Spectrum Awareness in Cognitive Radio Systems

M. Suchański, P. Kaniewski, K. Kosmowski, M. Kustra, and J. Romanik

Abstract—The paper addresses the issue of the Electromagnetic Environment Situational Awareness techniques. The main focus is on sensing and the Radio Environment Map. These two dynamic techniques are described in detail. The Radio Environment Map is considered the essential part of the spectrum management system. It is described how the density and deployment of sensors affect the quality of maps and it is analyzed which methods are the most suitable for map construction. Additionally, the paper characterizes several sensing methods.

Keywords—sensing, Electromagnetic Environment, Radio Environment Map, Situational Awareness

I. INTRODUCTION

The dynamic development of wireless communication, an increasing number of radio devices and occurring spectrum deficiencies have necessitated the implementation of more and more effective Electromagnetic Spectrum Management (ESM) methods. This is particularly important when it comes to designing new systems that require higher data rates and excellent Quality of Service (QoS) to meet growing needs of customers. In such a situation, the existing static ESM methods, which use the static allocation of frequencies in the form of appropriate licenses for some users, are not able to provide frequency allocation for all of them. Therefore, the concept of Dynamic Spectrum Access (DSA) has been created. Its main proposal is to share the spectrum between different systems. It has materialized as Cognitive Radio (CR) that autonomously coordinates the use of frequency bands.

The concept of CR appeared for the first time in the dissertation by Joseph Mitola III [1]. It was proposed to use the potential for self-organization, independent planning and self-regulation in order to increase the efficiency of the use of available spectral resources [2,3]. There are currently various definitions of CR depending on the context in which they are used, including the definition adopted by the American Federal Communications Commission (FCC) [4], Software Defined Radio (SDR) Forum or IEEE [5].

CR implements the so-called Opportunistic Spectrum Access (OSA) based on the use of temporarily unoccupied spectrum slots while avoiding interference between systems operating in the same frequency range. This philosophy entails additional challenges related to the need to constantly monitor the activity of other transmitters in order to turn off Secondary Users’ (SU) emissions when transmission appears from Priority (or Primary) Users (PU), i.e. those that have licenses to use these bands of the spectrum. In the described environment, CR must gather and use spectrum knowledge known as Electromagnetic Environment Situational Awareness (EESA).

The main body of this paper provides information on creating EESA using various methods, and then characterizes these methods in detail.

II. SPECTRUM AWARENESS

One of the basic functionalities of CR is gathering information about the radio environment and thus building the EESA for each element of the network [6]. In the literature, several methods are considered how to obtain information on the status of the spectrum usage. Access to the geo-location database [7-9] allows us to obtain information about the occupancy of specific frequency bands in a defined location as well as about operators providing services in a given region and their security requirements. The geo-location database access method is considered to be the most reliable method providing a stable but static source of information on the availability of radio resources. Whereas the Radio Environment Map (REM) can be a dynamic source of spectrum availability. The REM may support the Centralized Dynamic Spectrum Management (CDSM) [10-12] to increase the efficiency of spectrum usage. Another application of REM can be the enhancement of the local frequency management system, e.g. in CR networks.

Another tool to obtain information about the frequency band that can be used at a given time and place is the cognitive pilot channel [13]. It is a special broadcast channel that, in addition to information about spectrum usage, provides CR with data on radio access techniques, operators and radio communication systems working in a given location.

The next kind of methods to build EESA is sensing [14-17], i.e. recognition of the electromagnetic environment. It involves monitoring broad spectrum bands and detecting unused spectrum parts (spectrum holes / white spaces) as well as detecting PU, especially in ad-hoc radio networks. It is required that the decision about spectrum occupancy should be taken with high probability at a given, appropriately low level of the PU’s signal. It is noteworthy that in most cases SUs may have difficulty in distinguishing between PU and other SU signals. That is why all of them are collectively treated as a PU.
signal. Therefore, the signal detection process relies on checking the binary hypothesis:

\[ y(t) = \begin{cases} \eta(t) & H_0 \\ s(t) + \eta(t) & H_1 \end{cases} \]

where: \( y(t) \) – signal received by SU, \( s(t) \) – signal transmitted by PU, \( \eta(t) \) – Additive White Gaussian Noise (AWGN), \( H_0 \): the signal received by SU is a noise (AWGN) – no PU activity;

\( H_1 \): the signal received by SU is a radio (transmitted) signal plus noise (AWGN) – presence of PU signal.

Following the above, there are two possible decisions: stating or not stating the presence of the PU signal in the band under consideration and only one of them is true at a given time. For each of these hypotheses there are two possible results: the decision was made correctly or incorrectly, as presented in Table I.

<table>
<thead>
<tr>
<th>POSSIBLE OPTIONS OF DECISION MAKING PROBABILITY IN SENSING PROCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STATUS OF RADIO ENVIRONMENT</strong></td>
</tr>
<tr>
<td>Signal presence</td>
</tr>
<tr>
<td>Signal absence</td>
</tr>
</tbody>
</table>

Therefore, sensing efficiency is determined by probability of detection \( P_d \) and probability of false alarm \( P_{fa} \). It is desirable for \( P_d \) to be as high as possible, while \( P_{fa} \) should be as low as possible to prevent unused transmission possibilities.

The literature on the subject of sensing suggests a number of different methods to identify the presence of PU signal transmission. All these methods can be classified into two basic categories: non-cooperative and cooperative sensing [18], which is presented in Fig. 1.

**A. Non-cooperative sensing**

In the case of non-cooperative sensing, CRs independently monitor the spectrum and independently (autonomously) from other network users make a local decision about the presence of PU signal in the tested frequency band. The non-cooperative methods are presented in the third chapter.

**B. Cooperative sensing**

In the spectrum monitoring process, CR has to deal with numerous problems occurring in the real world of radio communication, such as: shadowing, inaccuracy of system parameters, or hidden node problem, which limit the effectiveness of non-cooperative sensing methods. In order to solve such problems, a solution called the Cooperative Spectrum Sensing (CSS) has been proposed [19,20,64]. In this case, a group of CRs share the data acquired from autonomous (non-cooperative) sensing. Therefore, a decision about the occurrence of PU transmission on a given frequency can be made, even though some of the CR network nodes are not able to detect the occurring transmission correctly. As a result, the reliability of the decision is improved.

Depending on the capabilities of a single CR, the decision made locally may be a soft decision or a hard decision. In the case of the hard decision method, CR sends a binary decision on the result of the local detection, i.e. the presence or absence of a signal transmission in the radio channel. However, in the soft decision method, the "soft" value of the detection result is sent (for example, the estimated energy level in a given band).

A global decision about the presence of a signal transmission in the radio channel is made basing on the fusion of data obtained from a group of CRs. Each of them sends (e.g. through a reporting channel, a dedicated time slot or using a different solution) the local detection result (the hard one or the soft one) to the Fusion Centre (FC) or the cluster head. In such a place, the results which have been gathered are used to develop a global decision by choosing the right strategy for making the decision about the presence or absence of the signal. Then, the decision which has been taken is sent back to each CR.

Depending on the type of decisions received from local CR (the hard one or the soft one), FC makes a global decision based on specific rules [21]. In the case of binary (hard) decisions sent by CR, the so-called hard decision rules can be used: AND (the band is occupied if each local CR has detected PU transmission), OR (the band is occupied if any CR has detected PU transmission) or the majority rule (the band is occupied if most of local CRs have detected PU transmission). In the second case, the soft decision rules can be applied: SLC (Square Law Combining) and MRC (Maximum Ratio Combining) [22].

According to [23], CSS can be classified into three categories based on how data is shared over the network: centralized CSS [24-27], distributed CSS [28-30], and relay-assisted CSS [31,32].

In the centralized CSS, FC collects local information on sensing from all the SUs in the network via a reporting channel. Then, FC determines the decision about the presence of PU transmission. After that, the decision is sent back to all the SUs. FC is also present in CSS with relay nodes, but in such a case, local data from SUs are sent not only directly, but also through other SUs to reduce transmission errors. In the case of distributed CSS, the decision is not made by FC. In this case, each SU simultaneously (via the reporting channel) sends and receives detection data to make a decision based on them using the local decision rule.
C. Radio Environment Maps

One of the ideas that is aimed at raising the EESA is described in the literature as the REM concept [33]. The main functionality of the REM is the ability to dynamically construct maps presenting the state of the radio spectrum for all interesting locations and frequencies. The maps are constructed on the basis of multi-domain information gathered in dedicated REM databases, e.g. the geographical data, the terrain model, the set of known transmitters, service providers, spectral regulations and policy. It is expected that REM database will enable us to analyse the current condition of the electromagnetic environment and to make some predictions on the basis of the knowledge from the past.

In the literature on the topic there are two main categories of map construction techniques: direct methods, known as spatial statistics based methods and indirect methods that are also described as transmitter location based methods [34].

Some papers recommend the indirect methods due to the fact that the maps constructed with their help are of higher quality if the propagation model has been selected properly [33]. It is worth mentioning that indirect methods require a set of input data, e.g. transmitter location, the TX power and activity pattern of the transmitter as a minimum.

On the other hand, the direct methods use measurement data provided by dedicated sensors. It is obvious that deploying sensors in all required locations is impracticable. For this reason, in direct methods interpolation techniques are applied to estimate the signal strength at required locations. The most promising techniques for REM construction are as follows: Nearest Neighbour (NN), Inverse Distance Weighting (IDW) and Kriging. These techniques are described in more detail in [36].

The main task of REM is the presentation of the electromagnetic environment. The problem of the quality metric of this presentation has been raised in the literature. The Root Mean Square Error (RMSE) has been proposed in [37] as a convenient metric calculated for all the locations within the area under analysis. In the case of REM, the RMSE indicates the similarity between real and estimated Received Signal Strength (RSS) values. When the direct methods are applied, such an approach assumes two kinds of sensors, namely sensors providing measurement results for the interpolation process and control sensors [39,40].

Our previous analysis led us to the conclusion that the quality of REMs created with the direct methods depends mainly on the following factors: the regularity of deployment of sensors, the distance between sensors, the propagation environment and the interpolation technique.

In the literature on the topic several methods of sensor placement are presented [41]: (a) simple random (b) systematic (grid-based), (c) stratified, where some regions are saturated more heavily than others and (d) a hybrid approach joining all the above. In [38] the authors describe the concept of deployment of sensors in zones with low saturation of sensors. The idea is to use a dedicated algorithm to identify the largest area without sensors in the existing network. In the next phase, another dedicated algorithm is applied to generate the best deployment of sensors for the identified area.

Our experience shows that the stratified approach to the deployment of sensors seems to be reasonable in diverse areas. In [39] we analysed the impact of the arrangement of sensors on the REM quality. We used the results of some field tests done for UHF range with different arrangements of sensors and various interpolation techniques to construct maps. In the next step, we calculated and compared RMSE for different sensor deployment. One of our conclusions was that the deployment of sensors plays an important role, particularly when the number of sensors is limited. We also showed that even a minor rearrangement of sensors can noticeably affect the map quality. In our experiment the rearrangement of 2 out of 13 sensors deployed on the area of 4 km² brought about a significant drop in the RMSE (by up to 2 dB). For the scenario with 13 sensors and a less favourable arrangement the RMSE reached 11.9 dB for NN, 10.95 dB for IDW p3 and 9.6 dB for Kriging, while for the most favourable arrangement of 13 sensors the RMSE reached 8.5 dB for NN, 8.7 dB for IDW p3 and 6.2 dB for Kriging.

Another essential issue for REM construction is the number of sensors in the network providing measurement results for the interpolation process and thus affecting the accuracy of maps. In most papers several dozens or hundreds of sensors per a few square kilometres are taken into account [43,44]. Although it is desirable to deploy as many sensors as possible, in real scenarios their number may be significantly smaller, e.g. in military networks with CRs.

In [40] we presented how the density of the sensor network affects the accuracy of REMs. In our research work we used measurement data from real field tests with 39 sensors deployed within the area of 4 km² (the size of the area similar to [43]). We analysed the results of the tests with various density of the sensor network applied for the interpolation process, namely 13, 20 and 26 sensors. It is worth noting that for each sensor network density we considered two tests with different arrangements of sensors. As the next step, REMs were created for the following interpolation techniques: NN, IDW and Kriging. Finally, in order to assess the quality of maps, the RMSE values were compared and analysed (similarly to [44]). Our general conclusion was that the growth of the density of the sensor network from 13 to 26 sensors brought about a noticeable improvement in the quality of REMs. We observed that for the most promising interpolation techniques the average RMSE values dropped from 8.7 dB to 6.3 dB for the Kriging and from 10 dB to 6.5 dB for the IDW p3 method respectively.

There are also new proposals aimed at obtaining higher accuracy of maps. In [42] the idea of the hybrid REM construction technique that combines direct and indirect methods was presented. This method seems promising, especially for networks with a limited number of sensors. Experiments confirmed that the quality of maps is much higher when compared to the results for direct methods.

III. SENSING METHODS

Sensing methods can be classified into several basic categories depending on the signal features used. The most common methods considered in the literature are energy detection, detection of cyclostationary features, the use of a matched filters as well as a wavelet transform.
A. Energy detection

Due to its low computational complexity and uncomplicated implementation the most common method used in the sensing process is Energy Detection (ED) [45, 48]. ED is also known as the semi-blind method since it does not require any prior knowledge about the PU signal. ED is also an incoherent method based on the assumption that the signal energy at the receiver’s location is greater than the energy of the noise. The detector decision rule is based on a comparison of the estimated value of signal energy with a detection threshold depending on the SNR value. As a result ED requires information on the spectral power density of noise, which makes it sensitive to the uncertainty of its estimation [74, 49]. For this reason the literature on the subject also analyses the usefulness of other techniques that do not require take this parameter into account. Most often, these methods use distinctive features of the useful signal to distinguish it from noise. However, such detection methods are not free of shortcomings as well.

B. Covariance detection

One of the methods insensitive to the uncertainty of spectral noise power density estimation is the CAV detector (covariance absolute value). This is a blind detection technique using time-space correlation to detect the signal. CAV is based on the difference in correlation between the received signal and the noise, as the values of signal and noise autocorrelation are usually different [50, 51]. In this case information on signal and/or noise levels is not required [52, 53]. Although the complexity of CAV is much higher than the one of ED, its effectiveness in low SNR cases deserves attention.

C. Eigenvalues of the covariance matrix

In order to detect the PU signal one can use a method based on eigenvalues of the covariance matrix, EGD (eigenvalue-based detection). Similarly to CAV, it is a blind detection method. Numerous types of this method have been proposed in the literature, such as MME (Maximum-Minimum Eigenvalue) or EME (Energy with Minimum Eigenvalue). The MME compares the detection threshold with the ratio of the maximum eigenvalue to the minimum eigenvalue of the covariance matrix. EME compares the detection threshold with the ratio of the average signal strength to the minimum eigenvalue of covariance matrix. In contrast to ED, the detection threshold is not based on spectral density of noise power since it is estimated on the basis of the number of signal samples, smoothing factor and probability of false alarm [54].

D. Cyclostationary detector

The cyclostationary detector (CD) exploits the fact that statistical parameters of modulated PU signals change periodically as a function of time [55]. The main advantage of this type of solutions is that they make it possible to distinguish whether the energy comes from a deterministic signal or from noise. In addition, due to the fact that the signals exhibit cyclostationary features for different shift values and different cyclic frequencies, it is also possible to distinguish the signals from different emitters. The detector of cyclostationary features is based on the use of the autocorrelation function and power spectral density. However, AWGN noise is a wide sense stationary process without correlation and that is why this method is resistant to interference and can be used at low SNR values. Nevertheless, it should be noted that CD is characterized by a high degree of computational complexity and a long detection time [56-59].

E. Matched filter detector

When using the methods presented above one assumes the lack or only partial information on the PU signal. In contrast, the detection using a matched filter is based on the assumption that the description of the PU signal in the time domain is completely known [60, 61]. This type of detection is a coherent one. The detector maximizes SNR in the presence of Gaussian noise. This effect is achieved as a result of an operation equivalent to a correlation of a known signal waveform with a received signal. The mutual correlation function is a measure of the statistical relationship of two different signals. The detector determines the value of the correlation of the signal received with the locally stored replica of the signal being detected. This is the optimal method of spectrum sensing [62], which compared to other methods provides greater efficiency and shorter detection time [63]. However, its basic disadvantage is the need to have full prior information on the parameters of the signal transmitted by PU. It should be noticed that various PU transmissions with different signal characteristics may occur in a given area. Thus, the considered detector would have to consist of numerous individual detectors designed for separate detection of individual signals. Therefore, the complexity of implementation of such a detector is impractical [64]. Another disadvantage of this method is the high energy consumption resulting from the fact that at the same time different algorithms must be performed in order to decide on the presence or absence of PU transmission.

F. Wavelet transforms

In [65] the use of wavelet transform (WD) is considered to detect discontinuities or edges in power spectral density (PSD) of a broadband channel that corresponds to a transition from an occupied band to an empty one or vice versa. By detecting irregularities in PSD it is possible to determine which sub-bands are not occupied. Spectrum sensing based on a wavelet detector is primarily aimed at rough estimation of the spectrum. The advantage of this method is the relatively short detection time. On the other hand, it is characterized by high computational complexity.

G. Hybrid methods

Each of the sensing methods described above has some advantages and disadvantages. Hence, studies on the effectiveness of hybrid sensing, which is a combination of different signal detection methods, have been undertaken.

Various proposals for HD architecture have already appeared in the literature, depending on the types of systems for which they would be dedicated [66-70]. HD methods have been precisely defined and described and researched in the PhD thesis [71]. The most commonly considered solution is a two-phase detector [72, 73], in which energy detection is used in the first phase. However, ED is sensitive to the uncertainty of the spectral noise power density estimation. For this reason, researchers consider combining ED with methods resistant to this uncertainty, which will increase the reliability of the detection. Therefore, in the second phase of HD the most commonly proposed methods are: cyclostationary feature detection, matched filter, covariance detector or detection using the eigenvalues of covariance matrix.
CONCLUSION

The article concerns EESA in CR systems. It presents the reasons for the emergence of such solutions and the reasons for building situation awareness in the frequency domain. The authors have thoroughly reviewed the literature. Based on this and their own research, they have identified the most promising methods for carrying out this task, both autonomously and cooperatively. They have also presented the potential of the REM, which is currently proposed in the military communications environment as a solution facilitating the coordination of activities of various entities using the spectrum. It is worth noting here that the usefulness of the REM is currently being investigated within NATO by the Research Task Group 069, in which the authors participate. The IST-146 project aims to evaluate the operational benefits for NATO in line with the Electromagnetic Spectrum Strategy and to evaluate the REM technology.

The article also discusses the pros and cons of a large group of spectrum sensing methods, which are a tool used in nearly all the EM systems and that allows ESM systems to increase their efficiency, i.e. improve spatial and temporal spectrum reuse and even pre-empt spectrum conflicts.

REFERENCES


