



Research Article

A Comparative Analysis of Different Outlier Detection Techniques in Cognitive Radio Networks with Malicious Users

Arshed Ahmed,¹ Muhammad Sajjad Khan ,^{2,3} Noor Gul,^{1,3} Irfan Uddin,⁴ Su Min Kim,² and Junsu Kim ²

¹Department of Electronics, University of Peshawar, Pakistan

²Department of Electronic Engineering, Korea Polytechnic University, Republic of Korea

³Department of Electrical Engineering, International Islamic University, Islamabad, Pakistan

⁴Department of Information Technology, Superior University, Lahore, Pakistan

Correspondence should be addressed to Junsu Kim; junsukim@kpu.ac.kr

Received 4 August 2020; Revised 12 November 2020; Accepted 22 November 2020; Published 9 December 2020

Academic Editor: Fawad Zaman

Copyright © 2020 Arshed Ahmed et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In a cognitive radio (CR), opportunistic secondary users (SUs) periodically sense the primary user's (PU's) existence in the network. Spectrum sensing of a single SU is not precise due to wireless channels and hidden terminal issues. One promising solution is cooperative spectrum sensing (CSS) that allows multiple SUs' cooperation to sense the PU's activity. In CSS, the misdetection of the PU signal by the SU causes system inefficiency that increases the interference to the system. This paper introduces a new category of a malicious user (MU), i.e., a lazy malicious user (LMU) with two operating modes such as an awakened mode and sleeping mode. In the awakened mode, the LMU reports accurately the PU activity like other normal cooperative users, while in the sleeping mode, it randomly reports abnormal sensing data similar to an always yes malicious user (AYMU) or always no malicious user (ANMU). In this paper, statistical analysis is carried out to detect the behavior of different abnormal users and mitigate their harmful effects. Results are collected for the different hard combination schemes in the presence of the LMU and opposite categories of malicious users (OMUs). Simulation results collected for the error probability, detection probability, and false alarm at different levels of the signal-to-noise ratios (SNRs) and various contributions of the LMUs and OMUs confirmed that out of the many outlier detection tests, the median test performs better in MU detection by producing minimum error probability results in the CSS. The results are further compared by keeping minimum SNR values with the mean test, quartile test, Grubbs test, and generalized extreme studentized deviate (GESD) test. Similarly, performance gain of the median test is examined further separately in the AND, OR, and voting schemes that show minimum error probability results of the proposed test as compared with all other outlier detection tests in discarding abnormal sensing reports.

1. Introduction

Radio spectrum is considered the backbone for wireless communication. The unique characteristic of the wireless sensor networks (WSNs) makes it distinguishable from the traditional networks [1]. In WSNs, a number of small sensor devices distributed spatially are allowed to cooperatively sense environmental and physical conditions. The WSN nodes have limited resources in terms of power, computational complexity, and memory [2]. Recently, the WSNs are employed in civilian applications, such as home appliance control, traffic control, checking environmental conditions,

Internet of things (IoT), and robotic games [2]. The frequency spectrum assigned to the WSNs and other communication devices is not efficiently utilized that results in spectrum scarceness issues. The CR network (CRN) is a promising technology in the field of WSNs to tackle the spectrum scarcity [3].

The idea of CRN was presented for the first time by Mitola in [4]. As demand to the frequency spectrum resources is increasing with the increased number of wireless devices, therefore, static spectrum allocation (SSA) policy is considered to have limitations to meet these requirements [5]. The 300 GHz bandwidth that once seems to be sufficient

is now becoming congested [6–9]. CRN is an intelligent wireless communication technology that has the ability to sense the radio environment and act accordingly. The CRN has two main objectives: reliable communication at any time and place and efficient use of the radio spectrum [10]. As static spectrum allocation is not the solution to meet with the increasing number of wireless communication devices, therefore to overcome this challenging problem, dynamic spectrum access (DSA) has been widely proposed as one of the most promising technologies to increase spectral efficiency [11]. The CRN is considered a feasible intelligent technology for 4G wireless networks or self-organization networks. In the CRN, unlicensed users or secondary users (SUs) periodically sense the spectrum band of the PU network. The SUs utilize vacant channels in the VHF and UHF frequency bands, allocated to TV broadcasting between 54 and 862 MHz frequency range [12]. The PU spectrum availability is inspected by applying various spectrum sensing techniques [6]. The SU performs local sensing by adopting sensing techniques such as energy detector (ED), matched filter detector (MFD), and cyclostationary [7]. When statistics of the PU are not available, then the ED technique is more suitable that requires only power of the PU channel. The received energy of PU is compared with a fixed threshold value in the ED technique. In case the received energy is greater than the threshold, the presence of PU is confirmed; otherwise, the absence of the PU signal is declared [7, 8]. In the proposed work, we will follow the ED technique to sense the spectrum of the PU channel.

In CRN, individual SU is not sensitive enough to detect PU channel weak signals. The single SU sensing performance is further deteriorated by the multipath fading and shadowing effects as in [13]. In order to tackle individual user sensing issues, CSS is used to solve this problem. This allows local sensing users to forward their sensing results to the fusion center (FC), where the final decision is made about the PU status [14].

1.1. Related Work and Contribution. Information reported to the FC by the SUs through local sensing is divided into two major categories: hard decision fusion (HDF) and soft decision fusion (SDF). In the HDF, the SUs convert the sensing reports into binary decision to represent a PU signal. The HDF schemes not only reduce communication cost but also reduce the implementation complexity [15]. In the SDF scheme, the reports are in the form of energy values of the PU signal forwarded to the FC. There are many SDF schemes suggested in the literature, where soft energy information is reported to the FC [16]. Similarly, in the Bayesian model, users report probabilities to represent the confidence level of the users' local decision [17]. The FC then takes a global decision by combining all these probabilities. An SDF model proposed in [17] reports two-bit information to state the free and occupied status of the PU channel. The SDF scheme known as the likelihood ratio test (LRT) has attained a significant attention. In [12], a linear test statistic is applied based on an LRT detector at several PU conditions. In [18], the focus is on maximum eigenvalue-based LRT against different noise behaviors of the PU signal. In [19], the authors have

investigated a distributed LRT detector for sensing the spectrum of the PU spectrum where the channels are considered having random and Nakagami-lognormal mixture distribution. Similarly, in [20], some inspections against frequency-selective Nakagami channels using correlation in the frequency domain are investigated. In [21], authors have presented a collusion pattern of the attackers. These attackers usually form a collusive group that can boost the spectrum sensing data falsification (SSDF) attack power, resulting in falsification of the spectrum sensing data. These attackers are prevented by applying a trust mechanism technique, in which the reports of the SUs are examined by their historical sensing behaviors [22]. The less trusted SUs are given low weights, or even their reports are deleted during final decision. The collusive attackers are riskier as they improve their trust value, which results in increasing their attack power. The main contributions of this paper are as follows.

- (i) In this paper, a new behavior of MU, i.e., a lazy malicious user (LMU), is introduced in the CSS environment. The LMU reports PU information to the FC in two operating modes, i.e., an awakened phase and sleeping phase. The user acts as normal SU during the awakened phase with accurate sensing reports in this phase, while in the sleeping phase, the LU acts maliciously by reporting false sensing data randomly selected as AYMU and ANMU probabilistically. The OMU category of MUs senses the PU channel and reports sensing data to the FC that negate the channel actual status
- (ii) The proposed techniques in the paper detect LMUs and OMUs by applying outlier detection tests while reporting to the FC. During the sleeping phase of the LMUs, received sensing reports are detected as abnormal and discarded while making global decision at the FC. Similarly, as the awakened phase sensing reports of the LMUs are accurate, therefore, the outlier detection tests declare their sensing reports as normal and suggest for consideration in the global decision
- (iii) Simulation verifies that, out of the many outlier detection tests, the median test shows better detection results of MUs in CSS and produces minimum error probability. The results are further compared at low SNR values with those of the mean test, quartile test, Grubbs test, and generalized extreme studentized deviate (GESD) test

The proposed work limitation lies in the parameter selection of statistical tests. It is noticeable that whenever univariate data samples are selected less than a certain limit, outlier values near the upper and lower fence of the data distribution cannot be detected reliably. Hence, the number of SUs should be sufficient enough to get better sensing results.

The rest of the paper is organized as follows: Section 2 presents the system model. Section 3 gives a detailed description of the proposed MU detection model. Section 4 discusses the simulation results. The paper is concluded in Section 5.

2. System Model

All the participating SUs sense the PU status and report their decisions to the FC. The SUs decide the PU activity locally and inform FC about their binary decision findings for making a global decision. FC collects individual hard binary decisions of the cooperative users and employs HDF schemes to recommend the final decision about licensed user activity as shown in Figure 1.

The binary hypothesis about the presence and the absence of the PU channel is given as

$$x_j(l) = \begin{cases} H_0 : & n_j(l) \\ H_1 : & h_j s(l) + n_j(l) \end{cases}, \quad (1)$$

where $x_j(l)$ denotes the energy received by the j^{th} SU in the l^{th} time slot. H_0 and H_1 represent the absence and presence hypothesis of the PU signal. $n_j(l)$ is the AWGN and h_j is the channel gain between the PU and the j^{th} SU. $s(l)$ is the PU transmission at the l^{th} time slot [23, 24]. The energy statistic of the PU received by the j^{th} SU in the i^{th} time interval is given as

$$W_j(i) = \begin{cases} \sum_{l=i}^{i+(b-1)} |n_j(l)|^2, & H_0 \\ \sum_{l=i}^{i+(b-1)} |h_j s(l) + n_j(l)|^2, & H_1 \end{cases}, \quad (2)$$

In (2), b is the number of samples at the i^{th} time interval. The central limit theorem (CLT) shows that for binary hypothesis and large sample size, the energy reported by the participating SUs resembles Gaussian random variables. The normalized energy is written as

$$W_j \sim \begin{cases} N(\mu_0 = b, \sigma_0^2 = 2b), & H_0 \\ N(\mu_1 = b(n_j + 1), \sigma_1^2 = 2b(n_j + 1)), & H_1 \end{cases}. \quad (3)$$

In (3), n_j represents the noise received by the j^{th} SU. The mean and variance of the energy statistics are μ_0 and σ_0^2 , respectively, for the hypothesis H_0 . Similarly, for H_1 , the mean and variance of the energy statistics are μ_1 and σ_1^2 . The energy statistics collected at each SU locally decide the existence of the PU status. These statistics are further compared with the predefined threshold value to send the hard decisions 1 or 0 to the FC [15] as

$$Z_j(i) = \begin{cases} 1, & W_j(i) \geq \gamma_j \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where $W_j(i)$ is the energy statistic of the PU received by the j^{th} SU in the i^{th} interval. γ_j denotes the threshold value for the j^{th} reporting user.

2.1. Proposed MU Detection Model. A flow chart of the proposed CSS model is shown in Figure 2, where multiple SUs sense a spectrum band of the PU and report their observations to the FC. In the flow chart, simple AND, OR, and majority voting are the schemes where outlier tests are not applied and reports are collected from all SUs about PU activity. Similarly, the modified AND, OR, and majority voting are those schemes where outlier tests were used for the MU identification based on all users' reported information. The global decision is calculated under both simple HDF and modified HDF schemes separately, and results are compared.

Pseudocode 1 of the proposed Algorithm 1 is shown in Section 3.

3. Pseudocode 1 of Algorithm 1

A pseudocode of the proposed algorithm to solve the given problem in a stepwise manner is shown. Here, the users take their hard binary decisions and report the same information as 1 or 0 to the FC. FC tries to collect and stores user reports during the N sensing intervals and stores the same in its local database in Z . The FC takes its final decision normally using hard decision schemes before collection of enough reports from the sensing users. At the end of a required number of iterations, the results in Z are accumulated by finding each user total sensing data to form vector \mathbf{z} . The algorithm calls statistical outlier detection tests to detect any abnormality in \mathbf{z} results as outlier or malicious data. After the identification of MUs, modified HDF schemes are allowed to take decision based on the sensing reports of the normally declared users in the subsequent sensing intervals.

3.1. Hard Decision Schemes. A centralized CSS allows SUs to forward their local sensing results to a central unit where the final decision of the PU activity is made based on sensing reports. To categorize the information provided to the FC, local sensing schemes are divided into HDF and SDF. In HDF, the SUs convert the sensing reports into binary digits 1 and 0 that represent the PU signal. HDF schemes reduce both the communication cost and implementation complexity of the system. In the SDF scheme, reports are in the form of energy values of the PU signal that are forwarded to the FC. LRT has attained a significant attention out of the different SDF schemes.

3.1.1. AND Scheme. In the AND scheme, all SUs have to be consistent about the reports of PU:

$$G_d = \begin{cases} H_1 : & \sum_{j=1}^n Z_j(i) = n \\ H_0 : & \text{otherwise} \end{cases}, \quad (5)$$

$Z_j(i)$ consists of reports in the i^{th} interval by the SUs. The channel is declared occupied when all SUs reports the PU availability where H_1 is generated by the FC as a global decision G_d ; otherwise, decision H_0 is declared.

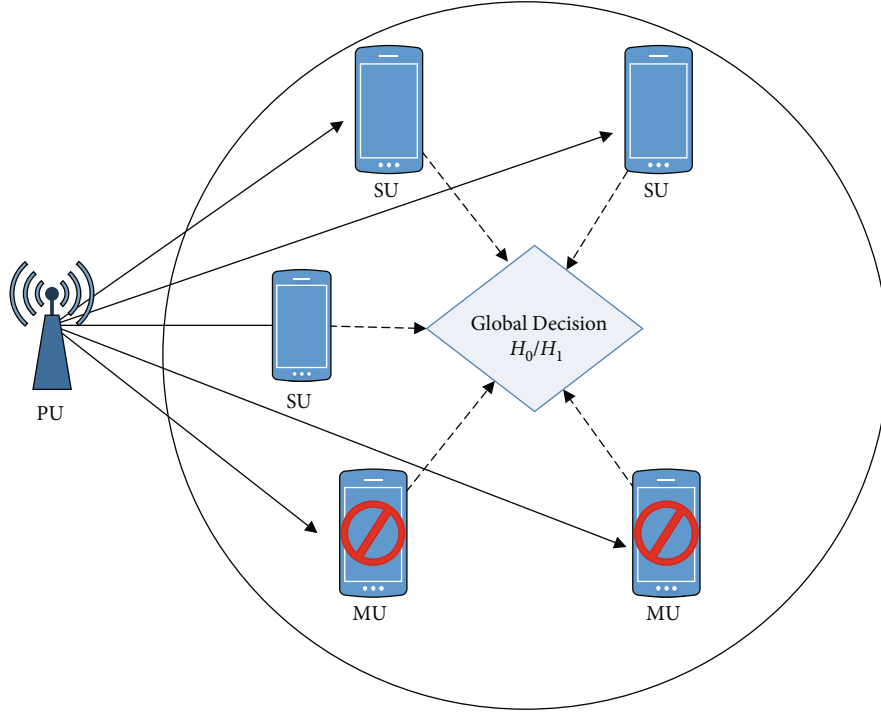


FIGURE 1: System model.

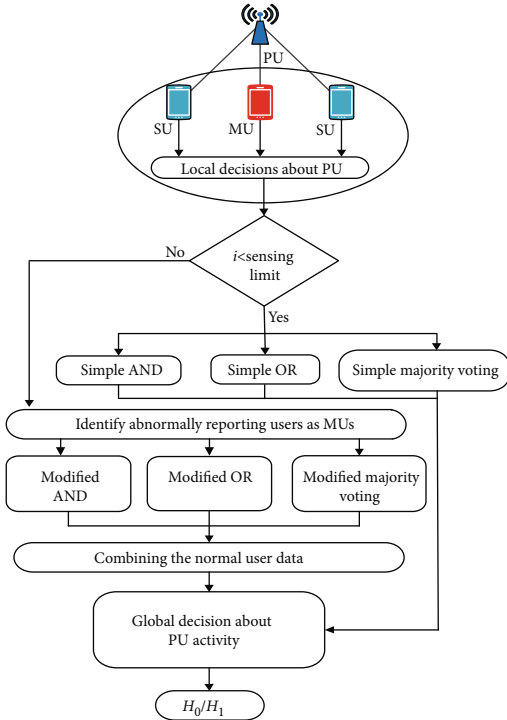


FIGURE 2: Flow chart for the proposed methodology.

3.1.2. *OR Scheme.* In the OR scheme, if any SU detects the PU signal, then FC takes it as a global decision and generates H_1 ; otherwise, the global decision is H_0 :

$$G_d = \begin{cases} H_1 : \sum_{j=1}^n Z_j(i) \geq 1 \\ H_0 : \text{otherwise} \end{cases}. \quad (6)$$

3.1.3. *Majority Voting Scheme.* The majority voting scheme is based on the voting of SUs. If majority users declare the PU availability, the decision is made in favor of majority voters:

$$G_d = \begin{cases} H_1 : \sum_{j=1}^n Z_j(i) \geq k \\ H_0 : \text{otherwise} \end{cases}, \quad (7)$$

where k is the number of SUs, declaring that PU has occupied the channel, and n is the total number of participating SUs. The majority voting scheme is the special case of global decision when $k = n/2$. The FC applies statistical analysis by combining the reports of all participating users to remove the nasty data from MUs in the local sensing.

3.2. *Statistical Outlier Tests.* The outliers in the data are dissimilar values to the rest of the data set. They are generated through different mechanisms in the CSS [25]. The outliers can also be defined as those observations that deviate from their members in the data sample [26]. In this work, the reports of the MUs are outliers because from the definition, outliers are the data samples generated by another

```

(1) For  $i=1$  to total iterations
(2)   For  $j=1$  to total SUs
(3)     if PU is available
(4)        $j^{th}$  user sensing in  $i^{th}$  interval  $E(i, j)$ 
(5)       if  $E_{(i,j)} > \text{threshold}$ 
(6)         local decision  $Z(i, j) = 1$  (Reporting 1 as hard decision),
(7)         else
(8)         local decision  $Z(i, j) = 0$  (Reporting 0 as hard decision)
(9)       end
(10)    end
(11)  end
(12)  if  $i < \text{Sensing limit}$ 
(13)  Compile the results for simple HDF schemes
(14)  Simple AND
(15)  Simple OR
(16)  Simple Majority voting
(17)  else
(18)  Identify  $z$ 
(19)  Run Outlier Tests using  $z$  to detect outliers
(20)  Compile results for Modified HDF schemes
(21)  end
(22)  if  $\sum_{\substack{j=1 \\ j \neq MU}}^n Z_j(i) = n$ 
(23)  AND global Decision as  $H_1$ 
(24)  else
(25)  AND global Decision as  $H_0$ 
(26)  end
(27)  Compile results for Modified HDF schemes if  $\sum_{\substack{j=1 \\ j \neq MU}}^n Z_j(i) \geq 1$ 
(28)  OR global Decision as  $H_1$ 
(29)  else
(30)  OR global Decision as  $H_0$ 
(31)  end
(32)  if  $\sum_{\substack{j=1 \\ j \neq MU}}^n Z_j(i) \geq n/2$ 
(33)  Voting global Decision as  $H_1$ 
(34)  else
(35)  Voting global Decision as  $H_0$ 
(36)  end
(37)  end of iterations

```

PSEUDOCODE 1

mechanism; hence, the reports from the MUs deviate from those reports which are generated by normal SUs [27].

In the proposed model, SUs sense the PU channel and report their hard binary findings to the FC, where it stores n SU sensing data reported in N sensing iterations to form matrix \mathbf{Z} as shown in

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{Nn} \end{bmatrix}. \quad (8)$$

At the end of the required number of iterations, each user contribution in sensing is determined by adding total hard decisions of the SUs to form vector \mathbf{z} as given in

$$\mathbf{z} = \sum_{i=1}^N (Z_j(i)), \quad i \in 1, \dots, N. \quad (9)$$

Outlier detection techniques are called by giving the result in equation (9) as an argument to declare the users as normal or abnormal using various detection tests:

$$\mathbf{Z} = [z_1 \quad z_2 \quad \cdots \quad z_n], \quad (10)$$

Finally, the detected outlier is declared as malicious and

taken out of the hard combination scheme in the following sensing intervals.

3.2.1. Proposed Outlier Median Test Scheme. This outlier detection scheme searches for anomaly in the normally distributed sensing data as in [28]. In the case of univariate data, the median absolute deviation (MAD) is the robust dispersion measure against outliers [28]. Therefore, outlier presence in the data needs to be properly detected and removed. Automatic analysis for the detection of these anomalies in the normally distributed data is mandatory. The traditional method of the mean plus-minus 3 test based on the standard deviation of the data follows normal distribution of the data, where 99.87% of the data type occurs within this range. Similarly, taking decision to remove the values occurring in 0.13% of all cases is not too conservative [28]. There are three problems when the mean is considered the central tendency in the data set. First, the data set has to be normally distributed when outliers are included. Secondly, the outliers in the data have a strong impact on the mean and standard deviation. At last, for any small data sample values, the outlier detection is not guaranteed. Due to these drawbacks, the mean test failed to detect outliers in data distributions when the data sample is limited in size.

Therefore, Miller proposed an outlier indication test using the median of the data set. This outlier test detects anomaly for the value of c : it is most conservative when c is 3, medium conservative when c is 2.5, and less conservative when c is 2 [29]. The constant value of 3 is used in this work. The limiting point against the users' total sensing reports in the \mathbf{z} vector is determined in

$$M - (c \times \text{MAD}) < \mathbf{z}_i < M + (c \times \text{MAD}). \quad (11)$$

The result in (11) is written in a more simplified form as

$$\frac{\mathbf{z}_i - M}{\text{MAD}} \geq |\pm 3|. \quad (12)$$

The results of the median test are compared further with those of other outlier tests such as the Grubbs test, GESD test, and quartile test such as the box and whisker plot and mean plus-minus 3 test.

3.2.2. Grubbs Test. Frank Grubbs in 1969 proposed an outlier test to verify some univariate data [30]. The Grubbs test is used to detect a single outlier in sampled data. This test analyzes the minimum/maximum values of the sample data and applies statistics to search outliers. The test statistic is

$$G = \frac{|\max \text{ value} - M|}{\sigma}, \quad (13)$$

where M is the sample mean and σ is the standard deviation given by

$$\sigma = \sqrt{\frac{\sum (\mathbf{z}_i - \bar{\mathbf{z}})^2}{S - 1}}. \quad (14)$$

In (14), S is the number of data values, \mathbf{z}_i is the maximum value of the row vector, and $\bar{\mathbf{z}}$ is the mean value of the vector \mathbf{z} . The following steps are used in the Grubbs test to detect suspicious report as an outlier.

- (i) Find the G test statistics using (13) for the users' sensing reports in \mathbf{z}
- (ii) State the null and alternative hypotheses about the existence of outliers in \mathbf{z}
- (iii) Find the G critical value from the table and select the confidence level. The default confidence level is 95%
- (iv) Compare the tested G statistic with the G critical value
- (v) The maximum value in \mathbf{z} is an outlier if the test statistics are greater than the critical value

The Grubbs tests can be used to detect and remove outlier values from the minimum values of the data sample as given in

$$G = \frac{|M - \min \text{ value}|}{\sigma}. \quad (15)$$

3.2.3. Generalized Extreme Studentized Deviate (GESD) Test. GESD is an iterative hypothesis test proposed by Rosner in 1983. It can spot one or more outliers in a data set. In this test, the upper bound or the total number of outlier values is given in the null hypothesis. After that, a separate test is performed by using the Grubbs statistics as given in [31]

$$T_k = \frac{\max |\mathbf{z}_i - M|}{\sigma}, \quad (16)$$

where M and σ denote the mean and standard deviations in the data. The observation corresponding to $\max |\mathbf{z}_i - M|$ is removed using Grubbs statistics, and T_2 is computed from the remaining sample. A sample mean and standard deviation are computed for the remaining $n - 1$ data values. This process is repeated until T_k is determined for a prespecified k . Here, k represents the number of outliers in the data set known as the upper bound specified in the null hypothesis [32].

3.2.4. Mean Test. This method is based on the characteristics of normal distribution of data. It is necessary for the outlier test to detect the presence of the outlier's data. In [33], the mean plus-minus 3 standard deviation scheme is formulated as

$$\bar{\mathbf{z}} - (a \times \sigma) < \mathbf{z}_i < \bar{\mathbf{z}} + (a \times \sigma), \quad (17)$$

where $\bar{\mathbf{z}}$ denotes the sample mean and σ denotes the standard deviation of \mathbf{z} . The constant parameter a is carefully selected which is 3 here to produce accurate results. The value \mathbf{z}_i is an outlier in the \mathbf{z} if it exceeds the upper boundary of the data sample such as $\mathbf{z}_i < \bar{\mathbf{z}} + (a \times \sigma)$ or if the data value exceeds the lower boundary, i.e., $\bar{\mathbf{z}} - (a \times \sigma) < \mathbf{z}_i$ [28]. It is guiding

the outlier detection test where the indicator itself is altered by the existence of outlying values in the data.

3.2.5. Quartile and Percentile Test (Box-Whisker Plot). Tukey in 1977 proposed a graphical outlier indication test to identify the skewness and unusual data points in the data distribution [34]. It can detect one or more outlier values in the data set and can also detect outliers in the upper and lower boundaries of the data samples [35]. The following are the steps of the quartile and percentile test.

- (i) Determine the first quartile of the data as the 25th percentile (Q_1) in \mathbf{z}
- (ii) Identify the third quartile of the data as the 75th percentile (Q_3) in \mathbf{z}
- (iii) Determine the interquartile range (IQR) of the \mathbf{z} vector as

$$\text{IQR} = Q_3 - Q_1. \quad (18)$$

- (iv) A data value is considered an outlier in the lower fence if it exceeds the results in

$$Q_1 - 1.5(\text{IQR}). \quad (19)$$

- (v) Similarly, a value is considered an outlier in the upper fence of the data set if it exceeds the results in

$$Q_3 + 1.5(\text{IQR}). \quad (20)$$

Figure 3 shows the diagram of the box-whisker plot. All parameters are indicated in the figure.

3.3. Modified HDF Schemes. After detecting the reports of MUs at the FC, global decision is made by FC in the modified form in the subsequent sensing intervals. In the modified AND scheme, sensing reports of the normally declared cooperative users are considered in the global decision. Similarly, the reports received from the detected outliers such as MUs are deleted in this combination. Hence, the modified equation of the AND decision scheme now takes the following form:

$$G_d = \left\{ \begin{array}{l} H_1 : \sum_{\substack{j=1 \\ j \neq \text{MU}}}^n Z_j(i) = n_{\text{Modified}} \\ H_0 : \text{otherwise} \end{array} \right\}. \quad (21)$$

In the modified AND scheme, only normal SU n_{Modified} sensing reports about the presence of PU are considered, whereas the reports of OMUs and LMUs are discarded.

The criteria for the decision of the OR scheme are modified as

$$G_d = \left\{ \begin{array}{l} H_1 : \sum_{\substack{j=1 \\ j \neq \text{MU}}}^n Z_j(i) \geq 1 \\ H_0 : \text{otherwise} \end{array} \right\}. \quad (22)$$

The modified majority voting scheme takes its global decision based on the reports of normally declared users. After the filtration and elimination of the MUs, the modified majority voting scheme decision is given as

$$G_d = \left\{ \begin{array}{l} H_1 : \sum_{\substack{j=1 \\ j \neq \text{MU}}}^n Z_j(i) \geq k_{\text{Modified}} \\ H_0 : \text{otherwise} \end{array} \right\}, \quad (23)$$

where k_{Modified} is the number of sensing reports received from the normal SUs that declare the presence and absence of the PU signal by H_1 and H_0 , respectively.

4. Simulation Results

In this section, we present simulation results of the proposed outlier detection-based HDF schemes and compared them with other statistical outlier schemes. In the simulation, the number of MUs varied in the cooperating environment to investigate the overall effect in the CSS. The simulation parameters are defined in Table 1.

4.1. Case 1: Median Test Results. In case 1, the results for the median test are plotted using HDF schemes. The median test results in Figures 4 and 5 are compared with those of the simple HDF schemes. Figure 4 shows the simulation results when there are no MUs in the cooperative environment. It is observed that when the proposed median test is applied, the error probability reduces than when the traditional HDF scheme is applied. In Figure 4, an increase in SNR from -30 dB to -15 dB results in an abrupt change in the error probability for proposed OR and AND schemes, where error probability reduces from 0.47 to 0.25 for the OR scheme. In the AND decision scheme, error probability starts at 0.50 that gradually reduces to 0.23 when SNR exceeds from -30 dB to -10 dB for the proposed outlier test. Similarly, the proposed majority voting error probability is reduced from 0.26 to 0.25, when SNR is increased from -40 dB to -10 dB. The proposed HDF schemes show better sensing results with minimum error probabilities, while the simple HDF schemes result in maximum error probability.

In Figure 5, the number of LMUs is increased to five with one OMU reporting with normal SUs. In this case, error probability remains high for the traditional HDF schemes while the proposed test has a reduced error probability. It can be observed that for the simple HDF schemes when the number of MUs is increased, the error probability remains high at all SNR values, i.e., 0.53 approximately for the simple

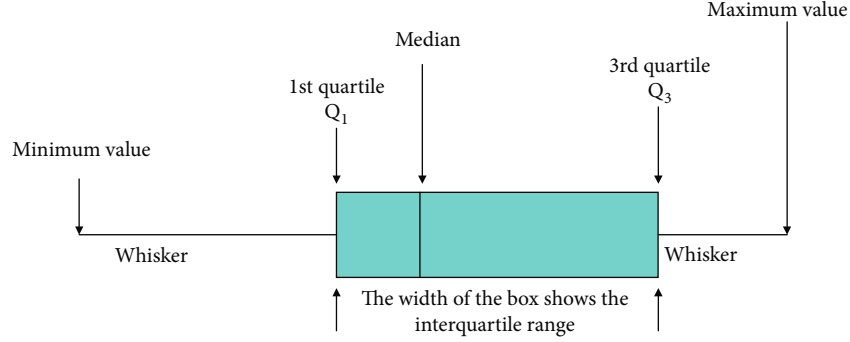


FIGURE 3: Box-whisker plot.

TABLE 1: Simulation parameters.

Parameter	Value
Total number of users	40
Malicious users	2
SNR range	-40 dB to -10 dB
Number of samples in each interval	270
Sensing iterations	1000
Iteration range for simple HDF	1-500
Iteration range for modified HDF	501-1000
Time consumption	1 msec

OR scheme and 0.5 for the simple AND scheme. Similarly, for the simple majority voting scheme, the error probability starts at 0.28 approximately and gradually reduces to 0.26. The proposed AND scheme has an error probability of 0.5 at the SNR value of -40 dB that sharply reduces to 0.26 when SNR is increased from -25 dB to -10 dB. The proposed OR scheme has an error probability of 0.48 at the SNR value of -40 dB that reduces to 0.25 at the SNR value of -10 dB. Similarly, in the case of the proposed majority voting scheme, error probability starts at 0.26 approximately and remains lower than that in the simple majority voting scheme.

The results of percent decrease in error probability of the modified and traditional HDF schemes at different SNRs values are illustrated in Table 2 in the presence of LMUs and OMU. The table result shows that when SNRs = -26 dB, the proposed voting scheme results in better sensing performance with 9.1% minimum sensing error probability compared with the simple voting scheme. Similarly, the proposed OR scheme obtained 8.9% reduction in error probability as compared with the simple OR scheme, while the proposed AND decision scheme has 0.4% reduction in the error probability results compared with the simple AND combination scheme. As the SNRs are increased to -10 dB, the percent decrease in error probability of the proposed HDF schemes is further improved for the proposed voting (9.8%), proposed OR (1.7%), and proposed AND (47.1%) schemes compared with the simple voting, simple OR, and simple AND decision schemes.

4.2. Case 2: Performance Comparison of the Proposed Scheme with Other Statistical Outlier Test Schemes. In this case, we present the performance comparison of the proposed median test with the other statistical outlier tests in Figures 6–11. HDF schemes are plotted separately and compared with outlier tests. The comparison is made for HDF schemes in the following scenario in CSS.

- (1) When no MU exists in the network
- (2) When five LMUs and one OMU exist in the network

4.2.1. Scenario 1: OR Scheme. A global decision of the OR scheme is made when a single SU detects the presence of a PU signal; hence, there is a chance of error in the sensing report. In Figures 6 and 7, the results for the OR scheme is investigated for all outlier tests along with the results of the simple HDF (OR) scheme in the global decision. These figures show that the proposed test scheme is outperforming other statistical outlier test schemes, when there is no MU in the network. The SNR varies from -40 dB to -10 dB. In the simple OR scheme, error probability starts at 0.51 that reduces after -20 dB and reaches a value of 0.38 approximately at -10 dB. All the statistical outlier test schemes have a starting error probability of 0.49 which is lower than that of the simple HDF OR scheme and abruptly reduces after -30 dB. The mean test results are with maximum error probability among all other outlier detection tests as SNR ranges from -30 dB to -10 dB which is followed by the GESD, Grubbs, and quartile tests. The proposed median outlier test scheme has minimum error probability from -30 dB to -10 dB.

In Figure 7, there are five LMUs with one OMU and 34 normal SUs reporting to FC for a global decision. The simple HDF scheme has an error probability of 0.51 approximately at -40 dB that is slightly reduced to 0.49 at -10 dB. Similarly, all the statistical outlier tests have the same probability of error up to -25 dB which is slightly reduced to 0.47 approximately for the mean test. For the GESD test, the P_e is 0.45 at -10 dB, and for the Grubbs test, the P_e is 0.43 approximately at -10 dB. These results further reduce to 0.26 approximately for the quartile test at the SNR value of -10 dB. Similarly, when the SNR value exceeds -25 dB, the proposed test scheme curve is skewed down to the error probability of

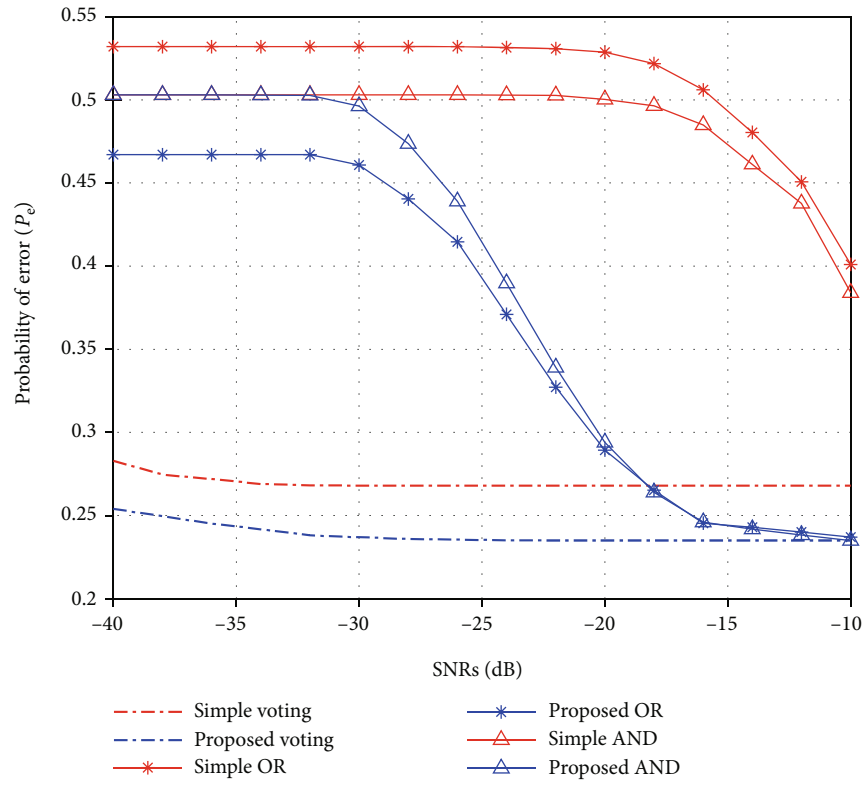


FIGURE 4: No MUs with 40 normal SUs in the network.

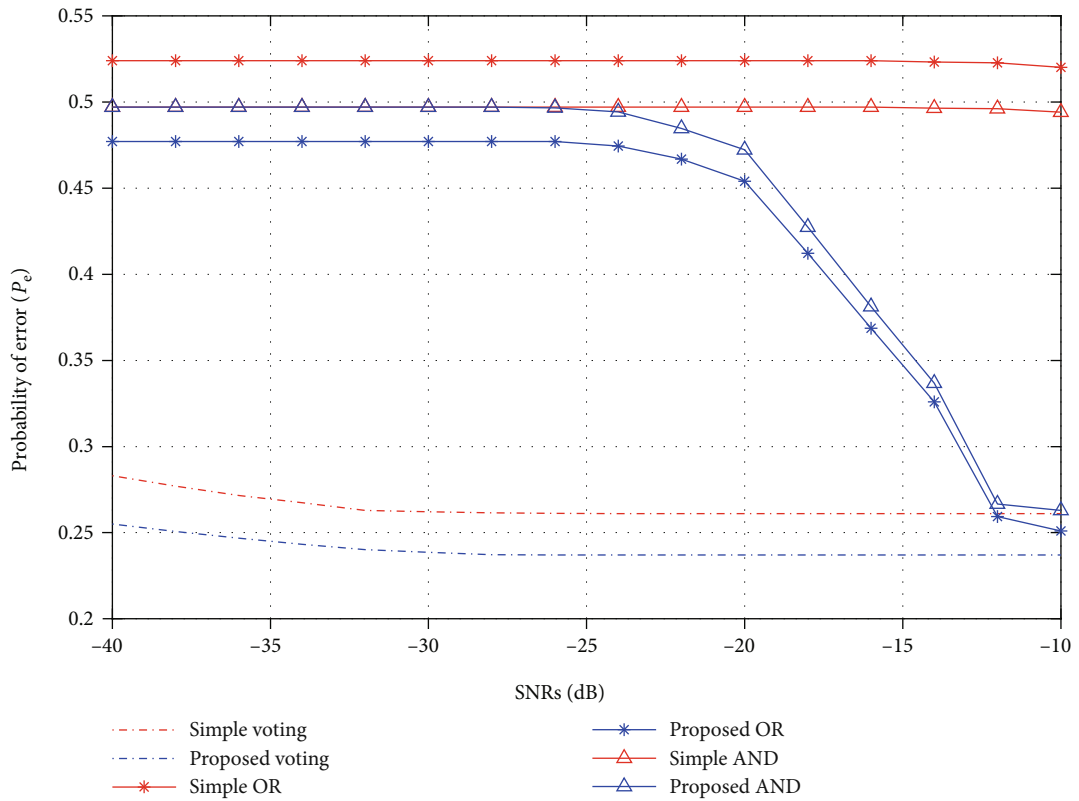


FIGURE 5: Five LMUs and one OMU with 34 normal SUs in the network.

TABLE 2: Percent decrease in the error probabilities for modified HDF schemes.

Decision schemes	SNR values				
	-26 dB	-22 dB	-18 dB	-14 dB	-10 dB
Simple vs. proposed voting	9.1%	9.1%	9.1%	9.1%	9.8%
Simple vs. proposed OR	8.9%	11%	21.3%	38.7%	51.7%
Simple vs. proposed AND	0.4%	2.6%	14%	32.2%	47.1%

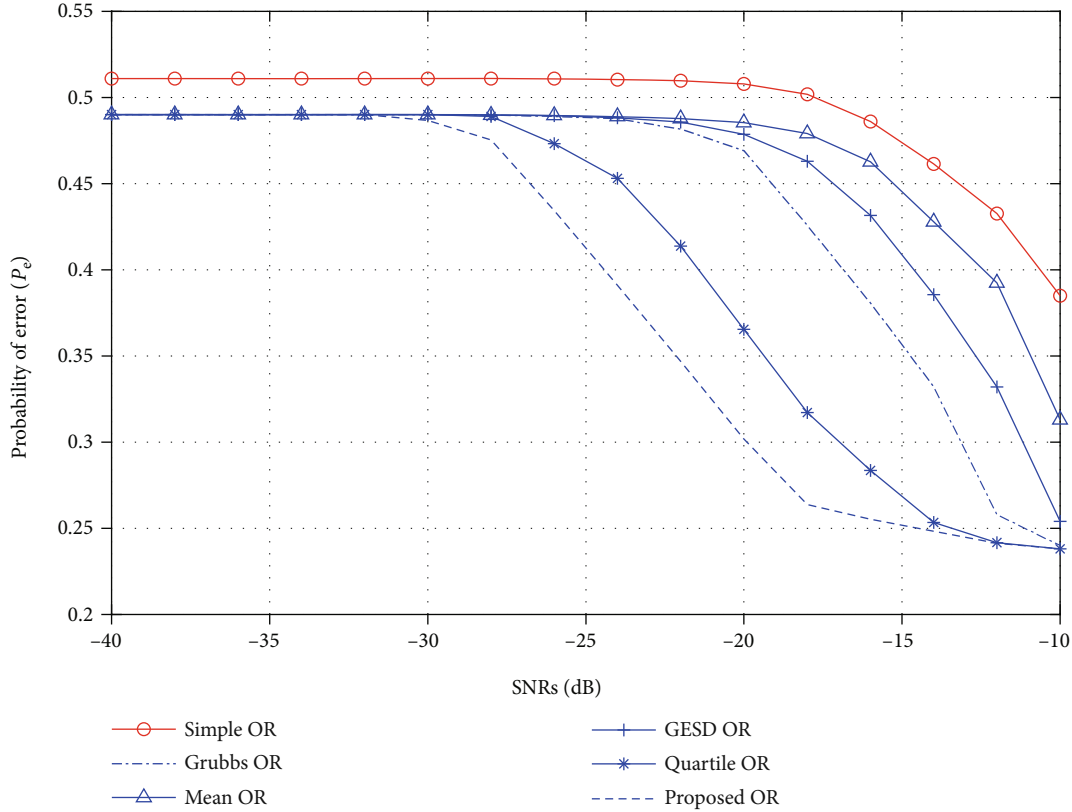


FIGURE 6: No MUs with 40 normal SUs in the network.

0.25 at -10 dB, which is the minimum error probability of all outlier tests.

The results in Table 3 illustrate the percent decrease in the error probability of the proposed median test-based OR HDF combination scheme as compared with other outlier detection tests in the presence of LMUs and OMU at various SNRs. The table shows that at -26 dB, the proposed OR HDF scheme results in better sensing performance with 1.67% reduction in sensing error probability compared with the mean OR, GESD OR, Grubbs OR, and quartile OR test schemes. Similarly, as SNRs are increased to -10 dB, the percent decrease in error probability of the proposed OR HDF scheme is further improved as compared with that of the mean OR (91.4%), GESD OR (81.7%), Grubbs OR (72.7%), and quartile OR (4.87%) test schemes.

4.2.2. Scenario 2: AND Scheme. When all the SUs confirm the presence of the PU signal, a global decision is made in favor of the AND scheme. This scheme is tested at the same SNR

values of -40 dB to -10 dB. In Figure 8, no MUs are included in sensing. Therefore, the simple HDF scheme performed moderately with an error probability of 0.50 which proceeded to 0.37 approximately at -10 dB. Similarly, other outlier tests have minimum error probabilities compared with the simple HDF scheme at -10 dB. The mean test has $P_e = 0.36$, the GESD test has $P_e = 0.3$, and the Grubbs test, quartile test, and proposed test have $P_e = 0.27$ approximately. It is observed that for the SNR values between -30 dB and -15 dB, the proposed median test scheme has the best performance with minimum error probability.

The simple HDF scheme has poor sensing performance when MUs appear in sensing with constant $P_e = 0.52$ approximately from -10 dB to -40 dB as observed in Figure 9. The same maximum P_e of 0.52 is observed for all outlier tests from -40 dB to -20 dB. These results are followed by the mean test that has $P_e = 0.46$, Grubbs and GESD tests with $P_e = 0.45$, and quartile test with $P_e = 0.34$ approximately at -10 dB. The proposed median test has $P_e = 0.28$ at -10 dB, which is

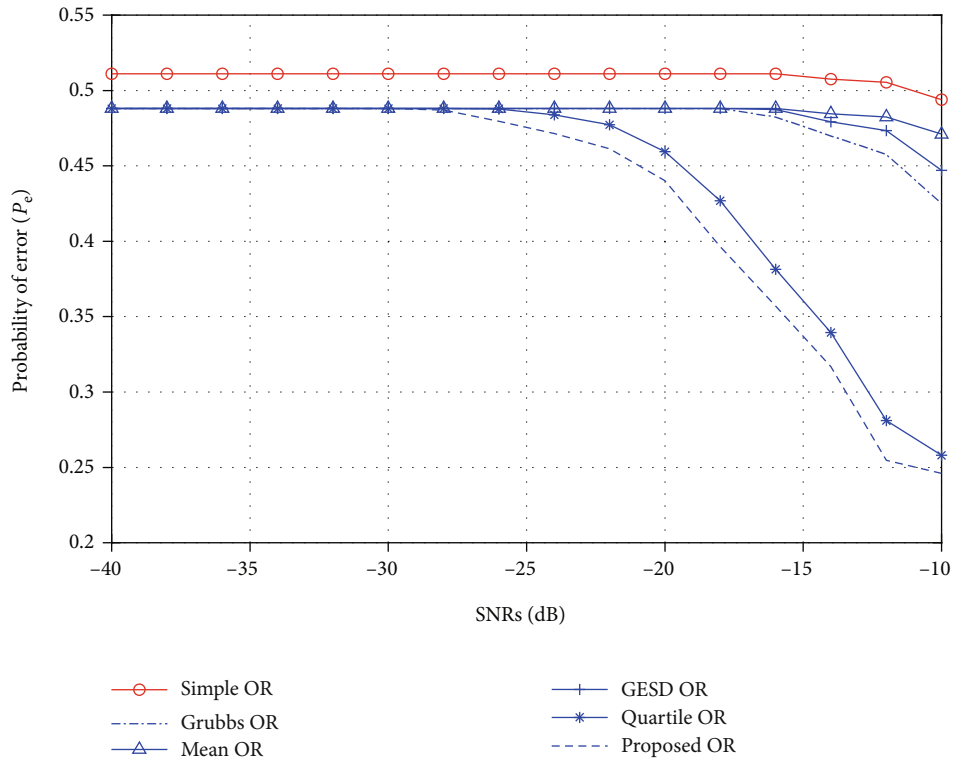


FIGURE 7: Five LMUs and one OMU with 34 normal SUs in the network.

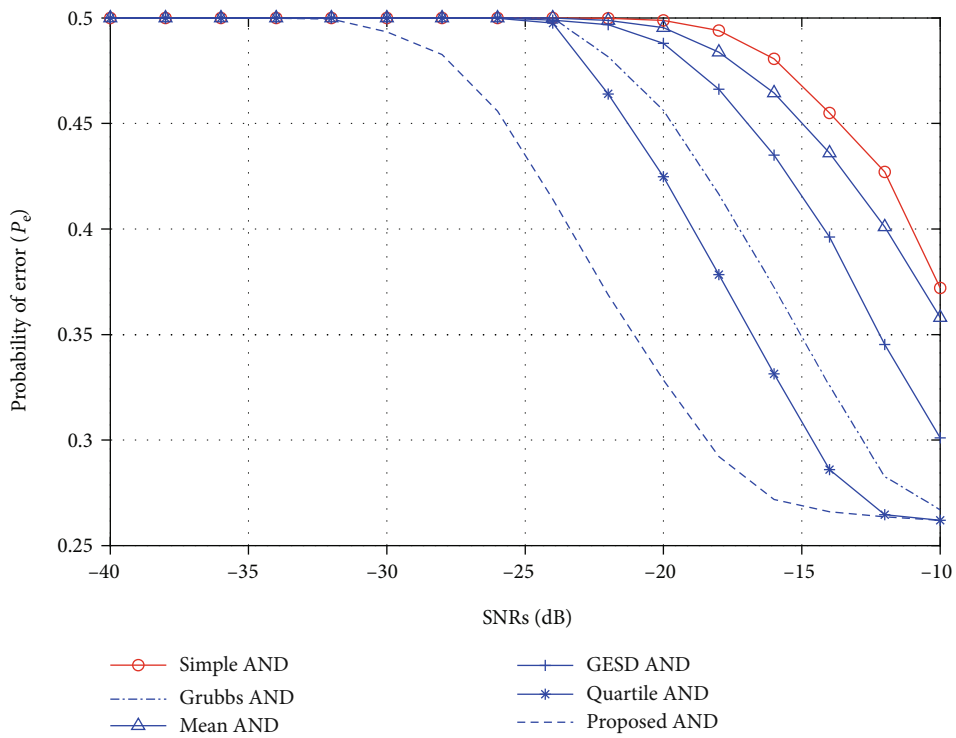


FIGURE 8: No MUs with 40 normal SUs in the network.

lowest in all the outlier test schemes. The proposed median test surpasses all other outlier tests at SNR values from -15 dB to -10 dB with the minimum P_e of 0.28 at -10 dB.

Table 4 shows the performance gain in terms of percent decrease in error probabilities at different SNR values for the proposed outlier detection test using the AND HDF

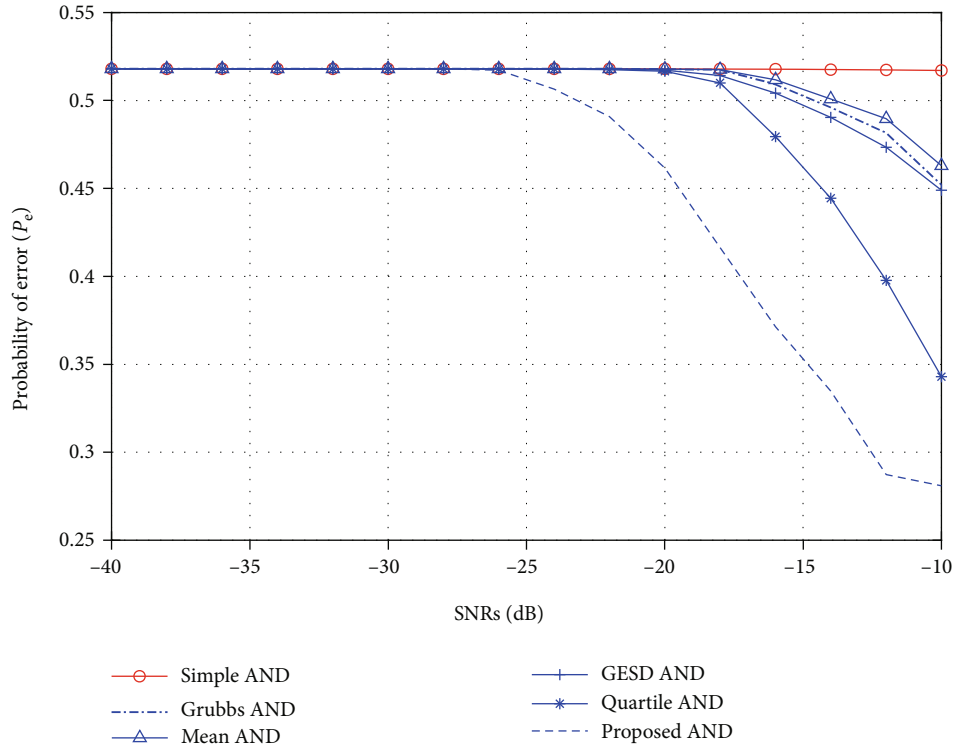


FIGURE 9: Five LMUs and one OMU with 34 normal SUs in the network.

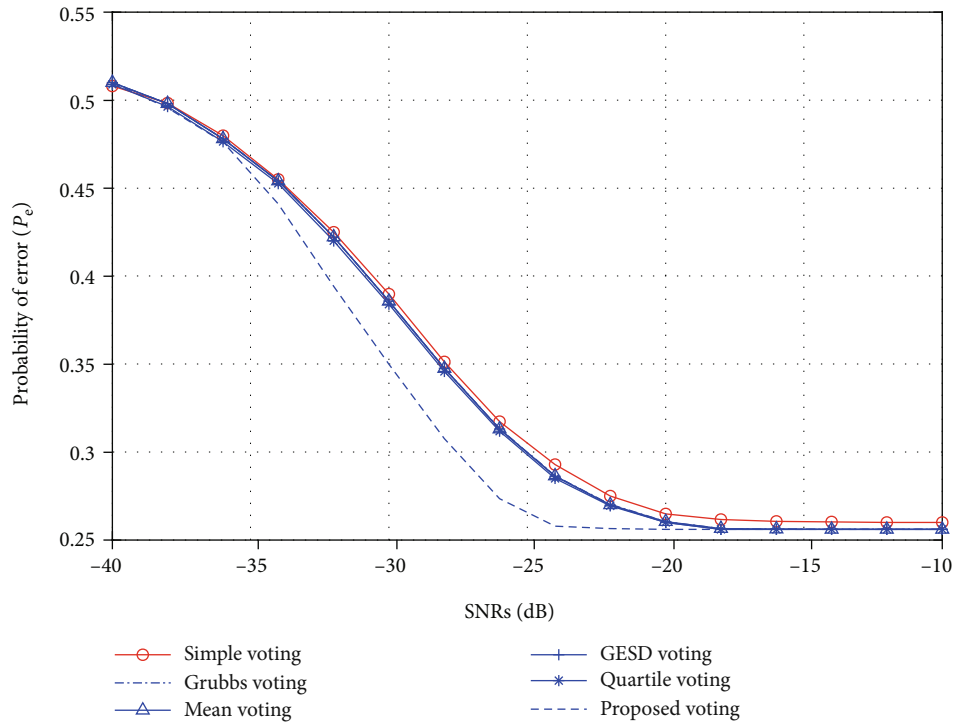


FIGURE 10: No MUs with 40 normal SUs in the network.

scheme containing LMUs and OMU. The table shows that at -26 dB, the proposed AND HDF scheme results in better performance gain with 0.19% decrease in sensing error probability compared with the mean AND, GESD AND, Grubbs

AND, and quartile AND schemes. Similarly, as the SNRs are increased to -10 dB, the percent decrease in error probability of the proposed AND HDF scheme is further improved as compared with the mean AND (64.7%), GESD AND

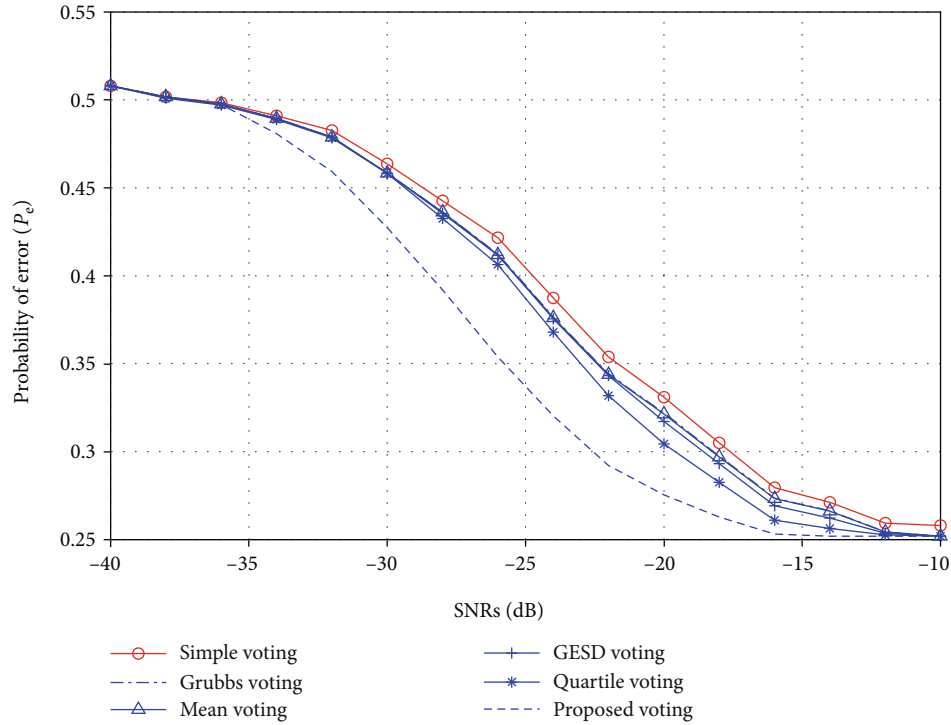


FIGURE 11: Five LMUs and one OMU with 34 normal SUs in the network.

TABLE 3: Percent decrease in the error probabilities for the proposed OR HDF scheme.

Decision schemes	SNR values				
	-26 dB	-22 dB	-18 dB	-14 dB	-10 dB
Proposed OR vs. mean OR	1.67%	5.8%	23.2%	53.1%	91.4%
Proposed OR vs. GESD OR	1.67%	5.8%	23.2%	51.5%	81.7%
Proposed OR vs. Grubbs OR	1.67%	5.8%	23.2%	48.4%	72.7%
Proposed OR vs. quartile OR	1.67%	3.47%	16.6%	7.27%	4.87%

TABLE 4: Percent decrease in the error probabilities for the proposed AND HDF scheme.

Decision schemes	SNR values				
	-26 dB	-22 dB	-18 dB	-14 dB	-10 dB
Proposed AND vs. mean AND	0.19%	5.71%	24.2%	49.7%	64.7%
Proposed AND vs. GESD AND	0.19%	5.71%	23.5%	46.7%	59.7%
Proposed AND vs. Grubbs AND	0.19%	5.5%	24.2%	48.5%	60.8%
Proposed AND vs. quartile AND	0.19%	5.7%	22.5%	32.9%	22%

(59.7%), Grubbs AND (60.8%), and quartile AND (22%) schemes.

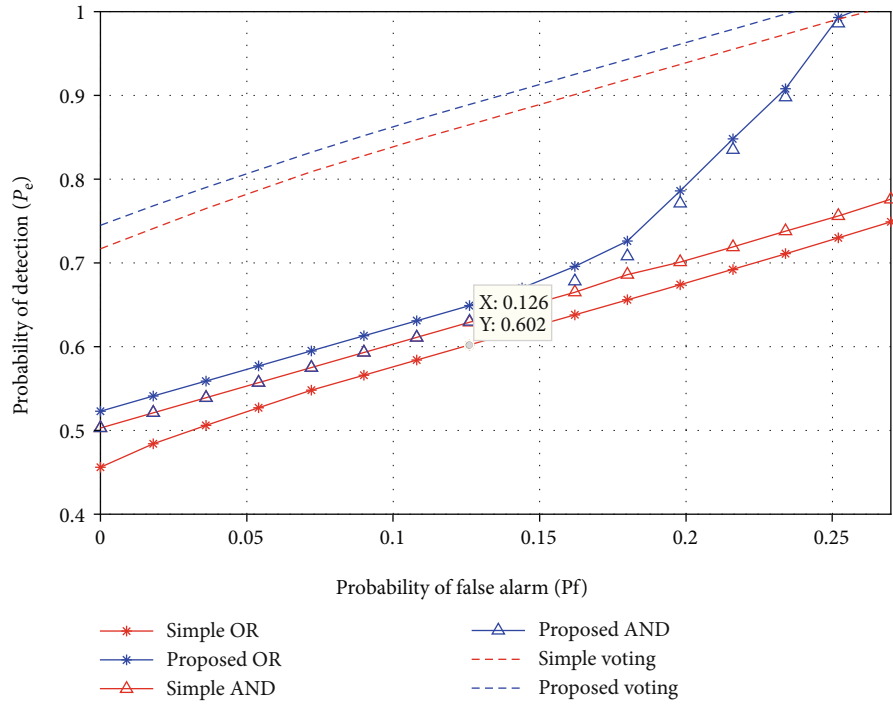
4.2.3. *Scenario 3: Majority Voting Scheme.* For simulation purposes, $k = n/2$ is selected for the majority voting scheme, where more than one SU has to declare the PU channel available to make its decision about the presence and absence of PU; otherwise, PU absence is declared. Figure 10 shows the results without any misbehaving users. The simple voting scheme has maximum P_e with a starting value of 0.51 at the

SNR value of -40 dB and reduces to a value of 0.26 at the SNR value of -10 dB. All outlier tests give similar P_e results at all SNR levels. The proposed median test results are slightly improved giving low P_e values from SNR values of -35 dB to -20 dB which are significantly lower than those of the simple HDF of the majority voting scheme.

The value of P_e is increased for the simple majority voting scheme when MUs transfer reports in the sensing interval. From Figure 11, it is observed that P_e of the simple majority voting scheme is 0.51 approximately at -40 dB and after

TABLE 5: Percent decrease in the error probabilities of the proposed voting scheme.

Decision schemes	SNR values				
	-26 dB	-22 dB	-18 dB	-14 dB	-10 dB
Proposed voting vs. mean voting	16.7%	17.4%	13.3%	5.5%	0.4%
Proposed voting vs. GESD voting	16.4%	17.4%	11.8%	3.9%	0.4%
Proposed voting vs. Grubbs voting	16.4%	17.4%	13.3%	5.5%	0.4%
Proposed voting vs. quartile voting	15%	13.3%	7.6%	1.6%	0.4%

FIGURE 12: Probability of detection (P_d) vs. probability of false alarm (P_f) for the simple and modified HDF schemes with normal SUs and MUs.

-30 dB it reduces further to 0.26. Outlier test schemes show that P_e is between -35 dB and -15 dB and is minimized slightly compared with that of the simple majority voting scheme. On the other hand, the proposed median test scheme shows significance upon all outlier test schemes from SNR values of -35 dB to -15 dB. It is observable from the simulation results that the outlier test scheme detects falsifying reports of MUs and removes them in the final decision that decreases the error probability.

The overall performance gain of the voting scheme is better than that of the AND HDF and OR HDF schemes by establishing minimum error probability. Table 5 illustrates the results of percent decrease in the error probabilities obtained by the proposed outlier detection test using the voting scheme in the presence of LMUs and OMU at different SNR values. At -26 dB, the proposed voting scheme results in better sensing performance with 16.7% decrease in sensing error probability compared with the mean voting scheme, 16.4% decrease compared with the GESD voting and Grubbs voting schemes, and 15% improvement compared with the quartile voting scheme. As the SNRs are increased to -10 dB, the percent decrease in error probability of the pro-

posed voting scheme is 0.4% as compared with that of mean voting, GESD voting, Grubbs voting, and quartile voting schemes.

The receiver operating characteristics (ROC) with probability of detection P_d vs. probability of false alarm P_f are collected in the presence of LMUs and OMUs in Figures 12–15. In Figure 12, the results are plotted for simple HDF and proposed (modified) HDF schemes. The simple OR HDF results are highly deteriorated by producing high P_f values with the contributions of MUs, whereas the proposed (modified) OR decision scheme gives better detection results. Similarly, AND combination scheme detection probability with the employment of the proposed median test has better detection results with minimum false alarm than the detection probability of the traditional AND decision scheme. Likewise, detection results of the proposed voting scheme remain superior and surpass those of all other HDF schemes in Figure 12.

The modified scheme ROC results are further compared with those of the other outlier detection tests to investigate the proposed test superiority. Figure 13 shows the result illustrations for the OR HDF scheme which was compared with the other outlier detection tests. It is observable from the

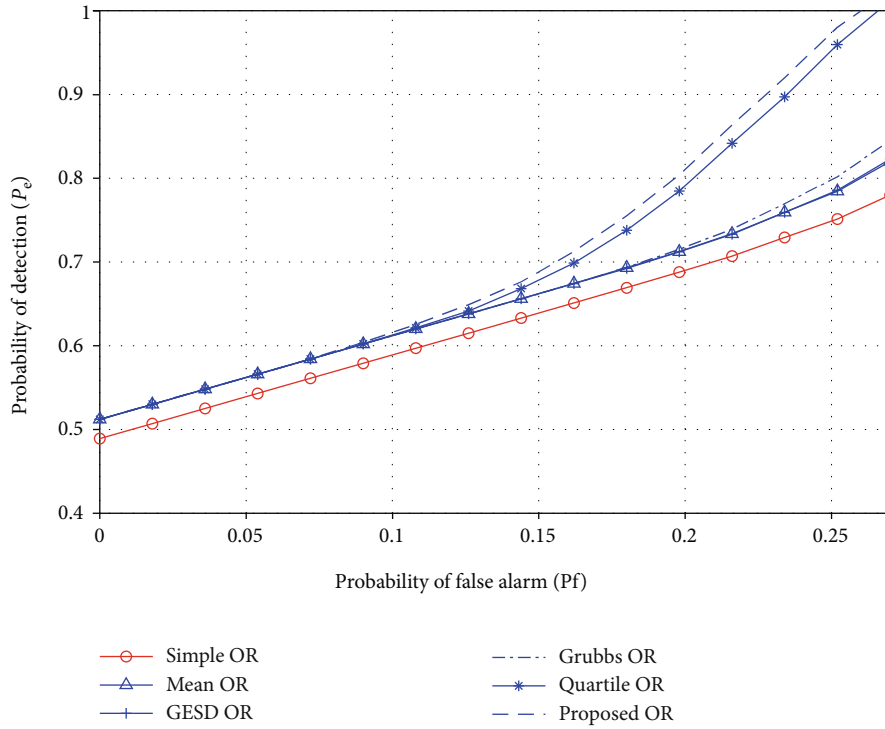


FIGURE 13: Probability of detection (P_d) vs. probability of false alarm (P_f) of the OR HDF schemes with the normal SUs and MUs.

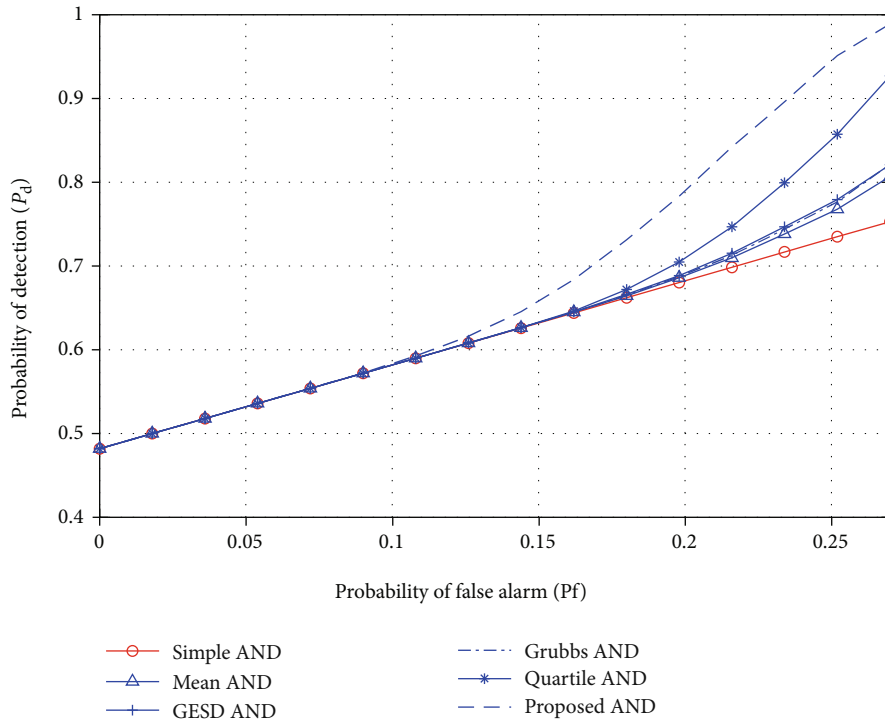


FIGURE 14: Probability of detection (P_d) vs. probability of false alarm (P_f) of the AND HDF schemes with the normal SUs and MUs.

results in Figure 13 that the proposed median outlier test has attained maximum detection probability, whereas the simple OR HDF scheme shows minimum detection probability. The median test performance is next followed by the quartile and

corresponding Grubbs tests. The mean and GESD tests produce similar detection results with their detection probabilities comparatively limited as compared with the detection probability of the proposed test.

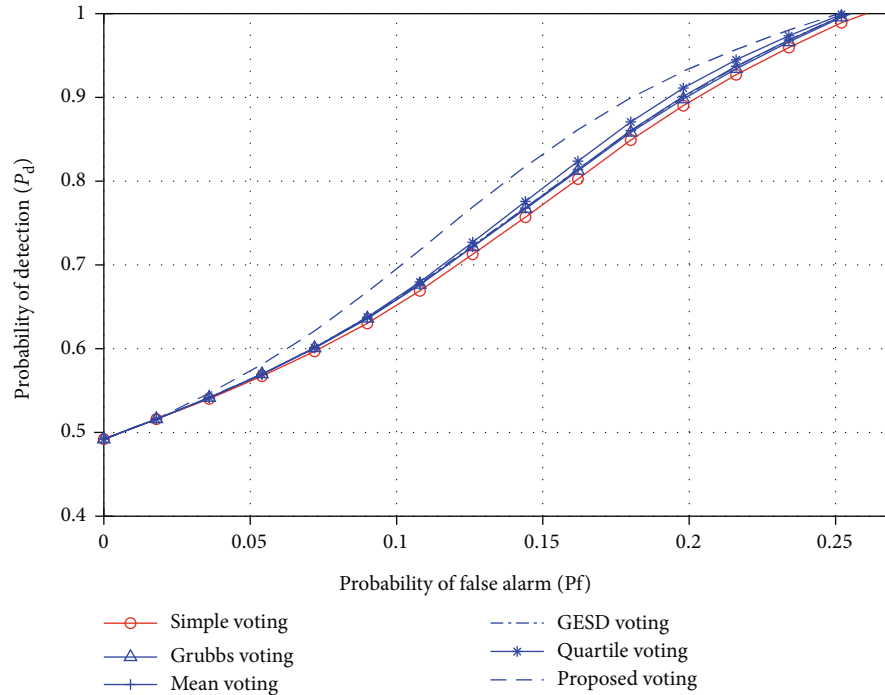


FIGURE 15: Probability of detection (P_d) vs. probability of false alarm (P_f) results of the voting HDF schemes with the normal SUs and MUs.

In Figure 14, it is noticeable that the proposed median test-based AND HDF scheme has high detection probability with minimum false alarm probability in comparison with the other outlier-based detection results. The proposed median test-based ROC results are followed by the quartile test, GESD test, and Grubbs test. The mean test has the minimum detection probability as compared with all other outlier detection tests.

The voting HDF scheme ROC results are shown in Figure 15 to compare the proposed and various outlier detection tests. In Figure 15, all other outlier detection tests give similar detection results, while the proposed median test is able to achieve significant improvement over all other outlier detection tests.

5. Conclusion

The CSS is reliable in detecting the presence and absence of the PU signal; however, the participation of the MUs in the CSS results in false report collection at the FC. This research work considered the involvement of the MUs in the CSS. An improved statistical analysis is employed for spectrum sensing in the CRN. The focus in this research work is to boost the performance of the traditional HDF schemes with some statistical analysis. The false reports of the MUs can be efficiently detected using different outlier tests. The results of the four outlier tests are compared and concluded that the median plus-minus 3 test outperforms other statistical outlier detection tests. The proposed outlier test is accurately detecting the behavior of the LMU and OMU in the CSS.

For future work, it is recommended that these outlier statistics should be further investigated by applying them to

detect the MU behavior of always yes, always no, and random opposite categories of the MUs. Similarly, other categories of outlier detection techniques such as density-based, depth-based, and cluster-based schemes can be employed.

Data Availability

The data used to support the finding of this study are included within the article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2018-0-01426) supervised by the IITP (Institute for Information and Communication Technology Planning & Evaluation) and in part by the National Research Foundation (NRF) funded by the Korea Government (MSIT) (No. 2019R1F1A1059125).

References

- [1] K. Ben Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, 2009.
- [2] S. M. Diamond and M. G. Ceruti, "Application of wireless sensor network to military information integration," in *2007 5th*

- IEEE International Conference on Industrial Informatics*, pp. 316–322, Vienna, Austria, June 2007.
- [3] N. Gul, I. M. Qureshi, A. Omar, A. Elahi, and S. Khan, “History based forward and feedback mechanism in cooperative spectrum sensing including malicious users in cognitive radio network,” *PLOS ONE*, vol. 12, no. 8, article e0183387, 2017.
 - [4] J. Mitola III, “Cognitive radio for flexible mobile multimedia communications,” *Mobile Networks and Applications*, vol. 6, no. 5, pp. 435–441, 2001.
 - [5] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, “Malicious user detection in a cognitive radio cooperative sensing system,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 8, pp. 2488–2497, 2010.
 - [6] T. S. Sundara and N. Padmaja, “Performance analysis of cognitive radio based on cooperative spectrum sensing,” *International Journal of Engineering Trends and Technology (IJETT)*, vol. 4, no. 4, pp. 821–827, 2013.
 - [7] N. Gul, I. M. Qureshi, A. Naveed, A. Elahi, and I. Rasool, “Secured soft combination schemes against malicious-users in cooperative spectrum sensing,” *Wireless Personal Communications*, vol. 108, no. 1, pp. 389–408, 2019.
 - [8] P. Pandya, A. Durvesh, and N. Parekh, “Energy detection based spectrum sensing for cognitive radio network,” in *2015 Fifth International Conference on Communication Systems and Network Technologies*, pp. 201–206, Gwalior, India, April 2015.
 - [9] O. León and K. P. Subbalakshmi, *Cognitive Radio Network Security BT - Handbook of Cognitive Radio*, W. Zhang, Ed., Springer Singapore, Singapore, 2017.
 - [10] M. Khasawneh and A. Agarwal, “A collaborative approach for monitoring nodes behavior during spectrum sensing to mitigate multiple attacks in cognitive radio networks,” *Security and Communication Networks*, vol. 2017, Article ID 3261058, 16 pages, 2017.
 - [11] S. Shobana, R. Saravanan, and R. Muthaiah, “Matched filter based spectrum sensing on cognitive radio for OFDM WLANs,” *International Journal of Engineering and Technology*, vol. 5, no. 1, pp. 142–146, 2013.
 - [12] N. Do and B. An, “A soft-hard combination-based cooperative spectrum sensing scheme for cognitive radio networks,” *Sensors*, vol. 15, no. 2, pp. 4388–4407, 2015.
 - [13] Y. L. Lee, W. K. Saad, A. A. El-Saleh, and M. Ismail, “Improved detection performance of cognitive radio networks in AWGN and Rayleigh fading environments,” *Journal of Applied Research and Technology*, vol. 11, no. 3, pp. 437–446, 2013.
 - [14] W. Han, J. Li, Z. Li, J. Si, and Y. Zhang, “Efficient soft decision fusion rule in cooperative spectrum sensing,” *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 1931–1943, 2013.
 - [15] N. Gul, I. M. Qureshi, A. Elahi, and I. Rasool, “Defense against malicious users in cooperative spectrum sensing using genetic algorithm,” *International Journal of Antennas and Propagation*, vol. 2018, Article ID 2346317, 11 pages, 2018.
 - [16] A. Haldorai and U. Kandaswamy, “Secure distributed spectrum sensing in cognitive radio networks,” *Intelligent Spectrum Handovers in Cognitive Radio Networks*, pp. 175–191, 2019.
 - [17] J. Luo and X. He, “A soft-hard combination decision fusion scheme for a clustered distributed detection system with multiple sensors,” *Sensors*, vol. 18, no. 12, p. 4370, 2018.
 - [18] R. Ujjimatad and S. R. Patil, “Spectrum sensing in cognitive radio networks with known and unknown noise levels,” *IET Communications*, vol. 7, no. 15, pp. 1708–1714, 2013.
 - [19] N. Reisi, S. Gazor, and M. Ahmadian, “Distributed cooperative spectrum sensing in mixture of large and small scale fading channels,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5406–5412, 2013.
 - [20] N. Reisi, S. Gazor, and M. Ahmadian, “A distributed average likelihood ratio detector for detection of signals in frequency-selective Nakagami channels,” *IEEE Wireless Communications Letters*, vol. 3, no. 3, pp. 245–248, 2014.
 - [21] S. Shrivastava, S. John, A. Rajesh, and P. K. Bora, “Preventing collusion attacks in cooperative spectrum sensing,” in *2018 International Conference on Signal Processing and Communications (SPCOM)*, pp. 90–94, Bangalore, India, July 2018.
 - [22] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, “A survey on security threats and detection techniques in cognitive radio networks,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 428–445, 2013.
 - [23] M. Abdelhakim, L. E. Lightfoot, J. Ren, and T. Li, “Distributed detection in mobile access wireless sensor networks under Byzantine attacks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 4, pp. 950–959, 2014.
 - [24] N. Gul, A. Naveed, A. Elahi, T. Saleemkhattak, and I. M. Qureshi, “A combination of double sided neighbor distance and genetic algorithm in cooperative spectrum sensing against malicious users,” in *2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 746–753, Islamabad, Pakistan, January 2017.
 - [25] D. Divya and S. S. Babu, “Methods to detect different types of outliers,” in *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, pp. 23–28, Ernakulam, India, March 2016.
 - [26] M.-j. Zhou and X.-j. Chen, “An outlier mining algorithm based on dissimilarity,” *Procedia Environmental Sciences*, vol. 12, pp. 810–814, 2012.
 - [27] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, “Outlier detection for temporal data: a survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2014.
 - [28] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, “Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median,” *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, 2013.
 - [29] C. C. Agarwal, *Outlier Analysis Second Edition*, vol. 24, no. 2, 2017.
 - [30] R. Kumar and A. Khadar, “A survey on outlier detection techniques in dynamic data stream,” *International Journal of Latest Engineering and Management Research*, vol. 2, no. 8, pp. 23–30, 2017.
 - [31] T. A. Cohn, J. F. England, C. E. Berenbrock, R. R. Mason, J. R. Stedinger, and J. R. Lamontagne, “A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series,” *Water Resources Research*, vol. 49, no. 8, pp. 5047–5058, 2013.
 - [32] Y. Guo, Q. Xu, S. Sun, X. Luo, and M. Sbert, “Selecting video key frames based on relative entropy and the extreme studentized deviate test,” *Entropy*, vol. 18, no. 3, p. 73, 2016.
 - [33] D. Cousineau and S. Chartier, “Outliers detection and treatment: a review,” *International Journal of Psychological Research*, vol. 3, no. 1, 2010.

- [34] A. Li, M. Feng, Y. Li, and Z. Liu, "Application of outlier mining in insider identification based on boxplot method," *Procedia Computer Science*, vol. 91, pp. 245–251, 2016.
- [35] V. Praveen, T. Delhi Narendran, R. Pavithran, and C. Thirumalai, "Data analysis using box plot and control chart for air quality," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, pp. 1082–1085, Tirunelveli, India, May 2018.