Improving Primary Statistics Prediction
Under Imperfect Spectrum Sensing

Ahmed Al-Tahmeesschi∗, Miguel López-Benítez∗†, Janne Lehtomäki† and Kenta Umebayashi‡

∗Dept. of Electrical Engineering and Electronics, University of Liverpool, United Kingdom
†Centre for Wireless Communications, University of Oulu, Finland
‡Graduate School of Engineering, Tokyo University of Agriculture and Technology, Japan

Abstract—Dynamic Spectrum Access (DSA) / Cognitive Radio (CR) systems utilize spectrum sensing to monitor spectrum status and decide transmission time in an opportunistic manner. This results in an increase in wireless spectrum efficiency. Spectrum sensing can also be used to monitor the statistics of primary users to gain information on occupation patterns and estimate the statistics of the primary traffic activity, a useful knowledge that can be exploited in many ways. In this research, three novel algorithms are proposed to enhance the estimation of primary user activity statistics under imperfect spectrum sensing given the knowledge of minimum transmission time. Simulation results show that the proposed methods enable accurate estimation for the primary user statistics. Moreover, the proposed methods are compared to previously proposed methods and it is shown they provide a significantly better estimation accuracy.

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

I. INTRODUCTION

Cognitive radio (CR) is proposed as a solution to increase the spectrum efficiency. The core concept of CR is based on dynamic spectrum access (DSA). CR is defined as a smart radio that is capable of changing its transmission parameters according to the surrounding environment [1]. In CR systems the secondary users (SU) access the primary users (PU) channel in an opportunistic and non-interfering manner during PU idle times (spectrum holes) [2]. SU have to quickly and accurately detect spectrum holes in PU bands.

The CR device could benefit from the knowledge of PU activity for opportunistic transmission, and to estimate the statistics of PU which could be used to access the spectrum more effectively. There are two main ways to obtain statistical information on PU activity: databases with relevant information and directly by sensing. There are several spectrum sensing methods [3], including the well-known energy detection (ED) method [4] which benefits from simplicity and low complexity with no requirements for PU signal characteristics.

It can be assumed that the sensing is nearly perfect when the PU signal has a sufficiently high power, while in the case of low power from PU sensing errors might occur. Given the opportunistic nature of CR, the knowledge of PU spectrum occupation is very important as it limits the operation time of SU. The statistical information of PU spectrum occupation can be used by SU to accurately predict the channel state, select appropriate transmission channel and increase the spectrum efficiency for CR. The calculation of PU spectrum occupancy in real life is far from being perfect as a result of sensing errors. The main target of this research is to design methods to correctly estimate the statistics of PU activity (e.g., distribution of period durations) in the presence of spectrum sensing errors. Multiple methods and algorithms have been proposed to reduce the effect of imperfect spectrum sensing including [5–8]. In this research, the errors and their effect are investigated and new algorithms to estimate the statistics and correct the imperfect sensed periods are proposed. The methods in [8] are the closest ones to this research and the obtained results will be compared to them. The computational cost/complexity analysis for all the considered methods is provided as well.

The rest of this work is organized as follows. First, Section II provides a formal description of the problem of estimating the PU activity statistics under imperfect spectrum sensing. The proposed algorithms to mitigate sensing errors are explained in detail along with previous proposed methods in Section III. The performance results of the proposed algorithms (obtained by simulation) are analysed and compared with other algorithms in Section IV. A discussion on the configuration of spectrum sensing based on the obtained results is provided in Section V. Finally, Section VI concludes the paper.

II. PROBLEM FORMULATION

A CR senses the channel at a finite rate with duration $T_s$ as shown in Fig. 1. In every sensing event $(T_s)$, a binary decision is made to result in either idle $(H_0)$ or busy $(H_1)$ state of the channel. A perfect spectrum sensing (PSS) is obtained when a CR receiver operates in a high PU transmission power. However under low SNR imperfect spectrum sensing (ISS) may occur with two types of errors: false alarm $(\hat{H} = H_1|H_0)$ where the signal is not present but announced as present because of the high noise level, and missed detection $(\hat{H} = H_0|H_1)$ where the signal is present but with power lower than the energy detection threshold.

Sensing errors have a significant impact on the performance of communication systems (both PU and SU). In the case of a false alarm, the SU transmitter will be silent in this slot when it should be transmitting, leading to a low CR transmission rate. While in the case of a missed detection interference with the PU signal will occur. Inaccurate detection leads to inaccurate estimation for PU activity statistics as well.

There are multiple factors that affect the estimation of PU activity pattern, including the probability of false alarm $(P_{fa})$, the probability of missed detection $(P_{md})$ and how the errors are distributed. Note that the sensing errors will affect the lengths of periods depending on their location as it can be appreciated from Fig. 2 where three types of errors can be identified: missed detection, false alarm and originally correct periods but with missed detection/false alarm that occurs at its beginning, middle or end, which results in making its length shorter than the minimum PU activity time $\mu_i$ ($i = 0$ for idle periods).
periods, $i = 1$ for busy periods) and therefore detectable if the value of $\mu_i$ is known. According to this, multiple errors could occur in a long period dividing it into multiple short periods which would have a significant effect on the calculation of PU statistics. Note that at low probability of error the case of single error is dominant, while at higher error probability the case of multiple errors is the dominant one.

The main objective of this research is to design algorithms that are capable of accurately estimating the PU activity statistics (such as the mean, variance, duty cycle and distribution of period durations) in the presence of sensing errors.

## III. METHODS PROPOSED TO OVERCOME THE EFFECT OF SPECTRUM SENSING ERRORS

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

1) Regional beacon signals with real-time information about minimum activity time for PU present in the geographical area. The main drawback of this method is the requirement to modify the primary network [9].

2) Offline method as in [10], where the SU utilize blind estimation methods in order to determine the minimum PU activity time.

3) Offline methods with prior knowledge on the operating standards. These methods can be applied in slotted systems where the frame duration is known (e.g., GSM [11]).

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

### A. Method 1

Fig. 3 shows the probability mass function (PMF) of the number of consecutive sensing events ($T_s$) affected by sensing errors (obtained from simulations). As it can be appreciated, in most cases sensing errors occur in individual isolated sensing events; only when the probability of error ($P_{fa}$) is higher (bottom of Fig. 3) bursts of two or more consecutive erroneous sensing events can occur. This observation can be exploited to identify the occurrence of sensing errors and reconstruct the original periods as follows. First, a threshold $\beta T_s$ (with $\beta \in \mathbb{N}^+$) is defined, which can be tuned based on Fig. 3. Starting from an initial estimated period $\hat{T}_{i,n}$ which has a duration less than the threshold and the minimum ($\hat{T}_{i,n} < \beta T_s$ and $\hat{T}_{i,n} < \mu_i$), where $n \in \mathbb{N}^+$ represents the index in the

![Fig. 1. Estimation of period durations from spectrum sensing decisions: (a) under perfect spectrum sensing, (b) under imperfect spectrum sensing [8]. $\hat{T}_i$ represents the estimated period under perfect spectrum sensing and $T_i$ represents the estimated period under imperfect spectrum sensing.](image)

![Fig. 2. Sensing errors according to their location: (a) Original period, (b) Single detectable incorrect period, (c) Two detectable incorrect periods, (d) Multiple detectable incorrect periods.](image)

![Fig. 3. PMF of the number of consecutive sensing events affected by sensing errors ($\Psi = 0.8, \mu_i = 10$ t.u., $E(T_s) = 50$ t.u.).](image)
sequence of observed periods, the durations of all subsequent periods (both busy and idle) are checked until a period of the opposite type with a duration greater than $\beta T_s$ (i.e., $\hat{T}_{i,n+N} > \beta T_s$) is found. A reconstructed period of the same type (idle/busy) as $\hat{T}_{i,n}$ is then estimated by adding the durations $\hat{T}_{i,n} = \hat{T}_{i,n} + \ldots + \hat{T}_{i,n+N-1}$ where $\hat{T}_{i,n}$ is the new reconstructed period of the original. The period $\hat{T}_{1,n+N}$ (which is of different type than the previously reconstructed period) is then taken as the starting point for a new reconstruction based on the same principle. This process is repeated over the sequence of estimated periods so that all periods shorter than $\beta T_s$ (which based on Fig. 3 can be assumed to be short periods resulting from sensing errors) will be added in an attempt to reconstruct the original sequence of busy/idle periods. After applying the method above, some of the reconstructed periods were observed to be shorter than $\mu_i$, thus indicating the presence of a few incorrectly reconstructed periods. Therefore a second step is performed after the reconstruction process above where all the reconstructed periods $\hat{T}_{i,n}$ shorter than $\mu_i$ are discarded and not included in the computation of the PU activity statistics.

B. Method 2

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

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

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

C. Method 3

As observed in Fig. 2, the occurrence of multiple sensing errors can lead to a pattern where two periods that are longer than $\mu_i$ (and therefore can be assumed to be observed in their original idle/busy state) contain a number of other periods shorter than $\mu_i$ (which cannot be classified as either correct or incorrect idle/busy observations). This method aims at reconstructing the original period by adding the durations of all the periods $\hat{T}_{i,n} + \ldots + \hat{T}_{i,n+N}$ between two consecutive periods longer than or equal to the minimum PU transmission time (i.e., $\hat{T}_{i,n−1} ≥ \mu_i$, $\hat{T}_{i,n+N+1} ≥ \mu_i$ and $\hat{T}_{i,n}, \ldots, \hat{T}_{i,n+N} < \mu_i$) to the next or previous period depending on the period types.

If the starting period $\hat{T}_{i,n−1}$ and ending period $\hat{T}_{i,n+1}$ are of the same type (busy/idle), then the reconstructed period will be of that same type. However, if $\hat{T}_{i,n−1}$ and $\hat{T}_{i,n+1}$ are of different types, then it is not possible to determine unambiguously the original period type since it depends on the particular order in which sensing errors occurred and it is not possible to determine which of the periods shorter than $\mu_i$ are correct/incorrect idle/busy observations. In this other case, the period type is randomly decided as busy with probability $\hat{\Psi}$ and idle with probability $1 - \hat{\Psi}$, where $\hat{\Psi}$ is an estimation of the PU channel duty cycle $\Psi$ obtained from past spectrum sensing observations. If the average busy duration is $\mathbb{E}\{T_1\}$ and the average idle time is $\mathbb{E}\{T_0\}$, then the duty cycle $\Psi$ can be estimated as:

$$\hat{\Psi} = \frac{\mathbb{E}\{T_1\}}{\mathbb{E}\{T_0\} + \mathbb{E}\{T_1\}}$$ (1)

Fig. 4 shows the estimated duty cycle $\hat{\Psi}$ quickly converges to the real value $\Psi$, meaning that the algorithm needs a short transition time before it can operate correctly.

D. Previous work

The proposed three methods will be compared with the ones presented in [8], which are the closest ones to this research. The methods proposed in in [8] are summarized below:

---

**Algorithm 1:** Method 1

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

**Algorithm 2:** Method 2

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

---
as in [8], which can be summarized as follows:

\[
\hat{\psi} \left\{ \mu_i, \hat{T}_i, \psi \right\} = 0
\]

while

\[
N = 0
\]

end

if type of \( \hat{T}_i \) = idle then

\[
\hat{T}_i = \hat{T}_i + \hat{T}_i + \ldots + \hat{T}_i
\]

else

\[
\hat{T}_i = \hat{T}_i + \ldots + \hat{T}_i
\]

end

end

\[
\hat{\psi} \left\{ \mu_i, \hat{T}_i, \psi \right\} = 0
\]

Generate uniform random number \( \zeta \) uniform random number \( \tilde{\zeta} \) 

Method 4: Every observed period with a length shorter than the minimum period is discarded.

Method 5: Every observed period with a length shorter than the minimum period is discarded along with the preceding and succeeding periods.

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

IV. SIMULATION RESULTS

The performance of the six methods considered in this work is evaluated by means of simulations following the same steps as in [8], which can be summarized as follows:

1) Generate idle/busy periods’ lengths \( T_i \) following a generalized Pareto distribution, which has been proven to provide the best fit to empirical spectrum data [12].

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

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

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

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

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

\[
D_{KS} = \sup_{T_i} \left| F_{T_i}(T_i) - F_{T_i}(T_i) \right|
\]

Note that the sequence of idle/busy lengths \( T_i \) generated in step 1 contains real positive values and its CDF \( F_{T_i}(T_i) \) has a continuous domain, while the sequences resulting from steps 4 and 5 contain values \( \hat{T}_i \) that are integer multiples of the sensing period \( \hat{T}_i = NT_s, N \in \mathbb{N}^+ \) and their CDFs \( F_{T_i}(\hat{T}_i) \) have a discrete domain. Since it is not possible to compare continuous and discrete sets, the discrete set \( F_{T_i}(\hat{T}_i) \) is interpolated to produce \( F_{T_i}(T_i) \), which is compared in (2) with the distribution of the original lengths \( F_{T_i}(T_i) \).

Fig. 5 shows the accuracy of the estimated CDF of PU idle periods in terms of the KS distance (with respect to the original periods of step 1) as a function of the sensing period, when no reconstruction method is used (i.e., after step 4) and when a reconstruction method is used (i.e., after step 5). Results are shown for method 1 (left column), method 2 (centre column) and method 3 (right column), for \( P_{fa} = P_{md} = 0.01 \) (top row) and \( P_{fa} = P_{md} = 0.1 \) (bottom row). Simulations were performed for duty cycle values of \( \psi = 0.2 \) (low PU channel load), \( \psi = 0.5 \) (medium load) and \( \psi = 0.8 \) (high load). Due to the lack of space, results are shown in Fig. 5 only for \( \psi = 0.8 \), however the results for \( \psi = 0.2 \) and \( \psi = 0.5 \) showed similar trends and provided the same conclusions.

As it can be appreciated in Fig. 5, the accuracy in all cases is noticeably better when the probability of error is lower (top row). However, all the methods proposed in this work can provide significant accuracy improvements with respect to the case where no reconstruction is performed (i.e., the PU statistics are computed based on the raw period lengths observed under ISS). In some particular cases the estimation is nearly perfect (i.e., \( D_{KS} \approx 0 \)).

For method 1, there is a wide range of (low) values of the sensing period \( T_s \) that can provide a nearly perfect estimation \( D_{KS} \approx 0 \). When the probability of sensing errors is lower.
sensing errors is lower, most sensing errors occur in individual

Fig. 5. Performance of the methods proposed in this work (details provided in the text).
The system designer can essentially control two parameters, namely the operating point of the employed spectrum sensing method and the employed sensing period $T_s$. The operating point of the employed spectrum sensing method determines its performance (i.e., the values of $P_{fa}$ and $P_{md}$) and can be tuned by modifying the parameters of the sensing algorithm; for example, in the case of energy detection the main design parameter is the detection/decision threshold. As suggested by the results of Fig. 5, the sensing algorithm should ideally be configured to operate in a point where $P_{fa}$ and $P_{md}$ are as low as possible since for lower values of $P_{fa}$ and $P_{md}$ there is a relatively wider range of values of the sensing period $T_s$ where the error is zero or very close to zero, which makes it relatively easy to provide a (nearly) perfect estimation of the PU activity statistics (i.e., $D_{KS} \approx 0$). If the operating point of the sensing algorithm cannot be modified (e.g., it is configured to provide specific performance targets required by a particular service) and the resulting $P_{fa}$ and $P_{md}$ are high, the proposed methods can still provide a (nearly) perfect estimation of the PU activity statistics, however in this case the value of the sensing period $T_s$ should be selected carefully. In particular, Fig. 5 shows that for method 1 it should take low values (ideally close to zero, $T_s \approx 0$), for method 2 it should be $T_s \approx \mu_s / 2$, and for method 3 it should take high values (ideally close to the minimum PU transmission time, $T_s \approx \mu_s$). Moreover, if the operating point of the sensing method and the sensing period are both constrained, the only degree of freedom for the system designer would be the selection of one of the three proposed methods. In such a case, the decision could be based on which method provides the best accuracy for the employed sensing period $T_s$ (e.g., method 1 when $T_s \approx 0$, method 2 when $T_s \approx \mu_s / 2$, and method 3 when $T_s \approx \mu_s$). In any case, the set of methods proposed in this work can provide a very accurate (nearly perfect) estimation of the PU activity statistics under imperfect sensing, even with a high probability of sensing errors, over a wide range of operating conditions.

**Fig. 6.** Performance comparison of the methods considered in this work.

**TABLE I**

<table>
<thead>
<tr>
<th>Method</th>
<th>Average computation time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1 ($\beta = 3$)</td>
<td>0.0062</td>
</tr>
<tr>
<td>Method 2 ($K = 10$)</td>
<td>0.1252</td>
</tr>
<tr>
<td>Method 3 (real-time est. $\Psi = 0.8$)</td>
<td>0.3811</td>
</tr>
<tr>
<td>Method 4</td>
<td>0.0004</td>
</tr>
<tr>
<td>Method 5</td>
<td>0.0011</td>
</tr>
<tr>
<td>Method 6</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

**VI. CONCLUSION**

CR systems can benefit from the knowledge of the PU spectrum activity statistics. The estimation of such statistics, however, can be very inaccurate when based on sensing observations due to the practical limitations of spectrum sensing and the presence of sensing errors. This paper has addressed the problem of accurately estimating the PU activity statistics under ISS by proposing three simple but effective methods to overcome the degrading effects of spectrum sensing errors on the estimated statistics. The performance has been evaluated by means of computer simulations and compared to other methods previously proposed in the literature. The obtained results show that the methods proposed in this work are able not only to provide significant accuracy improvements with respect to the existing methods but also, and more importantly, a very accurate (virtually perfect) estimation of the PU activity statistics. This paper has also discussed how the proposed methods should be configured in a practical system design in order to achieve the best attainable accuracy.

**ACKNOWLEDGEMENTS**

M. López-Benítez would like to thank the support of British Council under UKIERI DST Thematic Partnerships 2016-17 (ref. DST-198/2017). K. Umebayashi would like to thank the support received from the MIC/SCOPE #165003006, and JSPS KAKENHI Grant Numbers JP15K06053 and JP15KK0200.

**REFERENCES**