac{(|b - c| - 0.5)^2}{b + c}, \quad (22)$$

where b corresponds to the frequency (994 (852) from Table II) in the cell of channel 2500 (1700) being busy and channel 2501 (1701) being idle subsequently and c corresponds to the frequency (46 (167) from Table II) in the cell of channel 2500 (1700) being idle and channel 2501 (1701) being busy subsequently. Based on the computations, the statistic χ^2_{Iden} is equal to 862.3163 (452.89). The corresponding McNemar's test probabilities are less than $2.2e^{-16}$. For the test of marginal homogeneity, the McNemar's test probability of less than

TABLE III
OBSERVED AND EXPECTED FREQUENCIES OF SPECTRUM AVAILABILITY FOR AACHEN, GERMANY DATA

Time of the day	Frequencies				
	0.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
Observed(7 – 8am)	21	5	5	1	15
Expected(7 – 8am)	22.27	3.72	3.11	3.49	14.4
Observed(12 – 1pm)	23	4	5	8	9
Expected(12 – 1pm)	23.3	5.56	4.55	4.76	10.83
Observed(3 – 4pm)	16	6	3	6	11
Expected(3 – 4pm)	16.86	3.98	3.41	3.85	13.9
Observed(11p – 12a)	15	10	1	4	12
Expected(11p – 12a)	15.6	5.16	4.47	4.87	11.9

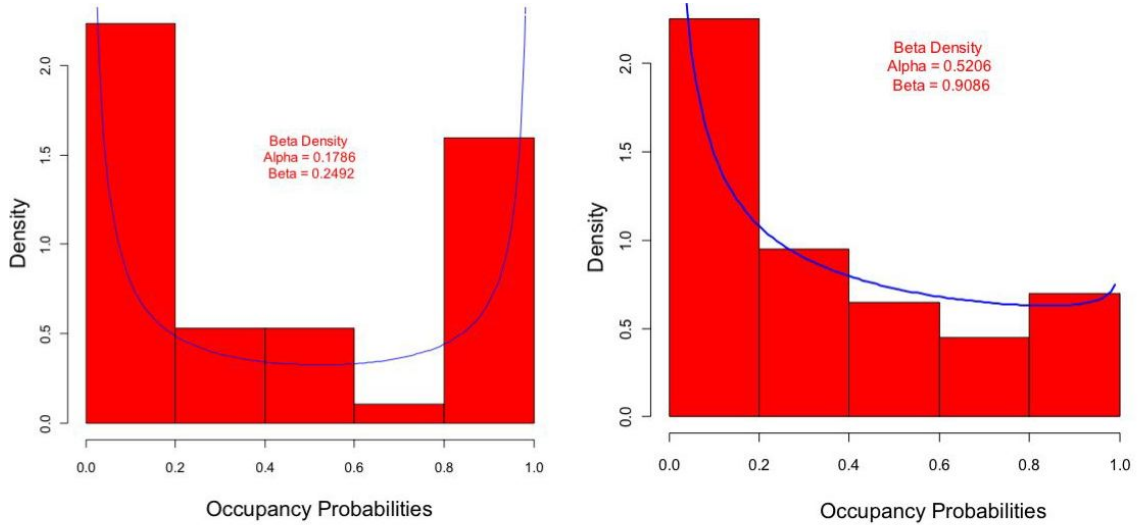
TABLE IV
OBSERVED AND EXPECTED FREQUENCIES OF SPECTRUM AVAILABILITY FOR MAASTRICHT, NETHERLANDS

Time of the day	Frequencies				
	0.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
Observed(4 – 5am)	45	19	13	9	14
Expected(4 – 5am)	40.97	18.27	14.62	13.02	13.12
Observed(1 – 2pm)	46	19	15	6	14
Expected(1 – 2pm)	43.11	17.73	14.0	12.43	12.74
Observed(4 – 5pm)	47	25	10	5	13
Expected(4 – 5pm)	48.54	16.37	12.52	10.99	11.58
Observed(9 – 10pm)	44	24	13	5	14
Expected(9 – 10pm)	42.91	18.64	14.56	12.49	11.39

or equal to 0.05 is commonly interpreted as justification for rejecting the null hypothesis. Therefore, the hypothesis of marginal homogeneity is rejected. Hence, the adjacent channels are non-identical in terms of occupancy or idleness. Similarly, we have performed the same analysis with randomly selected adjacent channels 1819 and 1820, 3155 and 3156, 1000 and 1001, and 3519 and 3520 and the hypothesis of non-identical distribution of occupancy or idleness were valid for both the spectra.

The Beta distribution model for non-identically distributed CIPs in a spectrum is validated here by using the above mentioned two sets of measurements. In other words, we have partitioned the measurements in *Spectrum 1* into four intervals of interest: (i) 7 to 8 am, (ii) 12 to 1 pm, (iii) 3 to 4 pm, and (iv) 11 pm to 12 am. Similarly, the measurements in *Spectrum 2* are grouped into four intervals: (i) 4 - 5 am, (ii) 1 - 2 pm, (iii) 4 - 5 pm, and (iv) 9 pm - 10 pm.

A total of 100 random channels are considered in each spectrum and the corresponding CIPs are evaluated by considering the following detection thresholds: -114 dBm for channels in *Spectrum 1* and -107 dBm for channels in *Spectrum 2* and normalizing the results over the one hour periods. Idleness of a channel is decided based on whether the received power is higher or lower than the detection threshold. At any given time instant, if the received power is greater than the detection threshold, the channel is said to be occupied and



Histogram of channel occupancy probabilities and Beta density for data between 7 to 8 am data at Aachen, Germany

Histogram of channel occupancy probabilities and Beta density for data between 4 to 5 am at Maastricht, Netherlands

Fig. 3. Plot of Beta density over the histogram of channel idleness probabilities in two different geographical locations

idle if received power is lower than the threshold. The mean availability μ_{av} and standard deviation σ_{av} are then computed over the obtained CIPs. The four sets of μ_{av} and σ_{av} are as follows:

Spectrum 1:

$$7 - 8 \text{ am: } \mu_{av} = 0.4176, \sigma_{av} = 0.1703, 12 - 1 \text{ pm: } \mu_{av} = 0.3756, \sigma_{av} = 0.1377, \\ 3 - 4 \text{ pm: } \mu_{av} = 0.4654, \sigma_{av} = 0.1617, \text{ and } 11 \text{ pm} - 12 \text{ am: } \mu_{av} = 0.4571, \sigma_{av} = 0.1396.$$

Spectrum 2:

$$4 - 5 \text{ am: } \mu_{av} = 0.3643, \sigma_{av} = 0.0953, 1 - 2 \text{ pm: } \mu_{av} = 0.3521, \sigma_{av} = 0.0961, \\ 4 - 5 \text{ pm: } \mu_{av} = 0.3202, \sigma_{av} = 0.0962, \text{ and } 9 - 10 \text{ pm: } \mu_{av} = 0.3462, \sigma_{av} = 0.0907.$$

Estimated $\hat{\alpha}$ and $\hat{\beta}$ parameters for the Beta distribution are computed using the following expressions [19]:

$$\hat{\alpha} = \mu_{av} \left(\frac{\mu_{av}(1 - \mu_{av})}{\sigma_{av}} - 1 \right), \\ \hat{\beta} = (1 - \mu_{av}) \left(\frac{\mu_{av}(1 - \mu_{av})}{\sigma_{av}} - 1 \right). \quad (23)$$

For *Spectrum 1* we have:

$$7 - 8 \text{ am: } \hat{\alpha} = 0.1786, \hat{\beta} = 0.2492, 12 - 1 \text{ pm: } \hat{\alpha} = 0.2642, \hat{\beta} = 0.4392, \\ 3 - 4 \text{ pm: } \hat{\alpha} = 0.2506, \hat{\beta} = 0.2879, \text{ and } 11 \text{ pm} - 12 \text{ am: } \hat{\alpha} = 0.3554, \hat{\beta} = 0.422.$$

Similarly, for *Spectrum 2* the estimated parameters are as follows:

$$4 - 5 \text{ am: } \hat{\alpha} = 0.5206, \hat{\beta} = 0.9086, 1 - 2 \text{ pm: } \hat{\alpha} = 0.4842, \hat{\beta} = 0.8909, \\ 4 - 5 \text{ pm: } \hat{\alpha} = 0.4046, \hat{\beta} = 0.8590, \text{ and } 9 \text{ pm} - 10 \text{ pm: } \hat{\alpha} = 0.5179, \hat{\beta} = 0.9780.$$

Based on the CIPs computed, we have classified them into five intervals of equal widths namely, 0.0-0.2, 0.2-0.4, 0.4-

0.6, 0.6-0.8, and 0.8-1.0. The frequency of idle channels in each interval is specified as ‘Observed’ in Tables III and IV for *Spectrum 1* and *Spectrum 2*, respectively. The ‘Expected’ frequencies are obtained using the Beta distribution in Eq. 1 with the corresponding $\hat{\alpha}$ and $\hat{\beta}$ parameters and specific intervals. The ‘Observed’ frequencies and the Beta distribution with the estimated parameters are shown in Figure 3 for both the spectra. The figure indicates that the Beta distribution is a good fit to the spectrum availability patterns in both *Spectrum 1* and *Spectrum 2* under investigation. Now we validate the Beta distribution assumption for channel availability modeling using KS and Chi-Square goodness-of-fit tests.

KS Test Validation: The null hypothesis under the KS test [19] is that the instantaneous CIPs follow a Beta distribution. Based on the computed CIPs, the corresponding p-values [19] for the KS test on *Spectrum 1* are as follows:

$$7 - 8 \text{ am: } p\text{-value} = 0.1882, 12 - 1 \text{ pm: } p\text{-value} = 0.766, \\ 3 - 4 \text{ pm: } p\text{-value} = 0.9103, \text{ and } 11 \text{ pm} - 12 \text{ am: } p\text{-value} = 0.7156.$$

The null hypothesis should be rejected if the p-value is ≤ 0.05 . It is evident that the p-value for all the four cases are > 0.05 . Hence the null hypothesis of Beta distribution assumption for the CIPs is accepted based on the KS test. Let us now focus on *Spectrum 2*. The corresponding p-values are:

$$4 - 5 \text{ am: } p\text{-value} = 0.06803, 1 - 2 \text{ pm: } p\text{-value} = 0.0991, \\ 4 - 5 \text{ pm: } p\text{-value} = 0.004139, \text{ and } 9 \text{ pm} - 10 \text{ am: } p\text{-value} = 0.1317.$$

The p-value of 0.004139 for the 4-5 pm interval suggests that Beta distribution is not a valid assumption. However, as observed from the rest of the scenarios, Beta distribution is a valid model for non-identically distributed CIPs.

Chi-Square Goodness-of-Fit Validation: From Tables III and IV, we have the ‘Observed’ and ‘Expected’ frequencies. Let us denote n_k for the observed frequency in the k^{th} interval and m_k for the corresponding expected frequency. The Chi-square statistic, χ_{av}^2 , is defined as the weighted difference

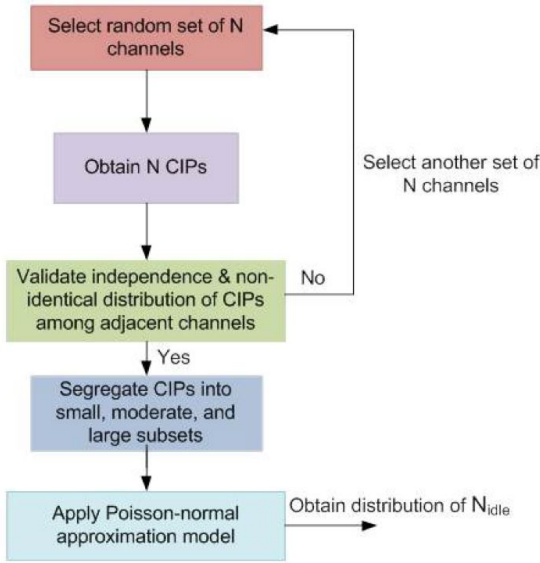


Fig. 4. Flowchart for computation of distribution of N_{idle} using Poisson-normal approximation method

between n_k and m_k and is expressed as:

$$\chi_{av}^2 = \sum_{k=1}^K \frac{(n_k - m_k)^2}{m_k}. \quad (24)$$

The 5% critical value of χ_{av}^2 for two degrees of freedom is 5.99. Referring to the ‘Observed’ and ‘Expected’ frequencies from Tables III and IV, χ_{av}^2 values for *Spectrum 1* are 3.46, 3.00, 2.92, and 5.41, respectively. For *Spectrum 2*, χ_{av}^2 values are 1.91, 3.81, 8.54, and 5.83, respectively. It is apparent that values of χ_{av}^2 for both the sets of measurements are less than the critical value, except for the measurements during 4-5 pm at Maastricht. Hence, we infer that the null hypothesis for the non-identical distribution of CIPs defined by a Beta distribution is accepted.

From both the KS and Chi-square goodness-of-fit tests, we infer that Beta distribution is a good fit in order to model the distribution of CIPs in an operating spectrum. However, it is advisable to examine this assumption for exclusions like the evening data in Maastricht prior to practical experimentations.

B. Distribution of N_{idle}

In this section, we compare the exact distribution of N_{idle} with its normal and Poisson-normal approximations in terms of *over-all approximation error* ϵ introduced in Section III.

As shown in Figure 4, we first consider a random N -set of 50 channels from both *Spectrum 1* and *Spectrum 2*. The p_i 's, $i \in \{1, 2, \dots, 50\}$, shown in Figure 5(a), are computed and normalized over an hour period (*i.e.*, 2000 time sweeps), a substantial number for statistical enumerations. Now, we validate the i.n.i.d. model among the computed p_i 's. If validated, the CIPs are then segregated into low, moderate, and large subsets. Otherwise, a new set of 50 channels has to be selected. We consider $p_{th1} = 0.03$ and $p_{th2} = 0.97$ as the lower threshold and upper threshold values, respectively. The CIPs greater than p_{th1} and lesser than p_{th2} will be in the moderate bin. Finally, we obtain the distribution of N_{idle} using

the proposed Poisson-normal approximation technique on the three subsets as illustrated in Section IV.

For comparative studies, we compute the exact distribution and the normal approximation of N_{idle} over the same set of 50 channels by using Eqs. (4) and (8), respectively, and plot Figure 5 for both *Spectrum 1* and *Spectrum 2* along with the proposed Poisson-normal approximation. As observed from the figures, the normal approximation deviates substantially from the exact distribution with over-all error ϵ , introduced in Eq. 7, of 0.1663 for *Spectrum 1* and 0.1829 for *Spectrum 2*. On the other hand, the Poisson-normal approximation follows the exact distribution closely with over-all approximation error ϵ of 0.0236 for *Spectrum 1* and 0.0830 for *Spectrum 2*. Here we consider $\theta = 0.0004$ and we neglect $Pr(N_{idle} = k) < 0.0004$ to obtain the essential support. Additionally, the normal approximation achieves a mode at $N = 37$ with modal probability 0.1493. On the other hand, the exact distribution achieves a mode at $N = 33$ while the Poisson-normal approximation achieves a mode at $N = 34$ with respective modal probabilities of 0.1472 and 0.1463. The essential support $\mathcal{S}_{N_{idle}}$ of the Poisson-normal approximation is $\{25, \dots, 43\}$, which exactly follows the essential support of the exact distribution. In contrast, the essential support of the normal approximation is $\{28, \dots, 44\}$. This implies that the normal approximation overestimates the number of idle channels in an operating spectrum at a given time period. This may prove to be detrimental for an adaptive spectrum sensing technique. However, our proposed Poisson-normal approximation method provides a close estimate of N_{idle} .

To advocate inefficiency of normal approximation and accuracy of our proposed model, we further consider another spectrum of 30 channels randomly selected from the real-time measurements at Aachen. The CIPs are computed over an hour period. The exact probabilities, approximate probabilities, and approximation errors for each k in $\mathcal{S}_{N_{idle}}$ are provided in Table V.

VI. CONCLUSION

In this paper, we have substantiated the observation that the probabilities of channel idleness in a spectrum follow an *independent and non-identical distribution* (i.n.i.d.) paradigm, providing a counterpoint to the existing research of hypothetical i.i.d channel idleness models. Later on, we have validated that the non-identically distributed *channel idleness probabilities* can be suitably modeled by a Beta distribution with appropriate values of α and β . The validation was performed based on two disparate data sets conducted at two different geographic locations. Based on this i.n.i.d. characteristics of channel idleness, we have proposed a unique analytical model for computing *spectrum availability*. We labeled this novel approach as *Poisson-normal approximation*. Comparison of results, using approximation error as the evaluation criterion, derived from real-time measurements prove the efficacy of our approximation model for *spectrum availability* and exhibits the inefficiency of the conventional normal approximation approach commonly used in the existing research work.

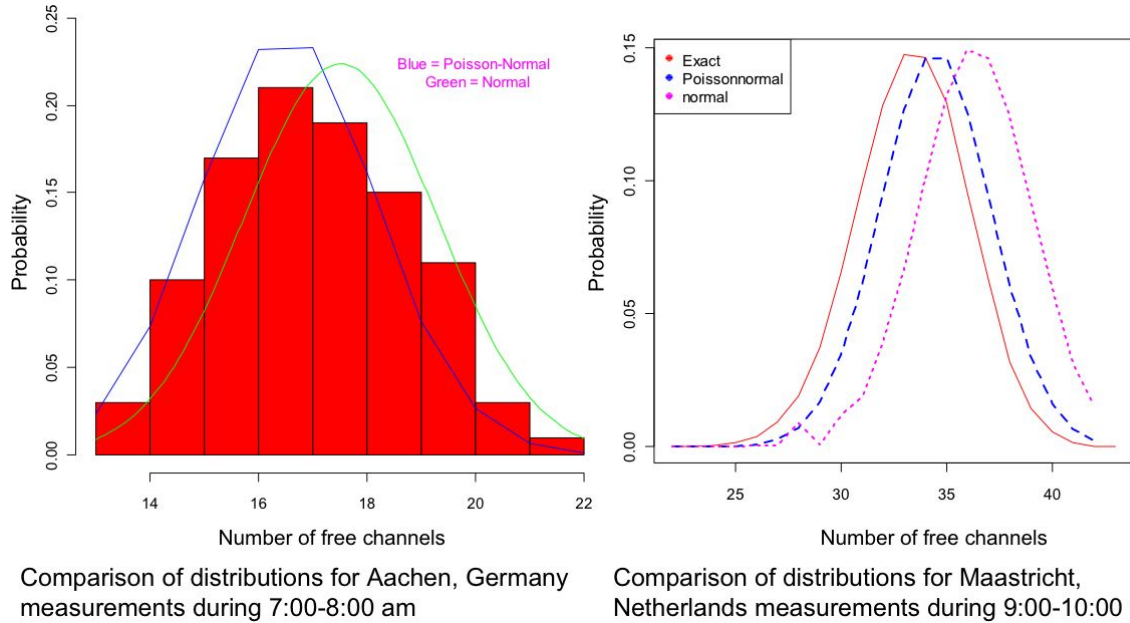


Fig. 5. Channel idleness probabilities and the comparison among the exact distribution, normal approximation, and our proposed approximation

TABLE V
COMPARISON AMONG POISSON-NORMAL APPROXIMATION, NORMAL APPROXIMATION, AND EXACT DISTRIBUTION WITH $\epsilon(k)$ FOR 30 CHANNELS

	$N_{idle} = k$									
	19	20	21	22	23	24	25	26	27	28
$Pr(N_{idle} = k)$	0.0004	0.0035	0.0199	0.0753	0.1876	0.2932	0.2664	0.1253	0.0262	0.0021
$PrNormal(N_{idle} = k)$	0.0023	0.0178	0.0784	0.1996	0.2928	0.2478	0.121	0.0341	0.0055	0.0005
$\epsilon(k)$	0.0019	0.0143	0.0585	0.1243	0.1052	0.0454	0.1454	0.0912	0.0207	0.0016
$PrPoi-Normal(N_{idle} = k)$	0.0012	0.0093	0.0341	0.0989	0.1963	0.2991	0.2876	0.1437	0.0581	0.0062
$\epsilon(k)$	0.0008	0.0058	0.0142	0.0236	0.0087	0.0059	0.0212	0.0184	0.0319	0.0041

ACKNOWLEDGEMENT

We would like to thank Prof. Petri Mahonen and Dr. Matthias Wellens from RWTH Aachen University for providing us with the spectral measurements and insights on the same. Additionally, we would also like to thank Prof. Preston Marshall from University of Southern California for his encouragement and comments.

REFERENCES

- [1] J. Mitola, III and G. Q. Maguire, Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Commun.*, Vol. 6, Issue 4, Aug. 1999, pp. 13 - 18.
- [2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, Vol. 23, Issue 2, Feb. 2005, pp. 201 - 220.
- [3] M. Wellens and P. Mahonen, "Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model," *Springer Mobile Networks and Applications*, August 2009. Available online: [http://dx.doi.org/10.1007/s11036-009-0199-9].
- [4] C. Ghosh, S. Roy, M. B. Rao, and D. P. Agrawal, "Spectrum occupancy modeling and validation using real-time measurements," *2nd ACM Sigmobility Workshop on Cognitive Wireless Networking, CoRoNet'10*, September 20, 2010.
- [5] P. F. Marshall, "Closed-form analysis of spectrum characteristics for cognitive radio performance analysis," *IEEE Dynamic Spectrum Access Networks, DySpan*, Chicago, IL, Sept. 2008, pp. 1-12.
- [6] C. Ghosh, S. Pagadarai, D. P. Agrawal, and A. M. Wyglinski, "Statistical spectrum occupancy modeling and validation against real-time measurements," *IEEE Trans. Wireless Commun.*, Vol. 9, No. 1, Jan. 2010, pp. 38-44.
- [7] L. Luo and S. Roy, "Analysis of search schemes in cognitive radio," *4th Annual IEEE Sensor, Mesh and Ad Hoc Communications and Networks, SECON'07*, 18-21 June 2007, San Diego, CA, pp. 647-654.
- [8] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, Vol. 2, Issue 1, Feb. 2008, pp. 28 - 40.
- [9] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part I: two user networks," *IEEE Trans. Wireless Commun.*, Vol. 6, Issue 6, June 2007, pp. 2204 - 2213.
- [10] G. Ganesan, Y. Li, B. Bing, and S. Li, "Spatiotemporal sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, Vol. 26, Issue 1, Jan. 2008, pp. 5 - 12.
- [11] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," *Proc. First International Workshop on Technology and Policy for Accessing Spectrum, TAPAS'06*, ACM International Conference Proceeding Series; Vol. 222.
- [12] M. A. McHenry and K. Steadman, "Spectrum occupancy measurements, location 2 of 6: Tyson's Square center, Vienna, Virginia, April 9, 2004," *Shared Spectrum Company Report*, August, 2005.
- [13] F. H. Sanders, B. J. Ramsey, and V. S. Lawrence, "Broadband spectrum survey at Los Angeles, California," *NTIA Report 97- 336*, May 1997.
- [14] D. A. Roberson, C. S. Hood, J. L. LoCicero, and J. T. MacDonald, "Spectral occupancy and interference studies in support of cognitive radio technology deployment," *1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks*, Virginia, 25 Sept. 2006, pp. 26-35.
- [15] R. B. Ash, *Real Analysis and Probability*, Academic Press, Orlando, Florida, 1972, pp. 337.
- [16] K. Lange, *Applied Probability*, Springer, New York, 2003.
- [17] S. Kandeepan, A. Giorgetti, M. Chiani, "M-Periodic Spectrum Sensing Performance and Requirements for Detecting Legacy Users with Temporal and Noise Statistics in Cognitive Radios," *IEEE GLOBECOM Workshops*, 2009, pp. 1-6.
- [18] Y-C. Liang, Y. Zeng, E. C. Y. Peh, A. T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, Vol. 7, Issue 4, 2008, pp. 1326-1337
- [19] A. Stuart, K. Ord, and S. Arnold. *Kendall's Advanced Theory of Statistics*, 2A. London, pp. 25.3725.43, 1999.



Chittabrata Ghosh is currently a Senior Researcher at Nokia Research Center (NRC), Berkeley, CA since July 2011. Before joining NRC, he was a Postdoctoral Researcher at University of Washington, Seattle since July 2009. He had received Ph.D. degree from Department of Computer Science, University of Cincinnati, OH in 2009. He is actively involved in IEEE standardization activities focusing on system design and architecture for low power Wi-Fi networks. Apart from this, his research interests include analysis of coexistence issues among various

cooperating wireless networks and MAC layer optimization techniques for wireless sensor networks. He is also leading Nokia's university research collaborations on multi-faceted research and policy issues in the TV White Spaces. His activities for the IEEE Communications Society (ComSoc) includes membership of several technical and conference program committees. He is also the Vice Chair and Co-Editor of Technical Committee on Simulations under the IEEE Computer Society.



M. Bhaskara Rao is currently a Professor of Statistical Genetics and Biomedical Engineering at the University of Cincinnati. He got his Ph.D. from the Indian Statistical Institute, Calcutta. He had faculty positions at the University of Sheffield, England, University of Pittsburgh, and North Dakota State University. He has over 150 publications and two books. His research interests span from topology to matrix algebra to abstract probability theory to computational statistics. On the biomedical engineering side he is drawn to tissue engineering on design

and data analysis issues. He guided 30 Ph.D. students and over 40 Masters students. He is a fellow of the American Statistical Association, Institute of Mathematical Statistics, International Statistical Institute, and American Association for the Advancement of Science. A national award Excellence in Teaching is bestowed on him by the Mathematical Association of America. He is currently working on Algebraic Statistics and its application to Statistical Genetics and Gene Environment Interactions.



Sumit Roy received the B. Tech. degree from the Indian Institute of Technology (Kanpur) in 1983, and the M. S. and Ph. D. degrees from the University of California (Santa Barbara), all in Electrical Engineering in 1985 and 1988 respectively, as well as an M. A. in Statistics and Applied Probability in 1988. Presently he is Professor of Electrical Engineering, Univ. of Washington where his research interests include analysis/design of wireless communication and sensor network systems with a diverse emphasis on various technologies: wireless LANs (802.11)

and emerging 4G standards, multi-standard wireless inter-networking and cognitive radio platforms, vehicular and underwater networks, and sensor networking involving RFID technology. His activities for the IEEE Communications Society (ComSoc) includes membership of several technical and conference program committees, notably the Technical Committee on Cognitive Networks. He currently serves on the Editorial Board for IEEE Trans. Communications and IEEE Intelligent Transportation Systems.