IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 22, 2019, accepted December 20, 2019, date of publication January 1, 2020, date of current version January 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2963512

Clustering Algorithm-Based Data Fusion Scheme for Robust Cooperative Spectrum Sensing

SHUNCHAO ZHANG[1], YONGHUA WANG[1,2], (Member, IEEE), PIN WAN[1,3], JIAWEI ZHUANG[1], YONGWEI ZHANG[1], AND YI LI[1]

[1]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[2]State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[3]Hubei Key Laboratory of Intelligent Wireless Communications, South-Central University for Nationalities, Wuhan 430074, China

Corresponding author: Yonghua Wang (wangyonghua@gdut.edu.cn)

This work was supported in part by the Special Funds from the National Natural Science Foundation of China under Grant 61971147, in part by the Central Finance to Support the Development of Local Universities under Grant 400170044 and Grant 400180004, in part by the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences under Grant 20180106, in part by the Foundation of Key Laboratory of Machine Intelligence and Advanced Computing of the Ministry of Education under Grant MSC-201706A, in part by the School-Enterprise Collaborative Education Project of Guangdong Province under Grant PROJ1007512221732966400, in part by the Foundation of National and Local Joint Engineering Research Center of Intelligent Manufacturing Cyber-Physical Systems and Guangdong Provincial Key Laboratory of Cyber-Physical Systems under Grant 008, and in part by the Higher Education Quality Projects of Guangdong Province and Guangdong University of Technology.

ABSTRACT In a centralized cooperative spectrum sensing (CSS) system, it is vulnerable to malicious users (MUs) sending fraudulent sensing data, which can severely degrade the performance of CSS system. To solve this problem, we propose sensing data fusion schemes based on K-medoids and Mean-shift clustering algorithms to resist the MUs sending fraudulent sensing data in this paper. The cognitive users (CUs) send their local energy vector (EVs) to the fusion center which fuses these EVs as an EV with robustness by the proposed data fusion method. Specifically, this method takes a Medoids of all EVs as an initial value and searches for a high-density EV by iteratively as a representative statistical feature which is robust to malicious EVs from MUs. It does not need to distinguish MUs from CUs in the whole CSS process and considers constraints imposed by the CSS system such as the lack of information of PU and the number of MUs. Furthermore, we propose a global decision framework based on fast K-medoids or Mean-shift clustering algorithm, which is unaware of the distributions of primary user (PU) signal and environment noise. It is worth noting that this framework can avoid the derivation of threshold. The simulation results reflect the robustness of our proposed CSS scheme.

INDEX TERMS Cognitive radio, robust cooperative spectrum sensing, sensing data fusion, K-medoids clustering algorithm, Mean-shift clustering algorithm.

I. INTRODUCTION

Cognitive radio is promising technology to boost utilization and alleviate the spectrum shortage. The basic ideal of cognitive radio (CR) is that licensed spectrum bands are allowed to be accessed by cognitive users (CUs) when primary users (PUs) are absent [1]–[4]. Under this regulation, the spectrum sensing is a crucial technique within CR, which senses the spectrum band to find spectrum holes. There are many single CU spectrum sensing methods such as energy detection, matched filter detection, and cyclostationary feature detection [5]–[7]. However, the detection performance of these spectrum sensing methods is susceptible to the impact of noise, hidden terminal, pass loss, shadowing and multi-path fading, which may cause incorrect sensing results provided by a single CU [8]. To solve this problem, cooperative spectrum sensing (CSS) methods have been attracted a lot of interesting, which have been verified to be more reliable than single CU spectrum sensing.

In the CSS, each CU independently collects sensing data or performs local spectrum sensing for a particular spectrum band. These CUs send their data or results to the fusion center (FC) periodically. The FC receives sensing data or results from each CU, fuses these data or results by

The associate editor coordinating the review of this manuscript and approving it for publication was Julien Le Kernec.

a fusion mechanism, and makes a global decision. However, due to the openness of low layers protocol stacks, the CSS is vulnerable to many severe attacks of malicious users (MUs). Spectrum sensing data falsification (SSDF) attack is considered in the paper. MUs tamper the locally collected sensing data and send these data to mislead the FC to make a wrong decision, which may seriously degrade the performance of CSS system.

To detect the MUs and eliminate the damage of its attacks in CSS, many techniques have been developed [9]–[11]. However, such existing techniques of against MUs are limit to some unrealistic assumptions that are easily violated in future or realistic spectrum sensing: 1) The attack ways are assumed to be identical and fixed in [9]. 2) The underlying distribution of PU signal and noise are assumed to be known [10]. Both of assumptions are easily violated in realistic spectrum sensing. 3) Most of existing techniques first identify MUs. Then, these reports of the MUs are prevented by a hard fusion mechanism [9], [11]. This may lead to the detection performance of CSS system degradation as the reports of MUs may be normal with a possible.

In this paper, we consider a centralized CSS system with soft fusion mechanism, where each CU independently collects sensing data and calculates energy vector (EV) for a particular spectrum band. Then, the CUs send their EVs to the FC periodically. The FC receives EV from each CU, fuses these EVs by a sensing data fusion scheme, and makes a global decision. In order to resist the attack of MUs, two sensing data fusion methods of soft fusion mechanism based on clustering algorithm are proposed. In these sensing data fusion methods, they fuse EVs from CUs as an EV with robustness for representing the status of PU. It is noted that the number of MUs is less than the honest users (HUs) should be assumed. We also propose a CSS framework which considers spectrum sensing as a two-class classification problem in clustering algorithm. The proposed CSS scheme does not need any prior information related to the attack strategy of MUs. Contributions of this paper can be briefly summarized as follows.

1) Two sensing data fusion methods are proposed. One method is based on K-medoids clustering algorithm, namely DF-medoids method. Another method is based on Mean-shift and K-medoids clustering algorithms, namely DFMS-medoids method.
2) In order to avoid threshold derivation, two robust CSS methods based on fast K-medoids and Mean-shift clustering algorithms are developed, respectively.
3) In simulation section, the performance of DF-medoids and DFMS-medoids methods is verified and the detection performance of these robust CSS methods is analyzed. The simulation results show that the proposed robust CSS methods improve the performance of spectrum when MUs attack the CSS system.

The rest of this paper is organized as follows. Section II introduces the related works. Section III introduces the system model of CSS. Two sensing data fusion methods are proposed in Section IV, which are DF-medoids and DFMS-medoids, respectively. Based on fast K-medoids and Mean-shift clustering algorithms, two achieving spectrum sensing methods are proposed in Section V. Section VI simulates the proposed rubost CSS methods and proves that these methods can improve the robustness of spectrum sensing. Section VII summarizes the full text.

II. RELATED WORKS

MUs attack is one of the threats for CSS system in CR networks, because they may cause the CSS system instability, such as increasing the probability of a false alarm and decreasing the the probability of detection. Many approaches have been reported to defense MUs attack based on weight assigned in recent literature. In [10], according to the difference between the energy of CU and the average energy of all CUs, each CU was given a Kullback Leibler divergence (KLD) score to assign weights for the sensing reports of CUs before sensing data fusion at the FC. The MUs were assigned low weights, whereas the HUs were assigned high weights. In [12], based on the assistance of trusted nodes, a reputation-based CSS method was proposed. The CUs were divided into three states, i.e., reliable, pending, and discarded. The decision of each CU was weighted by their reputation. If the CU was divided into discard, it was removed in the CSS. Generally, the defense framework includes two processes, i.e., defense reference establishment, and reputation evaluation.

There are many outlier detection techniques based robust spectrum sensing schemes proposed in [13]–[16]. In [13], a modified version of Grubb test was used for detection of a single outlier in a normally distributed data. In [14], the authors investigated outlier detection techniques to identify the MUs, in which outlier factors were assigned to each CU for distinguishing MUs from CUs. However, this scheme needs the maximum number of the MUs. In [15], the outlier detection technique was introduced to pre-filter abnormal sensing data. Then, a trust factor utilized as the weight to be given each CU. This method mere considers simple attacks such as 'always yes' or 'always no' without coping with more reality MUs attacks. In [16], to mitigate SSDF attack, the support vector data description (SVDD) was applied in sensing procedure. The SVDD algorithm distinguishes MUs from HUs and omits outliers from global decision process. The SVDD algorithm is a kind of one-class classification which considers sensing data as a target class and the rest of sensing data as the outliers. However, the method mere considers sample MUs ('always yes' or 'always no').

Many different metrics are introduced to distinguish MUs and HUs [11], [17], [18]. In [11], based on double-sided neighbor distance and frequency check, a robust CSS was proposed to detect MUs. In [17], the maximum mean discrepancy (MMD) was used as a metric of distance for the sensing reports to distinguish MUs and HUs. The genuine reports from MUs may still be used for sensing data fusion.

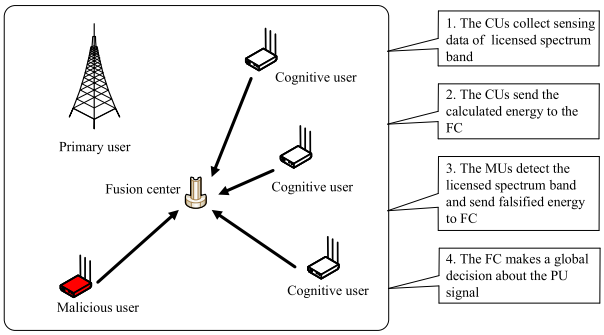


FIGURE 1. The centralized cooperative spectrum sensing scenario.

A Kruskal-Wallis test based MUs detection scheme was proposed in [18]. However, these methods are built upon the assumption that a small fraction of CUs are MUs.

To the best of our knowledge, the data analysis ideal based on clustering algorithm for sensing data fusion still remain open issues that require further investigation.

III. SYSTEM MODEL

In this section, the system model is given and introduced. Furthermore, the attack model used in this paper is given.

A. SYSTEM MODEL OF COGNITIVE RADIO NETWORK

The centralized CSS scenario is illustrated in Fig. 1. In this paper, consider C CUs with multi-antenna take part in CSS related to a licensed spectrum band. In each time slot, each CU collects sensing data from a specified channel, calculates energy, and sends its local energy to a FC. Note that all CUs need to be time-synchronized. Then, the FC makes a global decision and informs CUs whether the licensed spectrum can be accessed. The communication channels between CUs and the FC are assumed to be dedicated and reliable.

We assume that there are M MUs and the rest of CUs are HUs in the CSS system model. It is reasonable to assume that $M \ll C$, because studying a full of MUs network is meaningless. The FC does not know any information related to MUs, such as the number of MUs and the identities of MUs. For the MUs, we assume that they collect sensing data, make local decisions by their local detection method (such as energy detection), and send falsified energy to the FC for misleading the FC to make a wrong global decision.

Assume that the PU is either idle or active throughout the sensing period. Thus, the received signal by the lth antenna of the ith CU can be formulated as a binary hypothesis [19], such that

$$\begin{cases} \mathcal{H}_0 : x_i^l(k) = n_i^l(k), \\ \mathcal{H}_1 : x_i^l(k) = h_i^l(k)p(k) + n_i^l(k), \end{cases} \tag{1}$$

where $k = 1, 2, \ldots, N$, N is the number of sampling points, $l = 1, 2, \ldots, L$, L represents the number of antenna, $i = 1, 2, \ldots, C$, C represents the number of CUs. Some necessary assumptions are defined as

1) $x_i^l(k)$ is the received signal at the l antenna of the ith CU in the time k, $p(k)$ is the signal from PU, $h_i^l(k)$ is the channel gain between the l antenna of the ith CU and PU.
2) $n_i^l(k)$ represents the Gaussian white noise (WGN) with zero mean and a variance σ^2.
3) Each CU has certain computing ability and can simply perform data processing.

According to (1), in the condition of $\mathcal{H}_0$ and $\mathcal{H}_1$, the observed energy at the lth antenna of the ith CU can be represented as

$$\begin{cases} \mathcal{H}_0 : e_i^l = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} |n_i^l(k)|^2, \\ \mathcal{H}_1 : e_i^l = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} |h_i^l(k)p(k) + n_i^l(k)|^2. \end{cases} \tag{2}$$

Therefore, the EV of ith CU can be represented as

$$e_i = [e_i^1, e_i^2, \ldots, e_i^L]^{\mathsf{T}}, \tag{3}$$

where $e_i \in \mathbb{R}^{L\times 1}$.

In spectrum sensing, the detection probability, missed detection probability, and the false alarm probability are three key index to evaluate the performance of spectrum sensing schemes [20]. The detection probability is defined as

$$P_d = P[\hat{\mathcal{H}}_1|\mathcal{H}_1], \tag{4}$$

where $\hat{\mathcal{H}}_1$ is the measured state of the PU that is using authorized spectrum, $\mathcal{H}_1$ represents the actual status of the PU that is busy. Furthermore, the missed detection probability is given by

$$P_m = P[\hat{\mathcal{H}}_0|\mathcal{H}_1], \tag{5}$$

where $\hat{\mathcal{H}}_0$ is the measured state of the PU signal that is absent. Moreover, the form of false alarm probability is given as

$$P_f = P[\hat{\mathcal{H}}_1|\mathcal{H}_0], \tag{6}$$

where $\mathcal{H}_0$ represents the actual state of the PU that is absent.

B. ATTACK MODEL

The presence of MUs may severely degrade the detection performance of CSS system [21]. There are two purposes for MUs to attack the CSS system, which are to destroy the spectrum sensing system and obtain its own benefits, respectively [22], [23]. If the target of the MUs is to destroy the CSS system, when the PU is using the licensed spectrum, the MUs send the idle information of the PU to the FC, which cause the FC to make a mistake decision. Furthermore, The normal communication of the PU is interfered, and the trust between the PU and CUs are distrusted. If the target of MUs is to obtain its own benefits which is to monopolize the spectrum through abnormal ways, MUs send the information that the PU is using the licensed spectrum when the authorized spectrum is actually idle to the FC. Thereby, the FC considers that the PU

is active. Unfortunately, the HUs cannot access the licensed spectrum.

In this paper, the HUs send the original local EVs to the FC. However, the MUs send the falsified EVs to the FC. For example, the MUs find that the PU is absent by local judgment when the PU is absent. Then, they send falsified and higher EVs than real EVs to the FC. If the FC incorrectly makes a judgment about the states of PU, the licensed spectrum can be accessed by the MUs. When the PU is active, the MUs find that the PU exists by its local judgment, they will send a falsified and lower EVs than real EVs to the FC. If the FC incorrectly believes that the licensed spectrum is idle, allowing the CUs access, which will interfere the normal communication of the PU.

IV. SENSING DATA FUSION BASED ON CLUSTERING ALGORITHM

In this section, to defense MUs attack, we propose two sensing data fusion methods based on clustering algorithm.

The mean fusion method is very susceptible to interference by the falsified data from MUs, which causes the merged data to not accurately reflect real states of the PU. Inspired by [24], we propose a sensing data fusion method based on K-medoids clustering algorithm, which is called DF-medoids method. The medoids is more robust than the mean. Furthermore, we further propose a sensing data fusion approach based on DF-medoids method and Mean-shift clustering algorithm, namely DFMS-meioids, which uses the medoids as an initial value. Then, it performs sensing data fusion by iterative. It is noted that FC does not need to find out MUs in the whole data fusion process. The sensing data fusion methods proposed in this paper only needs to fuse the EVs received by the FC as an EV.

After each CU uploads its EV to the FC, let $\overline{E}$ be a set of EVs, i.e.,

$$\overline{E} = \{e_1, e_2, \ldots, e_C\}. \tag{7}$$

It is noted that $\overline{E}$ includes the honest EV from HUs and the falsified EV from MUs. Define a ideal EV, which is the average EV calculated by EVs from all HUs, such that

$$e_{ideal} = \frac{1}{H} \sum_{i \in \text{HUs}} e_i, e_i \in \overline{E}. \tag{8}$$

The mean of EVs from all CUs is calculated by

$$e_{means} = \frac{1}{C} \sum_{i=1}^{C} e_i, e_i \in \overline{E}. \tag{9}$$

From (8) and (9), we can conclude that the e_{means} will deviate the e_{ideal} ($e_{means} \neq e_{ideal}$) when MUs send falsified sensing data. Therefore, the e_{means} can not reflect the real states of PU, the mistaken decision may be made by FC. Thus, a effect sensing data fusion method is necessary, two sensing data method are proposed in the following section, respectively.

A. SENSING DATA FUSION BASED ON K-MEDOIDS CLUSTERING ALGORITHM

In this subsection, a DF-medoids method for sensing data fusion is proposed. Different with the mean of all EVs, the K-medoids [26] is used to find a existing EV in the set $\overline{E}$, which is called medoids. The medoids has this principle that the sum of distances between the medoids and each EV in $\overline{E}$ is shortest.

Define the object function $\mathcal{D}(\cdot)$ of DF-medoids as

$$\mathcal{D}(e_i) = \frac{1}{C} \sum_{i=1}^{C} \sum_{j=1}^{C} \|e_i - e_j\|. \tag{10}$$

By minimizing the objective function $\mathcal{D}(e_i)$, we can obtain a robust medoids of all EVs from CUs, such that

$$e_{medoids} = \arg\min_{e_i \in \overline{E}} \mathcal{D}(e_i). \tag{11}$$

Then, the $e_{medoids}$ is used as a feature vector for CSS.

B. SENSING DATA FUSION BASED ON MEAN-SHIFT CLUSTERING ALGORITHM

The Mean-shift algorithm is a density-based clustering algorithm, which is robust for falsified data [27]. In this subsection, the Mean-shift clustering algorithm is studied for sensing data fusion. Based on the Mean-shift clustering algorithm, a DFMS-medoids fusion algorithm is proposed, which obtain the fused data iteratively by using medoids $e_{medoids}$ as the initial value.

Based on (11), the $e_{medodis}$ is used as the initialized EV $e_{meanshit} = e_{medodis}$. The DFMS-medoids method searches for a center with relatively large density and updates a new center based on the original center.

Define a neighborhood $S_h(e_i)$ with $e_{meanshit}$ as the center and r as the bandwidth, such as

$$S_h(e_i) = \{y | (y - e_{meanshit})(y - e_{meanshit})^{\mathsf{T}} < r\}, \tag{12}$$

where

$$r = \frac{1}{C} \sum_{i=1}^{C} \|e_i - e_{meanshit}\|. \tag{13}$$

By calculating the distance between the $e_i, e_i \in S_h$ with the $e_{meanshit}$, we can get the mean-shift vector u by

$$u = \frac{1}{B} \sum_{e_i \in S_h} (e_i - e_{meanshit}), \tag{14}$$

where B represents the number of EVs in S_h.

According to the mean-shift vector calculated by (14), the new center $e_{meanshit}$ can be obtained by

$$e_{meanshit} = e_{meanshit} + u. \tag{15}$$

Then, the $e_{meanshit}$ is used as a feature vector for spectrum sensing.

The data fusion procedure based on Mean-shift is shown in **Algorithm 1**.

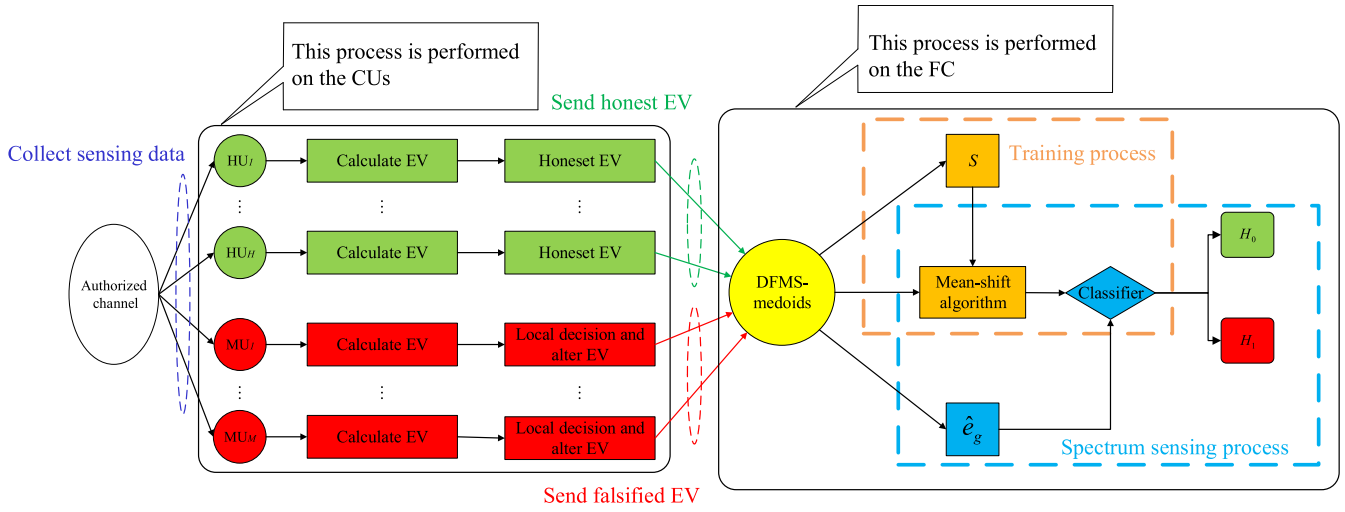


FIGURE 2. The proposed CSS framework based on clustering algorithm.

Algorithm 1 Data Fusion Method Based on Mean-Shift Clustering Algorithm

Step 1: Input $\overline{E} = \{e_1, e_2, \ldots, e_C\}$ from CUs.
Step 2: Let $e_{meanshit} = e_{medoids}$.
Step 3: Calculate the bandwidth r by (13).
Step 4: Calculate the neighborhood S_h with $e_{meanshit}$ as the center and r as the bandwidth.
Step 5: Calculate the mean-shift vector u by (14).
Step 6: Update $e_{meanshit}$ by (15).
Step 7: If the algorithm converges, then go the Step 7; otherwise, go to Step 3.
Step 8: Output $e_{meanshit}$.

V. ACHIEVING COOPERATIVE SPECTRUM SENSING BASED ON CLUSTERING ALGORITHMS

In spectrum sensing, samples mere need to be clustered into two classes [28], [29] by clustering algorithm in unsupervised learning. In this section, fast K-medoids and Mean-shift clustering algorithms are used to achieve spectrum sensing.

The detail flow of CSS based on clustering algorithm is shown in Fig. 2. It is divided into two parts, i.e., the training part, and the spectrum sensing part. In the training part, the CUs observe particular authorized spectrum, collect sensing data, and send their EVs to FC which uses data fusion method to obtain enough $e_g, g \in \{means, medoids, meanshift\}$. Assume that these e_g contain two status of PU. Then, these e_g can be clustered and trained by clustering algorithm. In spectrum sensing part, the CUs collect sensing data from a particular authorized spectrum in which the status of PU is unknown and send these data to the FC. Then, the FC performs data fusion by proposed data fusion method to obtain $\hat{e}_g$. Finally, the $\hat{e}_g$ is used as a input for the classifier which will give a decision about the status of PU.

By observing the licensed spectrum, the FC can get enough samples which are from the CUs after data fusion. Denote a set S which include all samples as

$$S = \{e_g^1, e_g^2, \ldots, e_g^J\}, \tag{16}$$

where $g \in \{means, medoids, meanshift\}$, e_g^j is the jth EV after data fusion. J represents the number of training EVs.

In CSS, the set S is clustered into two subsets. Denote the $\mathbb{S}_k$ is the kth class, such that

$$\mathbb{S}_k = \{e_g^j | e_g^j \in \text{Cluster } k, \forall j\}, \tag{17}$$

where $k = 1, 2$. Then clustering algorithms are used to cluster these samples into two classes.

A. CSS BASED ON FAST K-MEDOIDS CLUSTERING ALGORITHM

In previous work [28], [30], K-means clustering algorithm is introduced to achieve spectrum sensing. Unfortunately, K-means clustering algorithm is sensitive to outlier samples although it is quite efficient in the computational time. Thus, K-medoids clustering algorithm is introduced, which use the medoids instead of means. The medoids are the samples in its class, which are less sensitive to outlier samples. However, comparing with K-means clustering algorithm, K-medoids clustering algorithm is more complicated. Hence, a fast K-medoids clustering algorithm is introduced in this paper, which runs like the K-means clustering algorithm with the robust for outlier samples.

The dissimilarity measures between sample j and q is calculated by Euclidean distance, such that

$$d_{jq} = \|e_g^j - e_g^q\|, \; j, q = 1, 2, \cdots, J. \tag{18}$$

Then, a distance matrix can be obtained as

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1J} \\ d_{21} & d_{22} & \cdots & d_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ d_{J1} & d_{J2} & \cdots & d_{JJ} \end{bmatrix}, \tag{19}$$

where $\mathbf{D} \in \mathbb{R}^{J\times J}$, which is used for finding new medoids at each iterative step.

For each class $\mathbb{S}_k$, the representative medoids should be found, which has the smallest sum of distances to each sample in its class. The index of the medoids a_k is found by

$$a_k = \arg\min_j \sum_{j,q\in\mathbb{I}_k} d_{jq}, \tag{20}$$

where $\mathbb{I}_k$ is a set that contains the samples index of $\mathbb{S}_k$ in S. Then, the medoids Ψ_k is updated by $\Psi_k = e_g^{a_k}$.

The fast K-medoids clustering algorithm uses the Euclidean distance as a object function $\Theta(\cdot)$, such that

$$\Theta(\mathbb{I}_1, \mathbb{I}_2, \Psi_1, \Psi_2) = \sum_{k=1}^{2} \sum_{j,q\in\mathbb{I}_k} d_{jq}. \tag{21}$$

The object of the fast K-medoids clustering algorithm is to optimal the object function, i.e.,

$$\min_{(\mathbb{I}_1,\mathbb{I}_2,\Psi_1,\Psi_2)} \Theta(\mathbb{I}_1, \mathbb{I}_2, \Psi_1, \Psi_2). \tag{22}$$

The training procedure based on fast K-medoids clustering algorithm is shown in **Algorithm 2**.

Algorithm 2 Training Procedure Based on Fast K-Medoids Clustering Algorithm

Step 1: Input training data $S = \{e_g^1, e_g^2, \ldots, e_g^J\}$.
Step 2: Calculate distance matrix $\mathbf{D}$ by (19).
Step 3: Initialize medoids Ψ_1 and Ψ_2.
Step 4: Find two new medoids Ψ_1 and Ψ_2 based on the distance matrix $\mathbf{D}$, which are the samples minimizing the total distance to other samples in its classes.
Step 5: Assign each samples to the nearest medoid and obtain classes $\mathbb{S}_1$ and $\mathbb{S}_2$.
Step 6: Calculate the Θ by (21). If the Θ is not changed, then go the Step 7; otherwise, go to Step 4.
Step 7: Output Ψ_1 and Ψ_2.

B. CSS BASED ON MEAN-SHIFT CLUSTERING ALGORITHM

Mean-shift clustering algorithm [27], [31] is a hill climbing algorithm based on density estimation, which can be used for clustering, image segmentation, and tracking. In this section, the Mean-shift clustering algorithm is used to cluster EVs after data fusion by using the proposed sensing data fusion methods.

For each subclass $\mathbb{S}_k$, it has corresponding center which is defined as

$$\Psi_k = \frac{1}{num(\mathbb{S}_k)} \sum_{e_g^j\in\mathbb{S}_k} e_g^j, \tag{23}$$

where $num(\mathbb{S}_k)$ denotes the number of e_g^j belong to the class $\mathbb{S}_k$. Denote a neighborhood $\mathcal{S}_k(e_g^j)$ with Ψ_k as the center and r_k as the bandwidth, such that

$$\mathcal{S}_k(e_g^j) = \{y|(y-\Psi_k)(y-\Psi_k)^{\mathsf{T}} < r_k\}, \tag{24}$$

where

$$r_k = \frac{1}{num(\mathbb{S}_k)} \sum_{e_g^j\in\mathbb{S}_k} \|e_g^j - \Psi_k\|. \tag{25}$$

By calculating the distance between the e_g^j, $e_g^j \in \mathcal{S}_k$ with the Ψ_k, the mean-shift vector u_k is obtain as

$$u_k = \frac{1}{num(\mathcal{S}_k)} \sum_{e_g^j\in\mathcal{S}_k} (e_g^j - \Psi_k). \tag{26}$$

Then, the new Ψ_k can be updated by

$$\Psi_k = \Psi_k + u_k. \tag{27}$$

The training procedure based on Mean-shift clustering algorithm is shown **Algorithm 3**.

Algorithm 3 Training Procedure Based on Mean-Shift Clustering Algorithm

Step 1: Input training data $S = \{e_g^1, e_g^2, \ldots, e_g^J\}$.
Step 2: Initialize Ψ_1 and Ψ_2.
Step 3: Assign each sample to the nearest Ψ_1 and Ψ_2 and obtain the new classes $\mathbb{S}_1$ and $\mathbb{S}_2$.
Step 4: Calculate the bandwidth r_k by (25).
Step 5: Calculate the neighborhood $\mathcal{S}_k$ with Ψ_k as the center and r_k as the bandwidth.
Step 6: Calculate mean-shift vector u_k by (26).
Step 7: Update Ψ_k by (27).
Step 8: If the algorithm converges, then go the Step 9; otherwise, go back to Step 3.
Step 9: Output Ψ_1 and Ψ_2.

C. ACHIEVING CSS BASED ON TRAINED CLASSIFIER

After training, we can obtain a classifier to achieve spectrum sensing, the specific form is given as

$$Classifier(\hat{e}_g) = \frac{\|\hat{e}_g - \Psi_1\|}{\|\hat{e}_g - \Psi_2\|}, \tag{28}$$

where $\hat{e}_g$ represents the EV which is processed by the DFMS method. It is noted that $\hat{e}_g$ is unknown about the state of PU. If $Classifier(\hat{e}_g) > \varepsilon$ indicates that PU is using the authorized spectrum. Otherwise, the authorized spectrum can be accessed by CUs.

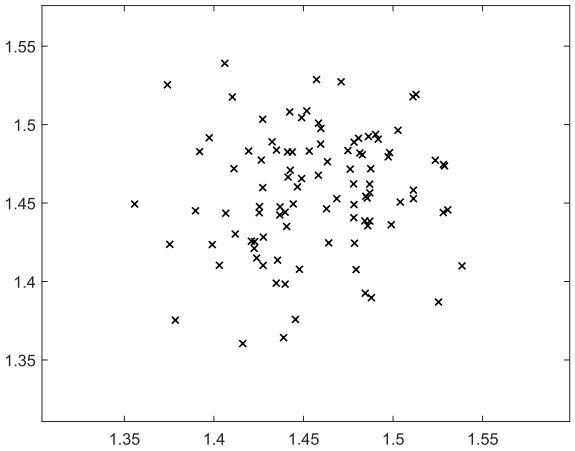


FIGURE 3. EVs from all CUs.

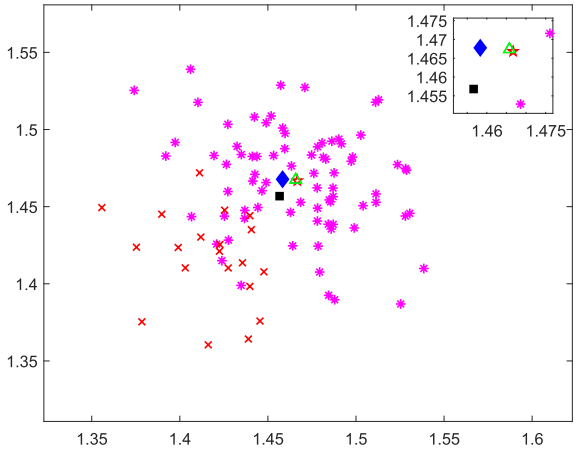


FIGURE 4. Data fusion based on Means, DF-medoids, and DFMS-medoids.

VI. SIMULATION

In this section, the performance of the proposed robust CSS methods is illustrated through computer simulation. The simulation platform is Matlab. The AM signal is chosen as the PU signal $p(k)$ in this paper. First, the performance of the developed sensing data fusion methods is presented. Then, the classification effect of clustering algorithms are analyzed. Finally, the performance of the proposed robust CSS method is given.

A. THE PERFORMANCE ANALYSIS OF THE DATA FUSION METHOD

The performance analysis of the DFMS method is presented in this section. Let the number of CUs be $C = 100$, the number of EVs received by FC be $REV = 100$, the number of honest EVs be $HEV = 80$ and the number of falsified EVs be $MEV = 20$.

Remark: The robust CSS methods proposed in this paper is based on the received EVs, which can not care related to the attack mode of MUs. It means that the proposed robust CSS method can defend "always no", "always yes", and smart attack from MUs.

At SNR $= -10$ dB, $N = 1000$, Fig. 3 shows that these 'x' points are the EVs from every HU and MU. As shown in Fig. 4, these 'x' points represent falsified local EVs from MUs, these '*' points represent honest local EVs from HUs. The pentagram represents the ideal mean value e_{ideal}, which is calculated by using honest EVs. The square point indicates the mean value e_{means} of EVs from HUs and MUs. The diamond represents the medoids $e_{medoids}$ of all EVs, which is calculated by DF-medoids method. The triangle represents the mean value $e_{meanshit}$ calculated by DFMS-medoids method. From Fig. 4, we can conclude that the DFMS-medoids method has better robustness than other data fusion method, such as Means and DF-medoids.

In the CRN, when MUs find that the PU is using the authorized spectrum, the MUs will report low EVs to FC for disturb the correct decision made by FC. When MUs find that the PU is not using the authorized spectrum, the MUs will report high EVs to FC which may make a incorrect decision related to the PU signal. Then, it informs the CUs that the authorized spectrum is busy and can not be accessed.

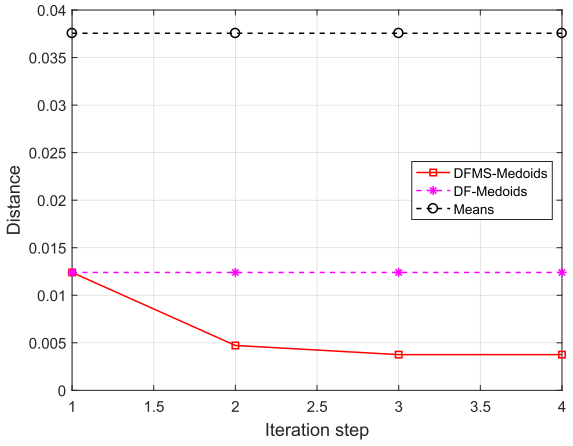


FIGURE 5. The distance between fused EV and ideal EV when MUs send many lower EVs.

As shown in Figs. 5 and 6, DFMS-medoids method has the short distance between $e_{meanshift}$ with the idea EV e_{ideal} comparing with Means and DF-medoids methods, which indicates the DFMS-medoids method can effect restrain the attack from MUs. In Figs. 5 and 6, the parameters are SNR $= -8$ dB, $HEV = 80$, $MEV = 20$, $L = 2$, and $N = 1000$.

B. CLASSIFICATION EFFECT ANALYSIS OF CLUSTERING ALGORITHM

In this section, we will present the clustering effect of fast K-medoids and Mean-shift clustering algorithms. The size of training set S is $J = 10000$. It is noted that the S include two states of PU, i.e., the PU is active, and the PU is absent.

Fig. 7 shows the unclassified samples collected at SNR $= -10$ dB, $HEV = 80$, $MEV = 20$, $L = 2$, and $N = 1000$.

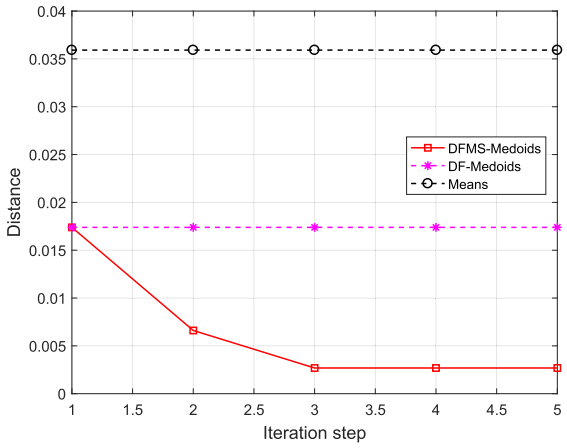


FIGURE 6. The distance between fused EV and ideal EV when MUs send many higher EVs.

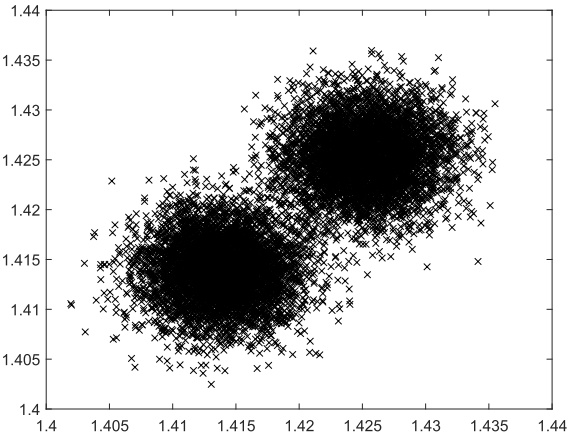


FIGURE 7. Unclassified samples.

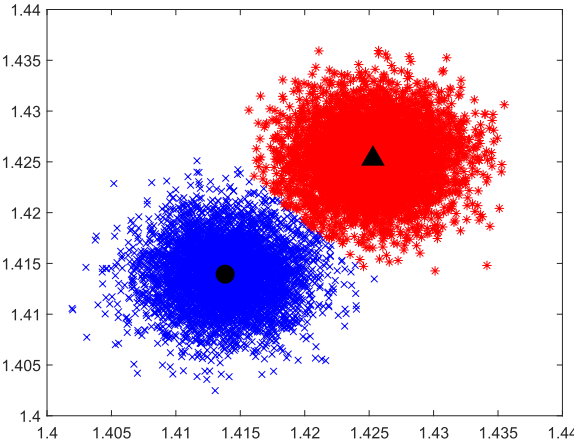


FIGURE 8. Classified samples by Mean-shift algorithm.

Fig. 8 shows the effect of unclassified samples clustered by the Mean-shift clustering algorithm. These red '*' dots are considered that the PU is active. These blue 'x' dots are considered that the PU is absent.

Fig. 9 displays the effect of unclassified samples clustered by the fast K-medoids clustering algorithm.

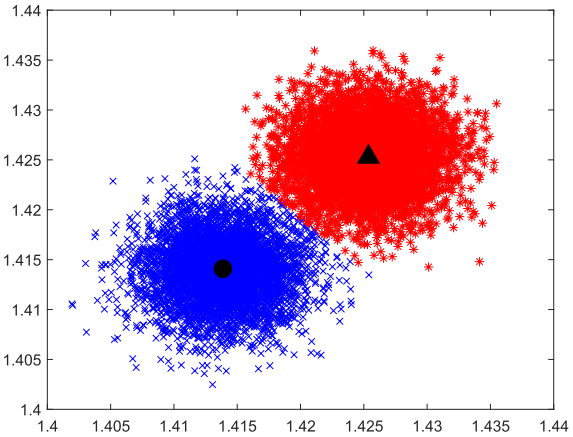


FIGURE 9. Classified samples by fast K-medoids algorithm.

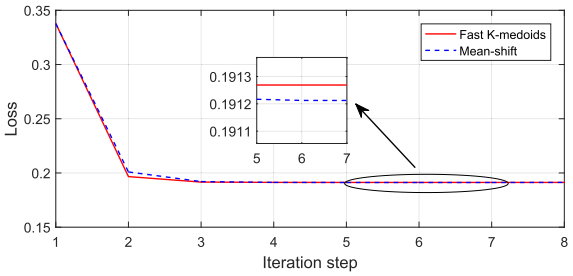


FIGURE 10. Loss of the fast K-medoids and Mean-shift algorithms.

It is difficult to distinguish the performance of the two clustering algorithms from Figs. 8 and 9. In Fig. 10, the loss function is presented. Comparing with the loss value of fast K-medoids and Mean-shift clustering algorithms, we can conclude that the Mean-shift has the better performance in this case. Thus, in the subsequent sections, Mean-shift clustering algorithm is chosen to achieve spectrum sensing.

C. THE PERFORMANCE ANALYSIS OF ROBUST CSS

In this section, the performance of spectrum sensing is verified. As shown in Figs. 11 and 12, the performance of different methods are presented. The parameters are set as $N = 1000$, $L = 2$, SNR $= -15$ dB.

In Fig. 11, the MUs send the lower EVs for interfering the FC to make correct decision. In this case, the detection probability will decrease under a certain false alarm probability. From Fig. 11, we can observe that the DFMS-medoids and DF-medoids based robust CSS has better performance than Means method. Furthermore, the DFMS-medoids method has the best performance in these method.

In Fig. 12, the MUs send the higher EVs for misleading the FC to make a incorrect decision that the licensed spectrum is busy when is actually idle. Thus, MUs can avoid competing with other HUs and access the licensed spectrum by itselfs. In this case, the false alarm probability will increase under a certain missed detection probability. From Fig. 12, we can see that the DFMS-medoids and DF-medoids based robust CSS has better performance than Means. Furthermore,

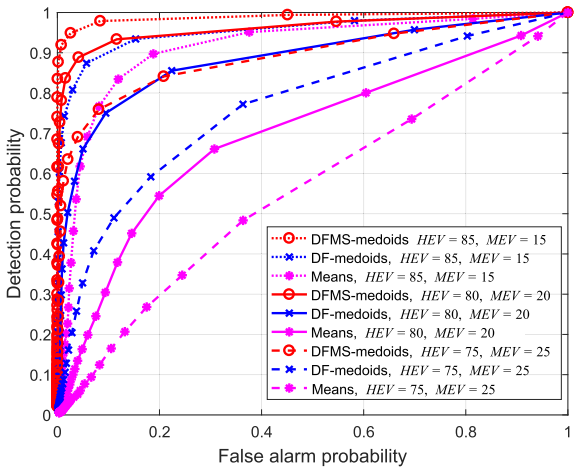


FIGURE 11. Comparison of ROC in different methods when facing the lower EVs send by MUs.

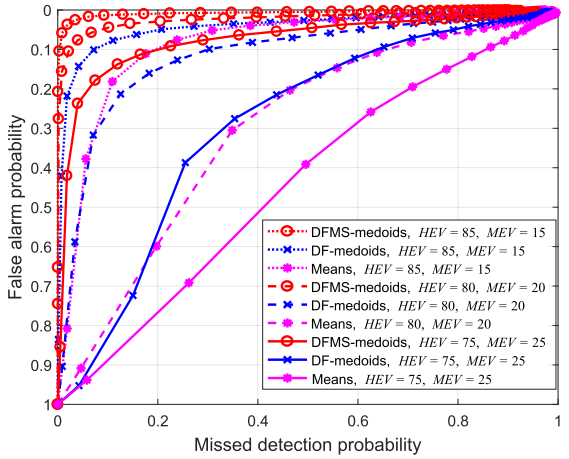


FIGURE 12. Comparison of ROC in different methods when facing the higher EVs send by MUs.

DFMS-medoids method has the best performance in these methods.

In general, the robust CSS method which combined DFMS-medoids and Mean-shift clustering algorithm has better sensing performance, since DFMS-medoids data fusion method is more robust than DF-mediods and Means method and the clustering effectiveness of Mean-shift is better than K-medoids. From Figs. 5, 6, and 10–12, it is easy to verity this conclusion.

VII. CONCLUSION

In this paper, we propose DF-medoids and DFMS-medoids data fusion methods to defend against SSDF attack for CSS in CRNs. These data fusion methods can effectively suppress malicious data to impact on decision made by the FC. Unlike existing defense schemes that need to find who are MUs and prohibit MUs from joining data fusion. Our methods can directly fuse sensing data from all CUs and the results are robust. In spectrum sensing, to avoid deriving the threshold, the fast K-medoids and Mean-shift clustering algorithms are adopted for obtaining a classifier to achieve spectrum sensing. In simulation, the result shows that the proposed CSS methods are robust when MUs attack the CSS system.

REFERENCES

[1] Y. Wang, Z. Ye, P. Wan, and J. Zhao, "A survey of dynamic spectrum allocation based on reinforcement learning algorithms in cognitive radio networks," *Artif. Intell. Rev.*, vol. 51, no. 3, pp. 493–506, Mar. 2019.

[2] X. Huang, H. Zhai, and Y. Fang, "Robust cooperative routing protocol in mobile wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5278–5285, Dec. 2008.

[3] L. Xiao, Y. Li, J. Liu, and Y. Zhao, "Power control with reinforcement learning in cooperative cognitive radio networks against jamming," *J. Supercomput.*, vol. 71, no. 9, pp. 3237–3257, Sep. 2015.

[4] L. Xiao, J. Liu, Q. Li, N. B. Mandayam, and H. V. Poor, "User-centric view of jamming games in cognitive radio networks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2578–2590, Dec. 2015.

[5] M. A. Fouda, A. S. T. Eldien, and H. A. Mansour, "FPGA based energy detection spectrum sensing for cognitive radios under noise uncertainty," in *Proc. Int. Conf. Comput. Eng. Syst.*, Dec. 2017, pp. 584–591.

[6] F. Salahdine, H. E. Ghazi, N. Kaabouch, and W. F. Fihri, "Matched filter detection with dynamic threshold for cognitive radio networks," in *Proc. Int. Conf. Wireless Netw. Mobile Commun.*, 2015, pp. 1–6.

[7] P. Sutton, K. Nolan, and L. Doyle, "Cyclostationary signatures in practical cognitive radio applications," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 13–24, Jan. 2008.

[8] J. Tong, M. Jin, Q. Guo, and Y. Li, "Cooperative spectrum sensing: A blind and soft fusion detector," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2726–2737, Apr. 2018.

[9] X. He, H. Dai, and P. Ning, "HMM–based malicious user detection for robust collaborative spectrum sensing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2196–2208, Nov. 2013.

[10] N. Gul, I. M. Qureshi, S. Akbar, M. Kamran, and I. Rasool, "One-to-many relationship based kullback leibler divergence against malicious users in cooperative spectrum sensing," *Wireless Commun. Mobile Comput.*, vol. 2018, no. 1, pp. 1–14, Sep. 2018.

[11] H. Li and Z. Han, "Catch me if you can: An abnormality detection approach for collaborative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3554–3565, Nov. 2010.

[12] F. Zhu and S.-W. Seo, "Enhanced robust cooperative spectrum sensing in cognitive radio," *J. Commun. Netw.*, vol. 11, no. 2, pp. 122–133, Apr. 2009.

[13] K. Arshad, "Malicious users detection in collaborative spectrum sensing using statistical tests," in *Proc. 4th Int. Conf. Ubiquitous Future Netw.*, 2012, pp. 109–113.

[14] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2488–2497, Aug. 2010.

[15] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Secure cooperative sensing techniques for cognitive radio systems," in *Proc. IEEE Int. Conf. Commun.*, May 2008, pp. 3406–3410.

[16] F. Farmani, M. Abbasi-Jannatabad, and R. Berangi, "Detection of SSDF attack using SVDD algorithm in cognitive radio networks," in *Proc. 3rd Int. Conf. Comput. Intell., Commun. Syst. Netw.*, Jul. 2011, pp. 201–204.

[17] S. Yuan, L. Li, and C. Chigan, "On MMD-based secure fusion strategy for robust cooperative spectrum sensing," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 504–516, Sep. 2019.

[18] F. Adelantado and C. Verikoukis, "A non-parametric statistical approach for malicious users detection in cognitive wireless ad-hoc networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2011, pp. 1–5.

[19] M. Jin, Y. Li, and H.-G. Ryu, "On the performance of covariance based spectrum sensing for cognitive radio," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3670–3682, Jul. 2012.

[20] S. Zhang, Y. Wang, and J. Li, "A cooperative spectrum sensing method based on information geometry and fuzzy c-means clustering algorithm," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 17–29, Jan. 2019.

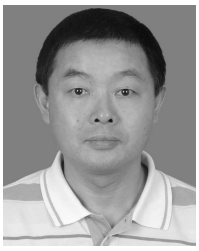
[21] J. Feng, S. Li, S. Lv, H. Wang, and A. Fu, "Securing cooperative spectrum sensing against collusive false feedback attack in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8276–8287, Sep. 2018.

[22] L. Zhang, G. Nie, G. Ding, Q. Wu, Z. Zhang, and Z. Han, "Byzantine attacker identification in collaborative spectrum sensing: A robust defense framework," *IEEE Trans. Mobile Comput.*, vol. 18, no. 9, pp. 1992–2004, Sep. 2019.

[23] R. Gao, Z. Li, P. Qi, and H. Li, "A robust cooperative spectrum sensing method in cognitive radio networks," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 1987–1990, Nov. 2014.
[24] T. Wang, Q. Li, D. J. Bucci, Y. Liang, B. Chen, and P. K. Varshney, "K-medoids clustering of data sequences with composite distributions," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2093–2106, Apr. 2019.
[25] J. Salt and H. Nguyen, "Performance prediction for energy detection of unknown signals," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3900–3904, Nov. 2008.
[26] Y. Wang, S. Zhang, Y. Zhang, P. Wan, and S. Wang, "A cooperative spectrum sensing method based on signal decomposition and K-medoids algorithm," *Int. J. Sensor Netw.*, vol. 29, no. 3, pp. 171–180, Mar. 2019.
[27] E. A. Castro, D. Mason, and B. Pelletier, "On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1487–1514, Apr. 2016.
[28] V. Kumar, D. C. Kandpal, M. Jain, R. Gangopadhyay, and S. Debnath, "K-mean clustering based cooperative spectrum sensing in generalized $\kappa - \mu$ fading channels," in *Proc. 22nd Nat. Conf. Commun.*, 2016, pp. 1–5.
[29] G. C. Sobabe, Y. Song, X. Bai, and B. Guo, "A cooperative spectrum sensing algorithm based on unsupervised learning," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Inform.*, Oct. 2017, pp. 1–6.
[30] Y. Zhang, P. Wan, and S. Zhang, "A spectrum sensing method based on signal feature and clustering algorithm in cognitive wireless multimedia sensor networks," *Adv. Multimedia*, vol. 2017, no. 4, pp. 1–10, Oct. 2017.
[31] H. Cho, S.-J. Kang, and Y. H. Kim, "Image segmentation using linked mean–shift vectors and global/local attributes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2132–2140, Oct. 2017.

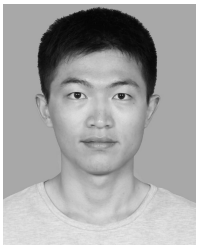
SHUNCHAO ZHANG received the B.S. degree from the Hunan Institute of Engineering, in 2016, the M.S. degree from the Guangdong University of Technology, Guangdong, China, in 2019, where he is currently pursuing the Ph.D. degree. His current research interests include spectrum sensing in cognitive radio, clustering algorithm, and adaptive dynamic programming.

YONGHUA WANG received the B.S. degree in electrical engineering and automation from the Hebei University of Technology, in 2001, the M.S. degree in control theory and control engineering from the Guangdong University of Technology, in 2006, and the Ph.D. degree in communication and information system from Sun Yat-sen University, in 2009. He is currently with the School of Automation, Guangdong University of Technology.



PIN WAN received the B.S. degree in electronic engineering and the M.S. degree in circuit and system from Southeast University, in 1984 and 1990, respectively, and the Ph.D. degree in control theory and control engineering from the Guangdong University of Technology, in 2011. He is currently a Professor with the School of Automation, Guangdong University of Technology.

JIAWEI ZHUANG received the B.S. degree from Jiaying University, in 2017. He is currently pursuing the M.S. degree with the Guangdong University of Technology, Guangdong, China. His research interest includes spectrum sensing in cognitive radio.



YONGWEI ZHANG received the B.S. degree from Jiaying University, in 2016. He is currently pursuing the Ph.D. degree with the Guangdong University of Technology, Guangdong, China. His research interests include spectrum sensing in cognitive radio and adaptive dynamic programming.

YI LI received the B.S. degree from the Hunan Institute of Engineering, in 2016. She is currently pursuing the M.S. degree with the Guangdong University of Technology, Guangdong, China. Her research interest includes spectrum sensing in cognitive radio.